UC Merced
UC Merced Electronic Theses and Dissertations

Title
The applications of GISystems to wilderness search and rescue, an overview within a GIScience context and examples from Yosemite National Park.

Permalink
https://escholarship.org/uc/item/0hq8z72c

Author
Doherty, Paul James

Publication Date
2013

Peer reviewed|Thesis/dissertation

eScholarship.org
Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, MERCED

The applications of GISystems to wilderness search and rescue, an overview within a GIScience context and examples from Yosemite National Park.

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Environmental Systems

by

Paul James Doherty

Committee in charge:
Professor Qinghua Guo, Chair
Professor Yihsu Chen
Professor Ruth Mostern
Professor Samuel Traina

2013

Copyright (or ©)
Paul Doherty, 2013
All rights reserved

Dedication

This text is dedicated in loving memory of the life of my Father, George Neal Doherty, who always encouraged me to open new doors, challenge myself, and to uphold principles of integrity at all times. His courageous fight with Colon cancer during the time of my PhD candidacy will continue to inspire me throughout life.

Table of Contents

List of Tables

List of Figures

Acknowledgements

This study was supported by the National Science Foundation (grant nos. BDI-0742986 and SBE-1031914).

I would express my appreciation and gratitude to all involved with this research initiative and will attempt to highlight some of those people here. First and foremost I would like to thank Yosemite Search and Rescue and all of the volunteers who risk their lives in service so that others may live.

I will be forever grateful to my fellow Rangers and friends, David and Dov Pope, Joy Marschall, Jon Dill, Merry and Werner Braun, and Keith Lober who provided a positive, challenging, exhilarating, and decidedly worthwhile experience working with YOSAR. Early discussions with the SAR-GIS Discussion Forum and now the MapSAR Team have helped motivate me and reassure that this research is worthwhile and will make a difference. Of these +400 members, George Durkee, Vanessa Glyn-Linaris, Tom Patterson, Donald Ferguson, Jon Pedder, Arnold Gaffrey, Art Fortini, Richard Laing, Loren Pfau, Marcus Kitchens, Bob Gerlt, Matt Baker, Edan Cain, Doug Morgenthaler, Liz Sarow, Charles Twardy, Monty Bell, and Robert Koester deserve recognition for their service and time spent sharing their insight. I also want to thank Wren Haaland, Sarah Nurit, Greg and Diane Ambrose, and Raj Bolla for tirelessly scanning, reading, and georeferencing many years of case incident reports.

The initial vision and pathway to this successful research initiative with UC Merced was provided by Stephen Shackelton, former Yosemite National Park Chief Ranger and for this I am truly thankful. A very special thanks to my colleague and close friend Jared Doke for co-authorship, dedication, research assistance, and overall support of our Search and Rescue GIS research. Without these two people, none of this research would be possible.

Thank you to the Yosemite National Park Division of Resource Management and Science for research permissions and constructive methodology suggestions. I would also like to acknowledge my peers and co-authors, John Wieczorek, Yu Liu, Otto Alvarez, and Wenkai Li for their contributions and partnership. This dissertation would not have been possible without the guidance of my committee members Sam Train, Ruth Mostern, and Yihsu Chen. Finally, thank you to my mentor and PhD Advisor Quinghua Guo who has given me a truly unique and most rewarding academic experience.

Curriculum Vita

University of California at Merced Merced, CA
Division of Graduate Studies May 2013
Doctor of Philosophy in Environmental Systems
Advisor: Dr. Qinghua Guo

Dissertation: “The applications of GISystems to wilderness search and rescue, an overview within a GIScience context and case-studies from Yosemite National Park.”

Hofstra University Hempstead, NY
College of Liberal Arts and Sciences December 2007
Master of Science Degree in Biology
Advisor: Dr. Julie Heath

Thesis: “Factors affecting Piping Plover (*Charadrius melodus*) parental care behavior and hatching success on Long Island, NY.”

SUNY College of Environmental Science & Forestry Syracuse, NY
Department of Environmental Forest Biology May 2005
Bachelor of Science Degree in Environmental Biology (135 credits)
Advisor: Dr. Guy Baldassarre

Publications

Alvarez, O., Q. Guo, R. Klnger, P. Doherty, and W. Li. *in press.* Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation. *International Journal of Climatology*

Doherty P.J., Q. Guo, and O. Alvarez. 2012. Expert versus machine: A comparison of two suitability models for emergency helicopter landing areas in Yosemite National Park. *Professional Geographer*

Doherty, P., Q. Guo, Y. Liu, J.Wieczorek, and J. Doke. 2011. “Georeferencing Incidents from Locality Descriptions and Its Applications: a Case Study from Yosemite National Park Search and Rescue.” Transactions in GIS 15 (6)

Doherty P.J. and J.A. Heath. 2011. The effects of hatchability and disturbance on Long Island Piping Plover hatching success. *Journal of Wildlife Management* 75

Dissertation Abstract

The process of searching for and rescuing people in distress provides an appealing spatial problem for geographers to support and for testing theoretical developments in the real-world. Essentially, the fundamental goals of wildland search and rescue (WiSAR) are to locate persons in need and extract them from dangerous situations. More recently, WiSAR researchers and professionals have also cited a need for proactive incident prevention as a critical responsibility, known as preventative search and rescue (PSAR). This research draws on my recently completed case-studies in Yosemite National Park and community development amongst GISystem and WiSAR professionals. Each of these components of WiSAR are inherently spatial and should be evaluated in light of emerging technology and theoretical advances in spatial sciences. Of particular interest are the real-world implications of time geography and probabilistic modeling of objects in space.

This dissertation is formatted using standalone chapters for publication. In the first chapter I discuss the overall need for GIS related research in search and rescue as well as a conceptual framework for doing so. In Chapter II I present research related to preventing incidents. This chapter features two papers with first describing methods for georeferencing text based locality descriptions and preliminary findings on spatial patterns of incident. The second paper presents a spatiotemporal method for mapping the probability of WiSAR in Yosemite National Park by month. Chapter III presents research related to searching for missing persons. I present a paper that describes two probability of area methods and findings related to applying global versus local models.

Finally, Chapter IV pertains to the rescue phase of WiSAR. Here I present a paper that compares an expert (weighted overlay) versus machine-learning technique for mapping suitability of helicopter landing areas. Following this I conclude the dissertation with final remarks and recommendations.

Chapter I: Introduction

Researching Wildland Search and Rescue GISystems: An Inherently Spatial Topic for Geographers and GIScientists

[Formatted and prepared for the *Annals of the Association of the America Geographers*]

Wildland search and rescue (WiSAR) is the process of searching for and rescuing people in distress. It provides an appealing spatial problem for geographers and allows for testing theoretical developments in the real-world. In this paper we aim to present a background on WiSAR, describe the operational components, identify current GISystem use-cases, and highlight areas of GIScience that should be explored within the context of WiSAR operations (Prevent, Search, and Rescue).

In the United States, the number of people and total time spent participating in one or more outdoor activities is believed to have increased steadily since 1999 and in 2011, 141.1 million Americans participated in some form of outdoor recreation (The Outdoor Foundation 2012). With this comes increased frequency of activities in and around wildland (wilderness / rural) areas, including hiking, climbing, skiing, etc., especially on public lands such as US National Parks. While this can be viewed as a positive impact on civic engagement in the protection of public lands – it places an increased demand on government agencies responsible for protecting the landscape from recreational impacts while at the same time, protecting people while they recreate in the landscape. When mishaps occur in undeveloped areas where modern emergency service equipment cannot access, a search and rescue response is typically required. In Heggie and Amundson (2009), a 15-year survey of US National Parks, it was found that the financial cost of WiSAR operations is between 3 and 4 million dollars per year and it was estimated that 1 in 5 incidents would have resulted in a victim fatality if a response was not initiated. Moreover, rescues require many hours of specialized training to prepare for and expose rescuers, even those that are unpaid volunteers, to great personal risk. Any technology that can reduce the frequency, risk, and / or resource cost would be of great benefit to the US National Park Service, volunteer teams, and other agencies responsible for WiSAR response.

WiSAR operations can typically be broken down into four phases: Locate, Access, Stabilize, and Transport (LAST; Hill & Gale, 1997). More simply, it can be described as the search (Locate) and rescue (Access, Stabilize, and Transport) phases. In some cases the process of locating a victim (search) takes much longer than the actual rescue. In other cases, particularly in extreme environments such as vertical terrain or high energy riverine environments the rescue phase will be more technical and complex, taking a longer period of time than locating the victim. In all cases time to completion and safety of rescuers are high priority and require the communication and understanding of relevant geographic elements (e.g. likely area of missing person, fastest route to the victim, coordinates of hazards in the area). Additionally, WiSAR researchers and professionals have also cited a need for proactive incident prevention as a critical responsibility for land management agencies, known as preventative search and rescue (Hung and Townes 2007a, Heggie and Amundson 2009, Heggie and Heggie 2009). This requires a high level of geographic understanding prior to an incident actually occurring.

Each component of WiSAR (Search, Rescue, and Prevention) is inherently spatial and should be evaluated in light of emerging technology and theoretical advances in spatial sciences. If Geographic Information Science (GIScience) is the theory behind the development, use, and application of geographic information systems (GISystems; Goodchild, 1992), then WiSAR is an ideal topic for GIScientists to study. However, to date, GISystems have not been integrated into WiSAR as they have been in other emergency services. One simple reason for this may be historically poor access to GIS technology. Most WiSAR teams in the US are volunteer (unpaid) and have limited or no exposure to advances in computer technology due to budget limitations and have limited access to GISystem training. Manager perceptions may be that the advantages of a true GISystem are be outweighed by the monetary or training-time cost of GISystem integration until proven otherwise. Yet the demands placed on these agencies for WiSAR continues to grow.

In S.L. Cutter's (2003) historic review of GIScience in disaster and emergency management, she highlighted challenges associated with the use of GISystems previously outlined by GIScientists (Alexander 1991, Goodchild 2003, Kwan 2003) and proposed core areas that were high priority for GIScience research. Since writing this paper, use of GISystems has expanded into many core workflows of emergency management agencies with dedicated software packages (e.g. HAZUS; Schneider and Schauer 2006; Tate, Cutter, and Berry 2010). Emergency Management and WiSAR are similar and therefore have similar challenges with utilizing GISystems. Today, the lack of GISystem utilization in WiSAR may be due to similar constraints previously faced by emergency management in the following manners: understandable user interfaces; data quantity, quality, and integration; real-time data and information. With advances in technology and more widespread availability of commercial-off-the-shelf GIS software, many of these constraints have been lifted and communities of WiSAR practitioners interested in GISystems have recently emerged. Recent research proposals and abstracts in *Annual Meeting of the Association of American Geographers* confirm the need to explore ways to make GISystem software more accessible, continue to raise GISystem awareness, and for spatial analyses to be incorporated into WiSAR operations (Doherty, Ferguson, *et al.* 2012).

In response to the needs described above we propose two related concepts; *WiSAR operations and prevention efforts can benefit from the incorporation of GISystems, and likewise, GIScience theory can benefit from the application GISystems to WiSAR.* Once integrated, new spatial analyses may emerge or converge as is seen at the interface of GISystems and other industries (e.g. crime mapping, wildland fire GIS, geomedicine). There have been some preliminary attempts to integrate GISystems to WiSAR, but our objectives here is to formally investigate prominent GIScience questions related to WiSAR and raise awareness in each respective community of professionals. We use the Prevention, Search, and Locate components of WiSAR to explain this synergistic relationship and illustrate the conceptual relationships of each with current topics in GIScience and Geography research. The case study we present here is an actual WiSAR incident in Yosemite National Park from 2004 (Figure 1.1). In this case, a 61-year old woman was reported missing near Half Dome surrounded by steep terrain.

Prevention

In order to intentionally prevent an event from occurring there are some prerequisites required. The incident must be preventable, one must know about the possibility of the incident ahead of time, and an action must occur at the right time and place to inflict some plausible change to prevent the situation (Lawson *et al.* 2006). In Preventative Search and Rescue (PSAR), the primary tool employed for prevention is to educate people at risk by providing information. In order to know what information to provide, PSAR personnel must be familiar with the when and where details of current hazards (weather, trail conditions, overall fitness of visitors) as well as hazards that have occurred in the past (spatio-temporal information). GISystems offer a robust solution to document and analyze the geographic elements that are factors controlling the presence or absence of events, such as disease outbreaks. Just as in classic spatial epidemiology studies (Dr. Snow's map of cholera in England, Koch and Denike 2009) that raised the notion of how geography might be used to understand disease transmission, this concept should be extended to SAR at a local level to provide a GIScience approach to risk assessment.

To date, little or no research has implemented the use of GISystems in the spatial epidemiology of search and rescue as suggested here. There appears to be a widespread interest in WiSAR risk assessment research , especially in national parks beginning over a decade ago with the paper Morbidity and Mortality in the Wilderness (Montalvo *et al.* 1998). This was the first of several retrospective studies to discuss how WiSAR incident data could be used to "guide future wilderness use, education, and management" and alluded to how a "standardized, computerized database would greatly facilitate future evaluations, decisions, and policies". Since then, however, findings in retrospective studies are limited to age, sex, recreational activity, mode of rescue, and contributing factors in WiSAR incidents (Wild 2008, Forrester and Holstege 2009, Heggie and Amundson 2009, Heggie and Heggie 2009). In order to know what information to provide however, PSAR personnel must be familiar with the when and where details of current hazards (weather, trail conditions, overall preparedness of visitors) as well as hazards that have occurred in the past (spatio-temporal information). Having knowledge of where and when incidents have occurred in the past can enhance our knowledge of where incidents will occur in the future and lead to prevention of incidents.

To begin understanding incident prevention we must first describe the locations in which previous incidents have occurred. In wilderness locations there is lack of building addresses for geocoding incident locations and coordinates are often not recorded at the time of the incident. Instead, reports are written and localities are described in text. Compared with geographic coordinates, textual locality descriptions can be seen as qualitative or semi-qualitative, representing a rough distribution range of the locality based on the understanding of the person who recorded the information and subject to the interpretation of the person reading it. Hence, uncertainty associated with these descriptions is inevitable (Goodchild 2004, Liu *et al.* 2009). However, georeferencing techniques such as the point-radius method (Wieczorek *et al.* 2004a) and probabilistic method (Guo *et al.* 2011a) have provided us with quantitative assessment tools that allow us to create spatial data that accounts for uncertainty in textual data. In particular, such text-based data of incidents can provide rich sources of cultural and geographic information if they are simply georeferenced for future use (Mostern and Johnson 2008).

In Doherty et al. 2011, 10-years of search and rescue incident reports in Yosemite National Park provided a unique opportunity to study the textual description and its uncertainty. It was found that while largely text based, the locality descriptions in historic reports were precise enough to be georeferenced. In addition, once incidents can be georeferenced they can be archived and also analyzed using spatial statistics to determine areas with significant amounts of WiSAR activity in the past. This was the first study where such georeferencing techniques were applied to incident data for analysis purposes. A follow up study showed that probabilistic models based on incident location and environmental layers alone could be used to determine areas of high incident probability in the future (P. J. Doherty et al. *in press*). This allowed for a novel test of forecasting techniques with presence-only data, or geographic one-class data (GOCD; Guo et al. 2011) with the presence and background learning algorithm (PBL; W. Li, Guo, and Elkan 2011). More importantly, the Yosemite dataset revealed the value of incorporating temporal factors in PBL algorithms – analyzing incidents by the time of year they occur reveals spatial patterns of probability that would not otherwise be seen. While this early research is very compelling, future research is needed in WiSAR Prevention and GIScience to focus on spatio-temporally enabled records management systems and further analyze incidents by specific attribute types (incident type, method of rescue, victim demographics) to direct and prioritize PSAR efforts.

Search

It is estimated that worldwide over 100,000 people will be reported missing in the wilderness each year and some are never found alive or recovered (National Association for Search and Rescue 2005). Search incidents are inherently spatial problems that need to be solved quickly. The search for a missing person is the Locate phase of a WiSAR incident as explained earlier (Locate, Access, Stabilize, Transport) and represents a true emergency. Finding a missing person in the wilderness or rural area is a precursor to a successful rescue.

In search incidents, time is a critical variable in determining patient survival, with survivability declining significantly after the first 51-hours of a search (Adams *et al.* 2007). Therefore search managers need to develop a search plan quickly and establish an area of containment. Once it is determined that someone is missing, the first step to finding these missing persons is to draw a boundary around the area where they are most likely to be lost or injured. Then the search area must be divided into feature-based polygons known as search segments. Finally, the search segments are prioritized in order of highest estimated probability and assigned to search teams.

There are concepts of probability in this process known as Probability of Area (POA), Probability of Detection (POD), and Probability of Success (POS) as described in equation 2.1 below.

$$POS = POA \; x \; POD \tag{1.1}$$

Search theory is completely dependent upon an accurate assessment of how well a search area was covered by a team (POD) and that the boundary of the area (polygon) being searched actually contains the missing person (POA).

$$POD = 1 - e^{-c} \tag{1.2}$$

Where *c* is the coverage and *e* is the base of the natural logarithm. Coverage is the ratio of two areas: search effort as track length multiplied by width (the polyline representing searchers movements and the buffer indicating where they can effectively search), divided by the total area searched. In search planning like most complex decision making processes, errors in judgment (underestimated POA, overestimated POD) early in the planning stages significantly hampers the search effort despite subsequent decisions.

The objectives of any search operation are to maximize POS as quickly as possible which is a function of increasing POA and POD. These concepts are grounded in Operations Research and were initially developed with maritime and aviation search techniques in mind (Koopman 1980, Stone 1989, 2007, Koester *et al.* 2004). At this time, most WiSAR operations utilize these concepts in a variety of ways but with little geospatial resources to accurately measure either POA or POD. With the integration of Global Position Systems (GPS) for searchers, a quantitative index of POD can be obtained by measuring coverage based on searchers' GPS receiver track-log, but the POA is still very much theoretical. Considerable effort in missing person research has been dedicated to analyzing lost person behavior and summaries of previous search incident outcomes to generate a POA value (Syrotuck 1977; R. J. Koester and Stooksbury 1995), and the best source of these data is the International Search and Rescue Incident Database (ISRID; R. Koester 2008). The ISRID database provides an excellent tool for comparing incident outcomes based on the missing person profile (e.g. age, gender, mental status) and activity (e.g. hiking, biking, gathering). However this database has limitations in its use as a search planning tool because of its lack of GISystem integration and the inability to infer global statistics onto local landscapes. For instance would a hiker in rolling topography in Great Britain be expected to behave the same as hiker on a mountainous trail in the Sierra Nevada of the United States?

In Doke et al. (in press), eleven years of missing person case incident reports were reviewed from Yosemite National Park beginning in the year 2000. The initial planning point (IPP; e.g. last known location or point last seen) and found locations for each of these incidents were georeferenced using the point radius method by decoding geographic data that were buried in narrative text (Figure 1.2). These two pieces of information were used to compare a local dataset to that of the ISRID. While many statistical summaries of the ISRID were similar to the Yosemite dataset, the localized dataset was significantly different in key areas such as distance between IPP and found locations (in general, a shorter total distance in Yosemite versus global incidents). This research also demonstrated the usefulness of other geographic data such as watershed boundaries (USGS 2012) for studying missing person search outcomes and developing POA strategy.

One way to improve on this statistical POA strategy would be to take into account the physical limitations to walking velocity in the landscape and combine it with the theoretical POA. Travel cost modeling (cost surface plus cost distance) integrated into GISystems can produce potential path areas (Kim and Kwan 2003, Lin and Goodrich 2010) and this has been cited as a critical need in recent years by both SAR (Koester 2008) and GIScience researchers (Miller and Bridwell 2009) for improving POA calculations. Here (Figure 1.3) we provide an example of such a tool used on an actual search incident. The result of this simulation is a display of classified isochrones by hour. While such isochrones products are a valuable visualization tool, they must be correctly interpreted by a trained search manager and should be further tested and refined by GIScientists.

While Search related applications of GISystems are used in Maritime environments (Search and Rescue Optimal Planning System Kratzke, Stone, and Frost 2010), a specific graphic user interface (GUI) is still lacking for wilderness search. Use of commercial-off-the-shelf software is beginning to increase in frequency (Ferguson 2008, Theodore 2009, Durkee 2010, Filipkowska *et al.* 2012) and remote sensing techniques have been applied to more recent high profile missing person incidents (e.g. Steve Fossett and Jim Gray). At the time of this writing, a community of practice has formed around WiSAR and GIS and they have made GISystem tools available, known as MapSAR[1]. In addition, government agencies are beginning to build requirements for GISystem integration into missing person search operations (*personal communication*; US National Park Service, California Emergency Management Agency). Therefore it is critically important that GIScience approaches are available to ensure the best use of GISystems for search operations. Future research in the Search component of WiSAR and GIScience should focus on quantifying POA and POD and incorporating time-geography into operational planning. This interesting coupling of theory and practice will benefit GIScientist and WiSAR practitioner alike.

Rescue

Once a person(s) is located, access has been established, and the situation has been stabilized, rescuers will need to make a transport decision. If the person is not ill or injured they will probably be guided out in the same fashion the rescuers arrived, by walking. If the person cannot leave the wilderness without support then they may be carried out on a wheeled-litter. This can require several people and a long period of time depending on the terrain and weather conditions. A critical piece of information for making this decision would be the amount of travel time needed to access the patient and then the total time needed to transport the patient based on the route (i.e. which trail to use) and mode of transportation (foot, horse, helicopter). This information is typically described by GIScientists as Location Science.

Location Science is the sub discipline of GIScience that seeks to solve problems based on the classic mathematical principles of maximizing benefits and/or minimizing costs in travel (Dijkstra 1959). GISystems provide a platform for solving these problems (Li and Yeh 2005, Murray and Tong 2007) by enabling mathematical analyses to be done in three dimensions and in a representative network of points, lines, and polygons (Miller 1996). Early Location Science research was actually conducted with the intent to solve the real-world problems that emergency service agencies face. For example, if we were to construct a new fire station in a growing city, where should we place it to handle fire response in the shortest time possible (Hogg 1968, Toregas *et al.* 1971, Church and ReVelle 1974)? Much of the initial theories behind such analyses were based on service areas defined as a simple radius extended outward from each station. These analyses however do not account for topography, barriers to travel, and travel corridor attributes. More recently researchers have approached

[1] www.mapsar.net

network analyses from a regional service coverage approach utilizing more advanced algorithms (Indriasari *et al.* 2010) in conjunction with a network dataset. These same techniques could be used to make decisions on what access points should be used for a ground-based rescue or for making decisions on how long a rescue may take (Figure 1.4a).

Finally, if a rescue is deemed extremely urgent and requires a helicopter, then a suitable site (landing zone) must be chosen within close proximity to the rescue scene. In addition, to land safely a landing zone must meet the terrain requirements of minimal slope and be free of hazards (vegetation and man-made structures). If a landing zone cannot be found, alternatives may include more dangerous helicopter rescue techniques or calling off a helicopter operation altogether. A landing zone suitability map (Figure 1.4b) would be a useful tool to assist in this decision making process. In some cases, a rescue team may have a set of coordinates for previously used landing areas, but incidents may occur in an unfamiliar location without a known landing location nearby. This is the problem researched in P. J. Doherty, Guo, and Alvarez 2012 where classic site-search analysis techniques (Malczewski 2004) were compared to machine learning algorithm models (Baldwin 2009). While this solved a practical problem of where to land helicopters, it also contributed a key GIScience finding that both expert and machine learning approaches can produce accurate suitability maps and be used to validate each other. In the future, GIScience research related to the Rescue component of WiSAR should focus on reducing the elapsed time between locating patients and rescue completion. This will have practical applications as well as provide innovative research topics in Location Science and beyond.

References

Adams, A.L., Schmidt, T. a, Newgard, C.D., Federiuk, C.S., Christie, M., Scorvo, S., and DeFreest, M., 2007. Search is a time-critical event: when search and rescue missions may become futile. *Wilderness & Environmental Medicine*, 18 (2), 95–101.

Adriaensen, F., Chardon, J.P., De Blust, G., Swinnen, E., Villable, S., Gulinck, H., and Matthysen, E., 2003. The application of "least-cost" modelling as a functional landscape model. *Landscape and Urban Planning*, 64 (4), 233–247.

Alexander, D., 1991. Information technology in real-time for monitoring and managing natural. *Progress in Physical Geography*, 15 (3), 238.

An, L. and Brown, D.G., 2008. Survival Analysis in Land Change Science : Integrating with GIScience to Address Temporal, (February 2013), 37–41.

Baldwin, R. a., 2009. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy*, 11 (4), 854–866.

Bateman, I.J., Garrod, G.D., Brainard, J.S., and Lovett, A. a., 1996. Measurement Issues in the Travel Cost Method: a Geographical Information Systems Approach. *Journal of Agricultural Economics*, 47 (1-4), 191–205.

Berger, A., Della Pietra, S., and Della Pietra, V., 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22 (1992), 39–71.

Bownds, J.W., Ebersole, M.J., Lovelock, D., O'Connor, D.J., and Toman, R.J., 2007. Win CASIE III: Computer Aided Search Information Exchange.

Van den Broek, M., Brederode, E., Ramírez, A., Kramers, L., Van der Kuip, M., Wildenborg, T., Turkenburg, W., and Faaij, A., 2010. Designing a cost-effective CO2 storage infrastructure using a GIS based linear optimization energy model. *Environmental Modelling & Software*, 25 (12), 1754–1768.

Church, R. and ReVelle, C., 1974. The maximal covering location problem. *Papers in regional science*, 32 (1), 101–118.

Clark, W., 1965. Markov chain analysis in geography: an application to the movement of rental housing areas. *Annals of the Association of American …*, 55 (2), 351 – 359.

Cutter, S.L., 2003. GI Science, Disasters, and Emergency Management. *Transactions in GIS*, 7 (4), 439– 445.

Dijkstra, E.W., 1959. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1, 269–271.

Doherty, P., Ferguson, D., Goodrich, M.A., Koester, R.J., and Doke, J., 2012. Wilderness Search & Rescue and GIScience. *In*: *Annual Meeting of the Association of American Geographers*. New York.

Doherty, P., Guo, Q., Liu, Y., Wieczorek, J., and Doke, J., 2011. Georeferencing Incidents from Locality Descriptions and its Applications: a Case Study from Yosemite National Park Search and Rescue. *Transactions in GIS*, 15 (6), 775–793.

Doherty, P.J., Guo, Q., and Alvarez, O., 2012. Expert versus machine: A comparison of two suitability models for emergency helicopter landing areas in Yosemite National Park. *Professional Geographer*.

Doherty, P.J., Guo, Q., Li, W., and Doke, J., n.d. Space-Time analyses for forecasting and understanding future incident occurrence: a case-study from Yosemite National Park using the presence and background learning algorithm. *International Journal of Geographical Information Science*.

Doke, J., 2012. Analysis of Search Incidents and Lost Person Behavior in Yosemite National Park.

Durkee, G., 2010. GIS Joins Search for a Missing Hiker on California's Mount Whitney. *ArcWatch*, Apr.

Esri, 2012. ArcGIS 10.1.

Federal Standards and Procedures for the National Watershed Boundary Dataset (WBD), 2012. Reston, Virginia.

Ferguson, D., 2008. GIS for Wilderness Search and Rescue. *In*: *Esri Federal User Conference*. Washington D.C., 1 – 11.

Fernandez, M.A., Blum, S.D., Reichle, S., Holzman, B., and Hamilton, H., 2009. Locality uncertainty and differential performance of four different common niche-modeling techniques. *Biodiversity Informatics*, 6 (1), 36–52.

Fielding, A.H. and Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24 (1), 38–49.

Filipkowska, E., Koester, R.J., Chrustek, R., and Zaród, M., 2012. Lost and Found in the Polish Carpathian Mountains. *ArcNews*, 34 (3).

Fink, D. and Hochachka, W., 2010. Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications*, 20 (8), 2131–47.

Forrester, J.D. and Holstege, C.P., 2009. Injury and illness encountered in Shenandoah National Park. *Wilderness & environmental medicine*, 20 (4), 318–26.

Frost, J.R., 1999. Principles of search theory. *Response*, 17 (2), 1 – 23.

Getis, A. and Ord, J.K., 1992. The Analysis of Spatial Association. *Geographical Analysis*, 24 (3).

Goodchild, M., 1992. Geographical information science. *International Journal of Geographical Information …*, 6 (March 2013), 37–41.

Goodchild, M.F., 2003. Geospatial data in emergencies. *In*: S.L. Cutter, D.B. Richardson, and T.J. Wilbanks, eds. *The Geographical Dimensions of Terrorism*. New York: Routeledge, 99–104.

Goodchild, M.F., 2004. GIScience, Geography, Form, and Process. *Annals of the Association of American Geographers*, 94 (4), 709–714.

Green, P.J. and Richardson, S., 2002. Hidden Markov Models and Disease Mapping. *Journal of the American Statistical Association*, 97 (460), 1055–1070.

Guo, Q., Li, W., Liu, Y., and Tong, D., 2011a. Predicting potential distributions of geographic events using one-class data: concepts and methods. *International Journal of Geographical Information Science*, 25 (10), 1697–1715.

Guo, Q., Li, W., Liu, Y., and Tong, D., 2011b. Predicting potential distributions of geographic events using one-class data: concepts and methods. *International Journal of Geographical Information Science*, 25 (10), 1697–1715.

Guo, Q., Liu, Y., and Wieczorek, J., 2008. Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *International Journal of Geographical Information Science*, 22 (10), 1067–1090.

Heggie, T.W. and Amundson, M.E., 2009. Dead men walking: search and rescue in US National Parks. *Wilderness & environmental medicine*.

Heggie, T.W. and Heggie, T.M., 2009. Search and rescue trends associated with recreational travel in US national parks. *Journal of Travel Medicine*, 16 (1), 23–7.

Hill, K. and Gale, R., 1997. *Managing the lost person incident*. Managing. Chantilly, VA: National Association for Search and Rescue.

Hirzel, A.H., Hausser, J., Chessel, D., and Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, 83 (7), 2027–2036.

Hogg, J.M., 1968. The Siting of Fire Stations. *Journal of the Operational Research Society*, 19, 275–287.

Huang, B., Wu, B., and Barry, M., 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24 (3), 383–401.

Hung, E.K. and Townes, D. a, 2007a. Search and rescue in Yosemite National Park: a 10-year review. *Wilderness & environmental medicine*, 18 (2), 111–6.

Hung, E.K. and Townes, D. a, 2007b. Search and rescue in Yosemite National Park: a 10-year review. *Wilderness & Environmental Medicine*, 18 (2), 111–6.

Imhof, E., 1950. *Gelaende und Karte*. Zurich, Switzerland: Rentsch.

Indriasari, V., Mahmud, A.R., Ahmad, N., and Shariff, A.R.M., 2010. Maximal service area problem for optimal siting of emergency facilities. *International Journal of Geographical Information Science*, 24 (2), 213–230.

Jaynes, E.T., 1990. Notes on present status and future prospects. *In*: W.. Grandy and L.H. Schick, eds. *Maximum entropy and Bayesian methods*. Dodrecht: Kluwer.

Kim, H.-M. and Kwan, M.-P., 2003. Space-time accessibility measures: A geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration. *Journal of Geographical Systems*, 5 (1), 71–91.

Koch, T. and Denike, K., 2009. Crediting his critics' concerns: remaking John Snow's map of Broad Street cholera, 1854. *Social science & medicine (1982)*, 69 (8), 1246–51.

Koester, R., 2008. *Lost Person Behavior*. Search. Charlottesville, VA: dbS Productions.

Koester, R.J., Cooper, D.C., Frost, J.R., and Robe, R.Q., 2004. Sweep Width Estimation for Ground Search and Rescue. Washington.

Koester, R.J. and Stooksbury, D.E., 1995. Behavioral profile of possible Alzheimer's disease patients in Virginia search and rescue incidents. *Wilderness & Environmental Medicine*, 6, 34–43.

Koopman, B.O., 1980. *Search and Screening*. New Yorrk: Pergamon Press.

Kratzke, T.M., Stone, L.D., and Frost, J.R., 2010. Search and Rescue Optimal Planning System (SAROPS). *In*: *13th Conference on Information Fusion*. Edinburgh, Scotland: IEEE, 1 – 8.

Kwan, M.-P., 2003. Intelligent emergency response systems. *In*: S.L. Cutter, D.B. Richardson, and T.J. Wilbanks, eds. *The Geographical Dimensions of Terrorism*. New York: Routeledge, 111–116.

Lawson, A., Gangnon, R., and Wartenberg, D., 2006. Developments in disease cluster detection. *Statistics in Medicine*, 25 (5), 721.

Li, W. and Guo, Q., 2010. A maximum entropy approach to one-class classification of remote sensing imagery. *International Journal of Remote Sensing*, 31 (8), 2227–2235.

Li, W., Guo, Q., and Elkan, C., 2011. Can we model the probability of presence of species without absence data? *Ecography*, 34 (6), 1096–1105.

Li, X. and Yeh, A., 2005. Integration of genetic algorithms and GIS for optimal location search. *International Journal of Geographical Information Science*, 19 (5), 581–601.

Lin, L. and Goodrich, M. a., 2010. A Bayesian approach to modeling lost person behaviors based on terrain features in Wilderness Search and Rescue. *Computational and Mathematical Organization Theory*, 16 (3), 300–323.

Liu, D. and Cai, S., 2011. A Spatial-Temporal Modeling Approach to Reconstructing Land-Cover Change Trajectories from A Spatial-Temporal Modeling Approach to Reconstructing Land-Cover Change Trajectories from Multi-temporal Satellite Imagery. *Annals of the Association of American Geographers*, 102 (6), 1329–1347.

Liu, Y., Guo, Q., Wieczorek, J., and Goodchild, M.F., 2009. Positioning localities based on spatial assertions. *International Journal of Geographical Information Science*, 23 (11), 1471–1501.

Malczewski, J., 2004. GIS-based land-use suitability analysis: a critical overview. *Progress in Planning*, 62 (1), 3–65.

Miller, H., 1996. GIS and geometric representation in facility location problems. *International Journal of Geographical Information …*, 10 (7), 37–41.

Miller, H.J. and Bridwell, S. a., 2009. A Field-Based Theory for Time Geography. *Annals of the Association of American Geographers*, 99 (1), 49–75.

Moffett, A., Shackelford, N., and Sarkar, S., 2007. Malaria in Africa: Vector Species' Niche Models and Relative Risk Maps. *PLoS ONE*, 2 (9), e824.

Montalvo, R., Wingard, D.L., Bracker, M., Davidson, T.M., and Diego, S., 1998. Conferences and Reviews Morbidity and Mortality in the Wilderness. *Wilderness and Environmental Medicine*, 168 (4), 248–254.

Mostern, R. and Johnson, I., 2008. From named place to naming event: creating gazetteers for history. *International Journal of Geographical Information Science*, 22 (10), 1091–1108.

Murray, A. and Tong, D., 2007. Coverage optimization in continuous space facility siting. *International Journal of Geographical …*, (March 2013), 37–41.

National Association for Search and Rescue, 2005. *Fundamentals of Search and Rescue*. Sudbury, MA: Jones and Bartlett Publishers.

Nikolakaki, P., 2004. A GIS site-selection process for habitat creation: estimating connectivity of habitat patches. *Landscape and Urban Planning*, 68 (1), 77–94.

Ostfeld, R.S., Glass, G.E., and Keesing, F., 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution*, 20 (6), 328–36.

Pearson, R., Dawson, T., and Liu, C., 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land- cover data. *Ecography*, 3, 285–298.

Phillips, S.J., Dudik, M., and Schapire, R.E., 2004. A maximum entropy approach to species distribution modeling. *In*: *Proceedings of the Twenty-First International Conference on Machine Learning*. Banff, Alberta, CA, 655–662.

Robertson, C., Nelson, T.A., MacNab, Y.C., and Lawson, A.B., 2010. Review of methods for space–time disease surveillance. *Spatial and Spatio-temporal Epidemiology*, 1 (2–3), 105–116.

Schneider, P.J. and Schauer, B.A., 2006. HAZUS—Its Development and Its Future. *Natural Hazards Review*, 7, 40–44.

Shannon, C., 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423.

Sherrill, K.R., Frakes, B., and Schupbach, S., 2010. Travel time cost surface model: standard operating procedure. Natural Resource Report NPS/NRPC/IMD/NRR—2010/238. Fort Collins, Colorado.

Stone, L.D., 1989. What's happened in search theory since the 1975 Lanchester Prize? *Operations Research*, 37 (3), 501.

Stone, L.D., 2007. *Theory of Optimal Search*. 2nd ed. Mathematics in Science and Engineering. New York: Academic Press.

Suárez-Seoane, S., García de la Morena, E.L., Morales Prieto, M.B., Osborne, P.E., and De Juana, E., 2008. Maximum entropy niche-based modelling of seasonal changes in little bustard (Tetrax tetrax) distribution. *Ecological Modelling*, 219, 17–29.

Syrotuck, W.G., 1976. *Analysis of Lost Person Behavior*. Mechanicsburg, PA: Barkleigh Productions, Inc.

Tate, E., Cutter, S.L., and Berry, M., 2010. Integrated multihazard mapping. *Environment and Planning B: Planning and Design*, 37 (4), 646–663.

The Outdoor Foundation, 2012. Outdoor Recreation Participation Report.

Theodore, J., 2009. When every second counts. *ArcNews*, (March), 66 – 69.

Tobler, W., 1965. Non-Isotropic Geographic Modeling. *Information Systems*.

Tobler, W., 1991. Non-Isotropic Geographic Modeling. *In*: W. Tobler, ed. *Geographic Information Systems in the Social Sciences*. Santa Barbara.

Toregas, C., Swain, R., ReVelle, C., and Bergman, L., 1971. The Location of Emergency Service Facilities. *Operations Research*, 19 (6), 1363–1373.

Whitley, T. and Hicks, L., 2003. A geographic information systems approach to understanding potential prehistoric and historic travel corridors. *Southeastern Archaeology*, (those 1994).

Wieczorek, J., Guo, Q., and Hijmans, R., 2004a. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18 (8), 745–767.

Wieczorek, J., Guo, Q., and Hijmans, R., 2004b. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18 (8), 745–767.

Wild, F.J., 2008. Epidemiology of mountain search and rescue operations in Banff, Yoho, and Kootenay National Parks, 2003-06. *Wilderness & Environmental Medicine*, 19 (4), 245–51.

Winter, S. and Yin, Z.-C., 2010. Directed movements in probabilistic time geography. *International Journal of Geographical Information Science*, 24 (9), 1349–1365.

Worsing, R.J., 1993. *Rural Rescue and Emergency Care*. 1st ed. Rosemont, IL: American Academy of Orthopaedic Surgeons.

Yee, K. and Iserson, K. V., 2008. The Epidemiology of Search and Rescue Incidents in the Grand Canyon National Park: Are Preventive Measures Making a Difference? *Western Journal of Emergency Medicine*, 9 (1), 3–5.

Figure 1.1

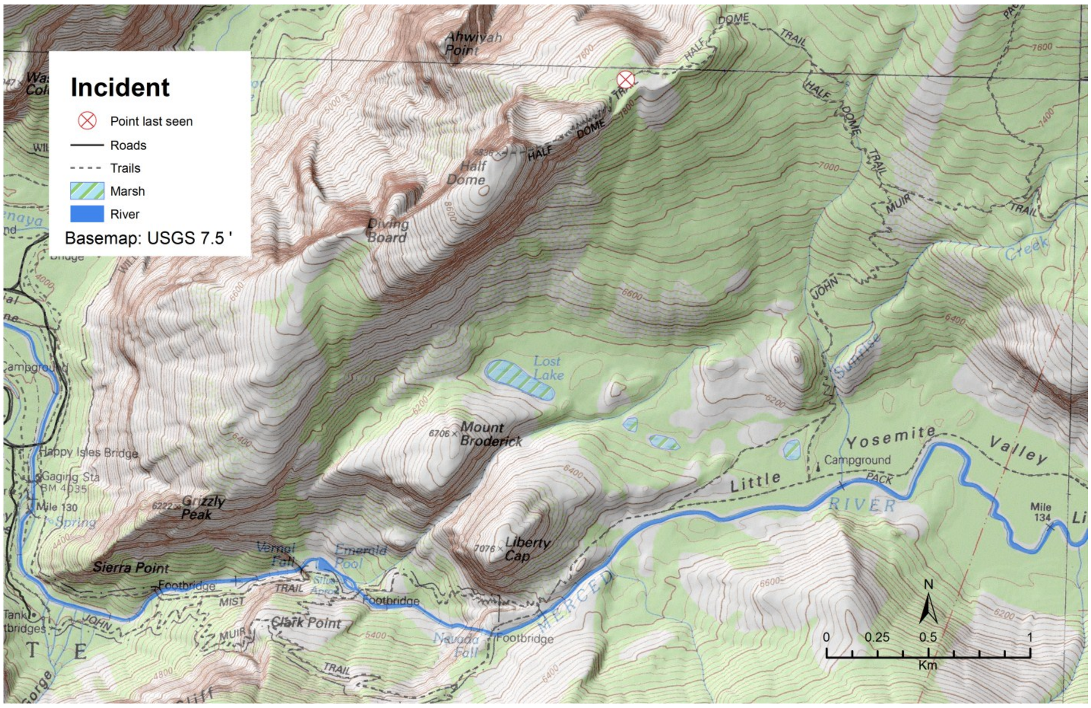

Incident
Point last seen
Roads
Trails
Marsh
River
Basemap: USGS 7.5 '
Half Dome
Diving Board
Ahwiyah Point
Lost Lake
Mount Broderick
Liberty Cap
Grizzly Peak
Sierra Point
Clark Point
Vernal Fall
Emerald Pool
Nevada Fall
Happy Isles Bridge
Little Yosemite Valley
Campground
MERCED RIVER
N
0 0.25 0.5 1
Km

Figure 1.2

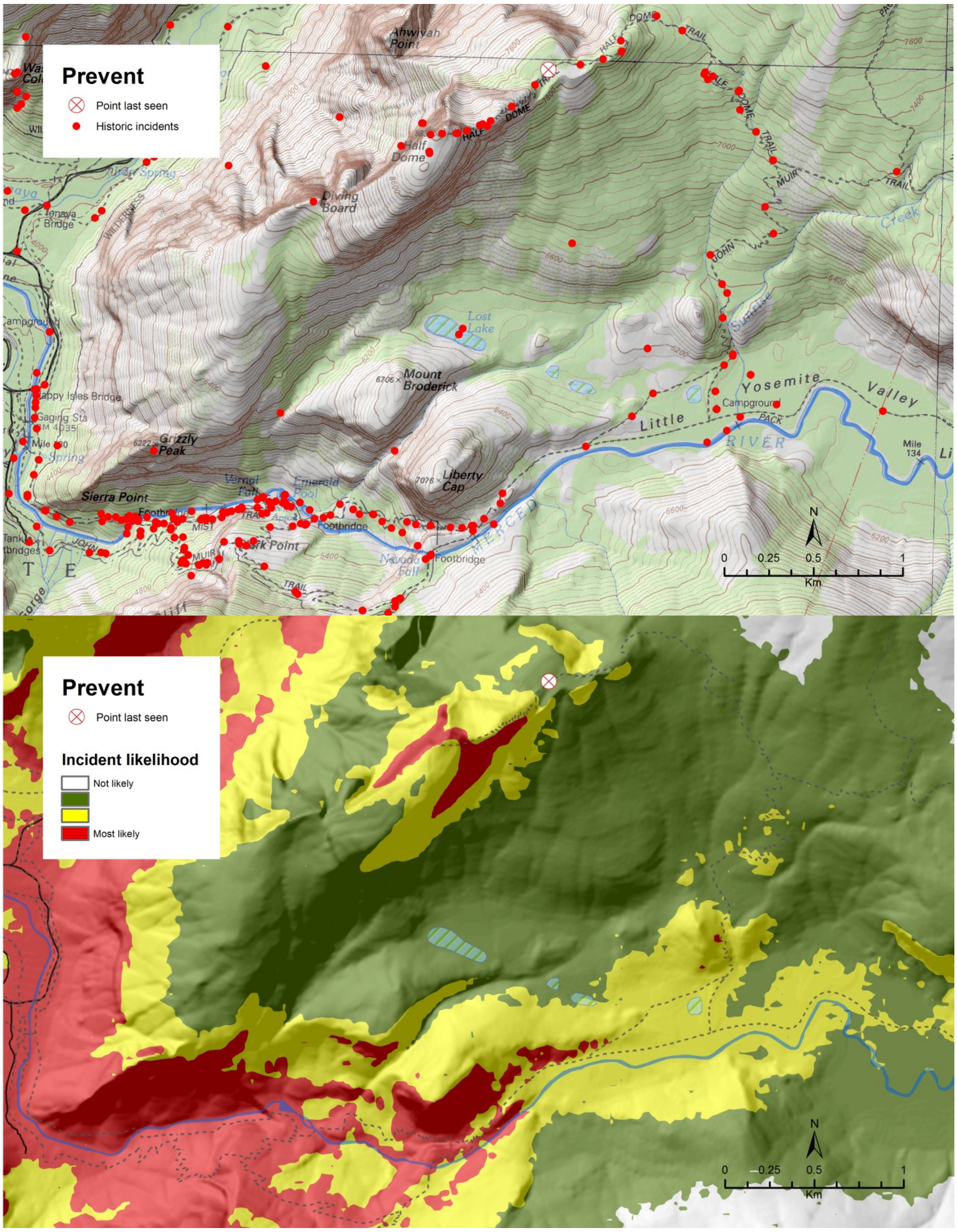

Prevent
Point last seen
Historic incidents
Half Dome
Diving Board
Lost Lake
Mount Broderick
Grizzly Peak
Sierra Point
Liberty Cap
Little Yosemite Valley
Happy Isles Bridge
Footbridge
N
0 0.25 0.5 1
Km
Prevent
Point last seen
Incident likelihood
Not likely
Most likely
N
0 0.25 0.5 1
Km

Figure 1.3

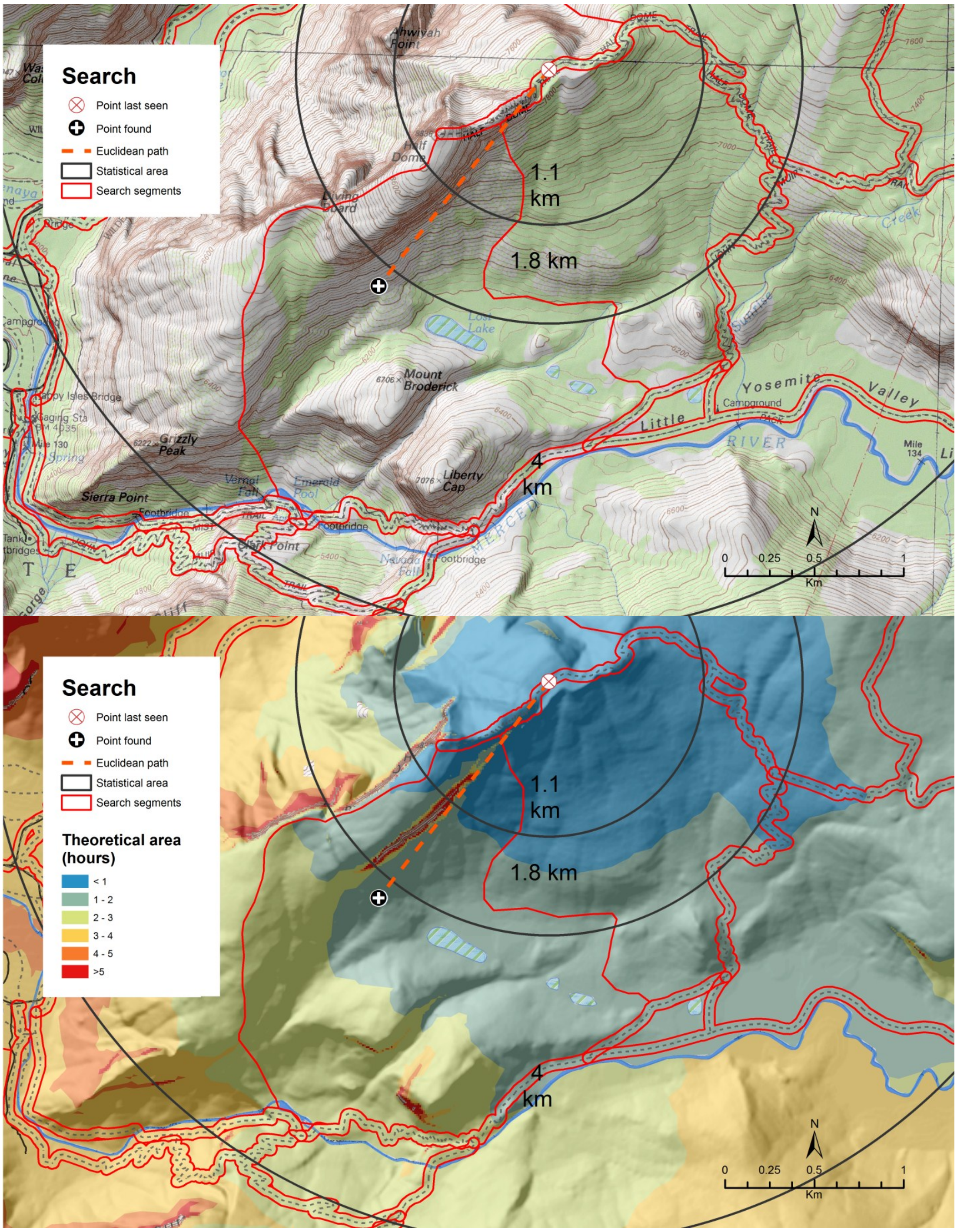
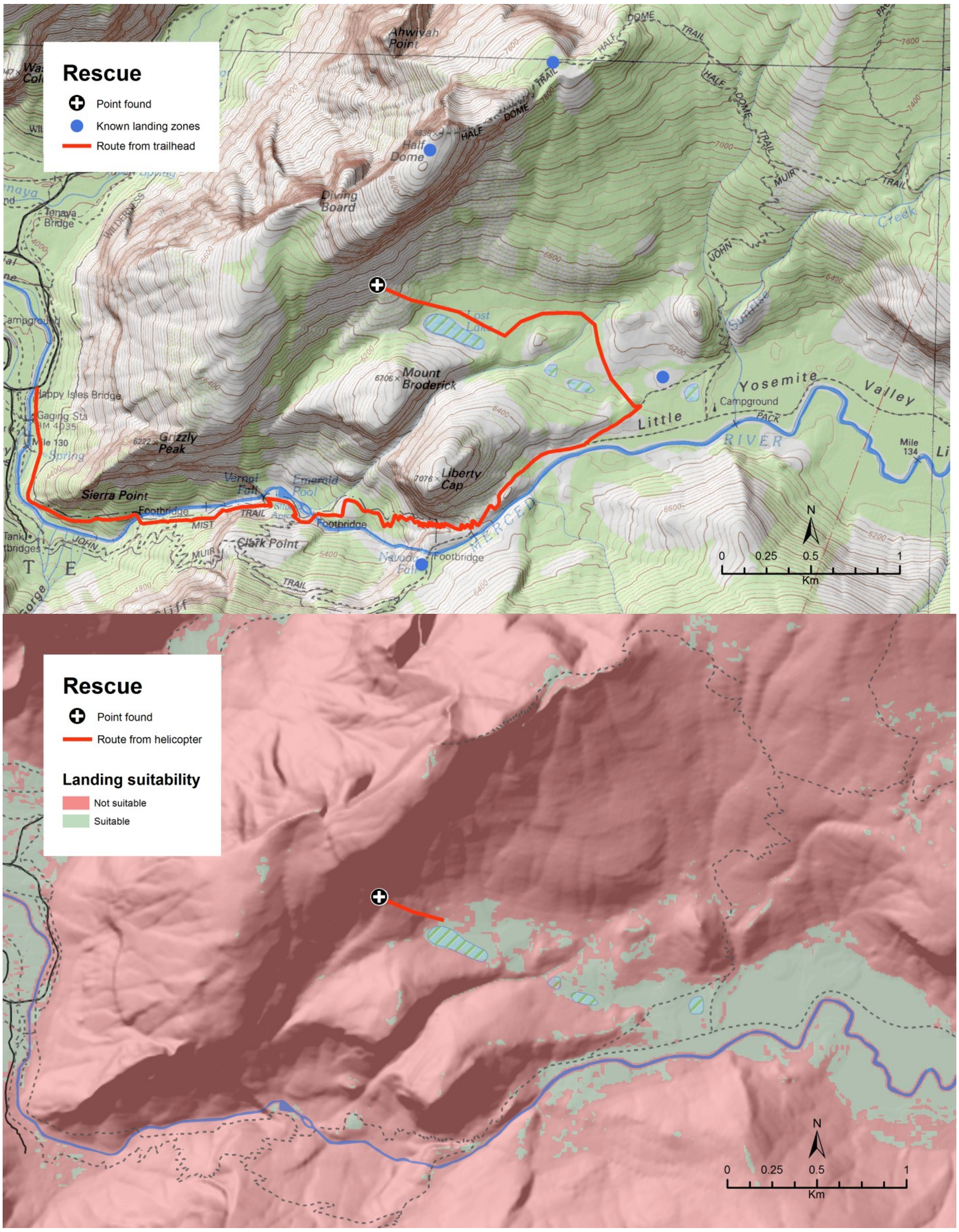

Search
Point last seen
Point found
Euclidean path
Statistical area
Search segments
1.1 km
1.8 km
4 km
Theoretical area (hours)
< 1
1 - 2
2 - 3
3 - 4
4 - 5
>5
0 0.25 0.5 1
Km
N

Figure 1.4


Rescue
Point found
Known landing zones
Route from trailhead
Half Dome
Diving Board
Mount Broderick
Liberty Cap
Grizzly Peak
Sierra Point
Happy Isles Bridge
Clark Point
Little Yosemite Valley
Campground
Footbridge
0 0.25 0.5 1
Km
N
Rescue
Point found
Route from helicopter
Landing suitability
Not suitable
Suitable
0 0.25 0.5 1
Km
N

Chapter II: Prevention

Georeferencing Incidents from Locality Descriptions and its Applications: a Case Study from Yosemite National Park Search and Rescue

[Formatted and previously published in *Transactions in GIS*]

Abstract
The Search and Rescue (SAR) of individuals who become lost, injured, or stranded in wilderness presents a unique and worthwhile spatiotemporal challenge to investigate. Once incidents are georeferenced they can be spatially queried and analyzed. However, one major challenge for evaluating SAR in a spatial context is the lack of explicitly spatial data (addresses or coordinates) for historic incidents; they must be georeferenced from textual descriptions. This study implemented two established approaches for georeferencing incidents, the 'Point-Radius' and 'Shape' methods. Incorporating uncertainty measurements into a spatial database allows for more appropriate analyses of spatial dependence and the spatial distribution of incidents. From 2005–2010, 1,271 of 1,356 Yosemite Search and Rescue YOSAR incidents (93.7%) could be georeferenced using the Point-Radius Method, with a mean uncertainty radius = 560 _ 51 m and mean uncertainty area of 3.60 _ 0.840 km2. However, when the Shape Method was applied to six case studies by considering the reference object shape, the uncertainty areas were reduced considerably (by up to 99.5% of the uncertain area generated by the Point-Radius Method). This is the first spatially-explicit study of SAR incidents and yields valuable insights into the role of georeferenced data in emergency preparedness.

1 Introduction

The lack of spatially explicit descriptions in text for natural and cultural information is a widespread challenge for the management and analysis of retrospective data. The process of spatially referencing localities or historic incidents from text is typically known as georeferencing and is a fundamental component of spatial analysis (Goldberg et al. 2007). Geographic Information System (GIS) technology has greatly enhanced our ability to process, store, and analyze georeferenced data as spatial data points. For instance, when textual descriptions already match those in a spatial database with coordinates, such as address lists, this process is known as geocoding. However, when these textual descriptions cannot be found in an existing geocoded database or do not have a physical address, localities must be georeferenced using a gazetteer (Mostern and Johnson 2008) of named places. This applies directly to the Yosemite Search and Rescue (SAR) data used for this research. SAR incident reports have been written or typed in case incident report format from 1967 to 2010 with narratives that describe where incidents occurred. Aspects of retrospective georeferencing from historical descriptions have been evaluated closely by ecologists (Hill et al. 2009) for spatially referencing museum specimens, but the techniques have not been applied directly to a wide variety of other potentially suitable case studies.

For example, an incident may have been described as occurring "3 miles North of Tuolumne Meadows". This description attempts to identify where the incident has occurred with only words. The only geographic information given involves a feature found in a topographic base map (Tuolumne Meadows), a heading (North), and a distance (3 miles). It is important to consider that retrospective georeferencing is done with the best information available and that much of the historic data were recorded without best practices in mind. While this is unfortunate from an analytic perspective, many of the accounts still contain valuable geographic information. Without a protocol for georeferencing these reports, these data will be underutilized and actionable spatial information will be unavailable for future studies on epidemiology or planning for emergency preparedness.

In historic text, most locality descriptions are based on at least one specific named place, which acts as the Reference Object (RO) for positioning a locality. The RO may be a point, linear (path), or areal feature, such as a junction, highway, or city. For simplicity, these objects are sometimes represented using a bounding box, by which the actual shape of the RO is circumscribed. The final shape containing the described locality is called the Target Object (TO). The objective of the georeferencing process is to estimate the TO based on the positions, shapes, and uncertainties of its ROs and spatial relationships (Guo et al. 2008, Liu et al. 2009).

While an estimated location based on this information may not be as specific as coordinates given from a modern GPS, it still provides valuable information. However, in traditional methods, the uncertainty involved with each frame of reference creates barriers for understanding spatial patterns of occurrence and causative factors (Chapman and Wieczorek 2006). In other words, the incident referenced will have occurred in one distinct location, but if it is represented by a single point or large indiscriminate bounding box, then isolated causative factors will not be identifiable. Spatial analysis without consideration of spatial uncertainty cannot be trusted in the same way as those computed with quantitative uncertainty values (Fisher 1999). New georeferencing techniques account for spatial uncertainty using hypothesis-driven approaches to overcome the shortcomings associated with traditional techniques. Two methods will be examined in this research; the "Point-Radius" and the "Shape" methods. The concepts of both methods have previously been described, but have not been compared explicitly, in the peer-reviewed literature.

The main objectives of this study are to: (1) review the conceptual basis for the Point-radius and Shape methods; (2) highlight several case-studies to compare the Point-radius and Shape methods; (3) present the integration of georeferencing with spatial pattern analyses; and (4) describe the implications of our findings on GIScience. Finally, the SAR georeferencing problem has real-world implications worth solving and we intend to lay the foundation for novel epidemiology research.

2 The Point-Radius and Shape Methods

The Point-Radius Method goes beyond describing a location using a single coordinate pair ('Point Method') by indicating an area of uncertainty using a circle with a well defined radius. The radius of each georeferenced incident represents a distinct hypothesis that combines multiple uncertainties into one metric attribute, the uncertainty radius. This method arose in the discipline of biodiversity informatics as a means to capture location information for historical accounts of the occurrences of species in nature in such a way that consumers of that information would have an easy way to assess each record's potential fitness for an intended use (Chapman 2005). The method is well documented (Wieczorek et al. 2004) and has six steps that can be performed relatively quickly using explicit documentation:

1. Classify the locality description;
2. Determine coordinates;
3. Calculate uncertainties;
4. Calculate combined uncertainties;
5. Calculate overall error; and
6. Document the georeferencing process

Once all of these steps have been completed for a particular georeference, the result is a set of unprojected coordinates (which can be projected by local datum) and a radius is typically given in meters for most locality descriptions. The Point-Radius Method was designed to be easy to construct and to store, providing a least common denominator to express georeferences that includes all of the associated uncertainties.

The Shape Method has the same overall goal as the Point-Radius Method – to capture all sources of uncertainty in a textual description of place in a spatial description. The main difference between the two is that in the Shape Method the geometries may be more complex than a circle and seek to exclude areas that are not within the original description. An easy way to illustrate the difference is to consider a textual description "along Strawberry Creek". Using the Point-Radius Method, the resulting georeference would be a circle with the center located on the creek in a position that minimized the distance to both mouth and head. With the Shape Method, the resulting georeference would be a buffered poly-line that included the full length of the creek with all of its twists and turns from head to mouth. Concrete examples of the practical differences between the two methods are illustrated in association with the case studies in later sections. Though the Shape Method is, in general, more time-consuming to produce than a Point-Radius, it excludes areas that do not fit the original description.

There are other methods for capturing georeferences and associated uncertainties that will not be considered in detail in this article. These include the Bounding Box method and the Probability Method (Guo et al. 2008). The Bounding Box Method is conceptually similar to the Point-Radius Method. Instead of a single central coordinate and a radius as a distance, the Bounding Box Method describes the area as two coordinate pairs. The disadvantage of the Method in comparison to the Point-Radius Method is that it requires a complex latitude- and datum-dependent calculation to determine the linear "size" of the georeference – it has no simple scalar with which to assess or filter georeferences.

The Probability Method is similar to the Shape Method in that it is used to constrain the resulting georeferences to only those areas consistent with the original description. The Probability Method goes further, to describe the area within the boundaries of non-zero probability with a raster of points having values representing the normalized probability that the point represents in the original description.

3 Case Studies

In order to compare the conceptual model and results from the two georeferencing methods, we selected representative case studies from the YOSAR case incident reports in Yosemite National Park (Figure 2.1). These case studies cover the type of descriptions used in the majority of incident accounts in the 2005–2010 dataset. As for reference objects, we used a variety of "map services" available online using commercial-off-the shelf GIS software and plug-in web-mapping applications. In addition to these authoritative and free sources of spatial data, we also used colloquial data such as digitized park base maps and the Yosemite Placenames Database (YPD) that are used internally by the National Park Service Emergency Communications Center. The YPD was created by park staff to combine the U.S. Geologic Survey (USGS) Geographic Names Information System and geocoded colloquial names for geographic localities within Yosemite National Park (USGS 1981). This database is dynamic and is updated when new localities are entered by park staff (at a scale no less than 1:24,000). The YPD is an example of volunteered geographic information (VGI) that is useful from a georeferencing and historic cultural data perspective.

Each georeference conducted using the Point-Radius Method was done so in accordance with Wieczorek et al. (2004) using the Georeferencing Calculator[2]. The georeferencing procedure utilized either the USGS Quadrangle maps or the YPD to place the TO. The Quadrangle maps have been converted to raster graphics that render at a 1:24,000 scale. In addition, all TOs were georeferenced using NAD 1983, and the latitude and longitude for each TO was measured in decimal degrees using GIS software. The maximum uncertainty was calculated as the sum of individual error components:

$$\textit{maximum uncertainty} = \sum u_i + \sum u_d \qquad (2.1)$$

where *u* is the maximum uncertainty for independent (i) or dependent (d) sources of error (Wieczorek et al 2004). In addition, the entire georeferencing process was recorded.

We successfully georeferenced six case studies, using both the Point-Radius and Shape methods. The output of each georeference can be visually compared as vector polygons in a GIS leading to a more robust conceptual understanding of the textual descriptions themselves. The vector polygon outputs from both methods can be stored in a spatial database along with other attribute information. More importantly, both forms of explicitly spatial data can be analyzed for autocorrelation, patterns in distribution, and spatial association with related factors. The Point-Radius Method generated much larger uncertainty areas (range 2.66 ¥ 10-3–2.17 ¥ 101 km2) than the Shape Method (range 5.70 ¥ 10-4–3.20 ¥ 100 km2; Table 2.1). Although the area of uncertainty was greater for the Point-Radius Method, the processing time per incident for the Point-Radius Method (five to 15 minutes) required less time than the Shape Method (15 to 90 minutes).

3.1 Case Study 1 – Junction

Report Description – "Incident occurred at the confluence of the Merced River and Tenaya Creek"

In this case (Figure 2.2), the locality description comes from two separate hydrological units (streams) that have been defined in polyline vector format (Simley and Carswell 2010).

Point-Radius Method – From this account, we classified the locality description as a junction and georeferenced the incident following the guidelines set forth for a feature or named place (Chapman and Wieczorek 2006). The poly-lines representing the Merced River and the Tenaya Creek, the two specified ROs, were identified from the USGS National Hydrography Dataset, and we plotted the TO at the intersection of the two ROs. The

[2] http://manisnet.org/gci2.html

uncertainty associated with the precision of the map scale as well as the extent of the intersection was taken into account.

Shape Method – Since each line feature is associated with error, we generated two buffer zones of the two linear features and computed their intersection as the maximum distribution range of the TO. The buffer distance is determined according to the precision due to the map scale. In this incident, the map scale is 1:24,000, and thus we set the buffer distance to be 12 m.

3.2 Case Study 2 – Areal Feature

Report Description – "Incident occurred on the Subdome of Half Dome"

In this case (Figure 2.3), the locality description for the areal feature known as "Subdome" was written using the USGS Topographic Quadrangle maps as a reference map. This has been common practice for the report-writers in Yosemite since the 1960s. Today the USGS Quadrangle maps have been converted to digital raster graphics and are available in a seamless format as a map service at the 1:24,000 scales[3].

Point-Radius Method – We classified the locality description as an areal feature. The center of the RO, "Subdome of Half Dome", was identified using the YPD, and the TO was plotted at this point. The uncertainty associated with the extent of the Subdome was then estimated using the USGS Quadrangle digital raster graphics at a 1:24,000 scale.

Shape Method – For this case, we assumed that the possible location of the TO is uniformly distributed inside the areal feature. Hence, the geometry of the feature is directly used to represent the uncertainty.

3.3 Case Study 3 – Offset Along a Path

Report Description – "Incident occurred on the Four-Mile trail approximately 1.5 miles up from the Four-Mile Trailhead"

The locality description georeferenced here (Figure 2.4) is based on poly-line vector data for trails from the National Park Service GIS data store[4] and the USGS Geographic Names Information System (GNIS) Gazetteer as point vector reference objects for the locality description.

Point-Radius Method – We classified the locality description as offset along a path. The ROs were then identified as 1.5 miles (distance) along the Four-mile trail, starting at the Four-mile trail trailhead. To plot the TO, we measured 1.5 miles along the Four-mile trail from the trailhead using a measuring tool within a GIS. The TO was plotted on the trail at this distance and the latitude and longitude were recorded for this point. The default extent of 15 m was used for small features in the extent of the starting point (Four-mile trail trailhead), and the distance precision was one-half mile.

Shape Method – We took into account two types of uncertainty for this case. First, the error associated with the path can be modeled by a buffer zone. Second, the offset distance may be imprecise. For example, if the description is "9 miles from A along N trail", the error interval is [8.5, 9.5]. Note that the distance is measured along the linear feature instead of Euclidean distance. Based on this information, the actual shape formed for the TO's distribution is a buffer zone created by a subset of the path.

3.4 Case Study 4 – Between Two Features

Report Description – "Incident occurred between Young Lakes and Mount Conness"

To georeference this incident (Figure 2.5), we used a combination of digital raster graphics for the summit of Mount Conness (see above for USA Topo map service) and polygon data for Young Lakes from the USGS National Hydrography Dataset (NHD) (Simley and Carswell 2010).

[3] http://services.arcgisonline.com/ArcGIS/rest/services/USA_Topo_Maps/MapServer
[4] https://nrinfo.nps.gov/Map.mvc/GeospatialSearch

Point-Radius Method – The case incident report indicates that a hiker went missing on a hike from Young Lakes to Mt. Conness. From this information, the locality description can be classified as “between two features”. We identified the ROs in this incident as Young Lakes and the summit of Mount Conness. The distance was measured between the center of the two named features, and the TO was plotted at this location. The extent was then measured as one-half the distance between the centers of Young Lakes and the summit of Mount Conness.

Shape Method – Suppose the two features are S1 and S2, we created a convex hull *H* for them and used H-S_1-S_2 to represent the TO’s possible distribution. Note that we did not consider the heterogeneity of uncertainty inside the TO region.

3.5 Case Study 5 – Distance from a Feature

Report Description – “Incident occurred 200 yards away from the Wilderness Parking Area”

The Wilderness Parking Area is located in Yosemite Valley, the center of which has been entered into the YPD as geographic coordinates. The extent of the Wilderness Parking Area can be digitized using high-resolution imagery which for Yosemite National Park is 1 m U.S. Department of Agriculture (USDA) National Agriculture Imagery Program (NAIP) imagery.

Point-Radius Method – We classified this locality description as “distance from a feature”. The coordinates for the center of the RO, the wilderness parking lot, were identified from the YPD, and the TO was plotted at this point. The extent of the wilderness parking lot was determined by measuring NAIP imagery of the area within a GIS. The uncertainty due to the dependent variable of distance precision (“200 yards”) was also taken into account (Figure 2.6).

Shape Method – We adopted the method proposed by Guo et al. (2008). For each point inside the feature, we generated an annular zone considering the distance error. The final shape is the union of all annular zones corresponding to the points inside the feature. Suppose the feature is *S* and the distance is *d* with error δ, the resulting uncertainty area can be computed by:

$$T = \bigcup_{p \in S} \qquad < dist(t,p) < d + \delta\} \tag{2.2}$$

where *dist*(*t*,*p*) denotes the distance between two points *t* and *p*.

3.6 Case study 6 – Direction with a header

Report Description – “Incident occurred 1/4 miles North of Glen Aulin High Sierra Camp”

The Glen Aulin High Sierra Camp is located in Tuolumne Meadows, which is geocoded in the USGS GNIS gazetteer. The extent of the Glen Aulin High Sierra Camp was digitized using high-resolution imagery, which for Yosemite National Park is 1 m USDA NAIP imagery (Figure 2.7).

Point-Radius Method – The RO was identified by the writers as Glen Aulin High Sierra Camp and there was no information indicating that the incident occurred on a trail. Therefore we assumed the distance (1/4 Mile) is to be measured ‘by air’. This locality description is classified as “direction with a header”. The geographic coordinates for the center of Glen Aulin High Sierra Camp were determined using the YPD. The extent of the camp as well as the dependent variables of the distance precision and directional precision was taken into account. The Georeferencing Calculator was then used with the calculation type: ‘Coordinates and Error’. Finally we placed the TO at the newly calculated coordinates.

Shape Method – This case is handled using the same approach with that of case 5. Suppose the specified direction is R, the resulting uncertainty area is:

$$T = \bigcup_{p \in S} \qquad , p) = R\} \tag{2.3}$$

where the *dir*(*t*,*p*) is the cardinal direction relation between *t* and *p*.

4 Spatial Analysis of Georeferenced Data

The ability to detect clusters amongst spatial count data is fundamental to such fields as ecology, geography, and epidemiology (Ostfeld et al. 2005). With regards to epidemiology, the ultimate goal of spatial analysis ultimately aims to detect where incidents, such as disease or injury occur (Lawson et al. 2006). In order to intentionally prevent an event from occurring, the incident must be: preventable, one must know about the possibility of the incident ahead of time, and an action must occur at the right time and place to inflict some plausible change to prevent the situation. The georeferencing process is crucial to gathering retrospective incident data, archiving attributes, and all subsequent spatial analyses.

In Preventative Search and Rescue (PSAR), the primary tool employed for prevention is educating the people at risk. PSAR personnel must be familiar with the spatial and temporal details of current hazards (weather, trail conditions, overall fitness of visitors) as well as hazards that have occurred in the past, to provide effective instruction at the right place and time. Examples of spatial analyses of pedestrian injuries in urban centers, (LaScala et al. 2000; Hameed et al. 2010) have shown significant relationships between pedestrian injury and factors such as traffic flow, population density, and socio-economic state of the respective neighborhoods. This geographic approach to pedestrian injury in urban environments may also be used for wilderness areas and national parks if spatially explicit data existed. To date, no such research exists for SAR to utilize in their response calls.

In order to perform spatial analyses using historic incidents as the TO, uncertainty areas and values must be defined using geographic information, i.e., uncertainty radii or shapes. For instance, distance statistics such as the *G* and *I* values require specification of spatial resolution or provision of areal features (e.g., municipal boundaries) prior to performing analyses (Getis and Ord 1992; Anselin 1995). If we cannot define the area of uncertainty for each incident and specify an appropriate scale for the spatial analyses, the results cannot be properly interpreted. For example, if our resulting TOs have an average of 5 km uncertainty radius using the Point-Radius Method, a spatial analysis using 1 km grids would not be appropriate. Likewise, with regard to the Shape method, if our TO polygons are larger than the areal boundaries within which the incidents occur, then we cannot assess the spatial distribution at this scale, because we would not be able to calculate incident frequency per areal boundary.

Despite the tabular documentation of over 6,000 incidents in 30 years within the 3,100 km^2 area of Yosemite National Park, the incident data have not been georeferenced in any way. The time, place, and contributing factors of each incident have not been evaluated on a large-scale. This rich dataset could contain valuable information essential for the prevention of incidents, saving lives and a large amount of monetary resources that could be used in areas to help preserve the park for visitors. If historic incident locations are described using named places with known geographic coordinates, then they can be georeferenced in a manner that controls for uncertainty and provides a suitable dataset applying spatial analyses in future studies.

4.1 Spatial Analysis of Yosemite Search and Rescue Incidents

To demonstrate the usefulness of georeferencing historic localities, case incident reports from 2005 to 2010 have been reviewed and georeferenced using the Point-Radius Method as described above. The resulting vector point database with uncertainty radii values for each georeference were then converted to a vector polygon feature class using a buffer geoprocessing tool where each feature is given a circular buffer using the uncertainty radii as the buffer distance. The mean (± 95% Confidence Interval) and median radii for the georeferenced incidents were used to make decisions on scale and resolution of spatial analyses.

To prepare the data for distance statistics, a spatial join (point-in-polygon analysis) geoprocess was used to calculate the number of incidents found within a square grid vector polygon overlay on the study area ("study grid"). The study grid cell size was determined by the 95% percentile value for uncertainty radii of the 2005 – 2010 georeferenced data. This scale ensures that 95% of the uncertainty radii are smaller than the grid size chosen for the spatial analysis. Once incident counts were generated for both vector polygon feature classes from the spatial join, the spatial distribution of the incident points were assessed for randomness/clustering using Patrick A.P. Moran's I statistical geoprocess with inverse distance squared conceptualization and Euclidean distance method (Anselin 1995).

$$I = (n/S_0)\ \textstyle\sum_i \sum_j w_{ij} z_i z_j / \sum_i z_i^2 \tag{2.4}$$

or

$$I = \textstyle\sum_i I_i / \left[S_0\left(\sum_i z_i^2/n\right)\right] \tag{2.5}$$

Where $S_o = \sum_i \sum_j w_{ij}$

Finally, a Getis-Ord G* statistical geoprocess was performed with inverse distance squared conceptualization and Euclidean distance method on the study grid to determine if areas of significant high or low frequency were present and identify which areas exhibited these patterns ("hot and cold spots"; Getis and Ord 1992) .

$$G_i^* = \frac{W_i + (k_1 - k_2)c}{n} \tag{2.6}$$

These analyses are presented to highlight the need to address spatial uncertainty prior to spatial analysis and draw attention to a novel application of GIS to an inherently spatial problem.

4.2 Spatial Distribution of Yosemite Search and Rescue Incidents

From 2005 – 2010, 1271 of 1356 YOSAR incidents (93.7%) could be georeferenced using the Point-Radius Method, with mean uncertainty radius = 560 ± 51 m and mean uncertainty area of 3.60 ± 0.840 km^2. The 95th percentile value for the uncertainty radii of the dataset was 2026 m, therefore a 2000 m (2 km) grid was chosen for the spatial join, Moran's I, and Getis-Ord G* analysis. The spatial join resulted in a 2 km grid vector polygon (N = 1560; count range 0 – 226 incidents). The grid exhibited statistically significant spatial autocorrelation (Global Moran's index 0.310, Z = 24.5, P < 0.001), which means there is less than a 1% likelihood that this clustered pattern could be a result of random chance. Of the entire study grid, 10 had statistically significant G* Z-scores (P < 0.05). These grids were clustered primarily around Yosemite Valley vicinity hiking trails with one isolated grid near Glen Aulin High Sierra Camp hiking trails (Figure 2.7).

5 Discussion

The majority of SAR reports contained enough spatial information in the narrative text to be successfully georeferenced (94%). The Point-radius and Shape methods produced explicitly spatial data from textual locality descriptions for SAR incidents in Yosemite National Park. The Point-Radius Method produced TOs with a wide range of uncertainty radii and subsequent uncertainty areas (some of which were larger than 7.9 ± 104 km^2), but was flexible enough to handle a large number and wide variety of cases. Within the six case studies, the Shape Method provided a more specific representation of uncertainty area dimensions for each case along with much smaller uncertainty areas (99% smaller in one case).

Conceptually, the Shape Method is a simple and accurate way to georeference textual locality descriptions. Methodologically, however, the Shape Method presented in this article is very time consuming and perhaps not as suitable for archiving a large number of textual records. In this and other studies, the Point-Radius Method can be used to georeference locality descriptions at mean rates of 10 to 30 cases per hour (Wieczorek 2005, 2008). The Shape Method produced georeferenced TOs at a rate closer to 2 or 4 per hour. This may prove to be problematic over a large dataset, such as the 2005–2010 YOSAR incident data. Improvements to the Shape Method are needed if large datasets (>10,000 records) are to be considered. Furthermore, at this time, the Shape Method may be preferred in studies where more exact uncertainty areas of TOs
are needed. In some cases, it may be feasible to use the results of the faster Point-Radius method in conjunction with spatial operations in a GIS to get results similar to those in the Shape method. For example, one could clip a stream layer by the vector result of a Point-Radius and apply a buffer to it. Another example might involve known z coordinates (elevation or depth) or terrain type (slope) to refine the probable location of the incident. Though this technique will not work for all locality types, it may be part of a hybrid solution to increase overall georeferencing rates while retaining the highest possible level of specificity in the results.

The results of the Point-Radius Method can be stored as vector point (coordinates) and vector polygon (circular with radii equal to the uncertainty radius attribute). While vector polygon data may be a more thorough way to visualize the uncertainty area, vector point data are more easily stored in traditional non-spatial databases where x and y coordinate fields can be populated. The Shape Method results are vector polygon with a shape uncertainty. For archiving a historic incident database, any of the three outputs can be used (Point-Radius polygon, Point-Radius point, actual shape polygon) in a spatial database. The choice of how to store the results will depend on the types of spatial analysis questions that need to be answered. These data can then be used for archiving history in spatially enabled digital libraries (Mostern and Johnson 2008) which can be utilized for records management, season pre-planning and implementing effective mitigation measures.

In addition to archiving historic events, the spatial distribution of the georeferenced SAR data has interesting and potentially meaningful patterns. The Point-radius Method was successfully used on the 2005–2010 YOSAR incident SAR reports with a mean uncertainty radius of 556 m and the overall spatial uncertainty of these incidents were not normally distributed (median = 151 m). Overall, 95% of the cases suitable for georeferencing had an uncertainty radius smaller than 2,000 m or 2 km. This is an important finding for both GIScientists and epidemiologists interested in using historic textual data for georeferencing localities and events.

Therefore, we were able to conduct spatial analyses at this scale with confidence that our findings are based on valid assumptions and are pertinent to the real world. If we chose a smaller scale, such as 100 m, we could not be confident that the spatial join placed incidents in the correct grid, since our uncertainty radii were frequently greater than 100 m. If we want to conduct analyses at finer scales, such as examining the spatial distribution of incidents near a particular switchback or curve in the trail, more robust data is needed. However, researchers are attempting to create a more global relationship between comparisons, allowing for a “hierarchy based criterion” (Goldberg and Cockburn 2010). This alteration allows for local variances to be ignored, allowing for a more precise return in geocodes.

For analyzing retrospective data at a fine scale, the Shape or a more probabilistic method is needed. In the future, the implementation of GPS receivers will help to solve this problem. Our research findings support that this is a much needed and helpful addition to the current practice of incidence response. As a result of this study, GPS receivers will be required on all Yosemite SAR operations for recording coordinates on-scene and spatial information will be integrated into a newly implemented records management system. Meanwhile, the results of our analyses at the 2 km scale indicated two findings regarding

SAR incidents from 2005 to 2010 in Yosemite National Park: the spatial distribution of incidents was not random and there were multiple regions with statistically significant clusters of high frequency. A visual analysis of the data (Figure 2.8) shows that most of the statistically significant high frequency areas are in the vicinity of Yosemite Valley, especially near trails. Injury epidemiologists may want to know more about the clustering of incidents in these areas. For instance, “Are areas with steep trails or rushing water associated with higher incident frequency? Are more people being injured nearer or further from trailheads?”. Before answering these and other questions, the spatial distribution of natural (e.g. terrain, hydrology, vegetation) and anthropogenic (e.g. trail and road access, signage, number of people present) need to be known. For instance, before the risk to visitors in the Yosemite National Park can be assessed, actual visitation levels (by geographic area) in the park need to be recorded or else results may be misleading.

One implication of being able to utilize text-based information for successfully georeferencing historic incidents is that the same can be done for temporal referencing. In our case-study, the date, and in some cases, time of day, are explicitly stated for each component of the incident. Key intervals are delineated at the time an incident was reported, the time emergency response began, when rescuers arrived at the scene, and time of incident completion are available for some reports. These temporal data are similar to those used by municipal computer-aided dispatch centers to evaluate expediency and effectiveness of emergency response. Temporal data can clearly be utilized in a similar manner for search and rescue. Where these data are not available, a method that incorporates temporal uncertainty should be used (Schockaert and De Cock 2008, Yuan and Liu, 2010). In addition, the spatiotemporal context of these incidents could be used for operational pre-planning and even preventive search and rescue (PSAR). The where and when queries such as, “Where should we stage our resources on a Saturday afternoon in August?”, will depend on both the spatial and temporal resolution of our dataset. The network analyses that are available in GIS software ultimately utilize spatiotemporal components of both base and operational data. Moreover, if a modern records management system that integrates real-time spatial data and a GISystem is implemented, both historic incidents and potential trends should be evaluated to ensure a successful decision support system is designed. This cannot be done unless legacy data are preserved and exposed in a meaningful way (Zhou and Hripcsak 2007), where space and time are available for analysis.

6 Conclusions

The georeferencing issue is important to address, not just for historical data and digital libraries alone. While the advent of GPS, wireless communications, geocoding services, and GISystems may increase spatial precision for future incidents, textual spatial data is still abundant and relevant in our culture. Social media on the Internet and across wireless platforms continue to carry spatial information that cannot always be reliably geocoded without a best practice implementation (Zook et al. 2010). Today, the Internet itself may be a rich source of geographic information (Goldberg et al. 2009). In addition to the need to georeference text-based data during disasters, health care in developing nations may operate in environments where geocoded addresses do not currently exist. It is clear that there are a wide variety of issues where georeferencing text-based data is critical. The conceptual and methodological issues addressed in this research are prerequisites to understanding geographic information relayed in the human languages of the past, as well as developing new technologies for the future. The best practice may be derived from an ensemble approach where artificial intelligence, advanced computer programming, and agent-based modeling can be leveraged in a Web GIS platform[5].

This research also addresses the fundamental and previously overlooked notion that

[5] http://www.biogeomancer.org

Search and Rescue provides an inherently spatial problem-set for GIScientists to research. In turn, GISystems offer a robust solution to documenting and analyzing the geographic features that are factors controlling the presence or absence of events, such as disease outbreaks or medical geography (Gundersen 2000). Just as Dr. Snow's map of cholera in England raised the notion of how geography might be used to understand disease transmission, this concept should be extended to SAR at a local level to provide a GIScience approach to risk assessment (Koch and Denike 2009). From this application of GISytems to SAR, interesting and meaningful GIScience theories can be addressed.

References

Anselin L 1995 Local indicators of spatial association – LISA. *Geographical Analysis* 27: 93-115

Chapman A D 2005 *Principles of Data Quality, version 1.0*, Report for the Global Biodiversity Information Facility, Copenhagen.

Chapman A D and Wieczorek J (eds) 2006 *Guide to Best Practices for Georeferencing*, Copenhagen, Global Biodiversity Information Facility

Fisher P F 1999 Models of uncertainty in spatial data. *Geographical Information Systems (Volume 1):Principles and Technical Issues (Second Edition).* edited by P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind. New York, John Wiley and Sons: 191-205

Getis A and Ord J K 1992 The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24(3): 189-206

Goldberg D W, Wilson J P, Knoblock C A 2007 From text to geographic coordinates: the current state of geocoding. *URISA Journal* 19(1): 33-46

Goldberg D W, Wilson J P, Knoblock C A 2009 Extracting geographic features from the Internet to automatically build detailed regional gazetteers. *International Journal of Geographic Information Science* 23(1): 93 – 128

Goldberg D W, Cockburn M G 2010 Improving geocoded accuracy with candidate selection criteria. *Transactions in GIS* 14: 149-176

Gundersen L 2000 Mapping it out: Using atlases to detect patterns of health care, disease, and mortality. *Annals of Internal Medicine* 133 (2): 161-164

Guo Q, Liu Y, Wieczorek J 2008 Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *International Journal of Geographic Information Science* 22(10): 1067-1090

Hameed S M, Lord S E, Schuurman N, Bell N J, Simons R K, and Chir B 2010 Vulnerability to pedestrian trauma: Demographic, temporal, societal, geographic, and environmental factors. *BC Medical Journal* 52: 136-143

Hill W H, Guralnick R, Flemons P, Beaman R, Wieczorek J, Ranipeta A, Chavan V, and Remsen D 2009 Location, location, location: utilizing pipelines and services to more effectively georeference the world's biodiversity data. *BMC Bioinformatics* 10: 1-9

Koch T and Denike K 2009 Crediting his critics' concerns: Remaking John Snow's map of Broad Street cholera, 1854. *Social Science & Medicine* 69: 1246-1251

LaScala E, Gerber D, and Gruenwald J P 2000 Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accident Analysis and Prevention* 32 (5): 651–658

Lawson A, Gangnon R, and Wartenberg D 2006 Developments in disease cluster detection. *Statistics in Medicine* 25: 721

Liu Y, Guo Q, Wieczorek J, Goodchild M F 2009 Positioning localities based on spatial assertions. *International Journal of Geographic Information Science* 23(11): 1471-1501

Mostern R and Johnson I 2008 From names place to naming event: creating gazetteers for history. *International Journal of Geographical Information Science* 22(10): 1091-1108

Ostfeld R S, Glass G E, and Keesing F 2005 Spatial epidemiology: An emerging (or re-emerging) discipline. *Trends in Ecology and Evolution* 20: 328-36

Simley J D and Carswell Jr W J 2010 The National Map—Hydrography: U.S. Geological Survey Fact Sheet 2009-3054

United States Geologic Survey (USGS) 1981 Geographic Names Information System. WWW document, http://geonames.usgs.gov/pls/gnispublic/

Wieczorek J, Guo Q, Hijmans R J 2004 The Point-Radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science* 18(8): 745-767

Wieczorek J 2005 MaNIS Georeference Repatriation. WWW document, http://manisnet.org/Repatriation.html

Wieczorek J 2008 ORNIS Georeferencing Results. WWW document, http://www.ornisnet.org/georeferencing/results

Zook M, Graham M, Shelton T, and Gorman S. 2010. Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. *World Medical and Health Policy* 2(2): 7 – 33.

Table 2.1 – A comparison of six georeferenced case studies using the Point-radius and Shape methods.

Case study	Georeference type	Target area (km^2)		Percentage difference
		Point-Radius	Shape	
1	Path junction	5.36×10^{-3}	5.70×10^{-4}	89.4
2	Areal feature	2.91×10^{-2}	6.61×10^{-3}	77.3
3	Offset along a path	2.25×10^{0}	1.78×10^{-2}	99.2
4	Between two features	2.17×10^{1}	3.20×10^{0}	85.2
5	Distance from a feature	1.95×10^{-1}	2.75×10^{-2}	85.9
6	Direction with a header	1.25×10^{0}	1.97×10^{-1}	84.3

Figures

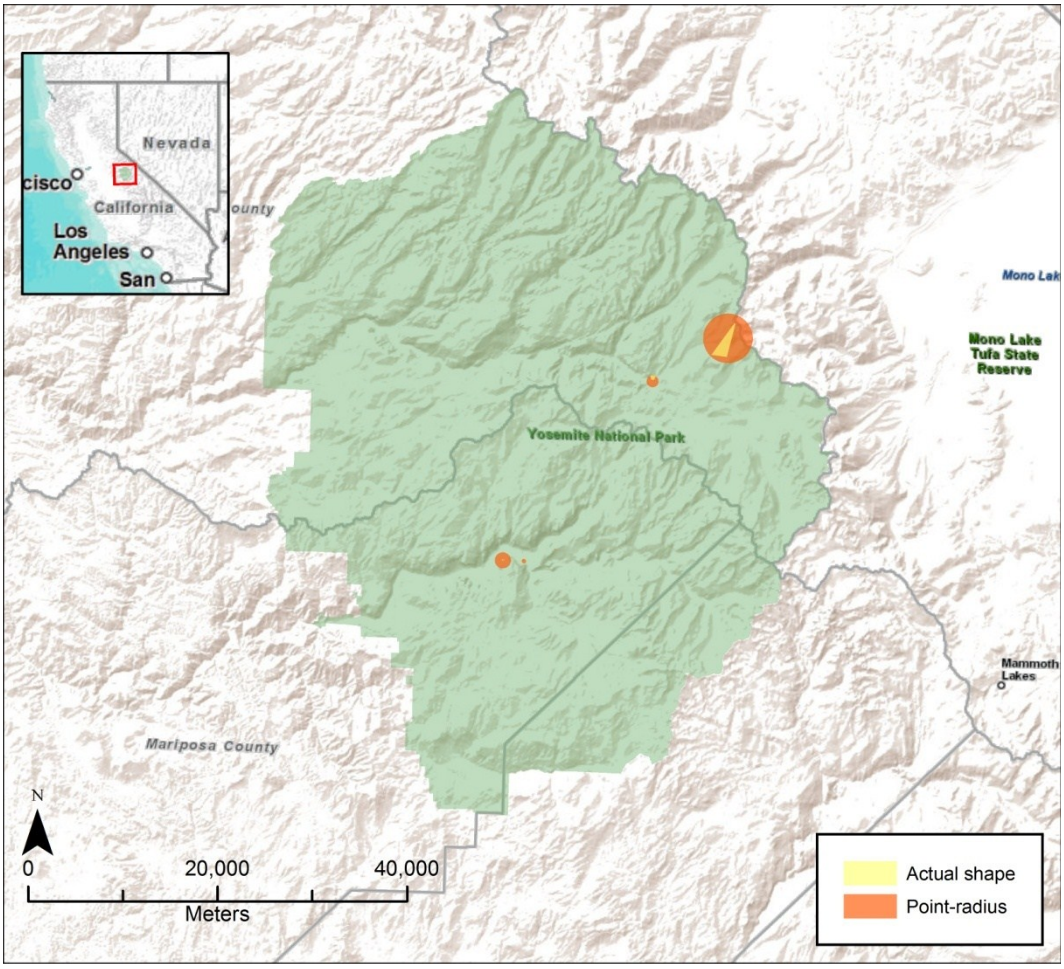


Figure 2.1 An overview map of the Yosemite Search and Rescue study area and case study locations.

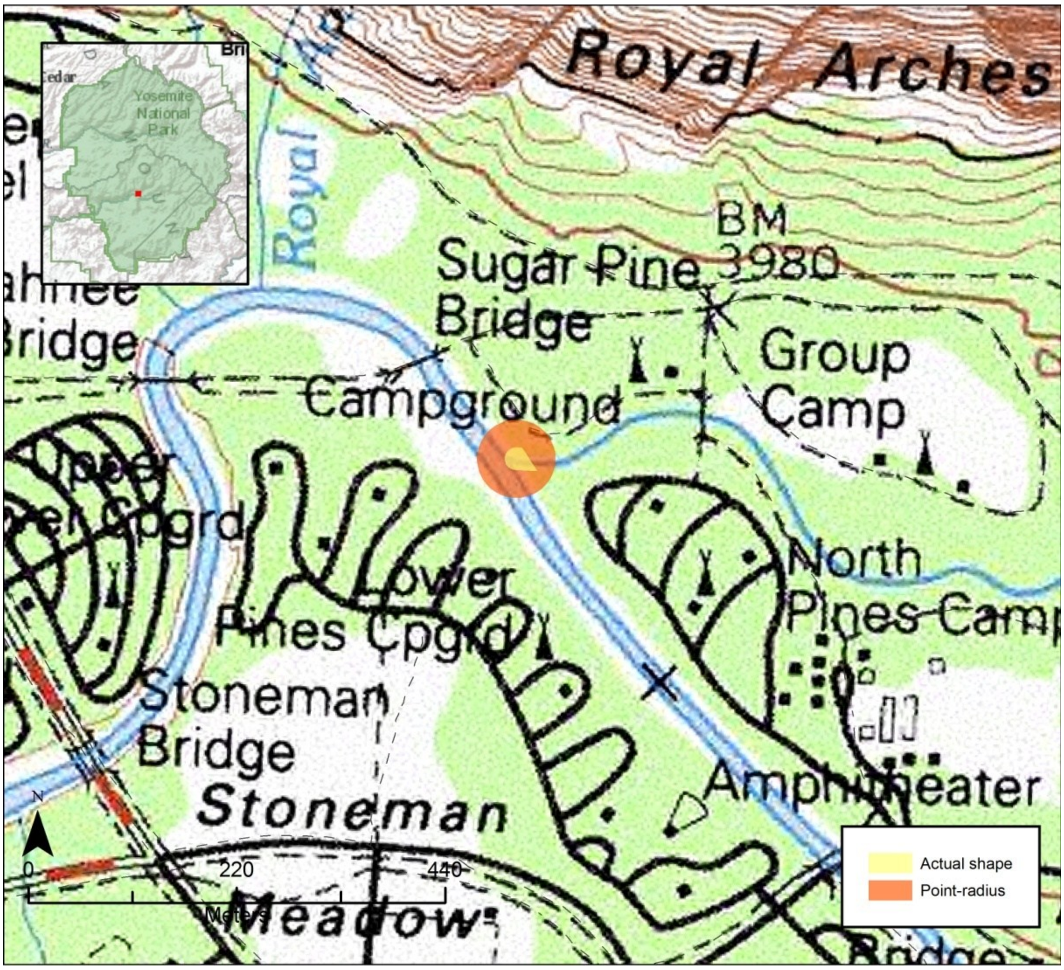


Figure 2.2 Case study #1 –Target object is the junction of two path reference object features.

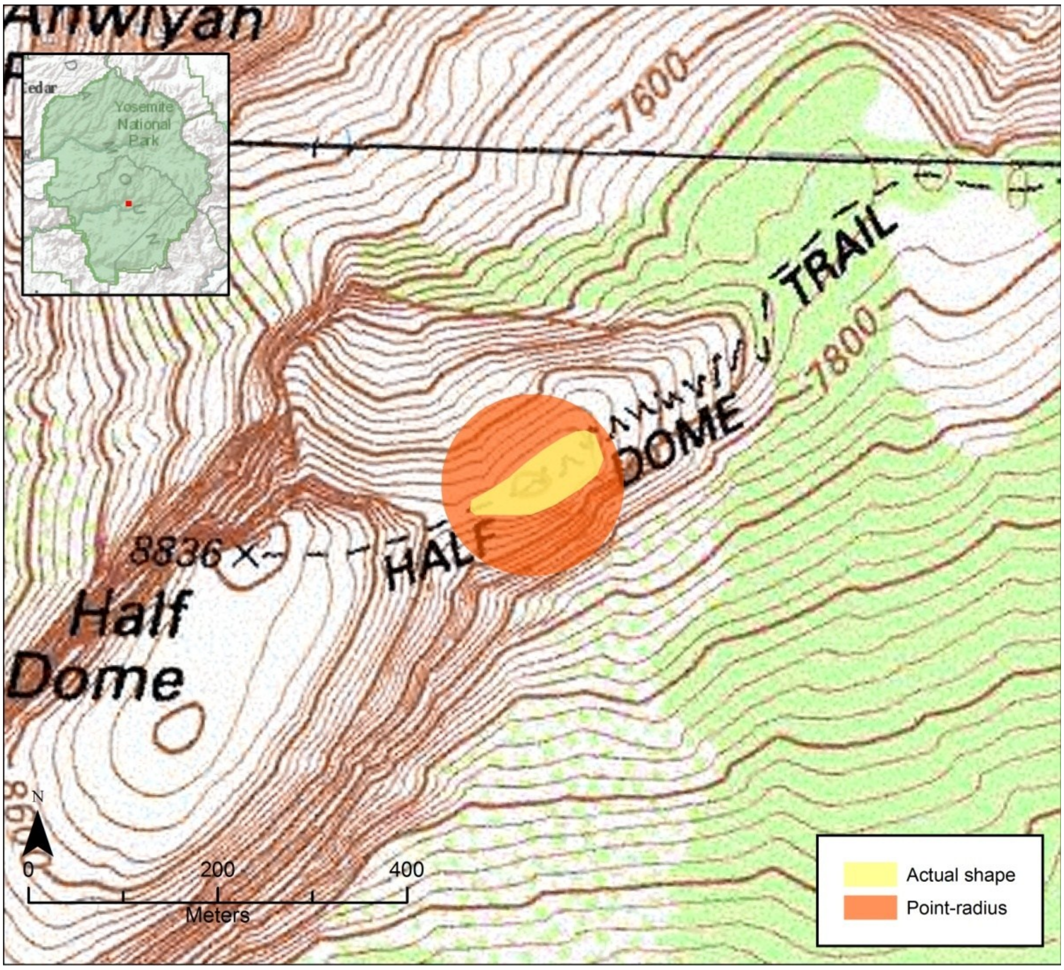


Figure 2.3 Case study #2 – Target object coincides with an areal feature reference object.

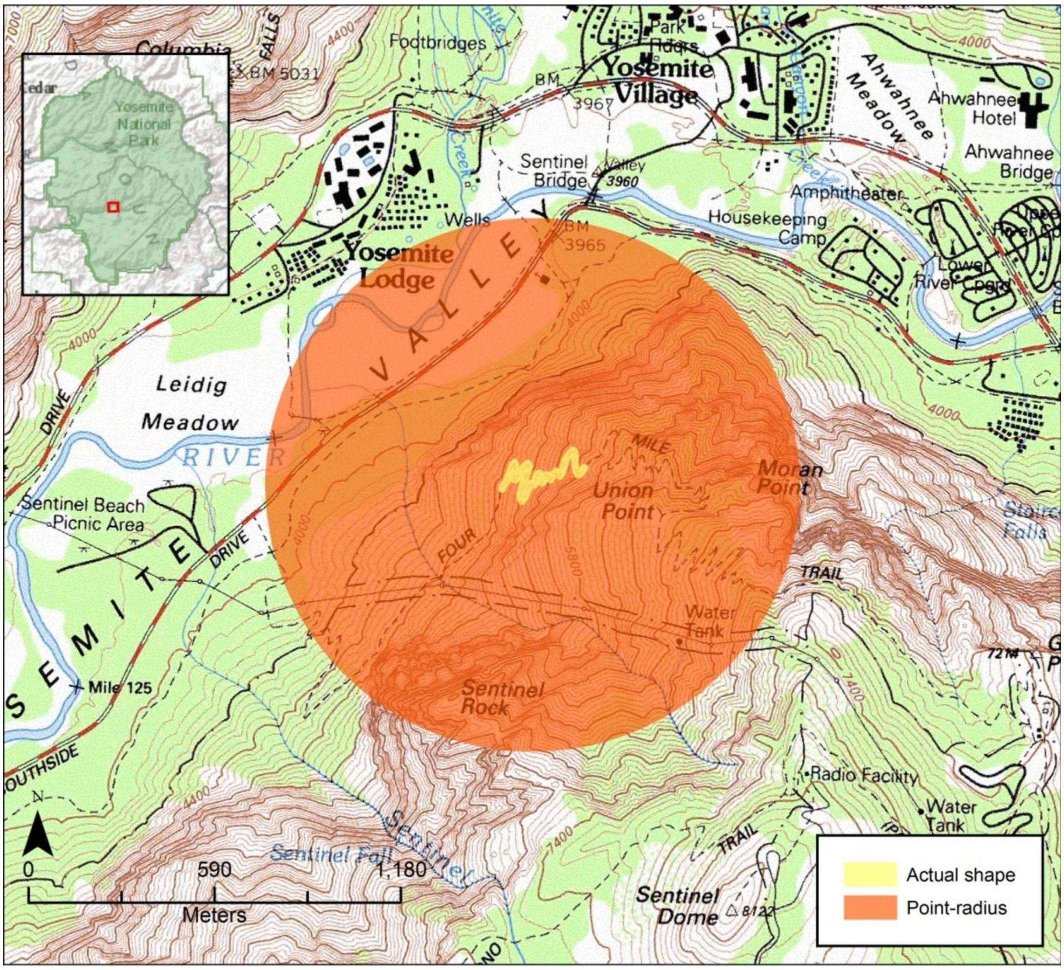


Figure 2.4 Case study #3 – Target object is offset along a path reference object an estimated distance from a feature reference object.

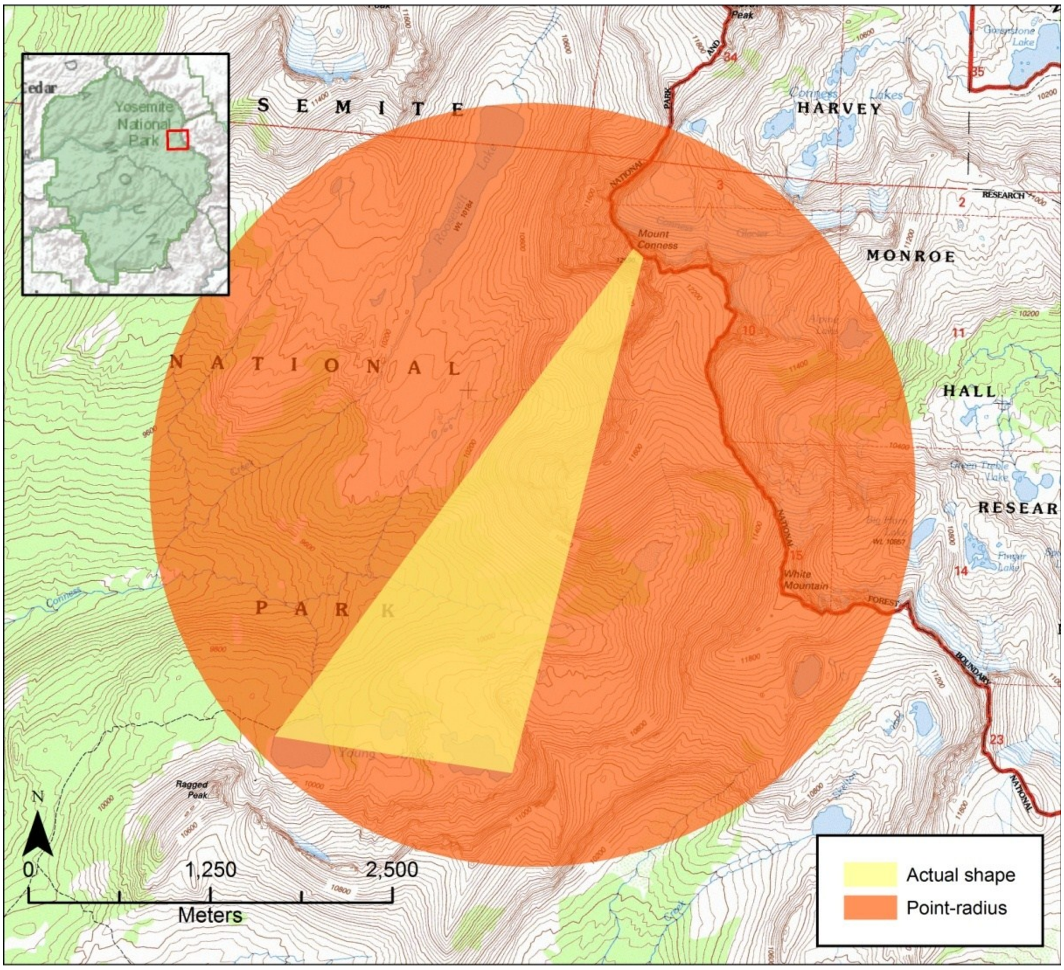


Figure 2.5 Case study #4 – Target object is an area between two reference object features.

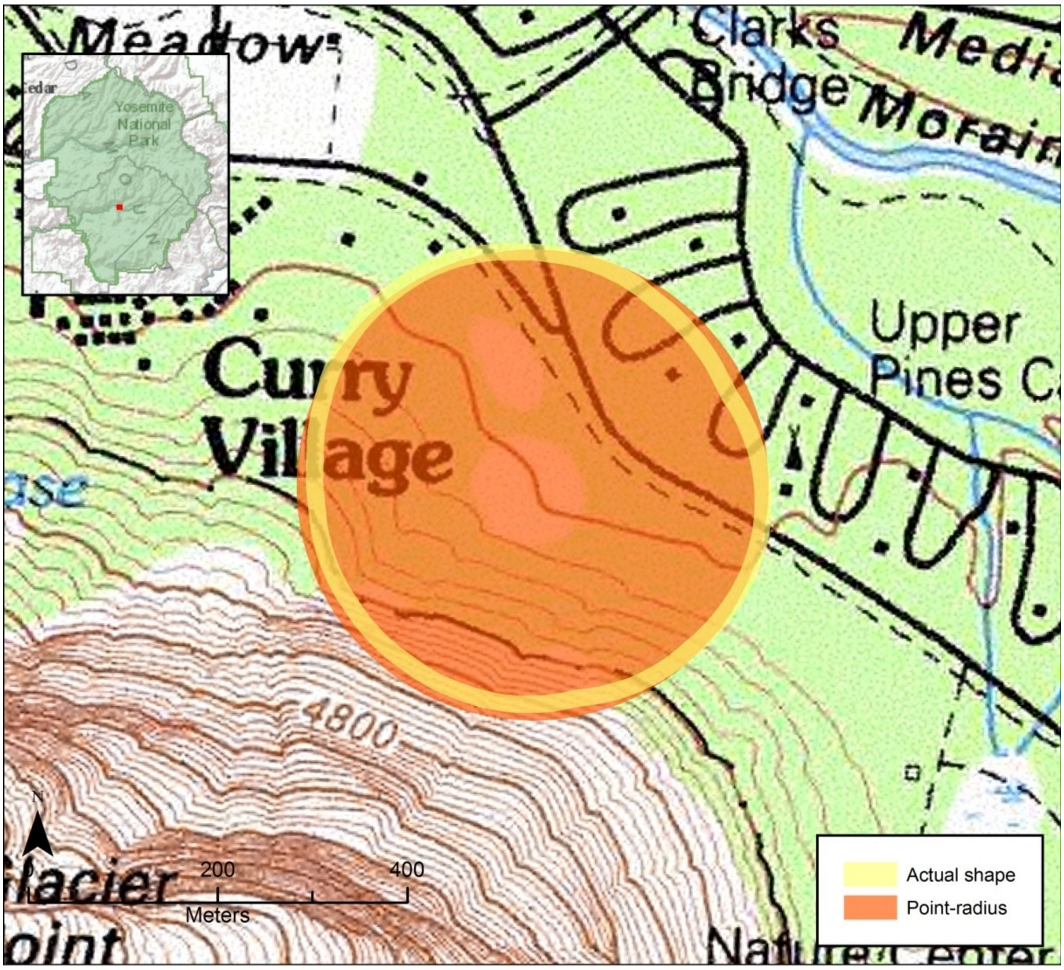


Figure 2.6 Case study #5 – Target object is an offset distance from an areal feature reference object.

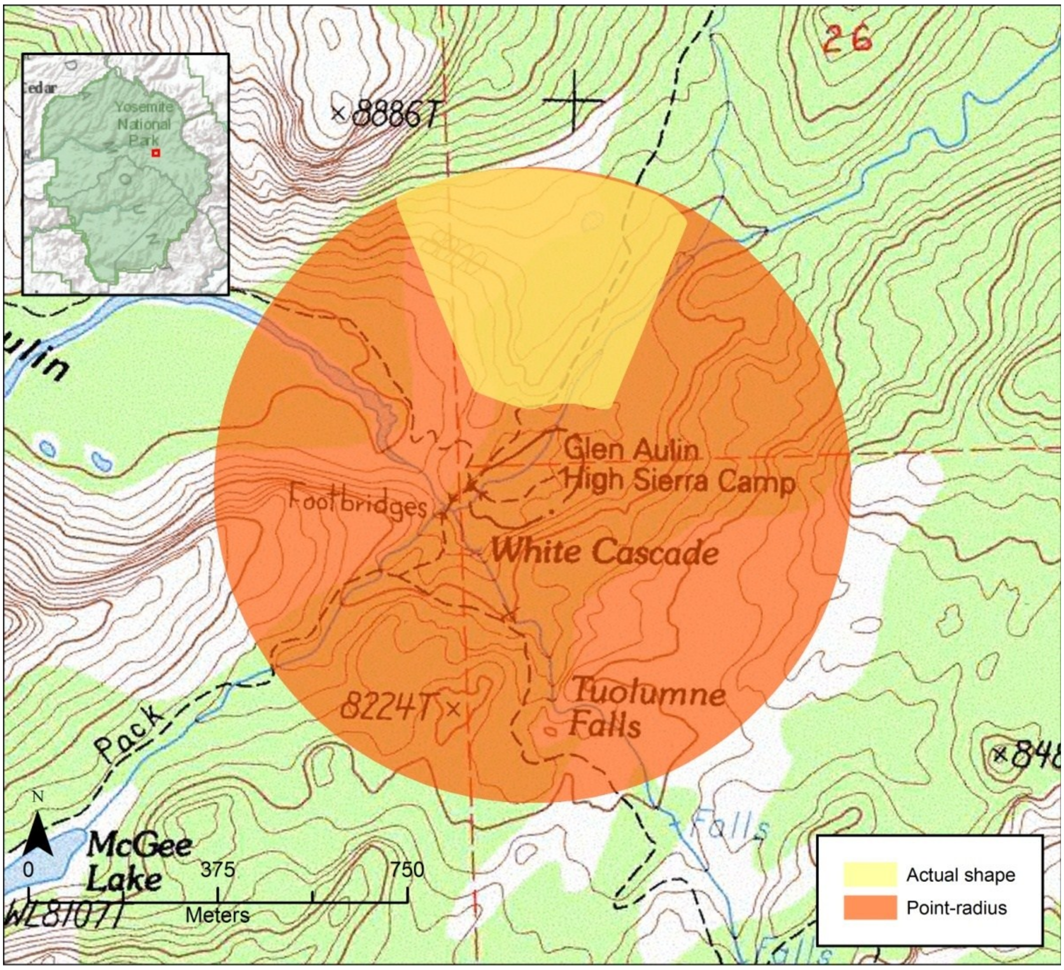


Figure 2.7 Case study #6 – Target object is a direction with a header from an areal feature reference object.

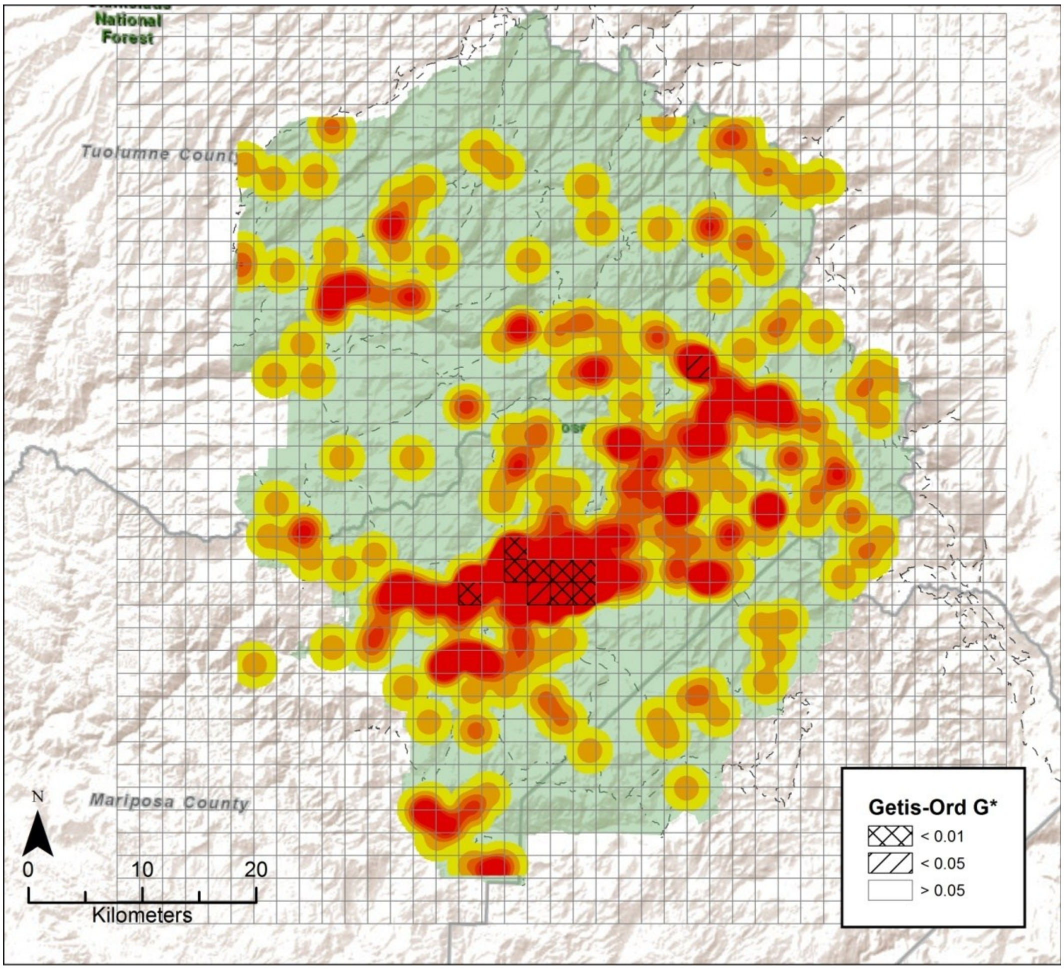


Figure 2.8 – A “hot spot” map where warm colors represent higher density cluster of incidents, the grids indicate statistical significance using the Getis-Ord G* spatial analysis.

Space-Time Analyses For Forecasting and Understanding Future Incident Occurrence: A Case-Study From Yosemite National Park Using the Presence and Background Learning Algorithm.

[Formatted and submitted to the *International Journal of GIScience*]

In order to address a spatio-temporal challenge such as incident prevention, we need information about the time and place where previous incidents have occurred in the past. Using geographic coordinates of incidents that occurred in the past in coincidence with spatial layers corresponding to environmental variables, we can produce probability maps in geographic and temporal space. Here we evaluate spatial statistic and machine learning approaches to answer an important space-time question: where and when are wildland search and rescue (WiSAR) incidents most likely to occur within Yosemite National Park in the future? We produced a probability map for the year 2011based on the presence and background learning algorithm (PBL) that successfully forecasts the most likely areas of future WiSAR incident occurrence based on environmental variables (distance to anthropogenic & natural features, vegetation, elevation, and slope) and the overlap with historic incidents from 2001-2010. This will allow decision-makers to spatially allocate resources where and when incidents are most likely to occur. In the process we not only answered questions related to a real-world problem, we also used novel space-time analyses that gives us insight into machine learning principles. The GIScience findings from this applied research have major implications for best-practices in future space-time research in the fields of epidemiology and ecological niche modelling.

Keywords: epidemiology**;** machine-learning**;** presence and background learning; search and rescue; spatio-temporal data

1 Introduction

The application of GISystems to solve real-world problems continues to expand from reactive, where we simply document and visualize where and when a phenomenon happened, to proactive, where we are able to reliably forecast event locations based on what we have learned from previous events. During this process we often test GIScience theories and techniques, leading to new scientific discovery. This is especially true in the field of spatial epidemiology, which merges spatial analysis with studies from public health (Ostfeld *et al.* 2005, Robertson *et al.* 2010). The objectives of such studies are to collect information about spatially varying factors that may contribute to the occurrence of disease, illness or injury and then attempt to map out likely areas of future occurrence, or locations, to apply preventative measures. In this paper we present a spatiotemporal presence and background learning algorithm and case-study to illustrate how this can be done using presence of incident locations and readily available environmental layers.

Within spatial epidemiology, datasets used for research often consist of incident coordinates, or other locality descriptions, that need to be georeferenced. Furthermore, most data describe locations where illness or injuries have previously occurred (presence) but not where they have not occurred (absence). Therefore, analyses have often been limited to descriptive analyses (density or "heat maps") and spatial statistics (hot spot or Getis Ord G* maps; Getis and Ord 1992)in lieu of predictive modeling because traditional modeling approaches require presence and absence data to derive relationships from underlying factors (Hirzel *et al.* 2002). This presence-only limitation is known as the geographic one-class data issue (GOCD; Guo *et al*. 2011) and requires a specialized approach to generate probability maps. The most promising methods for extrapolating meaningful information from GOCD have been machine learning algorithm techniques such as the presence and background learning algorithm (Li *et al.* 2011)that use background environmental variables and presence data to model suitability across the landscape. However, temporal variation between presence and environmental variables is not well studied(An and Brown 2008) and such analyses could reveal more meaningful relationships and useful distribution maps for decision making throughout the course of a year. A central goal for many GIScientists and their stakeholders is this; if given a data set with records of positive observations (e.g. presence of species, incidents, phenomena) then we can develop a model that forecasts presence/absence elsewhere at any specified time of day, week, or year in the future. This research highlights a novel approach for studying spatiotemporal relationships between incident occurrence and overlap with environmental variables that produces temporally explicit likelihood models.

2 Presence and background learning algorithm (PBL)

In this study we aim to model wildland search and rescue incident (WiSAR) likelihood that is conditional based on environmental covariates (denoted as x, Table 2.2). In other words, since previous WiSAR incidence in Yosemite is not randomly distributed across the landscape (Doherty *et al.* 2011), it is likely that there are landscape variables (environmental covariates) that can add probabilistic value to forecasts of where WiSAR will occur in the future. The incident locations themselves are GOCD. In the past, most presence and background modelling has relied on single time series data, and have done cross validation by splitting the data into training and testing data. In the current study, we use independent datasets from different time periods to train and test the modelling results, thus giving them temporal context.

2.1 The spatial probabilistic approach

This approach is described here using probabilistic equations where we denote presence of incidents as $y = 1$ and absence of incidents as $y = 0$. Hence, the desired model can be written as $P(y = 1 \mid x)$. Traditional statistical and supervised learning methods require both presence and absence data to model $P(y = 1 \mid x)$, but in reality it is difficult to obtain absence data.

Recently, Li et al. (2011) developed a presence and background learning algorithm (PBL) that is successful in modeling $P(y = 1 \mid x)$ without absence data. The model is trained by two completely separate sets: observed presence and background data. Note that background data contain presence and absence data but their labels are not known. The trained model is denoted as $P(s = 1 \mid x, \eta = 1)$, and hence

$$P(y=1 \mid x)=\frac{1-c}{c}\times\frac{P(s=1 \mid x,\eta=1)}{1-P(s=1 \mid x,\eta=1)} \tag{2.7}$$

By defining "prototypical presence" as locations where the incident likelihood is maximal, (Li *et al.* 2011) shows that the predicted value $P(s = 1 \mid x, \eta = 1)$ of any prototypical presence location can be used to estimate c. We selected half of the observed presence locations, those whose predicted probability $P(s = 1 \mid x, \eta = 1)$ are higher than 50^{th} percentile as prototypical presence locations. We then averaged their predicted probabilities to estimate the constant c and adjust the trained model to obtain the desired model using (1).

The PBL algorithm is a training framework and it can be implemented using any binary classifier that can estimate conditional probability. We used a discriminative Maximum Entropy (MAXENT) classifier (Berger *et al.* 1996) to implemented the PBL algorithm. Entropy is a fundamental concept in information theory, and it measures how much choice is involved in the selection of an event (Shannon 1948). The principle of maximum entropy indicates that the distribution model that satisfies any given constraints should be as uniform as possible (Phillips *et al.* 2004), which agrees with everything that is known, but carefully avoids assuming anything that is unknown (Jaynes 1990). In this study, we used the Matlab code for MAXENT that is freely available online[6].

2.2 The temporal probabilistic approach

Using machine learning algorithms such as PBL have been useful in understanding probable distribution patterns across geographic space; however little research has been done to study the usefulness of these techniques across temporal space. For example, the model in equation (1) assumes temporal stationarity which is not likely to be a valid assumption for incident data that varies seasonally due to weather or background activity patterns.

For situations where temporal variability limits the meaningfulness of such an approach, it is better to study $P(y = 1 \mid x)$ at defined time intervals represented as t. Given historical observation data, we can group the training samples by t and train the models individually, which will lead to $P(y = 1 \mid x)$ values that vary with time as in equation (2.8).

$$P_t(y = 1 \mid x),\ \mathrm{t} = 1,2,3\ldots.\mathrm{n} \tag{2.8}$$

For example, if we group the training data by month of the year, then $t = 1, 2, 3, ..., 12$. In this case, decisions on how to group the presence data by discrete time intervals t would be made prior to training the models based on expert knowledge about the data or specific research questions.

3 Case-study

Here we study a real-world problem: wildland search and rescue (WiSAR) incident prevention in Yosemite National Park. WiSAR is the process of locating, accessing, stabilizing, and transporting people in proximity to or within wilderness environments

---

[6] http://www.cs.grinnell.edu/~weinman/code/index.shtml

(Worsing 1993). Our objectives are to describe a unique methodology, explain results, and discuss the implications of our findings for WiSAR incident prevention, spatial epidemiology, and GIScience.

Every year millions of people will enjoy recreating on public lands in the United States and across the globe. Unfortunately, many of these visitors (estimated to be more than 100,000 per year) will experience an injury, medical issue, or lose their way in the wilderness requiring a WiSAR response (NASAR 2005). The process of searching for and rescuing victims can be dangerous, time-consuming, and costly (Heggie and Heggie 2009) for both the rescuee and the rescuer alike and therefore preventative efforts are warranted. Yet, in order to intentionally prevent any incident from occurring, there are some prerequisites required. Conceptually, incident prevention is simple (Lawson *et al.* 2006); the incident must be preventable, one must know about the possibility of the incident ahead of time, and an action must occur at the right time and place to inflict some plausible change to prevent the situation. In Preventative Search and Rescue (PSAR), the primary tool employed for prevention is to educate people at risk by providing information (or if necessary, enforcing restrictions and law enforcement measures). One study in Grand Canyon National Park (Yee and Iserson 2008) suggested that since implementing a PSAR program the WiSAR incident rate has decreased from 9.4 incidents per 100,000 visitors in the 10 years prior to 1998 to 7.6 per 100,000 visitors in the six years following ($p = 0.02$). In this study, park staff focused primarily on reducing heat-related illnesses and used summary statistics to support their conclusion on a park-wide scale.

There appears to be a widespread interest in WiSAR risk assessment research, especially in national parks, that began over a decade ago with the landmark paper, Morbidity and Mortality in the Wilderness (Montalvo *et al.* 1998). This was the first of several retrospective studies to discuss how WiSAR incident data could be used to “guide future wilderness use, education, and management” and alluded to how a “standardized, computerized database would greatly facilitate future evaluations, decisions, and policies”. Since then, however, findings in retrospective studies have been limited to age, sex, recreational activity, mode of rescue, and contributing factors in WiSAR incidents (Wild 2008, Forrester and Holstege 2009, Heggie and Amundson 2009, Heggie and Heggie 2009) and have not contained spatially explicit information. In order to know what safety information to provide to the public, PSAR personnel must be familiar with the when and where details of current hazards (weather, trail conditions, overall fitness of visitors) as well as hazards that have occurred in the past (spatio-temporal information). The United States, Canada, and other countries are currently re-evaluating their Search and Rescue data collection programs, therefore this research is timely. We intend to produce geographic and temporal knowledge from raw data to answer relevant questions related to PSAR in Yosemite National Park.

3.1 Study Site and dataset

Yosemite National Park (YNP; Figure 2.9) is located in the Sierra Nevada Mountains of California, USA and has over 1287 kilometers (800 miles) of hiking trails with over 200 WiSAR incidents each year (Hung and Townes 2007a). High visitation numbers, even in remote locations, lead to incidents that occur far from roads and in steep terrain that prevent emergency vehicle access. To transport patients, Yosemite Search and Rescue (YOSAR) use rescue teams with trail litter apparatus and/or a large helicopter to perform a significant proportion of their rescues. Therefore, WiSAR incidents result in not only injury or misfortune to park visitors, but also significant cost in terms of risk to rescuers and financial resources (Heggie and Amundson 2009).

Earlier researchers (Hung and Townes 2007) used an YNP case-study to document the types of injuries and illnesses visitors faced between the years 1990 - 2001 and also cited the need for incident prevention. This study is perhaps one of the most comprehensive of its kind. However, their description of YNP incidents did not make any specific geographic or temporal recommendations which may be why managers in YNP and elsewhere did not respond with spatio-temporally specific PSAR activities until later years (after preliminary results of our analysis were released).

The present research differs from these past studies in that the first step involved georeferencing WiSAR incidents from text using the point-radius method as described in (Doherty *et al.* 2011) this is a georeferencing approach that allows for generating hypothesis-based coordinates for incidents using reference objects (a known locality found in a gazetteer) and target objects (a location that has some spatial relationship to the reference object) that are common components to a text-based locality descriptions see (Wieczorek *et al.* 2004a). Essentially, without coordinate-based descriptions, we used a best practices approach to create spatial data with a calculated horizontal uncertainty so we could decide on the appropriate methodology for analysis. This method was chosen in place of the more costly and complex shape method (Guo *et al.* 2008) so that this study could be extensible to other WiSAR teams without specialized software. Also, the PBL analysis uses point data and techniques that are believed to account for spatial uncertainty (via local spatial autocorrelation), especially at moderate levels in the WiSAR incident reports. Preliminary research (showed that these data had uncertainty radii that were smaller (95% were between 269–866m) than acceptable limits described in previous studies (Fernandez *et al.* 2009).

We chose environmental variables (Table 2.2) based on expert knowledge from park staff and previous research(Hung and Townes 2007). Actual visitor activity is calculated temporally (number of visitors coming through entrance stations) but it has not been possible to measure spatially within the park. Therefore where visitors travel in the park is considered a latent variable strongly related to incident probability–we use variables related to accessibility (distance to trailheads, distance to road, distance to trails) as a proxy to actual spatial visitation data. Other variables were chosen based on their association with incident probability (distance to streams, slope, elevation, vegetation). This approach is similar to ecological niche modeling where variables related to habitat availability and preference would be considered. Moreover, for this study we are not interested in calculating risk (incidents / per capita) but rather spatio-temporal incident probability for decision making.

3.2 Spatial and temporal framework

With regards to spatial epidemiology, we are interested in hypotheses related to overall yearly WiSAR predictions and the seasonality of predictions (where and when). Therefore we used training and validation sets derived from a comprehensive dataset (all of the 2001–2010 incident data) and conducted analysis for global (entire year) and monthly forecasting periods.

The output of the PBL method gives an estimate of probability between zero and one, so we needed to utilize a threshold for generating probabilistic output to binary predictions. Here we randomly set aside 25% of the training set for validation and the probabilistic value corresponding to a 5% omission rate for the validation set was used as the threshold to make a binary prediction (Pearson *et al.* 2004, Li and Guo 2010). To evaluate model accuracy, or how well the PBL models classified true presence locations in the future (2011 data), we sampled the binary output for each of the models (global and monthly) against their respective temporal interval at the 2011 incident data locations. In essence, we used the 2001–2010 WiSAR data to forecast locations where and when WiSAR incidents were most likely to occur in 2011.

To evaluate the accuracy, or how well the PBL algorithm generated from presence points classified true presence locations we sampled the binary output from the model at these validated points. We then used the binomial test to compare the rate of classification at coordinates where WiSAR incidents occurred in 2011 but were not included in the training sample (2001 – 2010).

4 Results

The threshold for creating binary classification of PBL was a value of 0.021 which resulted in 22.9% of pixels in the Park being classified as “suitable” (1). The PBL model based on all of the 2001–2010 training and validation data predicted potential suitable “habitat” for WiSAR incidents with an overall high success rate (low omission rate) of 90.9% for the 2011 test data ($n = 186$, $p < 0.001$; see Figure 2.10).

According to a jackknife sensitivity analysis, the most important environment variables overall in determining which pixels receive the highest score were distance to trails and trailhead–although these variables varied spatio-temporally (Table 2.3). When we withdrew slope from calculations, the overall goodness of fit of the trained model decreased the most. This indicates that this variable provides unique relationships information not already contained in other variables.

PBL models varied considerably across temporal space and feature space (Table 2.4). In the example shown in Figure 2.11, the percentage of the Park classified as “likely” to have WiSAR activity varies from 5% (April) to 27% (August). Overall, 72% of incidents occur in May–August yet the spatial distribution of incidents varies considerably across these months.

5 Discussion and Conclusion

The PBL provided output maps that could successfully "forecast" future incidents from retrospective data year-round and by season. There were distinct spatio-temporal patterns exposed by probability maps and hypotheses between environmental variables. In general, we can conclude that incidents are most likely to occur relatively close to access points (trails, roads, and trailheads) and in steeper terrain and this can be predicted with a low omission rate.

The spatio-temporal approach, described in equation 2, conducted separate PBL analysis based on the monthly intervals. This approach yields maps that very clearly show temporal non-stationarity and reflect real-world patterns. For example, incidents are limited to the lowest elevations regions of the Park during the early winter months when roads are closed but snow conditions are not suitable for skiing. In late winter / early spring WiSAR incidents are distributed along roads where skiing is a popular activity. During the late summer / early fall months the Park is widely accessible and incidents are distributed across the trail network and further from roads than any other time of year. On the one hand these observations are consistent with general knowledge of how visitor use the Park throughout the year. Yet, the model outputs provide two distinct advantages over alternative methods for this type of study 1) the ability to study small-scale likelihood in absence of the ability to measure true background activity (spatially explicit visitation levels) 2) the ability to identify potentially suitable / probable areas that do not yet have a high frequency of occurrence, but could have greater activity in the future. For instance, trails along steep slopes around the Hetch Hetchy region of the Park have a high potential for WiSAR incidents based on environmental factors but do not have the same level of visitation. If visitation increases in those areas due to Park policies or other factors (e.g. climate change, shift in social norms and human geography) then we would expect WiSAR incidents to increase.

Since temporal non-stationarity is likely an issue with most datasets where the PBL method can be applied, we believe an analysis-by-interval approach to be most appropriate. Furthermore, we selected thresholds on the continuous model outputs based on a 5% omission error threshold on the validation dataset to produce Boolean probable-not probable maps. We did this for each monthly dataset and observed variability in the number of total pixels labelled as “probable”. As stated in earlier research studies (Fielding and Bell 1997, Suárez-

Seoane *et al.* 2008), it must be considered that as larger thresholds are selected, commission errors will decrease, but omission errors will increase. By accounting for temporal variation in probability modeling we can produce maps that more accurately reflect the size and distribution of probable areas (large in summer, small in winter) and alleviate the commission / omission error issue. For Park Staff, this helps describe seasonality of their incidents and justify resource allocation throughout the year. Videos[7] and time-enabled maps can also be used by non-GIS trained staff for planning PSAR operations.

The ability to exploit intra-annual cycles and spatial relationships is extremely important to a wide variety of scientists and practitioners. For instance, Finket al. (2010) introduced the concept of spatiotemporal exploratory models (STEM) for studying broad-scale survey data for bird species. This ecological research highlighted the advantages of spatiotemporal structure to studies over large geographic areas with parametric data. Other spatiotemporal techniques based on relatedness such as Markov Chain Analysis (Clark 1965), Hidden Markov Models (Green and Richardson 2002) and Markov Random Field Theory (Liu and Cai 2011) utilize relatedness between events or temporal trajectories to develop classifications for phenomena such as land cover. Yet, the presence and background learning algorithm approach described in this paper can be utilized on non-related small scale data that vary spatially with the seasons and do not need to meet the same assumptions / requirements of traditional statistical techniques.

Future research should investigate improvements to the spatiotemporal PBL approach. In this study we chose monthly intervals for model analysis because these time periods closely reflected park management principles (i.e. expert knowledge), but this led to small sample sizes in winter months and required some prior knowledge of the temporal nature of the data. There were months (e.g. December) where the incident sample size was not large enough to be used as a training sample dataset. This is likely to be a problem wherever incident likelihood varies across the temporal dimension, as in this study where overall visitation to park peaks during spring and summer months and access to the park changes due to weather conditions throughout the year. Due to limitations in sample size and decision making on temporal interval size for analysis, we suggest further research efforts be invested in a hybrid approach that allows for the model to decide on logical time intervals based on an adaptive weighting functions similar to those found in the Geographic-Temporally Weighted Regression methods described in Huang *et al.* (2010).This would allow for intervals of adequate sample size, reveal natural temporal clusters of probability, and remove the subjectivity involved with deciding on what time intervals to use for sampling. Overall, seasonality and machine learning techniques should be studied more closely when possible as this is apparent in habitat niche modeling for seasonally nomadic or migratory species (Suárez-Seoane *et al.* 2008), disease mapping (Moffett *et al.* 2007), and in this case search and rescue incidents whose distribution varies spatiotemporally.

For preventative search and rescue (PSAR) within YNP, our analyses have only just begun to provide meaningful conclusions related to where and when to stage resources. For instance, YNP staff members now have a map of WiSAR incident probability for each season and can use these to decide which trailheads or locations to implement PSAR interventions. Furthermore, a follow up study similar to Iserson and Lee (2008) Grand Canyon study could be conducted after new strategic PSAR efforts based on maps produced in this study have been implemented. In order to provide YNP staff with their own analysis capabilities we are designing a Web GIS platform that allows for spatial, temporal, and attribute queries of historic incidents. We expect the results and recommendations from this research to guide WiSAR records management systems and visitor protection policies.

[7] http://youtu.be/PloVyDxlI4o

References

Adams, A.L., Schmidt, T. a, Newgard, C.D., Federiuk, C.S., Christie, M., Scorvo, S., and DeFreest, M., 2007. Search is a time-critical event: when search and rescue missions may become futile. *Wilderness & Environmental Medicine*, 18 (2), 95–101.

Adriaensen, F., Chardon, J.P., De Blust, G., Swinnen, E., Villable, S., Gulinck, H., and Matthysen, E., 2003. The application of "least-cost" modelling as a functional landscape model. *Landscape and Urban Planning*, 64 (4), 233–247.

Alexander, D., 1991. Information technology in real-time for monitoring and managing natural. *Progress in Physical Geography*, 15 (3), 238.

An, L. and Brown, D.G., 2008. Survival Analysis in Land Change Science : Integrating with GIScience to Address Temporal, (February 2013), 37–41.

Baldwin, R. a., 2009. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy*, 11 (4), 854–866.

Bateman, I.J., Garrod, G.D., Brainard, J.S., and Lovett, A. a., 1996. Measurement Issues in the Travel Cost Method: a Geographical Information Systems Approach. *Journal of Agricultural Economics*, 47 (1-4), 191–205.

Berger, A., Della Pietra, S., and Della Pietra, V., 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22 (1992), 39–71.

Bownds, J.W., Ebersole, M.J., Lovelock, D., O'Connor, D.J., and Toman, R.J., 2007. Win CASIE III: Computer Aided Search Information Exchange.

Van den Broek, M., Brederode, E., Ramírez, A., Kramers, L., Van der Kuip, M., Wildenborg, T., Turkenburg, W., and Faaij, A., 2010. Designing a cost-effective CO2 storage infrastructure using a GIS based linear optimization energy model. *Environmental Modelling & Software*, 25 (12), 1754–1768.

Church, R. and ReVelle, C., 1974. The maximal covering location problem. *Papers in regional science*, 32 (1), 101–118.

Clark, W., 1965. Markov chain analysis in geography: an application to the movement of rental housing areas. *Annals of the Association of American …*, 55 (2), 351 – 359.

Cutter, S.L., 2003. GI Science, Disasters, and Emergency Management. *Transactions in GIS*, 7 (4), 439– 445.

Dijkstra, E.W., 1959. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1, 269–271.

Doherty, P., Ferguson, D., Goodrich, M.A., Koester, R.J., and Doke, J., 2012. Wilderness Search & Rescue and GIScience. *In*: *Annual Meeting of the Association of American Geographers*. New York.

Doherty, P., Guo, Q., Liu, Y., Wieczorek, J., and Doke, J., 2011. Georeferencing Incidents from Locality Descriptions and its Applications: a Case Study from Yosemite National Park Search and Rescue. *Transactions in GIS*, 15 (6), 775–793.

Doherty, P.J., Guo, Q., and Alvarez, O., 2012. Expert versus machine: A comparison of two suitability models for emergency helicopter landing areas in Yosemite National Park. *Professional Geographer*.

Doherty, P.J., Guo, Q., Li, W., and Doke, J., n.d. Space-Time analyses for forecasting and understanding future incident occurrence: a case-study from Yosemite National Park using the presence and background learning algorithm. *International Journal of Geographical Information Science*.

Doke, J., 2012. Analysis of Search Incidents and Lost Person Behavior in Yosemite National Park.

Durkee, G., 2010. GIS Joins Search for a Missing Hiker on California's Mount Whitney. *ArcWatch*, Apr.

Esri, 2012. ArcGIS 10.1.

Federal Standards and Procedures for the National Watershed Boundary Dataset (WBD), 2012. Reston, Virginia.

Ferguson, D., 2008. GIS for Wilderness Search and Rescue. *In*: *Esri Federal User Conference*. Washington D.C., 1 – 11.

Fernandez, M.A., Blum, S.D., Reichle, S., Holzman, B., and Hamilton, H., 2009. Locality uncertainty and differential performance of four different common niche-modeling techniques. *Biodiversity Informatics*, 6 (1), 36–52.

Fielding, A.H. and Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24 (1), 38–49.

Filipkowska, E., Koester, R.J., Chrustek, R., and Zaród, M., 2012. Lost and Found in the Polish Carpathian Mountains. *ArcNews*, 34 (3).

Fink, D. and Hochachka, W., 2010. Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications*, 20 (8), 2131–47.

Forrester, J.D. and Holstege, C.P., 2009. Injury and illness encountered in Shenandoah National Park. *Wilderness & environmental medicine*, 20 (4), 318–26.

Frost, J.R., 1999. Principles of search theory. *Response*, 17 (2), 1 – 23.

Getis, A. and Ord, J.K., 1992. The Analysis of Spatial Association. *Geographical Analysis*, 24 (3).

Goodchild, M., 1992. Geographical information science. *International Journal of Geographical Information …*, 6 (March 2013), 37–41.

Goodchild, M.F., 2003. Geospatial data in emergencies. *In*: S.L. Cutter, D.B. Richardson, and T.J. Wilbanks, eds. *The Geographical Dimensions of Terrorism*. New York: Routeledge, 99–104.

Goodchild, M.F., 2004. GIScience, Geography, Form, and Process. *Annals of the Association of American Geographers*, 94 (4), 709–714.

Green, P.J. and Richardson, S., 2002. Hidden Markov Models and Disease Mapping. *Journal of the American Statistical Association*, 97 (460), 1055–1070.

Guo, Q., Li, W., Liu, Y., and Tong, D., 2011a. Predicting potential distributions of geographic events using one-class data: concepts and methods. *International Journal of Geographical Information Science*, 25 (10), 1697–1715.

Guo, Q., Li, W., Liu, Y., and Tong, D., 2011b. Predicting potential distributions of geographic events using one-class data: concepts and methods. *International Journal of Geographical Information Science*, 25 (10), 1697–1715.

Guo, Q., Liu, Y., and Wieczorek, J., 2008. Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *International Journal of Geographical Information Science*, 22 (10), 1067–1090.

Heggie, T.W. and Amundson, M.E., 2009. Dead men walking: search and rescue in US National Parks. *Wilderness & environmental medicine*.

Heggie, T.W. and Heggie, T.M., 2009. Search and rescue trends associated with recreational travel in US national parks. *Journal of Travel Medicine*, 16 (1), 23–7.

Hill, K. and Gale, R., 1997. *Managing the lost person incident*. Managing. Chantilly, VA: National Association for Search and Rescue.

Hirzel, A.H., Hausser, J., Chessel, D., and Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, 83 (7), 2027–2036.

Hogg, J.M., 1968. The Siting of Fire Stations. *Journal of the Operational Research Society*, 19, 275–287.

Huang, B., Wu, B., and Barry, M., 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24 (3), 383–401.

Hung, E.K. and Townes, D. a, 2007a. Search and rescue in Yosemite National Park: a 10-year review. *Wilderness & environmental medicine*, 18 (2), 111–6.

Hung, E.K. and Townes, D. a, 2007b. Search and rescue in Yosemite National Park: a 10-year review. *Wilderness & Environmental Medicine*, 18 (2), 111–6.

Imhof, E., 1950. *Gelaende und Karte*. Zurich, Switzerland: Rentsch.

Indriasari, V., Mahmud, A.R., Ahmad, N., and Shariff, A.R.M., 2010. Maximal service area problem for optimal siting of emergency facilities. *International Journal of Geographical Information Science*, 24 (2), 213–230.

Jaynes, E.T., 1990. Notes on present status and future prospects. *In*: W.. Grandy and L.H. Schick, eds. *Maximum entropy and Bayesian methods*. Dodrecht: Kluwer.

Kim, H.-M. and Kwan, M.-P., 2003. Space-time accessibility measures: A geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration. *Journal of Geographical Systems*, 5 (1), 71–91.

Koch, T. and Denike, K., 2009. Crediting his critics' concerns: remaking John Snow's map of Broad Street cholera, 1854. *Social science & medicine (1982)*, 69 (8), 1246–51.

Koester, R., 2008. *Lost Person Behavior*. Search. Charlottesville, VA: dbS Productions.

Koester, R.J., Cooper, D.C., Frost, J.R., and Robe, R.Q., 2004. Sweep Width Estimation for Ground Search and Rescue. Washington.

Koester, R.J. and Stooksbury, D.E., 1995. Behavioral profile of possible Alzheimer's disease patients in Virginia search and rescue incidents. *Wilderness & Environmental Medicine*, 6, 34–43.

Koopman, B.O., 1980. *Search and Screening*. New Yorrk: Pergamon Press.

Kratzke, T.M., Stone, L.D., and Frost, J.R., 2010. Search and Rescue Optimal Planning System (SAROPS). *In*: *13th Conference on Information Fusion*. Edinburgh, Scotland: IEEE, 1 – 8.

Kwan, M.-P., 2003. Intelligent emergency response systems. *In*: S.L. Cutter, D.B. Richardson, and T.J. Wilbanks, eds. *The Geographical Dimensions of Terrorism*. New York: Routeledge, 111–116.

Lawson, A., Gangnon, R., and Wartenberg, D., 2006. Developments in disease cluster detection. *Statistics in Medicine*, 25 (5), 721.

Li, W. and Guo, Q., 2010. A maximum entropy approach to one-class classification of remote sensing imagery. *International Journal of Remote Sensing*, 31 (8), 2227–2235.

Li, W., Guo, Q., and Elkan, C., 2011. Can we model the probability of presence of species without absence data? *Ecography*, 34 (6), 1096–1105.

Li, X. and Yeh, A., 2005. Integration of genetic algorithms and GIS for optimal location search. *International Journal of Geographical Information Science*, 19 (5), 581–601.

Lin, L. and Goodrich, M. a., 2010. A Bayesian approach to modeling lost person behaviors based on terrain features in Wilderness Search and Rescue. *Computational and Mathematical Organization Theory*, 16 (3), 300–323.

Liu, D. and Cai, S., 2011. A Spatial-Temporal Modeling Approach to Reconstructing Land-Cover Change Trajectories from A Spatial-Temporal Modeling Approach to

Reconstructing Land-Cover Change Trajectories from Multi-temporal Satellite Imagery. *Annals of the Association of American Geographers*, 102 (6), 1329–1347.

Liu, Y., Guo, Q., Wieczorek, J., and Goodchild, M.F., 2009. Positioning localities based on spatial assertions. *International Journal of Geographical Information Science*, 23 (11), 1471–1501.

Malczewski, J., 2004. GIS-based land-use suitability analysis: a critical overview. *Progress in Planning*, 62 (1), 3–65.

Miller, H., 1996. GIS and geometric representation in facility location problems. *International Journal of Geographical Information …*, 10 (7), 37–41.

Miller, H.J. and Bridwell, S. a., 2009. A Field-Based Theory for Time Geography. *Annals of the Association of American Geographers*, 99 (1), 49–75.

Moffett, A., Shackelford, N., and Sarkar, S., 2007. Malaria in Africa: Vector Species' Niche Models and Relative Risk Maps. *PLoS ONE*, 2 (9), e824.

Montalvo, R., Wingard, D.L., Bracker, M., Davidson, T.M., and Diego, S., 1998. Conferences and Reviews Morbidity and Mortality in the Wilderness. *Wilderness and Environmental Medicine*, 168 (4), 248–254.

Mostern, R. and Johnson, I., 2008. From named place to naming event: creating gazetteers for history. *International Journal of Geographical Information Science*, 22 (10), 1091–1108.

Murray, A. and Tong, D., 2007. Coverage optimization in continuous space facility siting. *International Journal of Geographical …*, (March 2013), 37–41.

National Association for Search and Rescue, 2005. *Fundamentals of Search and Rescue*. Sudbury, MA: Jones and Bartlett Publishers.

Nikolakaki, P., 2004. A GIS site-selection process for habitat creation: estimating connectivity of habitat patches. *Landscape and Urban Planning*, 68 (1), 77–94.

Ostfeld, R.S., Glass, G.E., and Keesing, F., 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution*, 20 (6), 328–36.

Pearson, R., Dawson, T., and Liu, C., 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, 3, 285–298.

Phillips, S.J., Dudik, M., and Schapire, R.E., 2004. A maximum entropy approach to species distribution modeling. *In*: *Proceedings of the Twenty-First International Conference on Machine Learning*. Banff, Alberta, CA, 655–662.

Robertson, C., Nelson, T.A., MacNab, Y.C., and Lawson, A.B., 2010. Review of methods for space–time disease surveillance. *Spatial and Spatio-temporal Epidemiology*, 1 (2–3), 105–116.

Schneider, P.J. and Schauer, B.A., 2006. HAZUS—Its Development and Its Future. *Natural Hazards Review*, 7, 40–44.

Shannon, C., 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423.

Sherrill, K.R., Frakes, B., and Schupbach, S., 2010. Travel time cost surface model: standard operating procedure. Natural Resource Report NPS/NRPC/IMD/NRR—2010/238. Fort Collins, Colorado.

Stone, L.D., 1989. What's happened in search theory since the 1975 Lanchester Prize? *Operations Research*, 37 (3), 501.

Stone, L.D., 2007. *Theory of Optimal Search*. 2nd ed. Mathematics in Science and Engineering. New York: Academic Press.

Suárez-Seoane, S., García de la Morena, E.L., Morales Prieto, M.B., Osborne, P.E., and De Juana, E., 2008. Maximum entropy niche-based modelling of seasonal changes in little bustard (Tetrax tetrax) distribution. *Ecological Modelling*, 219, 17–29.

Syrotuck, W.G., 1976. *Analysis of Lost Person Behavior*. Mechanicsburg, PA: Barkleigh Productions, Inc.

Tate, E., Cutter, S.L., and Berry, M., 2010. Integrated multihazard mapping. *Environment and Planning B: Planning and Design*, 37 (4), 646–663.

The Outdoor Foundation, 2012. Outdoor Recreation Participation Report.

Theodore, J., 2009. When every second counts. *ArcNews*, (March), 66 – 69.

Tobler, W., 1965. Non-Isotropic Geographic Modeling. *Information Systems*.

Tobler, W., 1991. Non-Isotropic Geographic Modeling. *In*: W. Tobler, ed. *Geographic Information Systems in the Social Sciences*. Santa Barbara.

Toregas, C., Swain, R., ReVelle, C., and Bergman, L., 1971. The Location of Emergency Service Facilities. *Operations Research*, 19 (6), 1363–1373.

Whitley, T. and Hicks, L., 2003. A geographic information systems approach to understanding potential prehistoric and historic travel corridors. *Southeastern Archaeology*, (those 1994).

Wieczorek, J., Guo, Q., and Hijmans, R., 2004a. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18 (8), 745–767.

Wieczorek, J., Guo, Q., and Hijmans, R., 2004b. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18 (8), 745–767.

Wild, F.J., 2008. Epidemiology of mountain search and rescue operations in Banff, Yoho, and Kootenay National Parks, 2003-06. *Wilderness & Environmental Medicine*, 19 (4), 245–51.

Winter, S. and Yin, Z.-C., 2010. Directed movements in probabilistic time geography. *International Journal of Geographical Information Science*, 24 (9), 1349–1365.

Worsing, R.J., 1993. *Rural Rescue and Emergency Care*. 1st ed. Rosemont, IL: American Academy of Orthopaedic Surgeons.

Yee, K. and Iserson, K. V., 2008. The Epidemiology of Search and Rescue Incidents in the Grand Canyon National Park: Are Preventive Measures Making a Difference? *Western Journal of Emergency Medicine*, 9 (1), 3–5.

Table 2.2

Variable	Description	Hypotheses
Trails	Euclidean distance (m) to foot trails.	Visitors spend more time on or near hiking trails.
Trailheads	Euclidean distance (m) to foot access points.	Visitors spend more time near hiking trail access points.
Slope	Slope angle (degrees) of terrain, derived from 10m DEM.	Visitors spend more time in flat areas, but are also more likely to become ill or injured in steeper terrain.
Elevation	Elevation (m) above sea-level from 10m digital elevation model (DEM).	Visitors spend more time at elevations where the climate is moderate, but are exposed to extreme weather at the highest elevations.
Streams	Euclidean distance (m) to flowing water	Visitors spend more time near flowing water and these features themselves are hazards.
Open Water	Euclidean distance (m) to water bodies	Visitors spend more time near lakes and ponds and these features themselves are hazards.
Roads	Euclidean distance (m) to vehicle routes.	Visitors spend more time near roads where there is more access to amenities.
NDVI	An index of vegetation derived from aerial photography.	Visitors are more likely to encounter obstacles to their travel in vegetated areas when off-trail.

Table 2.3

			Percent contribution							
	Accuracy	N (2011)	Trails	Trailhead	Slope	Elevation	Streams	Roads	NDVI	Open Water
Overall	0.91	186	68.8	16.9	3.1	1.5	1.4	4.7	1.6	1.9
January	0.67	6	24.7	28.8	2.8	2.2	0.0	35.3	0.0	6.2
February	0.75	4	21.4	3.8	2.1	5.1	0.4	64.0	0.2	3.1
March	n/a	0	44.7	2.8	2.4	3.2	1.7	44.0	0.0	1.2
April	1.00	6	54.0	14.7	6.1	3.9	0.5	17.3	0.8	2.8
May	1.00	15	58.4	22.8	2.5	5.7	0.8	8.1	0.4	1.3
June	0.97	30	70.7	12.6	3.0	2.4	1.2	6.5	2.3	1.3
July	0.83	42	74.0	11.7	2.1	0.7	2.3	3.2	1.0	4.8
August	0.95	39	76.4	9.4	3.5	2.0	3.3	1.3	0.7	3.3
September	0.93	29	75.1	9.4	8.7	0.5	0.7	2.5	0.8	2.3
October	0.85	13	51.0	24.0	9.5	8.1	0.2	3.5	0.7	3.1
November	1.00	1	52.3	22.0	10.6	0.5	3.0	10.6	0.8	5.8
December	1.00	1	42.8	33.3	3.2	3.0	0.3	42.8	2.5	4.8

List of Figures

Figure 2.9 A reference map of Yosemite National Park with general location map as an inset.

Figure 2.10 Distribution of search and rescue incident likelihood in Yosemite National Park for all seasons. Left: probabilistic suitability map created from 2001–2010 incidents and environmental layers using the presence and background learning algorithm; right: binary suitability map created by applying threshold to probabilistic map. Overall, 91% of 2011 incidents fell on pixels classified as likely (1).

Figure 2.11 Side by side comparison of the monthly distribution of search and rescue incident likelihood in Yosemite National Park for April (left) and August (right).

Figure 2.9

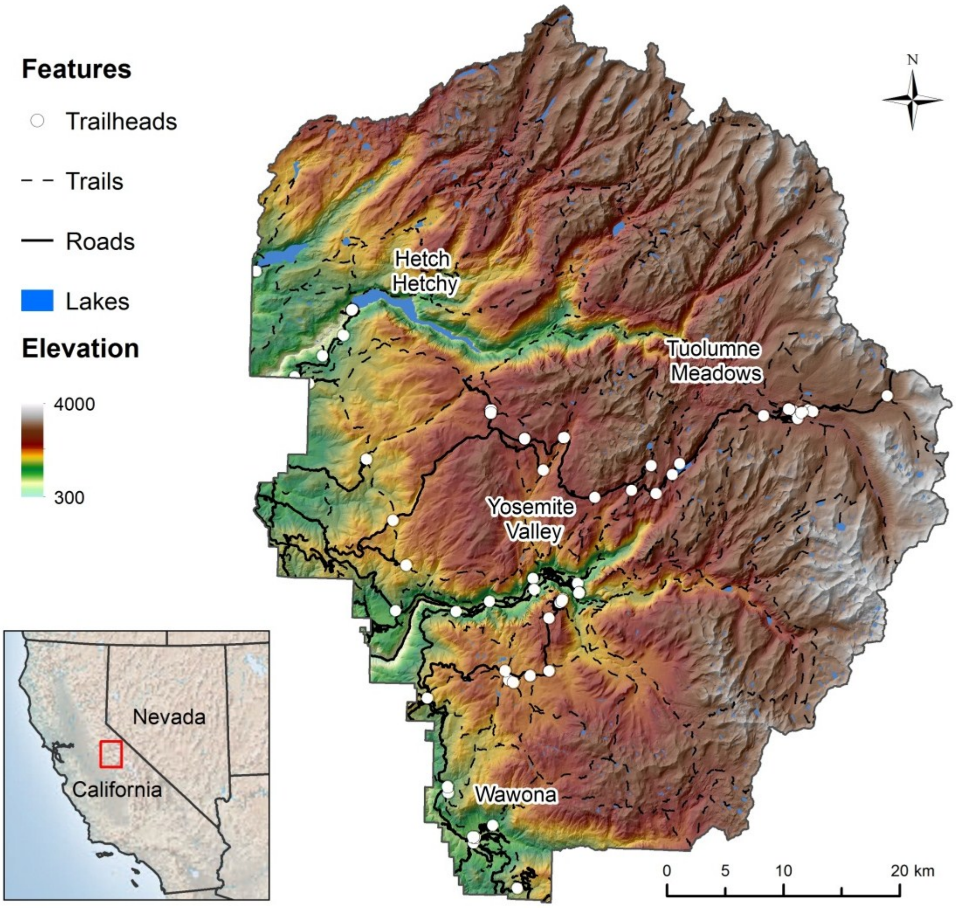
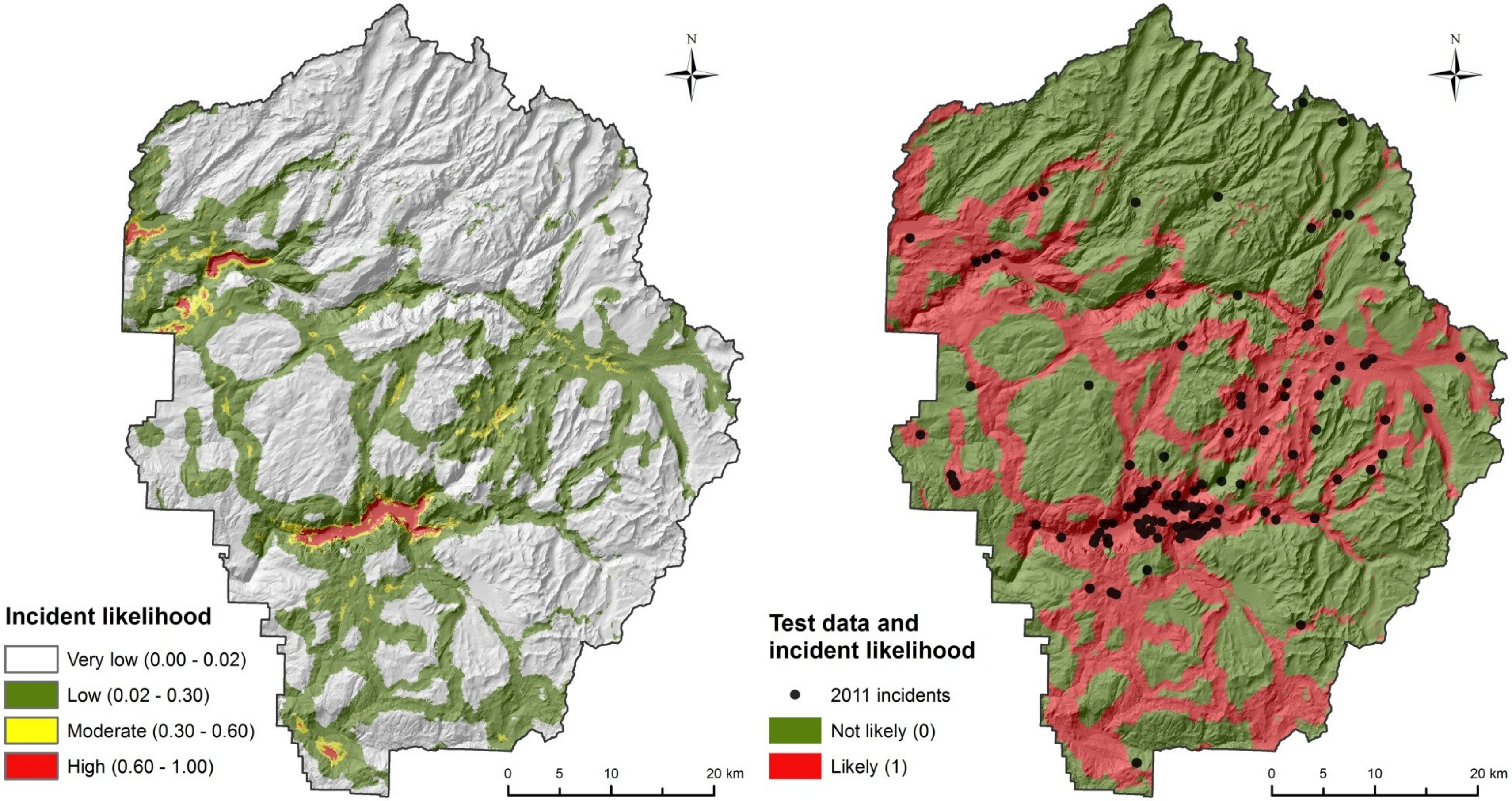
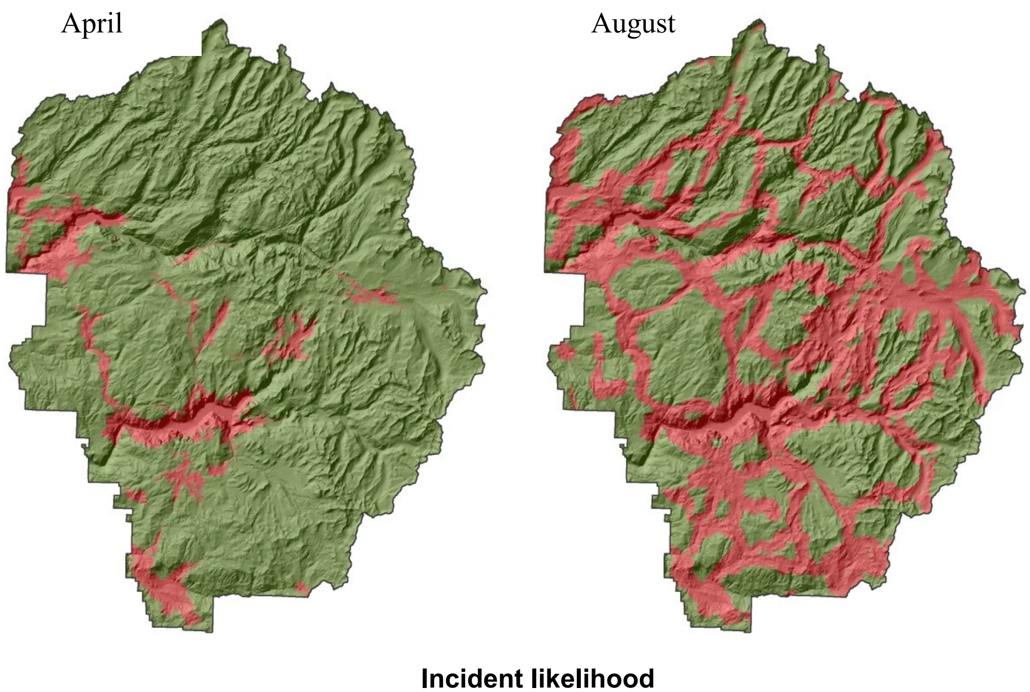

Features
Trailheads
Trails
Roads
Lakes
Elevation
4000
300
Hetch Hetchy
Tuolumne Meadows
Yosemite Valley
Wawona
N
0 5 10 20 km
Nevada
California

Figure 2.10



N
Incident likelihood
Very low (0.00 - 0.02)
Low (0.02 - 0.30)
Moderate (0.30 - 0.60)
High (0.60 - 1.00)
0 5 10 20 km
N
Test data and incident likelihood
2011 incidents
Not likely (0)
Likely (1)
0 5 10 20 km

Figure 2.11



Chapter III: Search

A Review of Probability of Area Techniques for Missing Persons in Yosemite National Park.

[Formatted for submission to *Applied Geography*]

The search for a missing person is assisted through the application of geographic knowledge from past events as general terrain features and subject demographics have been shown to prejudice locations where individuals are ultimately found. In this study, our objectives are to compare two complementary distance-based methods for allocating geographic probability of area in formal search theory and provide practical guidance for use and suggest future research direction. The first method uses historical Euclidean distance statistics to describe probability of where a person may be found based on where they were last seen, or known to be, resulting in a series of expanding rings. The second is a travel-cost model that accounts for the influence of anthropogenic and landscape features on subject mobility and travel time. To evaluate these methods we use actual missing person case data from years 2000 – 2010 for Yosemite National Park. The influence of localized terrain features and subject demographics are considered by comparing the Yosemite data to a large pool of internationally compiled cases consisting of similar subject profiles.

1 Introduction

The process of searching for and rescuing individuals who become lost, injured, or stranded in wilderness areas presents a unique geographic problem which provides a novel and largely unexplored testing ground for the spatial sciences. In fact, the most important question asked in wildland search and rescue (WiSAR) is “where is the missing person most likely to be right now?”, a classic time-geography (Winter and Yin 2010) challenge with real-world implications. Our goal is to help searchers look in the right place in order to find missing subjects more quickly. Here we present and evaluate two distance based elements of formal search theory: the Ring Model and the Mobility Model. We use missing person case studies that occurred in Yosemite National Park from 2000 – 2010 due to the high frequency of missing person cases in the area. The purpose of this research is to answer practical questions regarding the strengths and limitations of these methods using applied geographic techniques. The central research question is: can a globally derived dataset be used to predict the outcome of search incidents on a local scale using two commonly used models? This is the first study to apply GISystem techniques to analyze actual search incident outcomes in this manner.

1.1 Probability of area

In search incidents, time is a critical variable in subject survival, with survivability declining significantly after the first 51-hours of a search (Adams *et al.* 2007). Therefore search managers commonly rely on theory outlined in training manuals, largely based on *Lost Person Behavior* (Koester 2008), to quickly develop search strategies. These concepts are grounded in Operations Research that were initially developed with maritime and aviation search techniques in mind (Koopman 1980, Stone 1989, 2007, Koester *et al.* 2004). The formal study of search and rescue as a discipline grew out of military operations conducted during World War II. The mathematical theory of how to search for missing, lost, and hidden objects was used to search for enemy submarines as well as to recover lost allied ships and downed pilots (Frost 1999). These initial concepts have been adapted to the nuances of ground based search operations by Koester and others.

In WiSAR, grid-based searches are often not possible due to terrain features and the limited number of resources available. Based on this, the first step in allocating resources to find these missing persons is to begin searching the immediate vicinity of the places they were last seen (Point Last Seen - PLS) or last known to be (Last Known Point - LKP) based on the availability of substantial evidence. One of these locations is typically used as the Initial Planning Point (IPP) which is relevant in the application of standardized planning strategies such as those being discussed in this study. These strategies assist in defining the general extent of the search area which then must be divided into feature-based polygons known as planning areas where experts use their collective knowledge to prioritize resource allocation. Search planners will use previous experience from historic search operations and investigation techniques (e.g. clues, witness statements, behavioral profile) to further break down planning areas into searchable segments that can be prioritized and assigned to search teams. In larger searches, specialized software and consensus techniques such as the Mattson Consensus are used to assign quantitative values to segments based on combined scores from individuals involved in the consensus process (Bownds *et al.* 2007). Throughout this entire process, it is imperative that all of the geographic information be compiled in a way that allows search managers to easily access and understand it.

The search planning process incorporates several probabilistic concepts: Probability of Area (POA), Probability of Detection (POD), and Probability of Success (POS) as described in equation 1 below.

$$POS = POA \ x \ POD \tag{1}$$

The Probability of Success (POS) in search theory is completely dependent upon the boundary o f the area (polygon) being searched actually containing the missing person (POA) and an accurate assessment of how well a search area was covered by a team (POD). POD can be explained using the following equation.

$$POD = 1 - e^{-c} \tag{2}$$

Where c is the coverage and e is the base of the natural logarithm. Coverage is the ratio of two areas: search effort as track length multiplied by width (the polyline representing searchers movements and the buffer indicating where they can effectively search), divided by the total area searched. In search planning like most complex decision making processes, errors in judgment (underestimated POA, overestimated POD) early in the planning stages significantly hampers the search effort despite subsequent decisions. The objective of any search operation is to maximize POS as quickly as possible by increasing POA and POD. At this time, most WiSAR operations utilize these concepts in a variety of ways but with little geospatial resources to accurately measure either POA or POD. With the integration of Global Position Systems (GPS) for searchers, a quantitative index of POD can be obtained by measuring coverage based on searchers' GPS receiver track-log, but the POA is still very much theoretical.

Considerable effort in missing person research has been dedicated to analyzing lost person behavior and summaries of previous search incident outcomes to generate a POA value (Syrotuck 1977; R. J. Koester and Stooksbury 1995) with the current standard being the International Search and Rescue Incident Database (ISRID; R. Koester 2008). ISRID provides an excellent tool for comparing incident outcomes based on the missing person profile (e.g. age, gender, mental status) and activity (e.g. hiking, biking, gathering). Not all of these cases are truly lost persons; in fact many have been injured or suffered a medical condition but the profile comparison between missing person subject categories helps search planners make informed decisions based on historical accounts. Many factors have been evaluated with respect to the ISRID dataset and search theory, including direction of travel, elevation change, offset from linear features, and the type of habitat they are found in. These were also evaluated in a comprehensive manner for a localized dataset (Doke 2012) but for this research we focused on distance-based POA methods. The two most similar distance-based POA methods typically used in search operations are the Ring Model and Mobility Model.

1.2 Ring model

The Ring Model uses concentric circles or rings to delineate areas of probability based on distance. The probability rings (**Figure 3.1**, often called "crows flight distance") are quartile and 95% summaries for the Euclidean distance between the coordinates for missing persons' IPP locations (defined in 1.1) and the location they were found (Syrotuck 1976). This methodology can be applied to any search operation dataset where IPP and Found locations are available and Koester (2008) summarizes the results for hikers based on data from ISRID. This database and summary tables have known limitations in their use as a search planning tool because of their lack of GISystem integration and the risk of inferring global statistics onto local landscapes with extreme terrain features. Even though the ISRID separates lost person behavior by ecoregion it is unexpected that a hiker in rolling topography in Great Britain would behave the same as hiker on a mountainous trail in the Sierra Nevada of the United States. In this study we will compare the results of the Ring Model derived from a local dataset (Yosemite) to the global dataset (ISRID).

1.3 *Mobility model*

Mobility models have been proposed as a POA method, however to date, little GISystem integration has been employed in this process. Models of this type attempt to estimate subject travel based on a most likely path or period of mobility. Data concerning how long a subject will remain in motion is summarized in the ISRID for cases where this information is known, but applying this information can only be grossly estimated without modern GISystem techniques. Lin and Goodrich (2010) proposed a Bayesian model to automatically generate a probability distribution map from publicly available terrain feature data including topology, vegetation, and elevation data. This was the first study to incorporate GIScience techniques to evaluate mobility models for missing person search operations. However, this forecasting methodology requires actual path polyline data from previous incidents which is not readily available for most areas (if a person is found alive they still may not know exactly where they went without a GPS track log). Yet this does indicate a viable exploration into the use of terrain based models for developing mobility models for the purpose of estimating POA. In a GISystem, this is known as cost-distance modeling and has been used extensively for infrastructure development (Van den Broek *et al.* 2010), wildlife habitat analysis (Nikolakaki 2004), and anthropological studies (Whitley and Hicks 2003). In these models, costs are calculated by applying a least-cost path algorithm to a source raster and a resistance raster (Adriaensen *et al.* 2003). Tobler (1965) was the first to use the Imhof (1950) ''hiking function'' (equation 3) to calculate the cost associated with traversing a landscape. The ''hiking function'' estimates the velocity of travel for hikers across different slopes. According to Tobler (1991) in his overview of non-isotropic modeling between pedestrian movement and slope, travel speed can be estimated as:

$$v_s = 6e^{-3.5\, abs\, (Slope+0.05)} \tag{3}$$

Where *e* is the natural logarithm and v_s is pedestrian velocity defined by a mathematical function based on the slope of terrain in degrees. This yields a function that is asymmetric about zero slopes because it is generally faster to travel down hills than up hills. The effect of land cover on off-trail hiking velocity has been estimated by Tobler to be a reduction factor of 0.6***x***. However, this estimate can be refined by local knowledge of land cover and travel speeds as described in section 2.3 to produce a calibrated resistance raster that provides a predicted pedestrian velocity across a standard distance. Land cover impedance can be described as the area between the curve for Tobler's on and off-trail functions. This calibration of the impedance grid will help avoid using crude oversimplifications for slope and land cover impedance that often reduces the usefulness of such models (Bateman *et al.* 1996). If these methods represent actual pedestrian capabilities the impedance raster grid can then be used for travel-cost simulation where distance traveled is a function of estimated travel time (time since last seen), the coordinates of the IPP, and the impedance values of surrounding raster grids. The product would be a minimum potential path area.

Materials and methods

2.1 Georeferencing

The data obtained for this analysis were derived from Yosemite National Park's Search and Rescue Case Incident reports for the years 2000 to 2010. Access to these reports was granted by the National Park Service Division of Visitor Protection (Permit 1024-0236). During this eleven year span, Yosemite National Park responded to 2,308 total SAR Incidents. This includes both genuine searches for lost people as well as rescues in which the actual location of the individual was known. Of these SAR incidents, 2201 incident reports were available for review. Out of these reports, 393 true search incidents were identified. Each search incident was critically analyzed and must have met all of the following criteria in order to be retained for the current study:

1. The incident must have been a ground-based search incident. Searches for downed aircraft or searches involving bodies of water were not included.
2. The incident must have had a distinct PLS/LKP/IPP that can be georeferenced within the Yosemite National Park boundary.
3. The incident must have had a distinct found location that can be georeferenced within, or within walking distance of, the Yosemite National Park boundary.
4. An official SAR response must have been initiated by the National Park Service.

A total of 213 search incidents met these criteria and incidents were georeferenced from text based information as part of GIScience research described in Doherty et al. (2011) using the point-radius method (Wieczorek *et al.* 2004b).

2.2 Ring Model

We used the georeferenced locations to determine the Euclidean distance (*D*) between the IPP to the found location in kilometers, and then calculated the lower quartile, median, upper quartile, and 95th percentile of *D* for the hiker category. This category was the only one with a large enough sample to derive reliable statistics and is of the most interest to search managers because it is the most common missing person profile in Yosemite. This observed distribution of *D* for Yosemite was then compared with the expected distribution of D based on ISRID using a Chi-square Goodness of Fit Test with a significance level of 0.05. Standard geoprocessing and calculation of *D* was done using ArcGIS 10.1 (Esri 2012).

2.3 Mobility Model

The Mobility Model used in this study is based on the Travel Time Cost Surface Model (TTCSM) used by the National Park Service (Sherrill *et al.* 2010). The model provides an estimate of travel time using readily available geospatial products such as road, trail, and stream networks, digital elevation models and land cover data. The model consists of two basic components: speed surface and cost surface. First a simple speed surface based on Tobler's Hiking Function is defined using Equation 3 based on the slope as measured using the *Slope* function within ESRI ArcMap 10.1. This speed surface does not account for energy expenditure and assumes the subject would be able to maintain a nominal speed across a flat surface (slope = 0^{o}) of 5.0 kilometers per hour (kph).

The cost surface is defined as a function of the impedance to foot traffic that would be imposed by the presence of various geographical features compared to a nominal paved surface such as a side walk or roadway. Impedance values range from 0 – 100%, with 0% being no impedance (e.g. paved roadway or well-maintained trail) and 100% being absolute impedance (e.g. high fenceline or large body of water). A value of 25 would represent a feature that was 25% slower to cross (e.g. pasture / field) than the nominal surface (paved roadway).

The impedance raster is formulated using both vector and raster data layers. Vector layers include roads, trails, utility right-of-ways, fencelines (actual and virtual), streams and bodies of water. While maintained roadways offered 0% impedance, the maintenance level of the trail as recorded in the feature attribute table was used to assign impedance values with poorly maintained or unmaintained trails given a higher impedance value. A combination of hydrological vector data and DEM derived features are used to represent streams. This combination permits access to an estimate of stream order (Strahler) to account for the influence of the size of the stream on foot travel. The higher the stream order the greater the impedance.

Large bodies of water (lakes, ponds, etc) were considered absolute barriers since only foot traffic was being considered. Once these vector data was reclassified for impedance it was converted to a raster layer.

Slope derived from 10 meter digital elevation model (DEM) was not only used to define the speed surface across the entire area of interest but also for identifying locations with slope greater than 60^{o} which was classified as absolute barriers to foot travel. The underlying land cover was derived from the National Land Cover Dataset (2006) with features being reclassified similar to costs defined in Sherrill (Sherrill *et al.* classify cost with respect to Percent of Maximum Travel Speed – PMTS, where the model used in this study is defined as impedance, 1/PMTS).

The final impedance raster is compiled using a hierarchical overlay method where factors are listed in decreasing priority: fenceline, road, trail, impassable slope, bodies of water, streams (Strahler order), utility right-of-ways and land cover. This impedance layer is then combined with the speed layer to obtain a travel speed cost surface (km per hour) for the area of interest with foot travel speeds ranging from slightly above the nominal speed (maximum speed at a slope of -0.05) to zero. The inverse of the cost surface is used as input for the Path Distance function in ArcGIS 10.1. Starting at the IPP, the accumulated cost to travel across the cost surface is evaluated in an anisotropic manner in order to account for travelling up or downslope. While the Speed Surface has slope as the primary input, the resultant surface is isotropic as the direction of travel is yet unknown at the time the surface is tabulated. The Path Distance tool within ArcGIS provides a means of accounting for direction of travel across the Cost Surface which includes slope and travel restrictions. The resultant product is a travel time surface suggesting the minimal time required to reach a destination starting from the IPP.

The cost surface for Yosemite was used for analyzing individual search incident cases described in section 2.1. To begin, the IPP for a search incident is used to seed the travel cost surface to produce isochrones (**Figure 3.2**). The point found location is then sampled for the intersecting isochrones. This value is recorded and entered into a table for further analysis. This process was then run in batch mode for all search incidents for the hiker category. The isochrones value is essentially the predicted minimum amount of time (T_{min}) for the subject to reach the find location based on the cost surface. The model is not intended to take into account the fact that a person may wander, stay in one place, or leave an area and return to it but rather be an empirical model based on physical capabilities. Nor does it account for energy expenditure, as previously noted, which could potentially increase T_{min}. We summarized hiker mobility statistics in a format that could be directly compared to the ISRID as defined by Koester (2008). The lower quartile, median, upper quartile, and 95th percentile were calculated for each for the hiker category and T_{min} values are compared to the global dataset compiled in the ISRID using a Chi-square Goodness of Fit Test. We statistically compared the ring and mobility models using bivariate correlation of D and T_{min} for each hiker search incident.

3 Results and Discussion

3.1 Ring model

Of the 213 incidents georeferenced, 129 were cases involving lost hikers in Yosemite National Park and we included 130 lost hikers in the analysis since one group of two hikers were separated and were found in different locations at different times. Overall, the Euclidean distance (D) from each georeferenced IPP to its corresponding georeferenced found location was calculated (n = 130, mean = 3.34 km ± 0.25). When the local Yosemite hiker sample was compared to the global ISRID data, there were statistically significant differences (**Table 3.1;** n = 130, $X^2 = 15.4$, $P < 0.01$). This means that the distances that hikers travel, as the crow flies, from the IPP to the found location in Yosemite was significantly different than the same corresponding distances that hikers travel at the global scale. The values in the first quartile (25%) were equal, but were greater in the ISRID for 50%, 75%, and 95% values.

This would suggest that relying solely on the ISRID Ring Model may lead searchers to plan a larger POA than expected from locally derived data. This is not a critical error as you would rather slightly overestimate POA rather than underestimate and risk drawing a boundary that excluded the missing person. However, deriving ring models based on local data may refine POA techniques and help to better allocate resources. Once data has been collected and formatted, the ring model, is simple to apply using a GISystem (or even hand-drawn on a paper map) and can be considered a good starting point for defining a crude search boundary in the first minutes of a search. In all cases, we suggest that ring models should not be used alone without considering other sources of information such as mobility models discussed below. Overall, the ring model still provides an adequate starting point for areas that do not have georeferenced datasets and it is important that the 25% quartiles were similar since these should be the highest priority in the early phases of a search.

3.2 Mobility model

The average T_{min} value for the 130 missing hikers (T_{min} =1.5 hours) and the median (T_{min} = 1.0) were widely separated. In addition when compared to the ISRID, the Yosemite dataset differed significantly (**Table 3.2;** n = 130 $X^2 = 91.4$, $P < 0.01$). These differences may be due to differences in the way data were collected. In ISRID mobility values were taken from report narratives directly, whereas Yosemite T_{min} was derived from a modeling process (actual mobility values were not found in reports). The ISRID data is inherently biased in that mobility values were estimated and only available from hikers who were found alive (the victims found deceased on arrival cannot provide this information unless a GPS recorded such data). This suggests that users of the travel cost methods to produce T_{min} should not simply use mobility values from the ISRID mobility model to produce potential path area, instead local data should be summarized as described in this research. Alternatively, if no local historic data is available for mobility model analysis as in this study, isochrones can be evaluated in conjunction with investigative information (e.g. the amount of time a person has been missing, the time elapsed before a major snow storm would cease mobility). Our local analysis shows that in Yosemite, if we run a travel cost analysis, 50% of hikers should be found in the 1.0 hour isochrone. This means that prioritizing the search of a relatively small area close to the IPP should help resolve around half of all incidents. In addition, 95% of all hikers were found within the 6.0 hour isochrone, which should help make decisions regarding search boundaries and containment. Overall, the travel cost model does present a useful GISystem tool for visualizing probability areas and can be created using readily available base data.

3.3 Comparison of the ring and mobility model

When the ring model and mobility models were compared statistically, we found a significant correlation between D and T_{min} (**Figure 3.3**; n = 130, $R^2 = 0.96$, $P < 0.01$). This relationship between distance and time is expected, but the scatterplot (Figure 3) shows that there is a clear limit to the Euclidean distance (D) as a function of T_{min}. If we divide D by T_{min} for each incident, it yields an average value of 2.02 km / hour and maximum of 3.75 km / hour. This can most likely be explained by two factors: the physical limitations of hiking across varied terrain and the behavior of hikers. From a human geography standpoint, movement is limited by terrain and is represented by the travel cost model. D / T_{min} will be greatest and approach a maximum where the line between IPP and the location found is across flat terrain and on a trail. Conversely, D / T_{min} will be minimal when the line between IPP and location found is across steep terrain and off-trail. When we consider hiking behavior, D / T_{min} is maximized by a straight line and minimized by a circuitous path that closes in on itself over increasing elevation. We would expect the greatest variation of D / T_{min} in environments with varied terrain and circuitous trails. This is an important geographic phenomenon to understand relative to the use of the ring and mobility models. From a practical standpoint, in mountainous terrain the mobility model will provide greatest contribution of information, whereas in flat terrain with no vegetation the mobility model and ring model would be functionally identical.

The mobility model technique provides a more realistic tool for search boundary delineation because it is based on terrain and human geography. Essentially, if we know the point last seen, time of last sighting, and possible directions of travel, then we can create a potential path area for missing persons using a field-based approach to represent isochrones over the search area. Similar to the mobility models produced in this study, lines of equal travel time (isochrones) are used for studying accessibility in the city of Glasgow (O'Sullivan et al. 2000) and demonstrate a useful tool for visualizing space-time geography. Space-time accessibility is a field of research that is well studied in urban environments by geographers and GIScientists (Kim and Kwan 2003) and a similar approach has been suggested by other researchers (Miller and Bridwell 2009, Lin and Goodrich 2010) for use in WiSAR, yet little research has been done to expand this concept and the current study is the first of its kind to evaluate travel cost techniques for missing person mobility models using detailed accounts.

It is important to denote that travel cost techniques do not provide predictive models as they do not take into account the behavior of a person (i.e. wandering, returning to IPP, resting) but they do provide some guidance based on information that is known: the landscape the missing person is in and the amount of time they have been missing. Moreover, this model produces isochrones that allow for scenario building if provided in an interactive GISystem. For instance, if we have reason to believe a missing person would only be in motion for x hours (e.g. due to physical limitations or the arrival of a winter storm that would limit movement), we can indicate suggested probability area as indicated on the map by the isochrones equal to x hours. Mobility can be limited by other factors such as incoming weather, darkness, or the behavior of the person themselves (e.g. mental status). As suggested by Pingel (2013), we believe that travel costs methods can underestimate the cost of travel in hilly and mountainous terrain because these methods do not take into account human perceptions of slope and elevation change (Yang et al. 1999). However, this is an acceptable bias for developing POA in this use-case as we would rather overestimate than underestimate travel speeds. By using local data as described in this study we can verify the effect of this bias and future research should focus on agent-based techniques to calibrate these models. In the meantime, travel cost modeling is a tool with visually compelling results that can be used in conjunction with the ring model and other elements of search theory using GISystems. We demonstrate this with a historic case where these techniques could have been used to help develop a more well-informed POA if they were available to search planners at the time (**Figure 3.4**).

From this evaluation of the ring model and mobility model, we observed that other datasets can be used in a mountainous area such as Yosemite for forming functional planning areas and delineating the search boundary. One example are watershed boundaries derived from Federal Standards and Procedures for the National Watershed Boundary Dataset (WBD) 2012. Preliminary investigation (Doke 2012) suggests that missing persons rarely are found in watersheds further away than the one in which they were last seen. This fits well within the findings of our research since watersheds are determined by terrain models and are essentially bordered by ridgelines that greatly influence mobility models using travel cost techniques. Further research should be conducted to determine the usefulness of watersheds in search area planning since it is a readily available GIS dataset.

4 Conclusions

We have found that developing ring and mobility models from local data will give the best estimate of POA, but that in absence of these data global datasets will provide distance based models that can be used given one understands the constraints in mountainous terrain (a potentially larger POA than would be explained by local data). Overall, there is a great need for more research in the time geography of missing person searches. Efforts should be focused on georeferencing incidents up front in a records management system and then properly analyzing the data on a local level. Then software tools based on GISystems can be developed for assisting search managers. While we can learn from a global dataset, local information should be used for actually deriving meaningful probability of area and delineating search boundaries. Once more detailed behavioral data can be obtained, the ring model and mobility model concepts can be combined using more advanced GIScience techniques. In the meantime, the authors have begun integrating POA tools into desktop GISystem software to learn more about how these concepts apply to actual search incidents. We have identified techniques here that can be used for improving the search for missing persons; however it will require a community of experts in the field of geography working in conjunction with search and rescue personnel to answer the fundamental question: “where can the missing person be right now?”

References

Adams, A.L., Schmidt, T. a, Newgard, C.D., Federiuk, C.S., Christie, M., Scorvo, S., and DeFreest, M., 2007. Search is a time-critical event: when search and rescue missions may become futile. *Wilderness & Environmental Medicine*, 18 (2), 95–101.

Adriaensen, F., Chardon, J.P., De Blust, G., Swinnen, E., Villable, S., Gulinck, H., and Matthysen, E., 2003. The application of "least-cost" modelling as a functional landscape model. *Landscape and Urban Planning*, 64 (4), 233–247.

Alexander, D., 1991. Information technology in real-time for monitoring and managing natural. *Progress in Physical Geography*, 15 (3), 238.

An, L. and Brown, D.G., 2008. Survival Analysis in Land Change Science : Integrating with GIScience to Address Temporal, (February 2013), 37–41.

Baldwin, R. a., 2009. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy*, 11 (4), 854–866.

Bateman, I.J., Garrod, G.D., Brainard, J.S., and Lovett, A. a., 1996. Measurement Issues in the Travel Cost Method: a Geographical Information Systems Approach. *Journal of Agricultural Economics*, 47 (1-4), 191–205.

Berger, A., Della Pietra, S., and Della Pietra, V., 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22 (1992), 39–71.

Bownds, J.W., Ebersole, M.J., Lovelock, D., O'Connor, D.J., and Toman, R.J., 2007. Win CASIE III: Computer Aided Search Information Exchange.

Van den Broek, M., Brederode, E., Ramírez, A., Kramers, L., Van der Kuip, M., Wildenborg, T., Turkenburg, W., and Faaij, A., 2010. Designing a cost-effective CO2 storage infrastructure using a GIS based linear optimization energy model. *Environmental Modelling & Software*, 25 (12), 1754–1768.

Church, R. and ReVelle, C., 1974. The maximal covering location problem. *Papers in regional science*, 32 (1), 101–118.

Clark, W., 1965. Markov chain analysis in geography: an application to the movement of rental housing areas. *Annals of the Association of American …*, 55 (2), 351 – 359.

Cutter, S.L., 2003. GI Science, Disasters, and Emergency Management. *Transactions in GIS*, 7 (4), 439– 445.

Dijkstra, E.W., 1959. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1, 269–271.

Doherty, P., Ferguson, D., Goodrich, M.A., Koester, R.J., and Doke, J., 2012. Wilderness Search & Rescue and GIScience. *In*: *Annual Meeting of the Association of American Geographers*. New York.

Doherty, P., Guo, Q., Liu, Y., Wieczorek, J., and Doke, J., 2011. Georeferencing Incidents from Locality Descriptions and its Applications: a Case Study from Yosemite National Park Search and Rescue. *Transactions in GIS*, 15 (6), 775–793.

Doherty, P.J., Guo, Q., and Alvarez, O., 2012. Expert versus machine: A comparison of two suitability models for emergency helicopter landing areas in Yosemite National Park. *Professional Geographer*.

Doherty, P.J., Guo, Q., Li, W., and Doke, J., n.d. Space-Time analyses for forecasting and understanding future incident occurrence: a case-study from Yosemite National Park using the presence and background learning algorithm. *International Journal of Geographical Information Science*.

Doke, J., 2012. Analysis of Search Incidents and Lost Person Behavior in Yosemite National Park.

Durkee, G., 2010. GIS Joins Search for a Missing Hiker on California's Mount Whitney. *ArcWatch*, Apr.

Esri, 2012. ArcGIS 10.1.

Federal Standards and Procedures for the National Watershed Boundary Dataset (WBD), 2012. Reston, Virginia.

Ferguson, D., 2008. GIS for Wilderness Search and Rescue. *In*: *Esri Federal User Conference*. Washington D.C., 1 – 11.

Fernandez, M.A., Blum, S.D., Reichle, S., Holzman, B., and Hamilton, H., 2009. Locality uncertainty and differential performance of four different common niche-modeling techniques. *Biodiversity Informatics*, 6 (1), 36–52.

Fielding, A.H. and Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24 (1), 38–49.

Filipkowska, E., Koester, R.J., Chrustek, R., and Zaród, M., 2012. Lost and Found in the Polish Carpathian Mountains. *ArcNews*, 34 (3).

Fink, D. and Hochachka, W., 2010. Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications*, 20 (8), 2131–47.

Forrester, J.D. and Holstege, C.P., 2009. Injury and illness encountered in Shenandoah National Park. *Wilderness & environmental medicine*, 20 (4), 318–26.

Frost, J.R., 1999. Principles of search theory. *Response*, 17 (2), 1 – 23.

Getis, A. and Ord, J.K., 1992. The Analysis of Spatial Association. *Geographical Analysis*, 24 (3).

Goodchild, M., 1992. Geographical information science. *International Journal of Geographical Information …*, 6 (March 2013), 37–41.

Goodchild, M.F., 2003. Geospatial data in emergencies. *In*: S.L. Cutter, D.B. Richardson, and T.J. Wilbanks, eds. *The Geographical Dimensions of Terrorism*. New York: Routeledge, 99–104.

Goodchild, M.F., 2004. GIScience, Geography, Form, and Process. *Annals of the Association of American Geographers*, 94 (4), 709–714.

Green, P.J. and Richardson, S., 2002. Hidden Markov Models and Disease Mapping. *Journal of the American Statistical Association*, 97 (460), 1055–1070.

Guo, Q., Li, W., Liu, Y., and Tong, D., 2011a. Predicting potential distributions of geographic events using one-class data: concepts and methods. *International Journal of Geographical Information Science*, 25 (10), 1697–1715.

Guo, Q., Li, W., Liu, Y., and Tong, D., 2011b. Predicting potential distributions of geographic events using one-class data: concepts and methods. *International Journal of Geographical Information Science*, 25 (10), 1697–1715.

Guo, Q., Liu, Y., and Wieczorek, J., 2008. Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *International Journal of Geographical Information Science*, 22 (10), 1067–1090.

Heggie, T.W. and Amundson, M.E., 2009. Dead men walking: search and rescue in US National Parks. *Wilderness & environmental medicine*.

Heggie, T.W. and Heggie, T.M., 2009. Search and rescue trends associated with recreational travel in US national parks. *Journal of Travel Medicine*, 16 (1), 23–7.

Hill, K. and Gale, R., 1997. *Managing the lost person incident*. Managing. Chantilly, VA: National Association for Search and Rescue.

Hirzel, A.H., Hausser, J., Chessel, D., and Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, 83 (7), 2027–2036.

Hogg, J.M., 1968. The Siting of Fire Stations. *Journal of the Operational Research Society*, 19, 275–287.

Huang, B., Wu, B., and Barry, M., 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24 (3), 383–401.

Hung, E.K. and Townes, D. a, 2007a. Search and rescue in Yosemite National Park: a 10-year review. *Wilderness & environmental medicine*, 18 (2), 111–6.

Hung, E.K. and Townes, D. a, 2007b. Search and rescue in Yosemite National Park: a 10-year review. *Wilderness & Environmental Medicine*, 18 (2), 111–6.

Imhof, E., 1950. *Gelaende und Karte*. Zurich, Switzerland: Rentsch.

Indriasari, V., Mahmud, A.R., Ahmad, N., and Shariff, A.R.M., 2010. Maximal service area problem for optimal siting of emergency facilities. *International Journal of Geographical Information Science*, 24 (2), 213–230.

Jaynes, E.T., 1990. Notes on present status and future prospects. *In*: W.. Grandy and L.H. Schick, eds. *Maximum entropy and Bayesian methods*. Dodrecht: Kluwer.

Kim, H.-M. and Kwan, M.-P., 2003. Space-time accessibility measures: A geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration. *Journal of Geographical Systems*, 5 (1), 71–91.

Koch, T. and Denike, K., 2009. Crediting his critics' concerns: remaking John Snow's map of Broad Street cholera, 1854. *Social science & medicine (1982)*, 69 (8), 1246–51.

Koester, R., 2008. *Lost Person Behavior*. Search. Charlottesville, VA: dbS Productions.

Koester, R.J., Cooper, D.C., Frost, J.R., and Robe, R.Q., 2004. Sweep Width Estimation for Ground Search and Rescue. Washington.

Koester, R.J. and Stooksbury, D.E., 1995. Behavioral profile of possible Alzheimer's disease patients in Virginia search and rescue incidents. *Wilderness & Environmental Medicine*, 6, 34–43.

Koopman, B.O., 1980. *Search and Screening*. New Yorrk: Pergamon Press.

Kratzke, T.M., Stone, L.D., and Frost, J.R., 2010. Search and Rescue Optimal Planning System (SAROPS). *In*: *13th Conference on Information Fusion*. Edinburgh, Scotland: IEEE, 1 – 8.

Kwan, M.-P., 2003. Intelligent emergency response systems. *In*: S.L. Cutter, D.B. Richardson, and T.J. Wilbanks, eds. *The Geographical Dimensions of Terrorism*. New York: Routeledge, 111–116.

Lawson, A., Gangnon, R., and Wartenberg, D., 2006. Developments in disease cluster detection. *Statistics in Medicine*, 25 (5), 721.

Li, W. and Guo, Q., 2010. A maximum entropy approach to one-class classification of remote sensing imagery. *International Journal of Remote Sensing*, 31 (8), 2227–2235.

Li, W., Guo, Q., and Elkan, C., 2011. Can we model the probability of presence of species without absence data? *Ecography*, 34 (6), 1096–1105.

Li, X. and Yeh, A., 2005. Integration of genetic algorithms and GIS for optimal location search. *International Journal of Geographical Information Science*, 19 (5), 581–601.

Lin, L. and Goodrich, M. a., 2010. A Bayesian approach to modeling lost person behaviors based on terrain features in Wilderness Search and Rescue. *Computational and Mathematical Organization Theory*, 16 (3), 300–323.

Liu, D. and Cai, S., 2011. A Spatial-Temporal Modeling Approach to Reconstructing Land-Cover Change Trajectories from A Spatial-Temporal Modeling Approach to Reconstructing Land-Cover Change Trajectories from Multi-temporal Satellite Imagery. *Annals of the Association of American Geographers*, 102 (6), 1329–1347.

Liu, Y., Guo, Q., Wieczorek, J., and Goodchild, M.F., 2009. Positioning localities based on spatial assertions. *International Journal of Geographical Information Science*, 23 (11), 1471–1501.

Malczewski, J., 2004. GIS-based land-use suitability analysis: a critical overview. *Progress in Planning*, 62 (1), 3–65.

Miller, H., 1996. GIS and geometric representation in facility location problems. *International Journal of Geographical Information …*, 10 (7), 37–41.

Miller, H.J. and Bridwell, S. a., 2009. A Field-Based Theory for Time Geography. *Annals of the Association of American Geographers*, 99 (1), 49–75.

Moffett, A., Shackelford, N., and Sarkar, S., 2007. Malaria in Africa: Vector Species' Niche Models and Relative Risk Maps. *PLoS ONE*, 2 (9), e824.

Montalvo, R., Wingard, D.L., Bracker, M., Davidson, T.M., and Diego, S., 1998. Conferences and Reviews Morbidity and Mortality in the Wilderness. *Wilderness and Environmental Medicine*, 168 (4), 248–254.

Mostern, R. and Johnson, I., 2008. From named place to naming event: creating gazetteers for history. *International Journal of Geographical Information Science*, 22 (10), 1091–1108.

Murray, A. and Tong, D., 2007. Coverage optimization in continuous space facility siting. *International Journal of Geographical …*, (March 2013), 37–41.

National Association for Search and Rescue, 2005. *Fundamentals of Search and Rescue*. Sudbury, MA: Jones and Bartlett Publishers.

Nikolakaki, P., 2004. A GIS site-selection process for habitat creation: estimating connectivity of habitat patches. *Landscape and Urban Planning*, 68 (1), 77–94.

Ostfeld, R.S., Glass, G.E., and Keesing, F., 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution*, 20 (6), 328–36.

Pearson, R., Dawson, T., and Liu, C., 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, 3, 285–298.

Phillips, S.J., Dudik, M., and Schapire, R.E., 2004. A maximum entropy approach to species distribution modeling. *In*: *Proceedings of the Twenty-First International Conference on Machine Learning*. Banff, Alberta, CA, 655–662.

Robertson, C., Nelson, T.A., MacNab, Y.C., and Lawson, A.B., 2010. Review of methods for space–time disease surveillance. *Spatial and Spatio-temporal Epidemiology*, 1 (2–3), 105–116.

Schneider, P.J. and Schauer, B.A., 2006. HAZUS—Its Development and Its Future. *Natural Hazards Review*, 7, 40–44.

Shannon, C., 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423.

Sherrill, K.R., Frakes, B., and Schupbach, S., 2010. Travel time cost surface model: standard operating procedure. Natural Resource Report NPS/NRPC/IMD/NRR—2010/238. Fort Collins, Colorado.

Stone, L.D., 1989. What's happened in search theory since the 1975 Lanchester Prize? *Operations Research*, 37 (3), 501.

Stone, L.D., 2007. *Theory of Optimal Search*. 2nd ed. Mathematics in Science and Engineering. New York: Academic Press.

Suárez-Seoane, S., García de la Morena, E.L., Morales Prieto, M.B., Osborne, P.E., and De Juana, E., 2008. Maximum entropy niche-based modelling of seasonal changes in little bustard (Tetrax tetrax) distribution. *Ecological Modelling*, 219, 17–29.

Syrotuck, W.G., 1976. *Analysis of Lost Person Behavior*. Mechanicsburg, PA: Barkleigh Productions, Inc.

Tate, E., Cutter, S.L., and Berry, M., 2010. Integrated multihazard mapping. *Environment and Planning B: Planning and Design*, 37 (4), 646–663.

The Outdoor Foundation, 2012. Outdoor Recreation Participation Report.

Theodore, J., 2009. When every second counts. *ArcNews*, (March), 66 – 69.

Tobler, W., 1965. Non-Isotropic Geographic Modeling. *Information Systems*.

Tobler, W., 1991. Non-Isotropic Geographic Modeling. *In*: W. Tobler, ed. *Geographic Information Systems in the Social Sciences*. Santa Barbara.

Toregas, C., Swain, R., ReVelle, C., and Bergman, L., 1971. The Location of Emergency Service Facilities. *Operations Research*, 19 (6), 1363–1373.

Whitley, T. and Hicks, L., 2003. A geographic information systems approach to understanding potential prehistoric and historic travel corridors. *Southeastern Archaeology*, (those 1994).

Wieczorek, J., Guo, Q., and Hijmans, R., 2004a. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18 (8), 745–767.

Wieczorek, J., Guo, Q., and Hijmans, R., 2004b. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18 (8), 745–767.

Wild, F.J., 2008. Epidemiology of mountain search and rescue operations in Banff, Yoho, and Kootenay National Parks, 2003-06. *Wilderness & Environmental Medicine*, 19 (4), 245–51.

Winter, S. and Yin, Z.-C., 2010. Directed movements in probabilistic time geography. *International Journal of Geographical Information Science*, 24 (9), 1349–1365.

Worsing, R.J., 1993. *Rural Rescue and Emergency Care*. 1st ed. Rosemont, IL: American Academy of Orthopaedic Surgeons.

Yee, K. and Iserson, K. V., 2008. The Epidemiology of Search and Rescue Incidents in the Grand Canyon National Park: Are Preventive Measures Making a Difference? *Western Journal of Emergency Medicine*, 9 (1), 3–5.

Table 3.1 Comparison of Euclidean distances traveled from the IPP to the found location for hikers.

	Yosemite (km)	ISRID (km)
n	130	568
25%	1.1	1.1
50%	1.8	3.1
75%	4.0	5.8
95%	16.9	18.3

Table 3.2 Comparison of mobility values for hikers in Yosemite derived from isochrones versus a global dataset (ISRID) based on reported mobility.

	Yosemite (hours)	ISRID (hours)
n	130	232
25%	0.5	0.0
50%	1.0	3.0
75%	1.5	6.0
95%	6.0	14.0

Figure 3.1 Rings based on Euclidean distance (crow's flight distance) are often plotted from the Initial Planning Point (IPP). These distances correspond to the lower quartile, median, upper quartile, and 95th percentile of distance (measured in either miles or kilometers) of lost subjects collected by the ISRID.

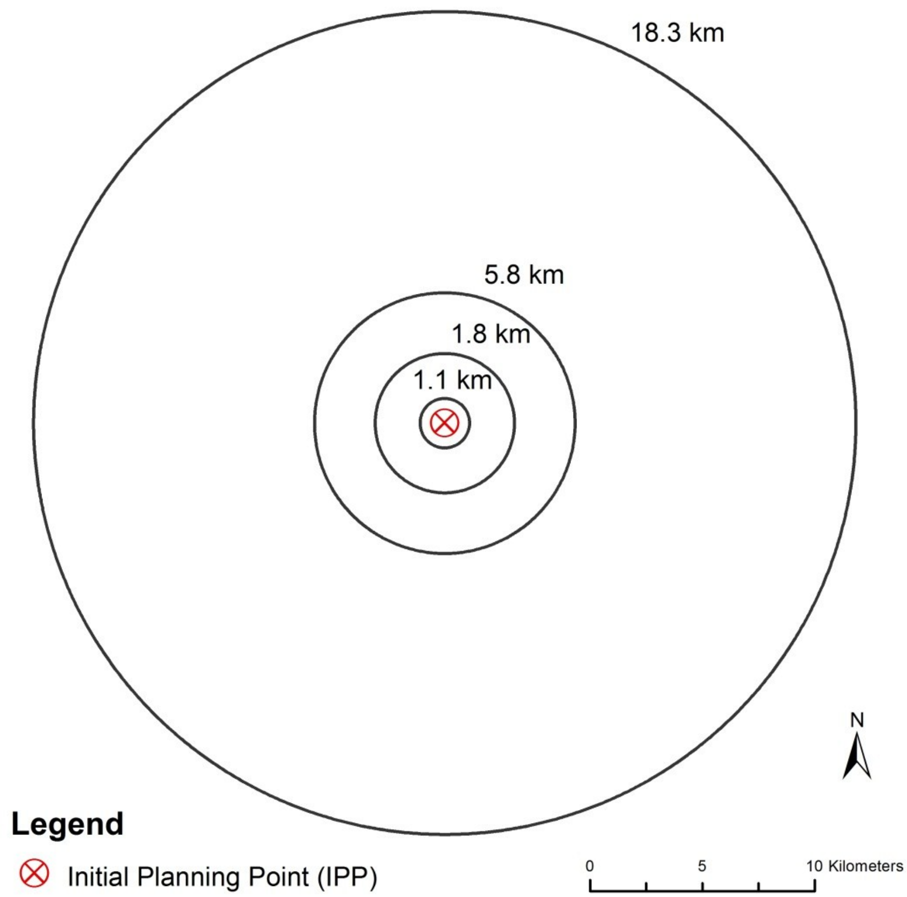


Figure 3.2 Lines of equal travel time given in hours (isochrones) for a hiker based on impedance due to geographic factors such as presence and absence of roads, trails lakes, and streams, terrain (slope, cliffs, etc.) and land cover layers. In this example we use the ISRID mobility table for the hiker category to classify isochrones corresponding to summary statistics for how long a missing subject had been actively moving before being found.

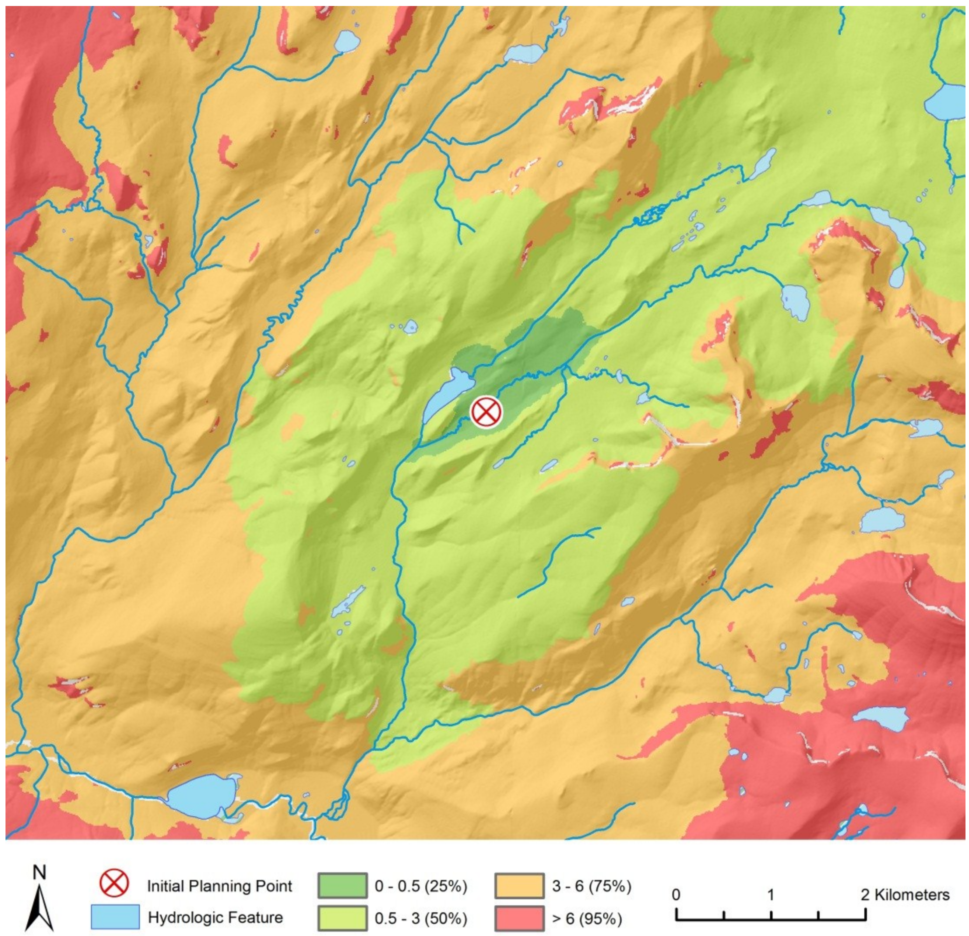


Figure 3.3 A scatterplot of isochrones (T_{min}) values at point found versus Euclidean distance (D) between point found and initial planning coordinates. The relationship showed a significant positive correlation (n = 130, R^2= 0.96, P < 0.01).

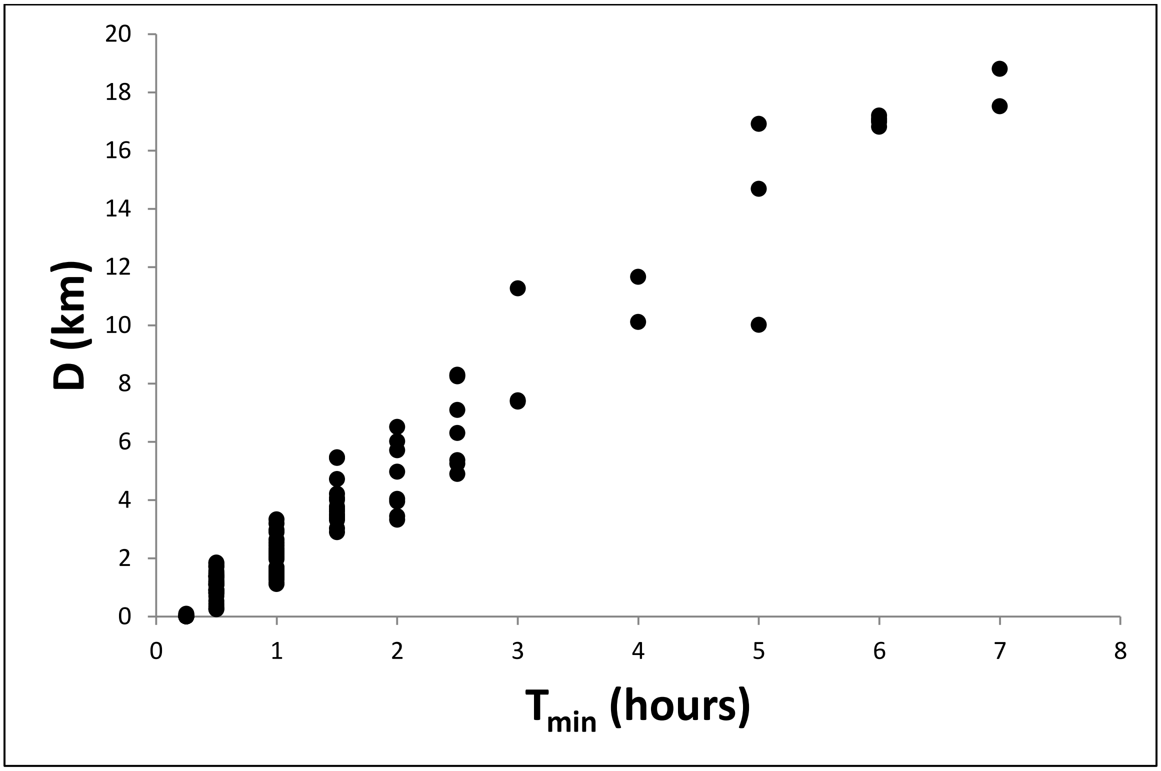
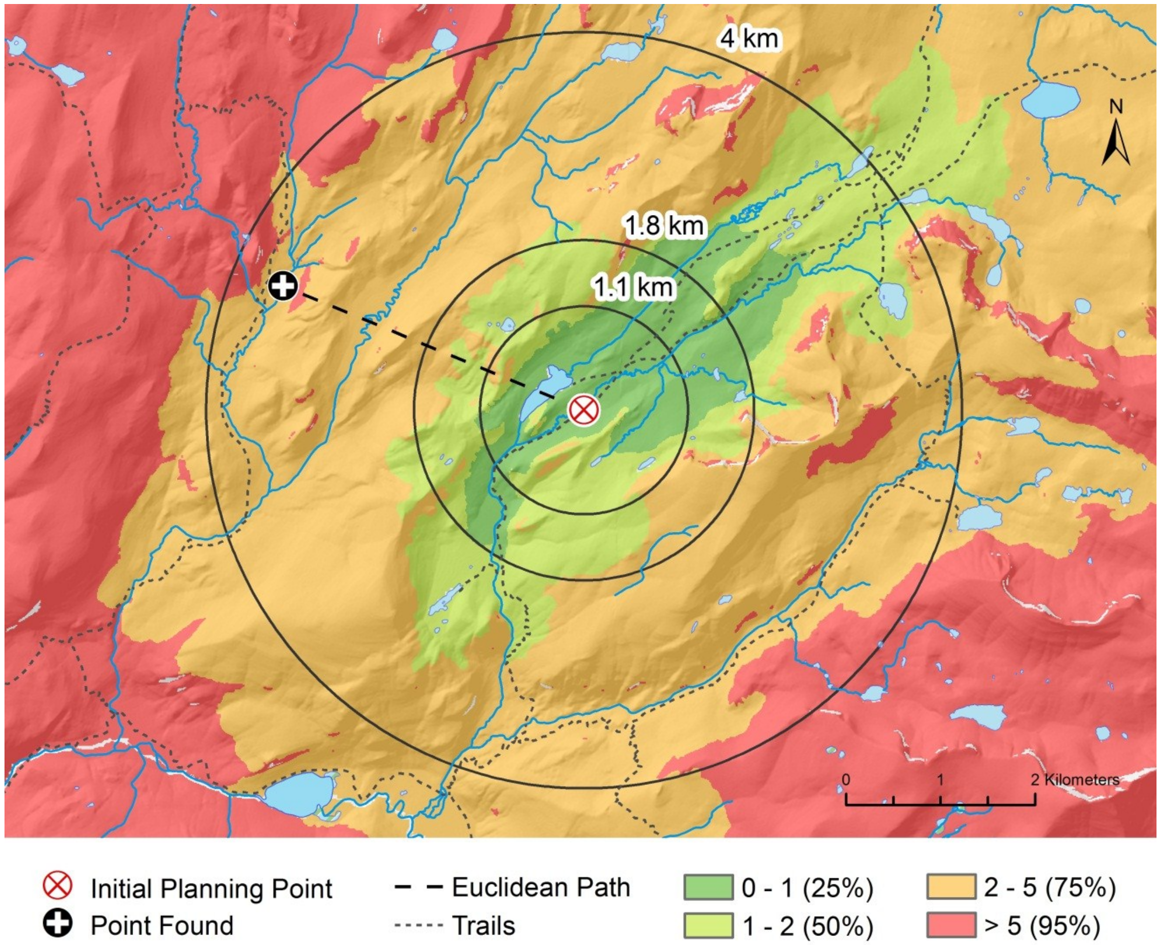


Figure 3.4 A map where both the ring and mobility models (isochrones in hours of travel time) derived from local data are incorporated for decision making on probability of area.



Chapter IV: Rescue

Expert versus Machine: A Comparison of Two Suitability Models for Emergency Helicopter Landing Areas in Yosemite National Park

[Formatted and previously published in *Professional Geographer*]

Landing a rescue helicopter in a wilderness environment, such as Yosemite National Park, requires suitable areas that are flat, devoid of tree canopy, and not within close proximity to other hazards. The objective of this study was to identify helicopter landing that areas are most likely to exist based on available geographic data using two GIScience methods. The first approach produced an expert model that was derived from pre-defined feature constraints based on existing knowledge of helicopter landing area requirements (weighted overlay algorithm). The second model is derived using a machine learning technique (maximum entropy algorithm, Maxent) that derives feature constraints from existing presence-only points, i.e. geographic one-class data. Both models yielded similar output and successfully classified test coordinates, however Maxent was more efficient and required no user-defined weighting that is typically subject to human bias or disagreement. The pros and cons of each approach are discussed and the comparison reveals important considerations for a variety of future land suitability studies, including ecological niche modeling. The conclusion is that the two approaches complement each other. Overall, we produced an effective GISystem product to support the identification of suitable landing areas in emergent rescue situations. To our knowledge, this is the first GIScience study focused on estimating the location of landing zones for a search and rescue application.

Key Words: *land suitability, geographic one-class data, Maxent, search and rescue, Yosemite National Park.*

Introduction

There are inherently spatial components underlying the process of rescuing people in distress or life-threatening situations (Search and Rescue or SAR), especially in remote locations such as in our National Parks. In National Parks, many SAR incidents occur each year and rescue operations can be extremely dangerous and costly (an average of 11 per day at a cost of 895 U.S. dollars each; Heggie and Amundson 2009). SAR operations usually consist of two parts: the search for missing or distressed persons and then the rescue operation. The helicopter is one tool rescuers may use to accomplish these tasks (up to 20 percent of rescues in some National Parks). If a rescue is deemed extremely urgent and requires a helicopter, then a suitable site (landing zone) must be chosen within close proximity to the rescue. In addition, to land safely a landing zone must meet the terrain requirements of minimal slope and be free of hazards (vegetation and man-made structures). If a landing zone cannot be found, alternatives may include more dangerous helicopter rescue techniques or calling off a helicopter operation altogether.

During helicopter operations, landing a helicopter and directly loading patients on the ground is generally the safest option because the other techniques such as short-haul or rappel require the helicopter to maintain a hovering position which greatly increases overall risk (Manwaring et al 1998). In addition, park managers attempt to reduce ecological impacts by finding suitable landing areas instead of manipulating vegetation. However, in the mountainous terrain of Yosemite National Park (YNP), our study area, finding suitable landing areas can be difficult. Yet, locating and identifying suitable land or habitat is a common workflow across many disciplines utilizing geographic information science (GIScience) and we propose that these techniques can be applied to find suitable helicopter landing areas. The process of employing these techniques will yield insight into the relationships between modeled geographic data, the real-world we are trying to study, and the analyses in question.

To identify suitable landing zones, an expert model can be generated using overlay analysis in which the layers are combined using constraints (Store and Kangas 2001). In this case we defined parameter constraints on feature space (slope, vegetation, aerial hazard layers) for suitable landing with best practice guides, standard operating procedures, and local knowledge. This is a subjective approach but can be an effective method if parameters being weighted adequately represent reality. There are several examples where this approach has been used to consider suitability against the spatial distribution of human features and environmental factors such as hazardous waste disposal (Jensen and Christensen 1986), road-siting (Li et al. 1999), Dengue Fever (Kolivras 2006) and geothermal exploration (Noorollahi 2008). With regards to helicopter landing areas, it is apparent in current guidelines (National Interagency Fire Center 2006) that the constraints on environmental layers (feature space) has been defined literally on feature space using expert knowledge (e.g. if slope at a location is less than 5 percent and no aerial hazards are present within 90 m, then our model should indicate increased suitability at the given coordinates). This can easily be re-projected onto geographic space. In this research we used expert knowledge to define criteria for several geographic layers and modeled landing suitability using a simple weighted overlay algorithm. However, the weighting process where layers are assigned values for their percent influence is subjective and it is likely that experts might disagree on how to weight parameters, which is a major concern for decision support systems. While many studies highlight the use of expert models or multi-criteria support systems (land use planning, transportation management, species habitat modeling, environmental systems; see Malczekski 2004 for a comprehensive overview, few report the accuracy or effect of bias in the output with regards to actual fitness for use. While land suitability analyses best practices utilize an evaluation of fitness of land for a particular use and the values of stakeholders in a region (Bojorquez-Tapia et al. 2001), in this research we evaluate an overlay model that defines fitness for use in emergency operations using an overlay technique based on fitness of land as defined by protocols.

Alternatively, we can model helicopter landing suitability using training data and environmental features without prior expert opinion. Machine learning models utilize an objective approach where only training data are used to examine parameters for relationships and relative contribution, on which the models interpolate or extrapolate the suitability of a geographic event. These methods primarily assume that the geographic phenomena of interest are determined by a set of environmental and non-environmental factors. For example, a species' distribution is controlled by environmental features such as temperature, precipitation, topography, soil and fires (Pearson and Dawson 2003). Another example is how landslide hazards are dependent on slope, aspect, elevation, soil moisture, and distance to major faults (Lan et al. 2004). Presence/absence models are frequently used for defining suitability, especially with regards to wildlife species distribution (Millspaugh 2009). In the past, predicting suitability areas has been challenging due to a lack of true absence data. For instance, in our study we only have coordinates for areas where helicopters have landed (suitable land), but not areas where they could not land (unsuitable land). This is a problem known as the geographic one-class data issue (Guo et al. 2011). Today however, powerful machine learning methods (Li et al. 2011) such as maximum entropy (Maxent; Phillips et al. 2006), utilize presence-only information to generate powerful predictive models with a relatively small training-sample collection effort (Li and Guo 2010). For example, in one recent study, 124 vulture presences or colonies were used to perform a Maxent analysis with strong results useful for predicting the future range of Griffon Vultures *Gyps fulvus* (Mateo-Tomás and Olea 2010). Maxent has been used in a variety of other species distribution studies for mammals (Baldwin and Bender 2008; DeMatteo and Loiselle 2008), birds (Yost et al. 2008), amphibians (Rödder and Weinsheimer 2009), lizards (Pearson et al. 2007) and endangered tree species (Kumar and Stohlgren 2009). Therefore, recent examples support the assumption that geographic data can adequately represent reality and relationships between these datasets can be characterized as meaningful environmental envelopes or niches. These environmental envelopes can then be used to estimate where a focal subject can or cannot exist in geographic or feature-space.

Machine learning algorithms require less human intervention during criteria selection and model coefficient determination than the expert model approach. Maxent is especially useful for this purpose, requiring only presence data to map species distributions. According to Baldwin (2009), "Maxent is relatively insensitive to spatial errors associated with location data, requires few locations to construct useful models, and performs better than other presence-only modeling approaches". In addition, machine learning algorithms such as Maxent, by design, can reveal the most compelling factors to explain some known spatial phenomena or determine the environmental envelopes of focal subjects. This is significant since assumptions made by experts can be incorrect or overemphasized (Alho et al. 1996). By matching locations of known landing locations within environmental envelopes, other similar locations can be predicted out with less bias than an expert model; however the machine learning approach is not interactive and the data processing is not always easily understood by those trying to utilize the output. In addition, while the algorithm itself is not biased by human opinion, the choice of which datasets to include is subject to human interpretation of what factors/layers are consequential to the distribution of the focal subject. The layer selection and sample collection process itself should still be based on scientific study and a machine learning algorithm is most useful if it generates informative models beyond what we already know (expert model). Thus, a comparison of the two techniques using a simple "species" (the helicopter) as the study subject is justified. Furthermore, to our knowledge, there is no previous research where expert knowledge or machine learning algorithms have been applied to helicopter landing suitability modeling.

We are testing the following assumptions in this study: the spatial layers we have chosen for evaluating helicopter landing suitability are appropriate, we can define parameter weights and coefficients using standard operation procedures, and a machine learning algorithm can autonomously extract meaningful relationships between training data and the base data provided. Therefore, we hypothesize the following; *if* the statements above are true *then* output from Maxent (1) should demonstrate similar relationships between base data and output as in the expert model and (2) Maxent should successfully classify habitat at test points. By testing these hypotheses we are essentially evaluating whether an expert model and a machine learning algorithm can have similar predictive abilities. This is a novel GIScience approach to locating suitable helicopter landing areas and comparing two techniques that are commonly used in land suitability analyses, but have not been compared directly.

Methods

Study Area and Data

YNP (Figure 4.1) is located in the Sierra Nevada Mountains of California and has over 1287.47 kilometers (800 miles) of hiking trails with over 200 SAR incidents each year (Hung and Townes 2008). The relatively high amount of visitation, even in remote locations leads to SAR incidents occurring far from roads that emergency vehicles can access. Therefore, Yosemite Search and Rescue (YOSAR) use a large helicopter to perform a significant proportion of their rescues each year (22 percent in 2009). Whether the rescued are retrieved by directly entering the helicopter at a landing zone or via short-haul (suspended rope) rescue, at some point during an operation the helicopter crew will need to locate a flat, open landing area. Therefore, it would be extremely helpful if land suitability analyses could be performed and generate raster data where helicopter pilots could be given coordinates for likely landing areas before departing for the mission.

Suitable landing areas are defined within institutional texts, such as the Interagency Helicopter Operations Guide (National Interagency Fire Center 2006). This source defined a suitable landing zone as 6.096 meters by 6.096 meters (20 feet by 20 feet) that is "reasonably level and clear of vegetation greater than 18 inches in height". This habitat is not easy to locate in the undeveloped areas of YNP (96 percent of the 3110 square kilometers is wilderness) where many hikers find themselves in trouble. In addition, to land a type II helicopter requires a safety circle of 27.432 meters (90 feet) where there is no open water, building or utility structure present.

Previously known landing areas were used as training data for the Maxent model against the same input layers used by the expert model. First, we extracted the YNP Landing Zone dataset (points, LZ) from the Park's geodatabase. The LZ Dataset was in vector format and contained 140 points where the Yosemite Helitack crew determined, using a GPS receiver (approximately 3 – 9 m accuracy), that a helicopter has landed in the past for rescue, fire, backcountry utility support, or snow-station monitoring. It is important to note that these sites were chosen based on field-decisions prior to any true spatial analyses and required little or no mechanical manipulation for landing.

We retrieved the following base data for model development: 10m USGS digital elevation model converted to slope in degrees (raster), 1m 2009 North American Imagery Program red and near infrared band digital numbers converted to pseudo-normalized difference vegetation index (raster), YNP buildings (polygon), and YNP open-water (polygon) from the publically-available NPS Data Store for vector data and the USDA GIS Server for imagery. In addition, we were provided with the location of YNP power lines (polylines). NDVI values in the Sierra Nevada have been derived from imagery to create vegetation classification for wildfire fuel models (Van Wagtendonk and Root 2003) and represent a high positive correlation with vegetation density with open areas having low NDVI values and densely vegetated areas (forest) having higher values. Therefore, lower NDVI values were given a higher ranking in the expert model. SLOPE was creating using the ArcGIS slope geoprocessing tool and NDVI was created using raster algebra. We were concerned about the level of spatial accuracy in YNP base data (horizontal and vertical uncertainty). We used a buffer technique to account for the horizontal accuracy of vector data (produced at the 1:24,000 scale with 12m uncertainty; USGS 2005) to err on the side of caution for hazards during the rasterization process. The highest optimal slope as indicated by the IHOG protocol (National Interagency Fire Center 2006) is 5 degrees and so any pixel 5 degrees or lower received a rank prior to weighting as a 3. During the expert model construction, we categorized using the vertical uncertainty of the digital elevation model derived slope (2 degree uncertainty; Wechsler and Kroll 2006) to account for overestimation of slope. Therefore, 5 to 7 degree slope pixels were ranked prior to weighting as a 2. All data were projected using UTM Zone 11N datum NAD 1983. All vector data were converted to raster format for analyses and model production. The extent was defined by the YNP boundary layer.

Expert Model

The first step of defining suitable habitat is known as site-search analysis (Malczewski 2004) in which we identify which factors affect landing suitability. We used literature, the Interagency Helicopter Operations Guide (IHOG; National Interagency Fire Center 2006) to identify what factors are important for landing a helicopter; slope, vegetation and the presence/absence of hazards. The IHOG states the following regarding helispots for the type of helicopters used in Yosemite (Type II) and we used these siting factors:

- Slope - Avoid slopes over 5 degrees or 11 percent (9:1) slope
- Water - Avoid open water
- Hazards (vegetation, buildings, power lines) - An adequate minimum width for an approach-departure path is the diameter of the safety circle (90 feet).

In the second step, we used the weighted overlay method to represent the above IHOG criteria in a deterministic manner, where each class within a map layer is given a score and each unique map layer is given a map weight to reflect the findings of subject matter experts and geographic information scientists.

$$\overline{S} = \frac{\sum_{i=1}^{n} s_{ij} w_i}{\sum_{i=1}^{n} w_i} \qquad (1)$$

Equation 1 (Bonham-Carter 1994) defines total suitability score as $\overline{S}$, where S_{ij} is the suitability ranking (from 1 to 3, 0 if the value restricts helicopter landing entirely) within individual layers (e.g. slope) and W_i is the weight assigned to individual map layers (from 0 to 100 percent, with the cumulative percentages equal to 100 percent). The output is a raster (10m resolution) rounded to ordinal values of 0 (0.00 – 0.49), 1 (0.50 – 1.49), 2 (1.50 – 2.49), or 3 (2.50 – 3.00). We defined the variables entered into this weighted overlay model using an object-oriented modeling approach where inputs could be selected from spatial layers (ESRI 2010). We constructed the model so that the presence of any "hazard", i.e. where distance to AIRHAZ and/or BUILDS < 30m or presence of open water, would nullify all other contributing values at that raster cell. So an output of 0 defines that raster cell as restricted, despite any other variable values because it cannot be suitable for landing (not suitable). Increasing suitability ranking values (S_{ij}) for each of the layers from 1 (Not likely), 2 (Possible), to 3 (Most likely suitable) represents the increased likelihood of the raster cell being an appropriate landing area based on expert knowledge (Table 4.1). The subjective weighting of each layer (W_{ij}) was based on how likely a layer was to impact a helicopter pilot's decision in YNP. The weighted overlay algorithm, therefore, produces a raster cell equal to 3 for an area that is flat, clear of tall vegetation, not open water, and clear of man-made hazards. Because Maxent produces a continuous probability output from 0 to 1, users of Maxent can choose a threshold for decision making (read below). To directly compare the decision making accuracy of the weighted overlay method (expert model) to the machine learning algorithm (Maxent) we needed to recode the 4 output categories into a binary output. For binary comparison with Maxent we reclassified the expert output by recoding as follows: IF 0 OR 1 THEN binary = 0 (not suitable); IF 2 OR 3 THEN binary= 1 (suitable).

Maximum Entropy Model

Entropy is a fundamental concept in information theory, which measures how much choice is involved in selection of an event (Shannon 1948). For a random variable, uniform distribute provides greater entropy than non-uniform one. Therefore, an equal distribution of probabilities has the maximum entropy (Bishop 2006). For example, given a discrete variable x with two possible values, 0 and 1, the best guess of x will be the uniform distribution model: p (x = 1) = 0.5 and p (x = 0) = 0.5. If we know that 80% of the samples are 1, then we updated the model as p (x = 1) = 0.8 and p (x = 0) = 0.2. This is based on everything we know and avoids making assumption that we do not know (Jaynes 1990).

In this study, the unknown probability distribution π is over a finite set X (the set of pixels in the study area) The distribution π assigns a non-negative probability $\pi(x)$ to each pixel x, with the sum of these probabilities equal to one. The constraints on the unknown probability distribution π are represented by a set of environmental features (e.g. NDVI, Slope, etc.) $f_1, \ldots, f_n$ on X. The information we know about π is the expectations (averages) of each feature f_j under π, which is defined as:

$$\pi[f_j] = \sum_{x \in X} \pi(x) f_j(x) \tag{2}$$

A set of sample pixels $x_1, \ldots, x_m$ is drawn independently from X. The corresponding empirical distribution is denoted as

$$\tilde{\pi}(x) = \frac{|\{1 \le i \le m : x_i = x\}|}{m} \tag{3}$$

The empirical average of f_j under $\tilde{\pi}$ is defined as

$$\tilde{\pi}[f_j] = \frac{1}{m} \sum_{i=1}^{m} f_j(x_i)$$

We use $\tilde{\pi}[f_j]$ as an estimate of $\pi[f_j]$. The goal is to seek the probability distribution $\hat{\pi}$, an approximation of π, subject to the constraint that the expectation of each feature f_j under $\hat{\pi}$ is the same as its empirical average, stated formally as

$$\hat{\pi}[f_j] = \tilde{\pi}[f_j] \tag{4}$$

Although there are many distributions satisfying these constraints, based on the maximum entropy principle, we choose the one that produces the maximum entropy. The entropy of $\hat{\pi}$ is defined as

$$H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x) \ln \hat{\pi}(x) \tag{5}$$

We used the 'Maxent' software version 3.3.3a that is freely available online[8]. The Maxent model received the base data layers in their native raster format (SLOPE, NDVI, HYDRO, AIRHAZ, BUILDS) and the coordinates of known landing zones to be used as training points for the learning algorithm. The output is a raster (10m resolution) with continuous values generated by a machine-learning algorithm based upon maximum entropy. For more detailed information regarding Maxent please refer to Phillips et al. (2006).

We used two output criteria for evaluating the Maxent method for identifying helicopter landing suitability; the estimate probability and binary prediction. We used the default output logistic that gives an estimate of probability between 0 and 1. Note the output is not the exact probability of being positive, but it is proportional to the conditional probability of being positive (Phillips et al. 2006). To investigate the ability of Maxent to be used for decision making and compare to the weighted overlay method we need to convert these probabilities into binary prediction values (0 or 1). Therefore, a threshold is required to convert the probabilistic output to binary predictions. To avoid over-fitting, we randomly set aside 40 of the 140 training set coordinates for validation. To avoid over-predictions, we allow for a small omission rate to account for outliers in the validation set. In this study, we chose the logistic value corresponding to a 5 percent omission rate for the validation set as the threshold to make a binary prediction of 0 for not suitable or 1 for suitable (Pearson et al. 2004). To evaluate the sensitivity of our analysis training dataset to sample size needed to train the model, we generated multiple model runs using sampling sizes starting from10 to 100 points with a step size of 10 (a total of ten tests). To control for different combinations of sampling selections, we ran for each of these ten configurations a Monte Carlo simulation that randomly selects point to train the model.

Statistical Analyses

To evaluate the level of agreement between the expert and Maxent model based on estimated probability (Hypothesis 1) we compared the per-pixel distribution of each models' output values. The expert overlay model initially has a categorical output (0, 1, 2, 3) for each pixel in the study area whereas Maxent's initial output are continuous values (0.000 to 1.000) for each pixel in the study area (N = 53,776,590 pixels). From the study area, we sampled 2000 random pixels stratified by the expert model's *i* factor-levels (four equally sample groups of n_i = 500). Then the continuous output of Maxent (values close to 0.000 indicate low suitability, 1.000 indicates high suitability) could be compared across the expert model discrete ordinal categorical values (0 indicates not suitably; 1 not likely suitable; 2 possibly suitable; 3 most likely suitable). To do so, we chose the non-parametric equivalents to the one-way ANOVA and Tukey test (Kolmogorov-Smirnov and Nemenyi Test for pairwise post-hoc comparison of ranks; Zar 1999; Tabachnik and Fidell 2007). Therefore, we compared the ranks of Maxent continuous data between the factor levels from the expert suitability model: not suitable, not likely, possible, most likely. Secondly, we compared the global per-pixel binary classification of all output pixels for percent agreement. Finally, we qualitatively compared the weighting determined by the Maxent machine learning algorithm to that defined by the expert model algorithm to identify possible sources of conflict.

To evaluate the accuracy, or how well the expert and Maxent models classified true-presence locations (Hypothesis 2), we sampled the binary output from each model at 40 test/validated points. We then used the Binomial test to compare this classification at coordinates where helicopters have landed but were not included in Maxent training sample. The test proportion is the random chance proportion of 0.5. Standard geoprocessing was done using ESRI

[8] http://www.cs.princeton.edu/~schapire/maxent/

ArcMap 10.0 and statistical analyses using SPSS Statistical Package (ESRI 2010; SPSS Inc. 2004).

Results

The output for both models had identical extent and pixel size (10m). The expert model output has four categories (not suitable, not likely, possibly, and most likely suitable) and the Maxent model output yielded continuous values from 0 to 1. When we compared the distribution of ranked-Maxent values between expert model groups at 2000 test points, there was an overall significant positive relationship (Kruskal-Wallis $\chi^2_3 = 3280.97$, $P < 0.001$) and significant pair-wise differences between each group (Figure 4.2).

The threshold for creating binary classification of Maxent was 0.0836 (< threshold = not suitable; > threshold = suitable). When we reclassified Maxent output and the expert model output into binary variables (0 = not suitable, 1 = suitable), there was an overall agreement of 90.2 percent (Table 4.2; Figure 4.3). To test the sensitivity of the sample size on the model outputs, we ran Maxent with the sample sizes of 20 to 100 with a step interval of 10, and found that as we increase the sampling density, there is little disagreement between suitability maps produced at each step (average of 0.78% change).

There is great similarity between the settings implemented in the expert model (percent influence, relationship) and the output from Maxent (relative contribution, relationship). Finally, both the expert and Maxent model output correctly classified the 40 validation points better than expected by chance alone (test proportion at 0.50; $P < 0.001$) and better than a distribution expected based on proportions found in the raster outputs (expert model expected from global distribution 0.23, $P < 0.001$; Maxent model 0.14 expected from global distribution, $P < 0.001$; Table 4.3).

Discussion

The expert model utilized a weighted overlay approach where reclassified variables are weighted by percent influence and added to give an integer suitability score (0 to 3), with 3 indicating a raster pixel that is most suitable (low slope, low NDVI score, not on open water, and at least 27m from buildings or aerial hazards). The Maxent machine learning algorithm used a maximum entropy approach to assigning suitability scores based off of sampling feature space at 100 training/presence coordinates. Both models indicate that areas of low slope, low NDVI, without standing water, and not in close proximity to hazards are suitable for landing helicopters.

Expert Model

The expert model, while subjective from an input standpoint, successfully classifies helicopter landing suitability areas (95 percent). While the weighted overlay approach can be limited by the quality of expert input and parameters, the spatial data available in this study adequately represented reality and the definitive real-world parameters defined in the IHOG document were appropriately transferred in the overlay process. In most cases a weighted overlay is a process where several raster layers are overlaid using a common measurement scale and each layer is weighted according to its importance. Our weighting scheme was determined based on some understanding of the spatial distribution of layer attributes. For example, we set the percent influence of hazards low (5 percent) because they are not spatially pervasive features in YNP, not because they are less important than slope. These weighting criteria may not transfer well to more developed environments where hazards are more spatially abundant.

Despite a drastic rise in publications related to expert models and GISystem based multi-criteria decision analysis, there has been a lack of research on conceptualization, validation, and use of these techniques in real-world spatial problems (Malczewski 2006). In light of this, it is important to note that we feel accounting for limitations due to spatial accuracy and spatial distribution (Goodchild and Gopal 1989; Zhang and Goodchild 2002) during the reclassification process led to increased model success. In addition to expertise on a focal subject's requirements (i.e. habitat), one must also understand and attempt to account for uncertainty regarding positions in geographic space within their dataset. Indeed, two sites of the previously used landing sites were misclassified by the expert model (Table 4.3) and appear to be within 10m of a small patch of suitable landing area, suggesting horizontal inaccuracy in the point coordinates themselves (due to GPS error) not a violation of helicopter landing criteria. This highlights the need to understand the horizontal accuracy of spatial data before allowing a machine learning model to produce distribution maps based off of training data. In this case, the performance (95% of points were correctly classified) of the expert model suggests that the majority of our test data were spatially accurate. Spatial data quality and its applicability to the real-world are not always fully understood prior to beginning a land suitability studies. Therefore the expert model approach is an appropriate step for testing prior assumptions and data quality prior to machine learning analyses.

Maximum Entropy Model

Based on the performance of the machine learning algorithm Maxent (95 percent on validated points), it appears that we can extract meaningful relationships between training data and the base data provided. Maxent was not sensitive to sample size in our study and produced consistent suitability maps between sampling densities) and this is consistent with sample size analysis of previous work (Guo et al. 2011). With regards to our first hypothesis, Maxent identified similar relationships between base data and landing suitability to the expert model and resulted in a global 90.2 percent agreement with expert model. With regards to our second hypothesis regarding model success, Maxent classified test points (not used in the training dataset) with success equal to the expert model. While Maxent had 90.2 percent agreement with the expert model, it was more conservative overall (Maxent classified 86 percent of the area as not suitable; expert model classified 77 percent as not suitable; see Figures 4.2 and 4.3). One explanation for this disparity may be that the expert model more heavily weighted NDVI than Maxent (Table 4.1). Indeed in post-hoc, if we set the percent influence for the expert model equal to the percent contribution determined by Maxent we get an even greater level of agreement (97 percent agreement). However this results in a lower classification score on the validated points (from 95 percent in the original to 85 percent) in the new and more conservative expert model. This highlights the complexity of subjective weighting criteria involved with the weighted overlay technique.

The way helicopter pilots perceive their environment is arguably no different from other species' habitat selection (mammals, Baldwin and Bender 2008; DeMatteo and Loiselle 2008; birds, Yost et al. 2008; amphibians, Rödder and Weinsheimer 2009; and lizards, Pearson et al. 2007). Both humans and other animals choose habitat based off of what is geographically available and appears desirable to them at the time of making the decision, yet we model large areas from the localized decisions made in the past for these species. We recognize pilots may not always choose suitable landing zones due to a variety reasons (error in judgment, emergency landing, etc.) and this will introduce some bias to the analyses. However, applying the 5% omission rate as a threshold for converting the probability map into the binary output, the resulting map will exclude some unrepresentative sample localities. Note that the machine learning algorithm used in this project was not created to find the most optimal landing zones; instead they are used to generate binary suitable and unsuitable landing zones for making decisions. Another concern regarding the use of Maxent in this study is the false classification of some open water pixels with zero degree slope values (which is desirable), but also a very low NDVI score (indicating the presence of open water). Whereas the expert can nullify the influence of other variables if one variable is restricted, Maxent could not always discern low vegetation land classes from open water due to the otherwise strong negative relationship between openness/suitability and NDVI. This might be corrected by utilizing a combination of vegetation indices (simple-ratio) or adding a mask to pixels that cannot possibly be physically (or biologically) possible. Nevertheless, the powerful ability of machine learning algorithms to correctly classify naïve points based on geographic one-class data was demonstrated without the challenge of selecting weighting criteria up front.

Theoretical Implications

The process of evaluating terrain for land suitability or site selection is typically an iterative process that has historically used a combination of GIScience and local or expert knowledge and opinion (Malczewski 2004). Challenges in the land suitability modeling process are typically constraints bounded by the quality of spatial data available, explicit knowledge relating to factors affecting suitability, lack of true absence data, and/or inability to adequately model the real world.

In this example, the weighted overlay model is not used in the classic multi-criteria evaluation scenario (such as discussed in Carver 1991). We are not concerned with conflicting policy or expert opinion at this stage. Weighted overlay was used in a deterministic sense to map out all of the areas where a helicopter may land based on physical parameters given in a helicopter operations guide. We could have used a simple map algebra approach, but the idea was to produce a model that can be validated by true one-class data and also compared to a machine-learning model output. The validation step is important because the outcome a model is only as valuable as the quality of input that feeds it. Now validated, this suitability model is stored in a usable format (ArcGIS toolbox customized model) that can be used to solicit opinions from other stakeholders using the multi-criteria evaluation method which is the strength of the weighted overlay technique over a machine-learning algorithm approach.

The deterministic or empirical data issues must be addressed before multi-criteria evaluation concerns regarding expert opinions are raised. In other words, if the very data values debated do not adequately represent reality, then policy discussions are meaningless in a modeling sense. In this study, slope and vegetation were assigned high percent influence values; not because they are more important than other factors, but because of they vary ubiquitously. The other factors, distance to hazards, vary in a linear fashion, and only a small percentage of the geographic area has data of importance (areas on or very near hazards). This was crucial to model success, mere opinion or logical conclusions without geographic understanding would have resulted in a model without any real-world meaning. Therefore, input based on both subject-matter expertise and geographic literacy is needed for expert model success. In the future, similar approaches to the expert model process such as public participation geographic information systems (PPGIS) are expected to increase with the advent of powerful web-based geoprocessing technology (Hall et al. 2010). This will likely have a great impact on the field of GIScience and society. Therefore, it is important to comprehend the fundamentals of spatial analyses based on human input and make advances in techniques (e.g. Fuzzy Set Theory – Thill and Sui 1993) and geographic awareness to address shortcomings in expert model approaches.

Alternatively, the power of machine learning algorithms to use presence-only data provides a tool that can complement expert models and PPGIS in a number of ways. The Maxent approach in this study required no subjective weighting criteria, used geographic one-class data (presence only points), and was efficient at creating a successful model. One advantage to utilizing both methods is to test basic assumptions of expert models in a competitive environment. If discrepancies exist between expert models and machine learning models then the areas where disparities occur might provide insight into a knowledge-gap or inadequate base data.

Conversely, constructing expert models prior to machine learning algorithms could provide justification and exploration for choosing which base data to provide the machine learning algorithms. This comparison is unique and constructive because in the expert model we define suitability up front using feature space constraints to map out geographic distribution (e.g. slope should be less than 5 degrees), whereas the Maxent model derives feature space constraints from geographic distributions of known points. In a sense, the expert makes a wager regarding geographic layer ranking and weighting, and bets against the machine-learning algorithm. Where there appears to be disagreement between the expert model and machine learning output we can gain insight into the inner workings of machine learning techniques and how to interpret results in future studies. This is significant because the comprehensive understanding of predictive modeling using geographic one-class data will likely lead to increased performance of machine learning techniques that matches or even exceeds expert models or any single machine-learning algorithm alone (Guo et al. *in press;* Li et al. 2011).

Expert model versus machine learning model

The primary benefit of using an expert modeling approach in conjunction with the machine learning approach is that the expert model validates simplistic assumptions regarding the quality of base data and our proposed relationships (flat areas with little or no hazards/vegetation are suitable). In this example, the expert model was essentially a map algebra approach where rules followed very clear standards stated in an operation guide regarding environmental constraints. The only subjective modification we (Geoscientists and domain experts) made was to address spatial uncertainty and spatial distribution in our weighting. Therefore, the expert model should represent a solid foundation for testing a machine learning model that is naïve to the rules until it has evaluated the distribution of environmental envelopes where training data exist. If this study's focus were a wildlife species whose distribution and behavior were observed for decades, would we not want to test our assumptions regarding its habitat selection? When the expert model and machine learning approach are used together we validate our assumptions and form hypotheses to test our machine-learning outcomes. Without this process, machine learning algorithms lend themselves to criticism for not being hypothesis driven. Likewise, the machine learning algorithm is a useful tool for modeling species distributions that are more complex than helicopter landing (many more factors, many more relationships) where fine-tuning each parameter relationship may be impossible. In this case, a machine learning algorithm could be used to identify the most pertinent factors and rank their importance, reducing the number of factors that need to be evaluated by experts. This is highlighted in our study (Table 1). We set slope and NDVI at 40% influence each. Open-water and hazards were each given 10% influence. Maxent however, identified slope as having a much higher influence than NDVI, and lower than 10% influence each. Maxent is obviously sensitive to the spatial distribution of datasets, whereas a domain expert may not be. An expert may overestimate the influence of a factor if they do not understand the spatial distribution of that factor. Together, these two approaches complement each other by balancing domain knowledge with geographic reality.

Real-world application

Both the expert and machine learning methods produced models that assigned higher suitability values for areas with little or no slope, low NDVI, and that were a minimum distance from water, buildings, and aerial hazards. This is consistent with our previous assumptions and industry standards regarding helicopter landing area criteria, however it should be noted that we did not have a spatially explicit decision and visualization tool prior to conducting this research. While the expert model (weighted overlay) method may be limited by decisions related to weighting the relative importance of geographic layers, Maxent efficiently produces a meaningful map without the challenge of subjective weighting criteria.

While we expect our land suitability analyses and modeling to be limited physio-geographically to YNP and surrounding Sierra Nevada management areas, the techniques described in this article will be informative to aviation programs that have access to geographic data. The expert model we produced is based exclusively on knowledge of the physical limitations of a helicopter (from the IHOG, National Interagency Fire Center 2006) and the spatial distribution of geographic layers (from GISystem data). In the future, natural resource advisors and land owners (experts and stakeholders) can use this process to work with aviation managers to minimize impacts on the environment through the Minimum Impact Suppression Technique for non-emergency operations (MIST; National Interagency Fire Center 2006). To do so modelers should consult established techniques for eliciting expert opinion, such as Delphi methods (Fink et al 1984) or Analytical Hierarchy Process (AHP; Saaty 1990) approaches, that can be used to solicit expert opinion from groups of experts. The mechanistic expert model we produced can be modified (and is available upon request) once additional expert opinions for determining factor weighting have been solicited for a geographic area.

The specific aim of this research has been accomplished. Coordinates for suitable landing areas near the rescue location can be derived from a suitability map in a command post and relayed to the pilots and rescuers before an uninformed decision is made to utilize a more dangerous helicopter rescue technique that involves hovering instead of landing. In essence this is a decision support tool to tell helicopter pilots what geographic area to consider and the likelihood of there being a suitable landing site near a rescue location, whereas before it was assumed that there was none unless they have previously landed in that location. The decision of exactly where to land will continue to be made on a local level by the helicopter pilot using experience and discretion, but narrowing down the areas of suitable terrain will be incredibly helpful for YNP emergency staff. We will continue to evaluate both the expert and Maxent suitability models as new landing zones are established in YNP.

Conclusion

Each year in YNP over 200 incidents will incite a Search and Rescue response and many (up to 20 percent) will require the use of a helicopter in order to be successfully resolved. Search and Rescue operations in National Parks save many lives each year, at great risk to rescuers and substantial financial cost. In this study, we compared two GIS methods for identifying suitable landing areas for helicopters and found that both the model produced by weighted overlay analysis (expert model) and Maxent (machine learning algorithm) had 1) a significantly positive relationship between suitability scores, 2) 90.2 percent agreement on binary decisions (suitable vs. not suitable), and 3) successfully classified 95.0 percent of test points. The expert model approach encourages researchers to test their assumptions which is an inherent component of the scientific method. The machine learning algorithm approach, in turn, uses an effective GIScience technique to explain the geographic distribution of events and validate our “expert” assumptions. This comparison of two successful GIScience methods for identifying suitable helicopter landing areas will be beneficial to rescuers and the lessons learned offer insight to researchers developing land suitability models across a variety of disciplines.

References

Baldwin, R.A. 2009. Use of maximum entropy modeling in wildlife research. *Entropy* 11:854-866.

Baldwin, R.B. and L.C. Bender. 2008. Den-site characteristics of black bears in Rocky Mountain National Park, Colorado. *Journal of Wildlife Management* 72:1717–1724.

Bishop, C.M. (2006). *Pattern recognition and machine learning*. Springer, New York

Bonham-Carter, G.F., 1994. Geographic Information Systems for Geoscientists, Modelling with GIS. (Pergamon, Oxford).

Bojorquez-Tapia, L.A., Diaz-Mondragon, S. and Excurra, E. 2001. GIS-based approach for participatory decision making and land suitability assessment. *International Journal of Geographical Information Science* 15(2):129-151.

Carver, S.J. 1991. Integrating multi-criteria evaluation with geographic information systems. *International Journal of Geographical Information Science* 5(3): 321-339.

DeMatteo, K.E. and B.A. Loiselle. 2008. New data on the status and distribution of the bush dog (*Speothos venaticus*): Evaluating its quality of protection and directing research efforts. *Biological Conservation* 141:2494–2505.

ESRI (Environmental Systems Resource Institute). 2010. ArcMap 10. ESRI, Redlands, CA.

Goodchild, M. and S. Gopal. 1989. The Accuracy of Spatial Databases. London: Taylor & Francis.

Fink, A., Kosecoff, J., Chassin, M., and R.H. Brook. 1984. Consensus methods characteristics and guidelines for use. *American Journal of Public Health* 74(9):979-983.

Guo, Q., W. Li, Y., Liu,, and D. Tong. 2011. Predicting Potential Distributions of Geographic Events Using One-class Data: Concepts and Methods. *International Journal of Geographical Information Science* 25(10): 1697-1715.

Hall, G.B., R. Chipeniuk, R. Feick, M. Leahy, and V. Deparday. 2010. Community-based production of geographic information using open source software and Web 2.0. *International Journal of Geographical Information Science* 24(5): 761-781.

Heggie, T.W. and B.S. Amundson. 2009. Dead Men Walking: Search and Rescue in US National Parks. *Wilderness and Environmental Medicine* 20:244-249.

Hung, E.K. and D.A. Townes. 2007. Search and Rescue in Yosemite National Park: A 10-Year Review. *Wilderness and Environmental Medicine* 18:111-116.

Jaynes, E.T., 1990. Notes on present status and future prospects. In: W.T. Grandy and L.H. Schick, eds. Maximum entropy and Bayesian methods. Dordrecht, The Netherlands: Kluwer.

Lan, H. X., Zhou, C. H., Wang, L. J., Zhang, H. Y., and R.H. Li. 2004. Landslide hazard spatial analysis and prediction using GIS in the Xiaojiang watershed, Yunnan, China. *Engineering Geology* 76 (1–2): 109–128.

Li, W., Guo, Q., and C. Elkan. 2011. A positive and unlabeled learning algorithm for one-class classification of remote sensing data. *IEEE Transaction on Geoscience and Remote Sensing* 49: 717 – 725.

Li, W. and Q. Guo. 2010. A maximum entropy approach to one-class classification of remote sensing imagery. *International Journal of Remote Sensing* 31(8):2227-2235.

Malczekski, J. 2006. GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographic Information Science* 20(7):703–726.

Malczekski, J. 2004. GIS-based land-use suitability analysis: a critical overview. *Progress in Planning* 62:3-65.

Manwaring, J.C., Conway, G.A., and L.C. Garrett. 1998. Epidemiology and prevention of helicopter external load accidents. *Journal of Safety Research* 29:107–121.

Mateo-Tomás, P. and P. Olea. 2010. Anticipating knowledge to inform species management: Predicting spatially explicit habitat suitability of a colonial vulture spreading its range. *PLoS ONE*, 5(8): e12374.
Millspaugh J.J. and F.R. Thompson. 2009. Models for planning wildlife conservation in large landscapes. Elsevier, San Diego, USA.
National Interagency Fire Center. 2006. Interagency Helicopter Operations Guide. NFES 1885. Interagency Management Council, Boise, Idaho.
Pearson, R.G. and T.P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, 12, 361–371.
Pearson, R.G., T.P. Dawson, and C. Liu. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography* 27:285–298.
Pearson, R.G., C.J. Raxworthy, M. Nakamura, and A.T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102–117.
Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modeling* 190:231–259.
Rödder, D. and F. Weinsheimer. 2009. Will future anthropogenic climate change increase the potential distribution of the alien invasive Cuban treefrog (Anura: Hylidae)? *Journal of Natural History* 43:1207–1217.
Saaty, T.L. 1990. How to make a decision – the analytic hierarchy process. *European Journal of Operational Research* 48(1):9-26.
Shannon, C.E., 1948, A mathematical theory of communication. *The Bell System Technical Journal,* **27**, pp. 379-423, 623-656.
SPSS Inc. 2004. SPSS Base 13.0 for Windows User's Guide. SPSS Inc., Chicago, IL.
Store, R. and J. Kangas. 2001. Integrating spatial multi-criteria evaluation and expert knowledge for GIS-based habitat suitability modeling. *Landscape and Urban Planning* 55(2):79–93.
Tabachnick, B.G. and L.S. Fidell. 2007. Using Multivariate Statistics, 5th Edition. Pearson Education, Boston, MA.
Thill, J. and D.Z. Sui. 1993. Mental maps and fuzziness in space preferences. *Professional Geographer* 45(3):264-276.
United States Geological Survey (USGS). 2005. National Mapping program Standards. Washington: USGS
Van Wagtendonk, J. and R. Root. 2003. The use of multi-temporal Landsat Normalized Difference Vegetation Index (NDVI) data for mapping fuel models in Yosemite National Park, USA. *International Journal of Remote Sensing* 24(8):1639-1651.
Wechsler, S.P. and C.N. Kroll. 2006. Quantifying DEM uncertainty and its effect on topographic parameters. *Photogrammetric Engineering & Remote Sensing* 72(9):1081-1090.
Yost, A.C., S.L. Petersen, M. Gregg, and R., Miller. 2008. Predictive modeling and mapping sage grouse (*Centrocercus urophasianus*) nesting habitat using Maximum Entropy and a long-term dataset from Southern Oregon. *Ecological Informatics*. 3:375–386.
Zar, J.H. 1999. Biostatistical Analysis. Prentice Hall, NJ.
Zhang, J. and M.F. Goodchild. 2002. Uncertainty in Geographical Information. London: Taylor & Francis.

List of Tables and Figures

Table 4.1 – This table represents layers used to build an expert model for helicopter landing suitability in Yosemite National Park, CA. The expert model was created using criteria from the Interagency Helicopter Operations Guide (National Interagency Fire Center 2006) and spatial uncertainties for slope (vertical accuracy 2°; Wechsler and Kroll 2006), NDVI (Yosemite NAIP vegetation classification; Van Wagtendonk and Root 2003), and USGS vector data (water and hazard data horizontal accuracy ± 12 m; USGS 2005). *S* represents scores assigned to layer categories and *W* represents weights (percent influence) assigned to each of the layers (Bonham-Carter 1994).

Table 4.2 – This table represents a confusion matrix comparing two helicopter landing suitability models per-pixel where helicopters for Yosemite National Park, CA (N = 53,776,590). The expert model was created using criteria from Interagency Helicopter Operations Guide and the Maxent model used 100 training points where helicopters have previously landed in Yosemite National Park, CA (unique from the validated points). Overall agreement (in bold) is 90 percent.

Table 4.3 – This table represents an error matrix comparing two suitability models at 40 validated points where helicopters have previously landed in Yosemite National Park, CA. The expert model was created using criteria from Interagency Helicopter Operations Guide and the Maxent model used 100 training points where helicopters have previously landed in Yosemite National Park, CA (unique from the validated points). Both the expert and the Maxent model classified 95% of the validated points correctly (38 of 40 points).

Figure 4.1 – A map of Yosemite National Park Boundary (green) overlain on a 10m shaded-relief map of the Sierra Nevada Mountains. Inset map shows a map of California, US (yellow) and surrounding states with map extent indicated by bounding box.

Figure 4.2 – A comparison of mean Maxent helicopter landing suitability scores by expert model suitability categories for stratified randomly distributed sample points in Yosemite National Park, CA. Random sample points were at a minimum of 10m from nearest neighboring sample points. Maxent scores, when ranked for non-parametric testing, varied significantly between expert model categories ($\chi^2_3 = 3308.016$, $P < 0.001$). Box plots are 25-75% percentiles, dots are outliers, and bars mark 95-5% ranges for Maxent scores. Medians are solid black lines and means are dotted lines. The dotted horizontal line represents the threshold required to convert the probabilistic output of Maxent to binary predictions (0.083).

Figure 4.3 - A comparison of helicopter landing suitability binary results for an expert model (**A**) and Maxent (**B**) in and near Yosemite National Park (black line indicates park boundary). Green indicates suitable and red indicates not suitable pixels. Blue circles represent pre-existing helicopter landing locations (N = 140). Overall agreement between the two models is 90.2 percent (N = 53,776,590 pixels).

Table 4.1 – This table represents layers used to build an expert model for helicopter landing suitability in Yosemite National Park, CA. The expert model was created using criteria from the Interagency Helicopter Operations Guide (National Interagency Fire Center 2006) and spatial uncertainties for slope (vertical accuracy 2°; Wechsler and Kroll 2006), NDVI (Yosemite NAIP vegetation classification; Van Wagtendonk and Root 2003), and USGS vector data (water and hazard data horizontal accuracy ± 12 m; USGS 2005). *S* represents scores assigned to layer categories and *W* represents weights (percent influence) assigned to each of the layers (Bonham-Carter 1994). *W'* represents the percent contribution assigned by the machine learning algorithm, Maxent. The scores defined by Maxent are Logistic Output in terms of probability of presence (0.0 – 1.0).

S		**Slope**	**NDVI**	**Distance to open water**	**Distance to air hazard or building**
	W	0.40	0.40	0.10	0.10
	W'	0.80	0.14	0.05	0.01
3		0 to 5^{o}	25 to 109	n/a	n/a
2		5 to 7^{o}	109 to 125	> 12 m	> 39 m
1		7 to 10^{o}	125 to 140	0 to 12 m	27 to 39 m
0		> 10^{o}	> 140	0 m	0 to 27 m

Table 4.2 – This table represents a confusion matrix comparing two helicopter landing suitability models per-pixel where helicopters for Yosemite National Park, CA (N = 53,776,590). The expert model was created using criteria from Interagency Helicopter Operations Guide and the Maxent model used 100 training points where helicopters have previously landed in Yosemite National Park, CA (unique from the validated points). Overall agreement (in bold) is 90 percent.

	Maxent		
Expert	Not suitable	Suitable	Total
Not suitable	**0.76**	0.01	0.77
Suitable	0.09	**0.14**	0.23
Total	0.86	0.14	

Table 4.3 – This table represents an error matrix comparing two suitability models at 40 validated points where helicopters have previously landed in Yosemite National Park, CA. The expert model was created using criteria from Interagency Helicopter Operations Guide and the Maxent model used 100 training points where helicopters have previously landed in Yosemite National Park, CA (unique from the validated points). Both the expert and the Maxent model classified 95% of the validated points correctly (38 of 40 points).

	Maxent		
Expert	Not suitable	Suitable	Total
Not suitable	n/a	2	2
Suitable	2	36	38
Total	2	38	40

Figure 4.1

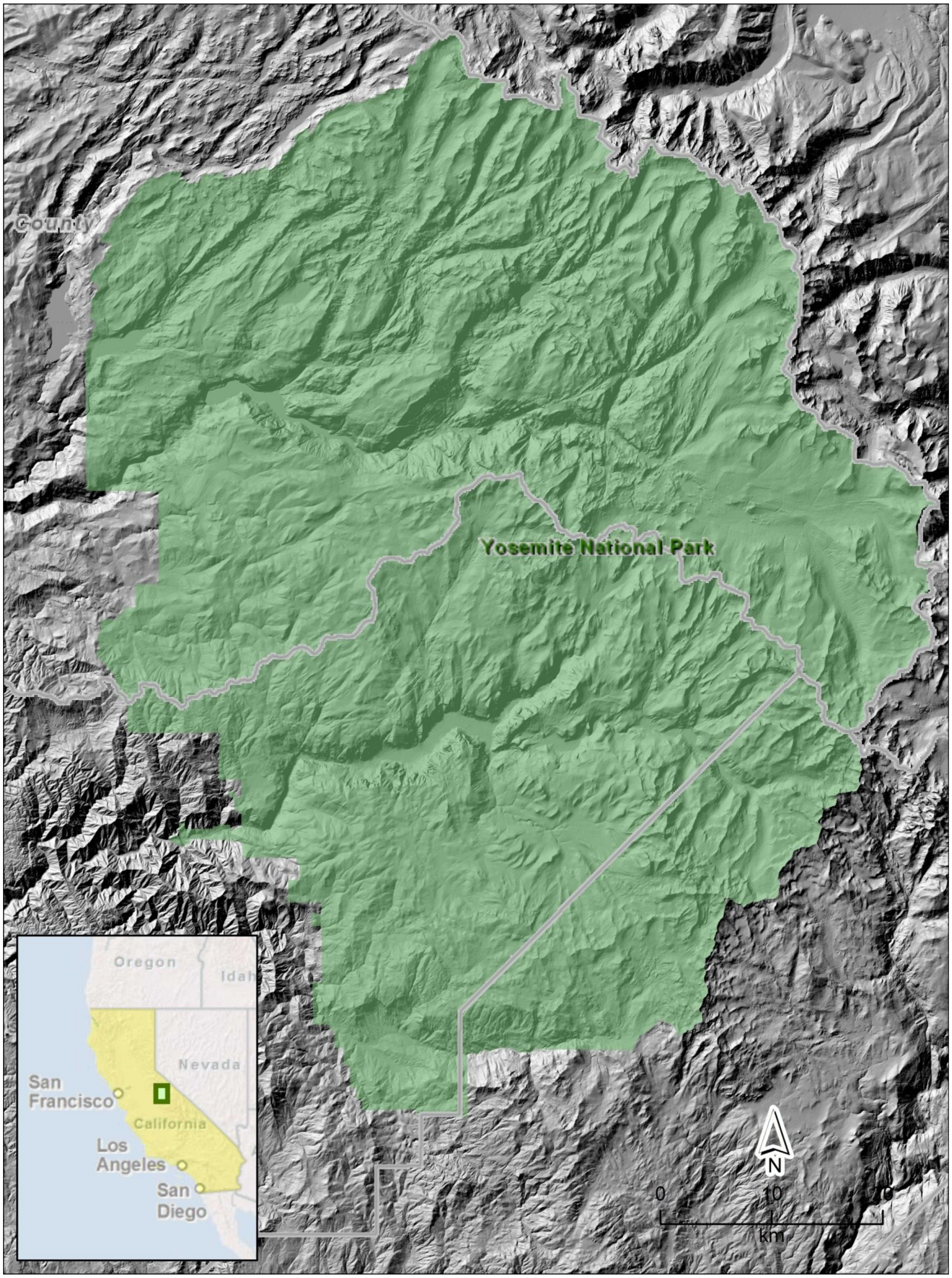

County
Yosemite National Park
Oregon
Idah
Nevada
San
Francisco
California
Los
Angeles
San
Diego
N
0
10
km

Figure 4.2

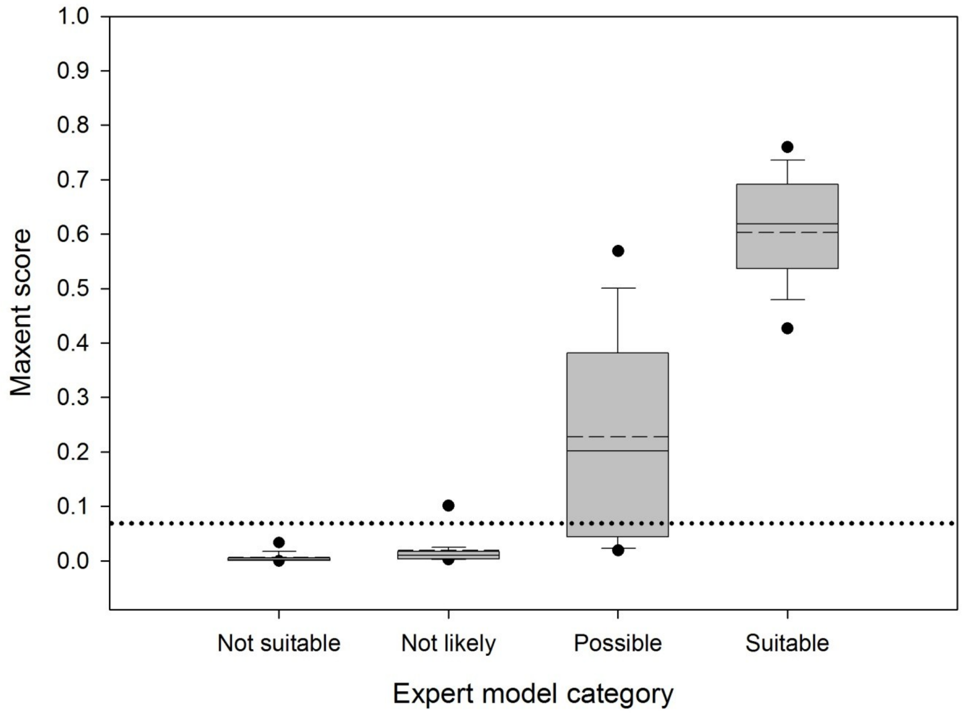
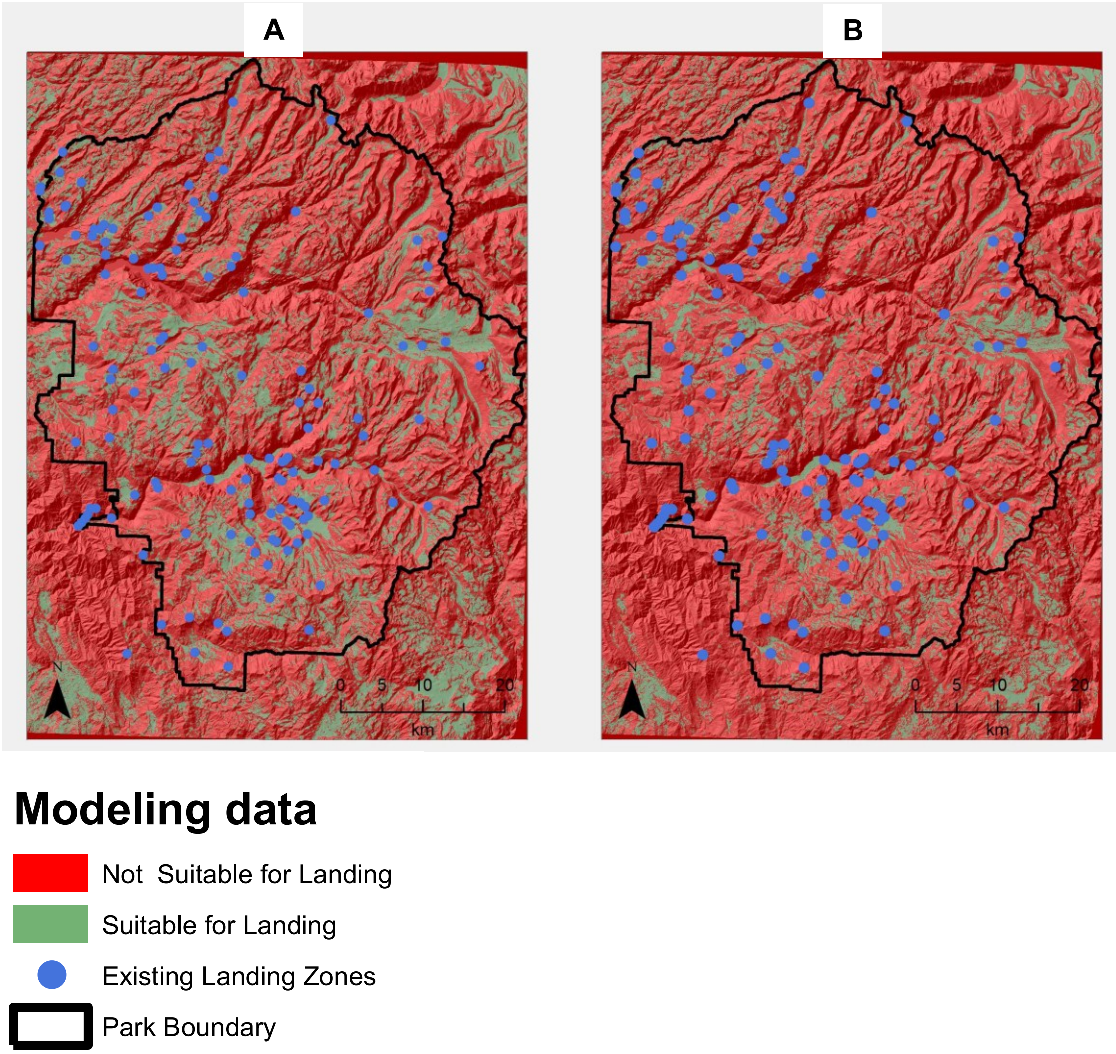

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Maxent score
Not suitable
Not likely
Possible
Suitable
Expert model category

Figure 4.3


A
B
0
5
10
20
km
N
Modeling data
Not Suitable for Landing
Suitable for Landing
Existing Landing Zones
Park Boundary

Overall Conclusion

The components of WiSAR operations are entirely spatial. Additionally, the prevention of WiSAR incidents requires close study of where and when incidents occur. I propose that the use of geographic information systems (GISystems) and spatial analyses could greatly enhance documentation and understanding of previous WiSAR incidents, as well as provide useful tools for saving lives in the future. This unique, but robust testing environment of WiSAR will allow for new discoveries within the spatial science disciplines. Based on case-studies in Yosemite alone, high-impacts topics such as Georeferencing from Text, Spatial Statistics, Time Geography, Search Theory, Location Science, Expert Modeling, and Machine Learning Algorithms were covered and presented to the GIScience and Geography communities.

If Geographic Information Science (GIScience) is the theory behind the development, use, and application of geographic information systems (GISystems), then WiSAR is an ideal topic for GIScientists to study. I have examined the spatial components of WiSAR, reviewed pertinent literature, used cutting edge GIScience techniques to solve WiSAR problems, and provided a framework for future research in WiSAR GISystems with broad implications for GIScience. This dissertation contains an Introduction to the central topic, and chapters that highlight in-depth research investigation in GIScience.

This dissertation presents preliminary, but compelling evidence that there is a universal need for GIScientists to address search and rescue problem solving and for WiSAR managers to adopt GISystems for apparent uses. Furthermore, I state that the future uses of GISystems in WiSAR are seemingly limitless. However, due to the nature of WiSAR as an emergency operation, technological efforts should be prioritized based on the potential for solving fundamental spatial questions that can be validated. For this reason I conclude that significant research efforts be directed towards the planning and operations related to searching for missing persons. The applications for time-geography, remote sensing, location science, and probabilistic modeling are very apparent in missing person search operations - and the process would greatly benefit researchers in these disciplines. Furthermore, I have found that GIS research related to preventing severe incidents through PSAR should be investigated further in collaboration with experts in spatial epidemiology. Finally, this dissertation is an evidence-based call to action for exploration of WiSAR in a GIScience context and for geographers of all specialties to get involved with their local search and rescue community.