SIMULATING UNCERTAINTY IN VOLUNTEERED

GEOGRAPHIC INFORMATION

by

David Nicosia, B.B.A., M.A.G.

A dissertation submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
with a Major in Geographic Information Science
December 2013

Committee Members:

F. Benjamin Zhan, Chair

Yongmei Lu

T. Edwin Chow

Robert K. Lyons

ACKNOWLEDGEMENTS

I would like to thank Dr. F. Benjamin Zhan for his insight and encouragement and the members of my dissertation committee, Drs. Yongmei Lu, T. Edwin Chow, Robert K. Lyons and James Kimmel for their thoughtful review and valuable suggestions. For sharing her deer expertise and assistance I would like to thank Dr. Susan Cooper. For the hours and hours of effort they contributed to this project I owe a debt of gratitude to the camera station hosts and many volunteer deer counters from the Sierra Circle and Tanglewood neighborhoods. I thank the Graduate College for funding through an immensely helpful dissertation research stipend. I am grateful to Remme Corporation for support during this process and for generously funding the volunteer incentive program.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

ABSTRACT

Facilitated Volunteered Geographic Information (VGI), crowdsourced data that includes a geographical reference that is solicited for a specific purpose, holds great promise for environmental monitoring, yet a major limitation of VGI is unknown information quality. Prior to initiating a VGI project, it is difficult to know if the collected data will be useful for the intended project purpose. This research explores the use of computer simulation to inform the design and implementation of a facilitated VGI project, specifically an urban neighborhood white-tailed deer survey. The project was conducted in two phases; first, a computer simulation phase, and second, a simulation validation phase including a VGI neighborhood deer count.

During the simulation phase of the project five different data collection methods were tested, each subject to various types of uncertainty, including observation location uncertainty, distance estimation uncertainty, and deer detection and classification uncertainty. Methods were tested under permutations of four levels of volunteer participation and three levels of deer density. Additional simulation refined and optimized the most promising data collection methods. Simulation results suggested a neighborhood counting protocol based on predefined observation areas with focused counting times to increase observed area, and the inclusion of zero-deer observations, that is, reports of areas searched that did not contain any deer.

During the simulation validation phase of the project, results from the simulation phase guided development of the facilitated VGI neighborhood deer count. The 28-day volunteer deer count was conducted in October, 2012 in two adjacent neighborhoods in San Marcos, Texas. Aggregate results from volunteer observations were used to estimate the neighborhood deer population. Concurrent with the volunteer deer count, an Infrared Triggered Camera (ITC) deer survey, a scientifically accepted survey method, was conducted in the same area. The VGI population estimate was 72% of the ITC population estimate. Although the volunteer population estimate fell outside of the targeted range of 75% - 125% of the ITC population estimate, simulation was nonetheless useful for testing alternative data collection procedures, optimizations to data collection procedures and the relative performance of those procedures under differing conditions of deer density and participation. Simulation results also informed interpretation of VGI results, but simulation was not useful for predicting volunteer behavior or participation level. This research introduces the use of computer simulation to inform and improve the design and implementation of facilitated VGI initiatives.

1. INTRODUCTION

Volunteered Geographic Information (VGI) is a topic of keen interest among Geographic Information Science (GIScience) researchers because it represents a departure from traditional patterns of geographic data collection. VGI is georeferenced information that is created and shared, typically on the Internet, by ordinary people with little or no formal training in geographic data acquisition and information production. It is the result of the combination of a shift in Internet and popular culture towards greater interactivity and a dependence on user generated content, generally known as Web 2.0, along with increasingly easy to use geolocation and map-making technologies including handheld GPS receivers, GPS enabled smart phones, digital Earth viewers like Google Earth, and “slippy map” mash-up libraries like the Google Maps API. This departure from traditional methods of geographic information creation either by authoritative entities like the U. S. Geological Survey or by commercial data producers like NAVTEQ has prompted a range of academic questions: What motivates people to do this? What kinds of information do people create and share? What does this mean for traditional geographic information providers and how they share data? What do these changes mean for individuals and for society? And, can VGI be trusted, and if so, for what purpose?

Answers to the questions of whether VGI can be trusted and for what purpose (Goodchild 2009, Goodchild and Glennon 2010, Haklay 2010 , Flanagin and Metzger 2008) are constrained by issues of uncertainty and spatial data quality within VGI. Uncertainty is widely studied within GIScience as well as in other disciplines. While there are several very good descriptions of uncertainty within the GIScience literature (van Oort 2006), in this work the definition and typology of uncertainty follows that of

McEachren et al. (2005). Uncertainty includes both objective measures of error including accuracy and precision but also other characteristics associated with spatial data quality, like completeness, consistency, lineage and currency. Particularly relevant for VGI, MacEachren et al. (2005) also includes characteristics of credibility, subjectivity and interrelatedness from the field of geospatial intelligence that reflect characteristics of the source of the volunteered information. If uncertainty can be managed, for example by screening data or by screening volunteers (Seeger 2008,Haklay 2010), VGI may offer a new and unprecedented resource for scientific investigation (Goodchild 2007). Research shows that VGI can be as accurate as data created by authoritative data sources (Haklay 2008, Girres 2010), but in these cases, a pre-existing reference dataset of higher quality already existed for comparison. For entirely new datasets, a prime opportunity for VGI, the absence of a pre-existing reference dataset remains a challenge.

One area of particular interest for VGI is environmental monitoring. It is theorized that as the number of volunteers increases, the potential spatial and temporal resolution of observations offers not only an opportunity to better understand our environment, but also may provide a timely indicator of emerging environmental problems (Goodchild 2007). "Citizen Science" programs like the Christmas Bird Count (National Audubon Society 2012), EPA Water Quality Monitoring (Environmental Protection Agency 2012), Sudden Oak Death (Connor et al. 2011) have engaged ordinary people in the collection of environmental data for scientific purposes for many years and Public Participation GIS literature often depict environmental monitoring as an example of public involvement in environmental resource management (Gouveia 2008), but these programs have relied on relatively limited groups of focused volunteers, usually with some formal training and occasionally with embedded data quality controls (Pfeffer and Wagenet 2007). Efforts have been made to integrate contributions of concerned citizens into Public Participation Geographic Information Systems (PPGIS), however, data quality has remained a primary concern, along with other issues of unequal access to technology or training (Gouveia 2004). Greater ubiquity of enabling technologies, for example devices capable of collecting geographic location with high accuracy and precision coupled with portable sensors, and larger numbers of observers with limited scope of required domain expertise and robust data screening procedures will help address the problems of uncertainty in volunteered data (Goodchild 2007, Gouveia 2008, Haklay 2010). All of these methods may be effective at improving VGI, but none help determine, a priori, that VGI will be useful.

Seeger (2008) differentiates "facilitated VGI" that is solicited, organized and shepherded by some controlling entity from other forms of VGI. Geotagged photos uploaded from a smartphone to Flickr and placed on a map are different than OpenStreetMap where there is a specific objective (albeit a broad one, "to map the whole world") (OpenStreetMap contributors 2012), a formal data model, elaborate editing rules, and sophisticated user interaction tools. An entity acting as facilitator offers a means of gate-keeping -- of controlling one or more aspects of compiled data quality, for example, through screening data values, screening contributors, or even using the community to screen contributions (Flanagin and Metzger 2008). One of the key elements of citizen science programs is having clearly stated objectives and research questions and a clear understanding of how collected data will help answer the questions or contribute to the objectives (Silvertown 2009). This concept is particularly relevant for environmental monitoring with VGI. The facilitator of VGI has a certain professional obligation to ensure an acceptable level of quality or usefulness of data, even if only for fitness of use (Harvey 2007). If one were to propose, for example, a wildflower spotting VGI application whose purpose is to map the emergence of wildflowers in the spring, then the facilitator must ensure that the collected data meet the objectives of the program, whatever they may be. If the primary objective is to raise awareness of local native wildflowers, then a strategy that enables the greatest number of people to participate is a good one, however, if the objective is to create an accurate spatiotemporal map of wildflower emergence, then some attention is required in validating the plant species, date, and location of each observation. If the later is the primary objective and it is unlikely that volunteers can or will produce useful information, then the program should be re-evaluated. How can one determine if a VGI initiative will produce useful results? One possibility is simulation.

Simulation is widely used to investigate scenarios or predict future outcomes particularly in situations where it is difficult or impossible to test alternatives in the real world (Ahola, et al. 2007). Fishery ecologists, for example, use simulation to evaluate alternative catch limits prior to the season in order to help ensure sustainable fish populations (Cooke 1999, NOAA 2011), and hazards researchers use simulation to investigate hurricane evacuation strategies to improve evacuation transportation networks (Chen and Zhan 2008). If uncertainty can limit the usefulness of VGI, can simulation be used, prior to collecting VGI, to explore potential uncertainty in order to promote useful information? Might simulation results help improve collection or screening of VGI or perhaps suggest abandoning the VGI project altogether? Within the context of planning a facilitated VGI initiative this work presents 4 specific research questions:

- Can simulation guide what geographic information is collected and how?
- Can simulation reveal the influence of potential uncertainty on usefulness and provide a method to reduce its impact?
- Can simulation show the effect of participation on VGI usefulness?
- Do simulation results correspond to actual VGI results?

While the applicability of simulation to investigate uncertainty in potential VGI is broad and theoretical, here it is examined in the context of an urban white-tailed deer (*Odocoileus virginianus*) population. Some residents enjoy the closeness to nature offered by large numbers of deer freely roaming through town; others find the ungulates to be a costly, destructive nuisance. The question of appropriate deer population density is somewhat subjective, but communities that choose to address the issue must first answer two difficult questions: How many deer are too many, and how many deer do we have? Wildlife ecology offers insight into both questions, but the second question is particularly important. Scientific methods of estimating deer population abundance are expensive and the most commonly used technique, spotlight survey, produce population estimates of questionable usefulness due to high variability and resulting broad confidence intervals. In this research, possible uncertainty and participation level in a facilitated VGI urban deer count initiative were examined in order to address the stated research questions. First, simulation was used to examine issues of uncertainty and participation in order to facilitate the design and implementation of an actual VGI initiative. Next, the VGI initiative was conducted concurrently with a scientifically accepted Infrared-Triggered Camera (ITC) deer survey to allow comparison of VGI results with ITC survey results.

The prevalence of VGI and facilitated VGI is expanding at a rapid rate, and VGI offers considerable potential for scientists and for society, but the ease with which VGI can be created or collected could be problematic if what is collected is not useful. This research addresses the widely reiterated question, "For what is VGI useful?" with a generalized and theoretical answer -- an approach and method to test if any specific potential VGI program can produce useful information. Although the term VGI is relatively new, the problem of uncertainty in data collected by ordinary users, for example in citizen science programs, is not. The approach presented here is equally applicable to the citizen science domain. This approach provides a baseline for anticipated quality in VGI which is a valuable step in research concerning VGI information quality. By applying a typology of uncertainty to VGI this research demonstrates an approach to the examination of VGI uncertainty by decomposing uncertainty into specific elements each of which can be investigated individually. Finally, by first verifying and further enhancing the usefulness of VGI through preliminary simulation, it is hoped that the promise of more informed environmental management is realized.

Literature Review

VGI Definition

The term "volunteered geographic information" was coined to define the emerging phenomenon of "the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information" (Goodchild 2007). VGI is important within the discipline of Geographic Information Science (GIScience) because it represents a departure from traditional forms geographic data creation. Budhathoki et al. (2010) offers the definition “a complex GI ecology resulting from different actions and interactions that actors engage in to serve their underlying motives” arguing that not all of the contributors are voluntary and some contributors could be GI experts, however, this definition is no more definitive of VGI and is much more ambiguous. It could as easily describe Participatory GIS, Public Participation GIS, emergency response GIS and other known forms of collaborative GIS. The widely noted elements of a VGI definition include ordinary users (Goodchild 2007, Sui 2008, Coleman et al. 2009, Tulloch 2008), creating their own geographic information (Goodchild 2007, Coleman et al. 2009, Tulloch 2008) sometimes independently, sometimes collaboratively (Sui 2008,Budhathoki et al. 2008) under their own authority (Goodchild 2007, Elwood 2010, Coleman et al. 2009, Tulloch 2008, Budhatoki et al. 2008).

Seeger (2008) differentiates geographic information created spontaneously by individuals from geographic information that is solicited by a facilitator from the public as part of a planning or design process. In the later case, solicited information is usually limited to a specific topic and one geographic extent and is intended to provide the public a means to comment on and participate in the design and planning processes. This structured collection of topical geographic information is called facilitated-VGI. While Seeger’s definition of facilitated-VGI is rather specific, the role of facilitation in VGI can be thought of along a continuum. At one end is completely spontaneous and independent geographic information perhaps enabled by technological coincidence, for example, a photo uploaded and mapped in Flickr (Yahoo! Inc. 2012) simply because the smartphone which captured the image also happened to geotag the photo, or a reference to a specific restaurant in a Twitter tweet. At the other end is a project like OpenStreetMap (OpenStreetMap contributors 2012) where there are specific objectives, a complex data model, a sizeable enabling infrastructure and sophisticated rules regarding contribution (Goodchild 2007). Here the term facilitated-VGI will be used in the spirit of Seeger's definition to mean VGI that is solicited for a specific purpose, even though the purpose is not strictly planning or design.

Enabling Context: Web 2.0, NeoGeography and Geolocation

The emergence of VGI is widely attributed to Web2.0 (O'Reilly), NeoGeography, and the widespread availability of location technologies like GPS receivers and other portable devices (Turner 2006). Central to the concept of Web2.0 is user generated content, the blending of user provided information into the web experience. Examples include user ratings on retail web sites, blogs (short for web logs), wikis and social media. VGI is user generated content that contains a georeference (Goodchild 2007).

NeoGeography springs from the user generated content culture of Web2.0 in the presence of easy to use mapping tools and technologies. There are a number of thorough treatments which discuss the subtle cultural and technical circumstances that produced the "GeoWeb" as it exists today and the implications of "ordinary" users being able to create and map their own geographic data (Turner 2006, Haklay, Singleton, and Parker 2008, Goodchild 2009 (NeoGeography), Crampton 2009, Hudson-Smith et al. 2009). GIScience, Geographic Information Systems (GIS) and mapmaking have traditionally been inaccessible to non-practitioners requiring expensive hardware and software and considerable training, however, advances in technology may mitigate required mapping expertise, cartographic skill, specialized equipment and subject matter expertise

(Goodchild 2008 whit). With user-friendly mapping technology within reach, novices are empowered to produce their own maps, an ability described as the "wiki-fication" of GIS (Sui 2008). What maps will be made and how they will be used are subjects of GIScience research.

Positioning technologies are important not only for providing a georeference for VGI, but also for Location Based Services (LBS) and other geographical studies (Lu 2012). Not only is positioning technology a common component of virtually every new mobile phone and many portable devices, the accuracy of positioning continues to increase (Zandbergen 2009,2011).

Use of web technologies is rapidly evolving. A number of patterns of use and a number of questions regarding ownership, use and reuse of data (both authorized and unauthorized), are emerging, yet, it is too soon to suggest a "best practice." (Batty 2010)

VGI Research Agenda

As a relatively new and rapidly evolving area of research, there are many facets of VGI to be examined, but among the most widely noted are, motivation – why people would engage in VGI creation and for what purpose, societal impact – what this activity means for society, and VGI uncertainty – the quality of VGI and its trustworthiness.

Motivation

Budhathoki et al. (2010) presents a framework for understanding VGI that is divided into Motivational, Action and Interaction, and Outcome arenas. Literature regarding "volunteerism" and "leisure" is applied to the Motivational arena and literature concerning "online social production of knowledge" is examined for insight into the

Motivation, Action and Interaction, and Outcome arenas. In short, motivation for participating in VGI could be complex and may be explained and better understood by looking at these other literatures.

Coleman (2009) draws parallels for VGI motivation from motivation documented from wikis, free and open source software and other user contribution systems. Among the enumerated motivations are: altruism, professional or personal interest, intellectual stimulation, protection or enhancement of personal investment, social reward, reputation, creative outlet, and pride of place. He also recognizes motivations such as mischief, alternative agenda, and malice and/or criminal intent.

Societal Impact

Elwood (2009) is concerned with the social and political impacts of the GeoWeb, the combination of Web2.0, VGI and NeoGeography. The ability of new actors to use previously inaccessible authoritative data and to collect and construct their own data creates the potential for new forms of activism. As an example, Elwood (2008b) points to NGO's and grassroots groups which until recently have not had equal access to geographic information for a number of reasons. Now that technology enables their use of information, available institutional information may not be suitable for their use; however, they are now enabled to produce their own geographic information. Early GeoWeb social issues include: who is included by these developments, who is excluded, and the way in which the authority of asserted geographic information is used to gain social or political advantage (Elwood 2010).

Budhathoki et al. (2008) suggests the role of user of spatial data infrastructure shifts in light of VGI from a consumer-only perspective to a producer/user or "produser" perspective requiring the spatial data infrastructure to become more flexible to accommodate the needs of the produser. Others recognize the potentially shifting role of authoritative data providers from strictly data creation to data creation and moderation (Coleman 2009).

Uncertainty

There are many potential sources of VGI, but mechanisms to ensure quality, detect and remove errors, and establish trust is needed (Goodchild 2007b). Empirical studies have examined the positional and attribute accuracy of VGI (Haklay 2010, Girres 2011) but uncertainty in VGI may take on additional quality characteristics, timeliness and credibility for example, when used in emergency response (Goodchild and Glennon 2010). In the same way that VGI is crowd-sourced, perhaps VGI quality can be determined by the users/producers of VGI (Goodchild 2008a, Grira 2009). An additional approach uses credibility and trust relationships between users and producers of VGI to measure reliability (Bishr and Mantelas 2008). Ultimately the quality of data must be sufficient for its intended use. The ability of an individual land owner to collect and use her own soil data illustrates at least one scale at which VGI can and is useful (Goodchild 2008b). Diversity of volunteers contributes to heterogeneity in uncertainty, so blanket statements about the quality of data could be less appropriate than local measures of quality. Rather than trying to eliminate uncertainty, Haklay (2010) suggests to work to understand and live with inherent uncertainty in VGI.

Environmental Monitoring

The practice of concerned citizens collecting environmental data in order to create change in environmental management is well documented (Gouveia et al 2004). In fact, it is argued that there may be no other alternative than VGI to collect certain types of environmental data at the temporal and spatial scales necessary for effective decision-making (Goodchild 2007, Whitelaw et al. 2003). Gouveia (2008) frames the emergence of VGI within existing community environmental monitoring initiatives, and provides a conceptual framework for implementing VGI initiatives based on past experiences with community environmental monitoring.

Citizen Science

Prior to professional scientists, all science was citizen science conducted by individuals engaged in some other occupation but with an interest in some area of science, like Benjamin Franklin and Charles Darwin. Contemporary citizen science is driven by technology, the viability and reliability of citizen collected data, and funding sources' increasing emphasis on outreach (Silvertown 2009). Connors et al. (2011) critically examines the intersection of VGI and citizen science in an application of environmental monitoring to first produce a conceptual model of the users, interaction between users and data, and the types of information produced, then refines an existing environmental monitoring system. For environmental monitoring, recommendations are made to include multiple sources of information, both from a technological standpoint using the web, mobile phone apps, Twitter feed and others, as well as a conceptual standpoint engaging multiple groups of trained and untrained individuals and expert scientists.

VGI and Simulation

Kuhn (2007) suggests VGI could be useful to inform simulation, for example, for validation of models, to determine initial conditions and to ground truth simulation results. Birkin (2011) illustrates the use of crowd-sourced attitudes about traffic congestion to calibrate transportation simulation models. Rinner (2008) first collects non-geographic user generated content, then georeferences it by key words to evaluate how a prototype map-based tool might have worked. Each of these uses presupposes the existence of VGI and proceeds with further modeling and simulation. What is examined here is a reversal of these methods, investigating the prospect of collecting useful VGI through simulation.

Research Method

This research was conducted in two phases, a simulation phase and a VGI phase. During the simulation phase computer simulation was used to investigate the influence of potential uncertainty and levels of participation on a planned VGI deer count initiative. Using a sequence of tests various methods of collecting and screening data were simulated along with various levels of participation resulting in a set of guidelines which were then used in the VGI phase to design and implement a facilitated VGI initiative. VGI results are compared with simulation results and with the results of an infrared-triggered camera (ITC) deer survey conducted concurrently with the VGI initiative.

This section continues with some background information regarding the origin of this project followed by a few details regarding the simulation phase and the VGI phase. The final section of this chapter discusses the contribution of this research to the discipline of GIScience. Subsequent chapters will describe in some detail simulation, the

VGI initiative and the ITC deer survey. The concluding chapter discusses the results of the preceding chapters in view of the stated research questions, highlights some additional observations and outlines possible future work.

Background

Several neighborhoods in San Marcos, Texas, USA had expressed concern to the city about the urban deer population. Previously the city had not taken formal action on the complaints, but in the summer of 2011, as the first step in addressing the problem, the city arranged for a professional biologist to conduct a scientific deer survey in order to estimate the total deer population and population structure quantified as the ratio of the number of bucks to does to fawns (Coolidge 2011). As an alternative to the scientific survey approach, would it be possible to simply ask residents to count and report deer, an example of Volunteered Geographic Information? Given this proposition, what kind of information should be solicited? How should the information be collected? How should it be verified and how can its quality be assessed? How many people must respond for the information to be useful? Perhaps most importantly, what can be done with the collected information? These questions are not unique to this situation; rather they are typical in deciding whether or not to solicit volunteered information through a facilitated VGI initiative. The focus of this work was to investigate these questions first through simulation.

Simulation Phase

During the simulation phase a computer model was used to simulate volunteer deer observation under various conditions. A number of different methods of capturing and reporting deer observations were tested each subject to various types of uncertainty.

Other factors tested included various levels of participation, data screening and filtering techniques and optimizations to reporting methods.

Simulation

Simulation is performed using computer modeling. The modeling carried out in this project is static – meaning situations are examined at a single point in time rather than over a sequence of times, and agent-based – meaning objects in the problem domain interact with each other to produce the result (Goodchild 2003). In the agent-based model produced in this project, volunteer objects report observations of deer group objects, an example of VGI. Volunteer objects could be stratified based on a number of demographic characteristics; however, here they are homogenous and differentiated only by reporting device, one of Desktop computer, Laptop computer, Tablet or Smartphone, and position reporting method, either Map interface or Geolocation, which are both determined based on fixed probabilities. Since all data include uncertainty, various types and amounts of uncertainty in volunteer observations are introduced in the model to help reveal the effect of uncertainty on observation results. The Monte Carlo Method (Metropolis and Ulam, 1949) of simulation is widely used across a diverse range of research to investigate computational models. By iteratively executing a model over a very large number of model input scenarios, the model output can be examined as a preliminary test of model fitness, to test hypotheses that could not otherwise be tested, for sensitivity analysis, and for uncertainty analysis (Goodchild 2003, Helton et al. 2006). Simulation provides an opportunity to examine potential outcomes given different assumptions, alternative agent behavior, and varying inputs all in a zero-risk environment. Here simulation is used to evaluate the usefulness of a potential VGI program given a range of uncertainty and participation inputs. Simulation provides a path towards a more effective facilitated VGI initiative.

Types of Uncertainty

It may not be possible to know the full range of uncertainty that could be present in proposed VGI, however, it is possible to identify many of the rather obvious and perhaps some of the less obvious types of errors or problems that could be encountered. MacEachren et al. (2005) presents a useful typology of uncertainty for VGI which serves well as a reference for brainstorming a wide variety of context-specific problems. The typology includes such components as Accuracy/Error, Precision, Completeness, Consistency, Lineage, and Currency but also Credibility, Subjectivity and Interrelatedness from the domain of intelligence information assessment. The Accuracy and Error type is described as the difference between the observation and reality and includes measurement and/or estimation errors. This category includes both positional-accuracy as well as attribute-value accuracy. The Precision uncertainty type includes uncertainty resulting from the exactness of the measurement or estimate, typically derived from the parameters of the device or procedure. Completeness describes the extent to which the phenomenon has been observed in totality. Consistency refers to the extent to which elements of information are in agreement. Lineage includes uncertainty as a byproduct of the sequence of processing steps to produce the information. Currency uncertainty results from the span of time between the time information is collected and the time it is used. Credibility refers to the reputation of the source of information. The Subjectivity type identifies uncertainty related to observer judgment and Interrelatedness refers to the independence of the source from other information.

Enumerating potential uncertainty guided by a formal typology of uncertainty helps ensure thoroughness. Not all types of uncertainty need to be included in simulation, but the ability to choose which types should be included from a comprehensive list is helpful. Starting with Accuracy and Error, specifically positional accuracy as well as attribute value accuracy, a volunteer could report an observation at the wrong location, or could get the location right, but the deer count or classification wrong. The volunteer could incorrectly report the area associated with an observation either by making an erroneous distance estimate or area assignment. The volunteer could associate an incorrect time or date with an observation. Deer detection probability changes by time of day as well as by season. Age and sex classification accuracy change by season. Deer counts could be intentionally inflated or deflated based on volunteers' attitudes about deer. Although the term precision is often associated with repeatability, the Precision uncertainty type discussed here includes uncertainty resulting from the exactness of the measurement or estimate. Observations in which volunteers are asked to report the distance to the observed group of deer to the nearest 10 yards, for example, may have greater precision uncertainty than if the distance were more exactly estimated. Precision uncertainty could also result from large groups of deer which are difficult to count and as a result turn into, "about a dozen" or from an inability to classify deer, for example, during a certain times of year. Temporal precision uncertainty could result from observations made at one time being reported at another time. Completeness describes the extent to which the phenomenon has been observed in totality. An example of this type of uncertainty in the present context might be failing to see every group of deer in an area for a specific time window or failing to report each category of deer.

Consistency includes situations where elements of information are not in agreement, for example incorrectly reporting 2 bucks, 4 does and 3 fawns as a total 12 deer in the group. Another possibility is reports of deer observations by a single observer at locations too far apart in space and too close together in time to be reasonably considered feasible. Lineage includes uncertainty as the result of the sequence of processing steps to arrive at a result. It is unlikely that observers will actually report observations in real-time, rather they will record them and later enter the information. The errors and uncertainty that result from this process like transposition errors and illegible writing are included in this category. Currency uncertainty results from the span of time between when information is collected and used. Delayed observations, that is, observations that are reported after VGI has been processed, analyzed and used, are of little value. Credibility refers to the reputation of the source of information and could appear in the current context as volunteers who consistently over report or under report deer. An extreme observation reported by an observer who routinely makes normal observations may be more believable than one made by a sporadic or new observer or a series of extreme reports may be less believable than a single extreme report. The Subjectivity type identifies uncertainty related to observer judgment, for example reporting 2 groups of 3 deer or 3 groups of 2 deer or the classification of a juvenile male deer as a fawn or buck. Interrelatedness refers to the independence of the source from other information, for example, a father's report of deer observed by his daughter is less certain than his report had he also seen the deer. While it would appear neighbors independently reporting the same group of deer would be problematic, it could actually reduce uncertainty which could be a great advantage if harvested. On the other hand, if all observations in one area are by a single observer, interrelatedness uncertainty must be high. Table 1.1 enumerates various types of uncertainty. Of these types of uncertainty, several were selected for inclusion in simulation.

Table 1.1 Uncertainty Types

Uncertainty	**Uncertainty Type**
Position error – deer	Accuracy/Error
Position error – volunteer	Accuracy/Error
Incorrect area association	Accuracy/Error
Incorrect distance estimate (distance magnitude)	Accuracy/Error
Distance rounding (10,20,30 yds)	Precision
Incorrect distance units	Consistency
Incorrect time/date	Accuracy/Error, Currency
Wrong number of deer	Accuracy/Error, Consistency, Subjectivity
Deer detection	Accuracy/Error
Deer classification	Accuracy/Error, Completeness, Consistency, Subjectivity
Deer grouping (2 sets of 3 vs. 3 sets of 2)	Subjectivity
Duplicate reports of group	Interrelatedness
Fail to observe	Completeness
Fail to report	Completeness
Time/space mismatch	Consistency
Report delay	Currency
Transcription, data entry errors	Lineage
Count inflation/deflation	Accuracy/Error, Credibility
Strong volunteer area affinity	Interrelatedness

VGI Phase

During the VGI phase, insight developed through simulation was used to inform the design and implementation of the VGI deer count, but there were a number of other considerations for the volunteer count.

Volunteer Count

Residents could provide both qualitative information regarding attitudes towards human-deer interactions, but also quantitative information including the density and structure of the deer population and some economic measures of the costs of, for example, property loss or damage due to deer. As a first step towards implementing the VGI program, the specific objectives of the program were enumerated and acceptability criteria established to provide a measure of usefulness. While qualitative information is important, this project focused on quantitative information.

Perhaps the most obvious approach for residents to provide volunteered deer observations is to simply report the date and time, number and location of deer they observe. There are many examples of this type of observation in citizen science programs, however, it represents a form of the presence-only dilemma, where although one may have observations where the phenomenon occurs, there is no information about areas that were not observed at all, or areas that were observed and did not contain the phenomenon. As Pearce and Boyce report, "We are unaware of any application explicitly modelling abundance given presence only. " (2006, p409) Since the objective of the VGI program is to come up with an estimate of deer population, care must be taken to ask residents about the search area in which deer were observed.

One approach to collecting both deer information and search area information is to use the distance between the volunteer and the observed deer or the volunteer's field of view distance as an indicator of search area. For example, if a volunteer were to report her position and the position of a deer, the distance between the two could be used to estimate a searched area. There could be many variations on this theme. A second approach might simply use pre-defined search areas with which volunteers associate deer observations.

Within these approaches, there may be alternatives for collecting certain types of information, for example, the volunteer's location could be determined by a location sensor, for example, the GPS receiver in the user's mobile phone, or it could be determined by the user marking a position on a web map. Each of these techniques is subject to different types and amounts of uncertainty.

Scientific Survey

The objective of the VGI program was to produce an estimate of population density with accuracy as good as or better than a scientifically accepted method. Three methods were considered, two recommended by Texas Parks and Wildlife Department (TPWD) for use by the general public, and a third approach based on distance sampling. The TPWD methods are based on scientifically accepted spotlight/cruise survey method and infrared-triggered camera method (ITC). In the infrared-triggered camera method (Oetgen, Lambert and Whiteside 2008)(Jacobson, et al. 1997), images from infrared-triggered wildlife cameras are examined in order to identify specific bucks within all images. The number of uniquely identified bucks relative to the total number of observed bucks provides a population estimate multiplier. The product of the population estimate multiplier and the total number of observations of does and of fawns provides population estimates for does and fawns respectively. The total population estimate can be divided by the surveyed area to provide an estimate of deer density. While this technique could be widely deployed by residents to provide a basis on which to conduct an urban deer population study, the amount of effort and expertise required to analyze the images quickly becomes daunting. The technique simply may not scale well to a broad general public. McKinley et al. (2006) shows population estimate accuracy as high as 90% from a 14 day survey with a camera density of 41 ha/camera.

In the spotlight/cruise method (Jester and Dillard n.d.), the more common of the two TPWD recommended methods, a specific driving route is selected through the study area. Periodically along the route, lateral visibility measurements are made in order to calculate the entire visible area of the route. The route is then driven a number of times at night and deer observed with a spotlight along the route are counted and categorized. Deer density is calculated as the number of deer counted divided by the visible area of the route. This technique is widely employed by wildlife managers on both public and private lands due to its low cost and simplicity even though its empirical accuracy has shown to be rather limited. In one study using simulated deer in two habitat types, on average only 67 - 72% of deer were observed depending on habitat type (Whipple et al. 1994). A later study supported these findings showing a deer detection probability in a spotlight survey of less than or equal to 0.66 and inter-observer variability as high as 30% (Collier 2007). While age and sex classification was better than 90% accurate using simulated deer (Whipple et al. 1994), McCullough (1993) showed that classification ratios of spotlight surveys vary dramatically throughout the year, and that males are generally underrepresented and fawns are “greatly underrepresented” (McCullough 1982, p 968). A comparison of the spotlight and infrared-triggered camera survey methods on key deer showed a significant difference between the two methods with the spotlight method estimating less than half the population estimated by the ITC method (Roberts et al. 2006).

Another method, distance sampling, could be an alternative, however, one assumption of the technique is that *all* targets are observed and the distance from the observer to the target is measured accurately. Based on this technique, Koenen (et al. 2002) calculated population estimates of 405 (.87 deer/km2) in the summer to 1162 (2.5 deer/km2) in the winter on a 46,540 ha National Wildlife Refuge in Arizona (Koenen et al. 2002). The range of the 95% confidence interval during summer was 205 to 795 and in winter, 423-3204. The usefulness of an estimate where the true population could be as little as 50% of the estimate or as high as 200% of the estimate seems questionable, particularly in light of the difficulty in meeting technique assumptions and other evidence regarding spotlight deer detection probability. The ITC survey method was chosen for this project.

Acceptability Criteria

While scientifically accepted deer survey methods may not be completely accurate, their limitations and uncertainty are understood. Much less is known about a potential facilitated VGI initiative to count deer. The success of the initiative lies in the usefulness of the information that is produced and in this case the stated objective was to produce a population estimate that is as good, or better, than a scientifically accepted ITC method. In order to quantify useful VGI results, a VGI population estimate between 75% and 125% of the ITC population estimate is considered useful.

Contribution

This work contributes to GIScience literature in several ways. First, it introduces a new VGI dataset. Much of what is known about VGI has come from very few VGI datasets and primarily the OpenStreetMap (OSM) project. Although its value for academic research cannot be overstated, OSM is but one example of aVGI dataset. Second, this work highlights an opportunity to expand literature concerning uncertainty and error associated with practical VGI approaches, for example the use of web map interfaces or device geolocation. Literature is scarce that looks experimentally at the results when volunteers engage these technologies. Third, this project introduces simulation as a strategy to improve any facilitated VGI project. It is hoped that with improved VGI more informed environmental management may follow.

2. SIMULATION

Uncertainty limits the usefulness of VGI. Prior to conducting a VGI initiative, one can only make educated guesses as to the types and amounts of uncertainty that may be present during the initiative, but by examining potential uncertainty through simulation, the VGI initiative may be refined, reworked or even abandoned. Simulation offers a low risk method to test ideas about a VGI initiative before using volunteer effort on an initiative that could fail to meet user expectations and needs. Simulation provides not only a way to understand potential uncertainty but also a way to examine various approaches to handling potential uncertainty. The purpose of the present simulation was to model the process of neighborhood volunteers counting white-tailed deer subject to potential uncertainty in order to test and improve counting process effectiveness. While considerable effort went into creating the simulation, the paramount goal was to prevent wasted volunteer effort on an ineffective counting process that achieves less than useful results.

The simulation project was comprised of four sets of simulations. Each simulation set was used for a slightly different purpose, with each step building on the previous towards an optimized counting process or abandonment of the VGI initiative. Simulation Set 1 investigated the comparative performance of five variations of deer observation collection and reporting methods. Simulation Set 2 looked at the performance of the same five observation methods given several combinations of volunteer participation and neighborhood deer density. Simulation Set 3 tested for changes in performance of observation methods when steps are taken to reduce uncertainty. Finally, Simulation Set 4 evaluated optimizations to selected observation methods to improve method performance, reliability and usability.

Simulation Mechanics

Each simulation run consists of observers (model agents) representing neighborhood volunteers who observe deer distributed throughout the simulated study neighborhood. The areas in which observers look for deer are called observation areas. When deer are encountered in an observation area, observers report an observation including both the number of deer and a representation of the observation area where the deer were found. Observations from all observers are aggregated to determine an observed deer density. The observed deer density can be compared to the actual density of deer introduced into the simulation. The ratio of observed to actual deer density (OADR) is used as a primary metric for simulation results. An OADR equal to 1 indicates exact agreement between observed and actual deer density. As OADR increases from 1 the population is increasingly overestimated and as OADR decreases from 1 the population is increasingly under estimated.

Simulation Frame and Study Area

The simulation frame was based on a real neighborhood in San Marcos, Texas which was targeted for the VGI initiative. The study area neighborhood was selected based on the size and location of the neighborhood within the city limits, the presence of nearby green space areas and anecdotal evidence, confirmed through inspection, of the presence of urban deer. In order to create a simulation environment that was similar to the study area, 32 road segments and 107 houses were digitized from orthorectified aerial imagery of the target study area neighborhood. A buffer of 75 meters around roads was used to create a defined simulation area. The buffer distance of 75 meters reflects the maximum depth of lots in the target neighborhood.

Spatial Dependence

During simulation observers and deer were positioned in the simulation frame in a spatially dependent way. It was assumed that all observers are homogenous, that observers behave according to a set of stochastic rules and that observations are made either from within or around certain houses (observer houses) or along a road segment. In terms of probability, observers are more likely to be nearby a subset of houses or along a road.

Likewise, it was assumed deer are likely to be found in groups nearby certain houses (deer houses) that perhaps provide food, water or cover and along roads. Observer houses and deer houses are not necessarily the same set of houses.

The process for allocating either observer locations or deer group locations was the same, although different subsets of houses are used.

Point Allocation Process

The point allocation process is outlined below:

1. Select a subset of houses
2. Create a surface *dist* representing the shortest distance from each cell to the closer of the nearest road or nearest house in the selected subset
3. Select a large number of points *m* randomly within the bounding box of the study area
4. Attribute each point with the value of *dist* at its location
5. Order points ascending by *dist* value and select *n* points from the top of the list
6. Repeat steps 3 to 5 until a complete set of *n* points are selected that are also contained within the study area boundary

The point allocation process is shown graphically in figures 2.1 through 2.4.

Figure 2.1 Study area with subset of houses selected

Figure 2.2 Surface *dist* calculated as distance from selected houses and roads

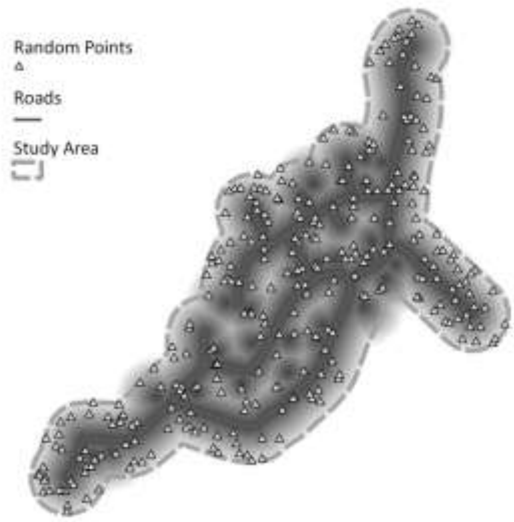


Figure 2.3 Large number of random points

Figure 2.4 Select *n* points with minimum *dist* value

To conserve time and compute resources during simulation runs, ten realizations of shortest distance surfaces *dist*, as described in steps 1 and 2, above, were calculated based on 10 random samples of houses. The probability of any single house being selected out of the total of 107 houses was P=.3. The samples included 37, 30, 31, 29, 30, 32, 30, 31, 35 and 25 houses. During a simulation run, one of these 10 probability surfaces was chosen at random to distribute observers, then, for each iteration of the simulation run, one of the remaining 9 surfaces was chosen at random to distribute deer.

Simulation Run

For each simulation run, first, observers are allocated over the simulation frame according to the selected distribution surface. Next, for each simulation iteration, the specified number of deer are allocated into groups and the groups are arranged in the study area according to the point allocation process described above. During each iteration, each of the observers "observes" deer and reports "observations" subject to various types and amounts of uncertainty. Observations are aggregated and the OADR, the ratio of observed deer density to actual deer density is reported. The set of OADRs for each iteration of the simulation is the result of a simulation run. Unless otherwise specified, each simulation run was comprised of 1,000 simulation iterations. During each simulation run, unless otherwise noted, the default number of observers was 40 and the default number of deer was 30. Deer were allocated in groups of 2 (P=.38), 3 (P=.18), 4 (P=.19), 5 (P=.16), 6 (P=.06), 7 (P=.02) and 8 (P=.01) deer. The buck to doe ratio was 1:3, and the doe to fawn ratio was 2:1.

Acceptance Criteria

VGI initiatives require volunteer effort and cooperation, and although many times these resources are free of charge, they are, in fact, finite resources and should be used wisely. Technology enables VGI projects, but the facilitator of a VGI initiative has a duty to volunteers to ensure the project has a high probability of producing useful results. What qualifies as useful is subjective but should relate directly to the purpose of the project. If the purpose of an invasive species VGI project is to raise awareness of local invasive plants, then a volunteer initiative that is easy and inclusive allowing a very large number of people to participate is useful. The usefulness of the result could be measured in terms of the number, demographic composition or spatial distribution of volunteers, or through changes in awareness perhaps measured through pre- and post-project questionnaires, or through project website usage statistics. On the other hand, if the purpose of a volunteer deer count is to produce a reliable estimate of deer population abundance, then a real-world benchmark for the performance of the population estimate must be established. Put simply, prior to the project there must be a qualification of what constitutes a "reliable" population estimate. For this project, acceptable performance for the VGI deer survey method is OADRs between 0.75 and 1.25 in 95 out of 100 cases.

Simulation Set 1: Input Methods and Uncertainty

Input Methods

The first question in designing a VGI initiative is determining what information the initiative is to produce and how it is to be used. The measure of the information's usefulness is its suitability for use for the intended purpose. The second question is establishing what data the volunteers will gather and how they will report it, and the third question is identifying how the reported data will be processed into the required information. In the present project, the desired output is an estimate of deer abundance in the study area neighborhood. If volunteers report both the number of deer and the area searched, the unifying representation of deer density (animals per unit area) can be used not only to aggregate observations but also to extrapolate a population based on the size of the study area.

A volunteer's observation area could be represented in a number of ways, each requiring slightly different input data. For example, a volunteer could, in addition to the count of deer, provide an estimated distance representing the average radius of the volunteer's field of view. The observation area could be calculated as the area of a circle of the provided radius. If the distance is provided along with the coordinates of the volunteer at the time of the observation, a spatially explicit representation of the observation area is possible. Alternatively, the volunteer could define spatially explicit observation areas and then report deer associated with the observation area containing deer. Which of these methods will be most effective is difficult to know prior to the start of the VGI project, but makes a suitable subject for simulation. Simulation Set 1 investigates five such input methods for collecting spatially explicit observation areas and deer counts from volunteers.

Method A

In method A, the volunteer identifies her location (in terms of coordinates), the location of the group of deer (in terms of coordinates), and the number of bucks, does and fawns. The observation area associated with the observation is computed as a circle whose center is collocated with the observer's coordinate and whose radius is equal to the distance between the observer coordinates and the deer group coordinates.

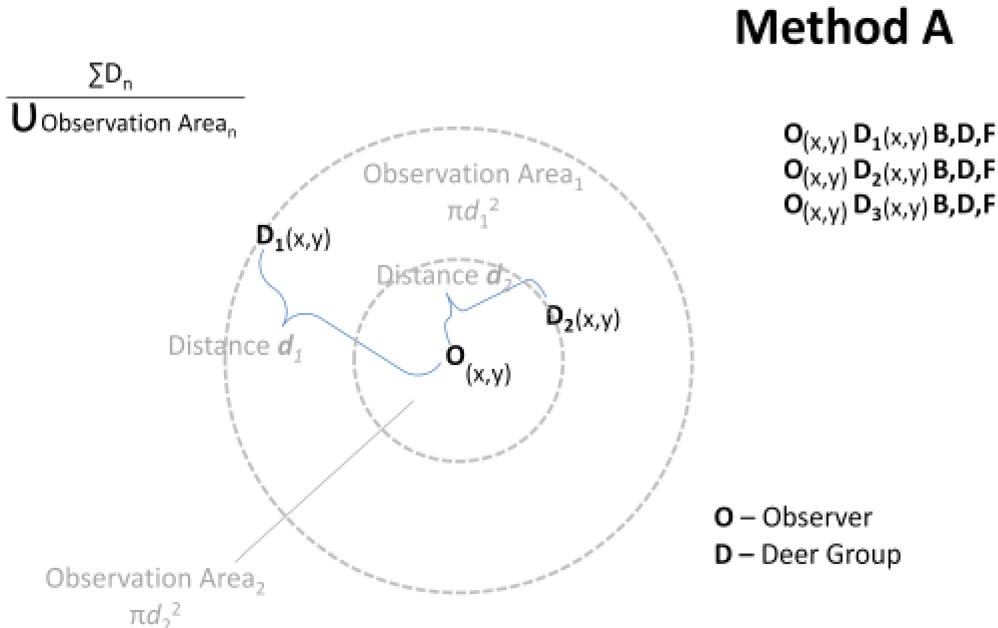


Figure 2.5 Input Method A

Observations are aggregated by summing the number of bucks, does and fawns over all observed groups and dividing by the area of the geometric union of all observation areas.

Method B

In method B, the volunteer identifies the location of the group of deer (in terms of coordinates) and provides an estimate of the distance of the deer group from the volunteer and the number of bucks, does and fawns. The observation area is calculated as a circle whose center is collocated with the deer group coordinate and whose radius is equal to the distance between the deer and the observer.

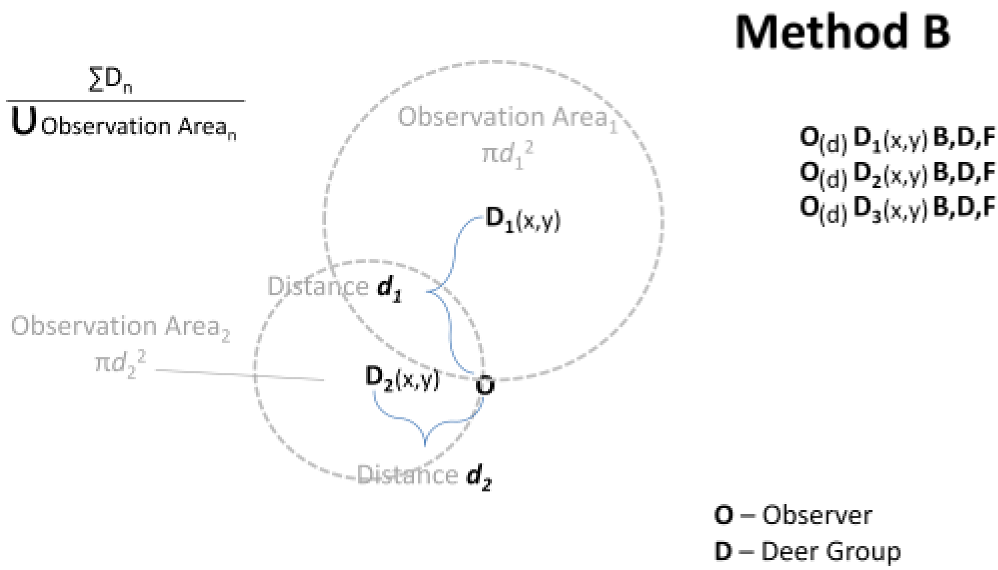


Figure 2.6 Input Method B

Observations are aggregated by summing the number of bucks, does and fawns over all observed groups and dividing by the area of the geometric union of all observation areas.

Method C

In method C, the volunteer identifies her location (in terms of coordinates) and reports the number of bucks, does and fawns. In addition, the volunteer also provides an estimate of the volunteer's average field of view distance, that is, the maximum distance the volunteer can see. The observation area associated with the observation is computed as a circle whose center is collocated with the observer's coordinate and whose radius is equal to the field of view distance.

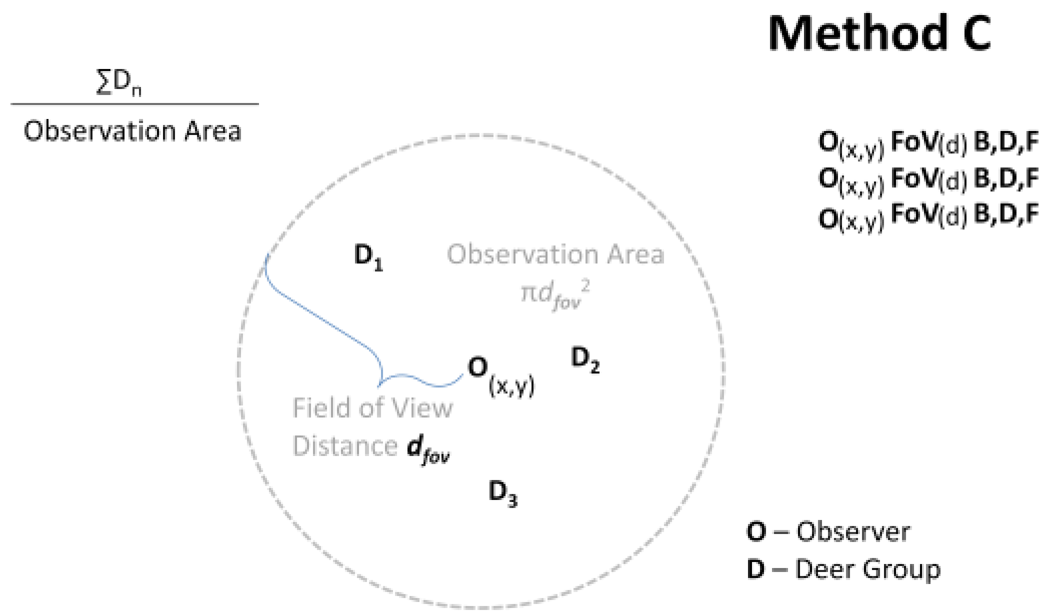


Figure 2.7 Input Method C

Observations are aggregated by summing the number of bucks, does and fawns over all observed groups and dividing by the area of the geometric union of all observation areas.

Method D

Method D is similar to method C except in method D in addition to field of view distance, volunteers also provide an estimate of the field of view width (in terms of degrees) and direction (in terms of degrees) to further refine the dimensions of the observed area. Field of view distance and width are used to calculate the pie-shaped observation area.

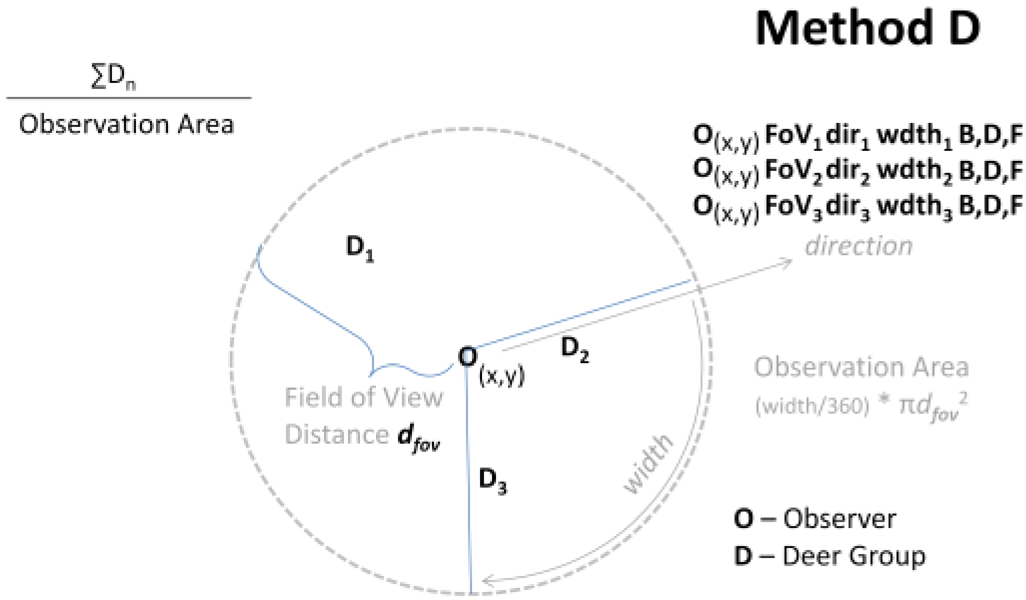


Figure 2.8 Input Method D

Observations are aggregated by summing the number of bucks, does and fawns over all observed groups and dividing by the area of the geometric union of all observation areas.

Method E

In method E, the observer reports deer groups by simply recording the number of bucks, does and fawns and associating the group with one of one or more predefined areas.

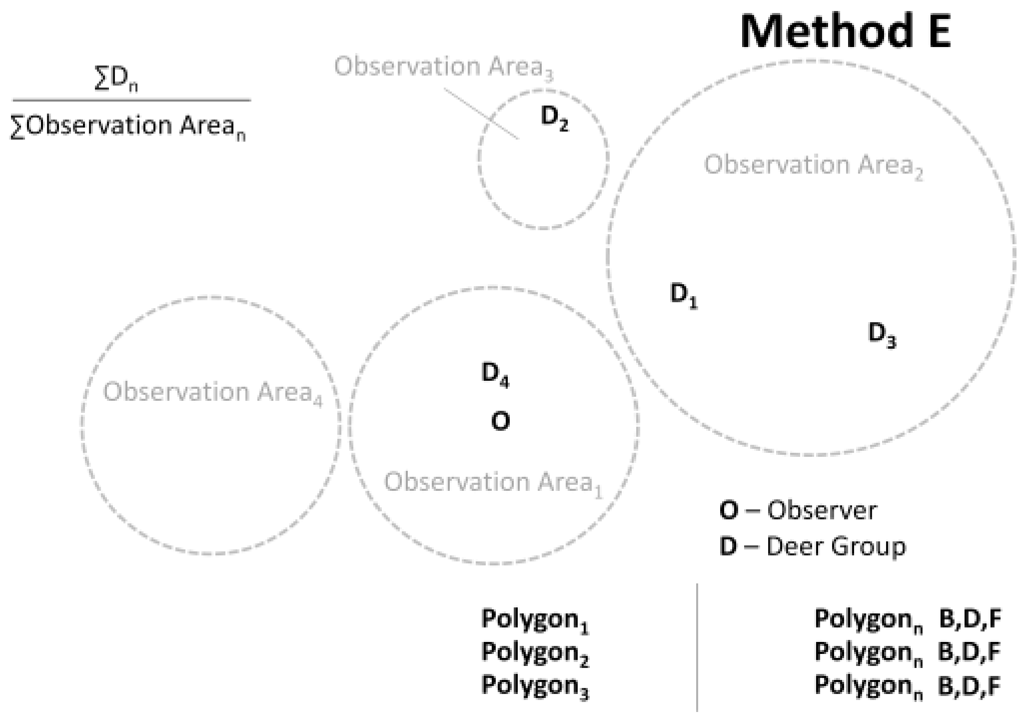


Figure 2.9 Input Method E

Observations are aggregated by summing the number of bucks, does and fawns over all observed groups and dividing by the area of the geometric union of the predefined observation areas.

Careful analysis of the above input methods could provide guidance on the best theoretical method, however, that approach would leave out an important reality, that of uncertainty. For example, Method C might be very effective with reliable field of view distance estimates, but if volunteers are likely to have poor or highly variable field of view distance estimates, then Method C may not be the best choice. It is the combination of the input method and possible uncertainty that is the focus of this simulation. Uncertainty types are detailed in the next sections.

Uncertainty Types

Each observation method requires two or more of the following data elements:

- Volunteer location (spatial coordinates)
- Deer location (spatial coordinates)
- Distance estimate
- Field of View width estimate (degrees, percent)
- Area specification
- Area association
- Number of deer by category (buck, doe, fawn, unknown)

These data elements are subject to various types of uncertainty. For example, a volunteer might use the geolocation capability of her smartphone to report her location. Alternatively, she may identify her location using a web-based map user interface. Both of these approaches to providing volunteer location are associated with different types and amounts of uncertainty. Geolocation provides the location of the observer (observers's supporting geolocation device) rather than the location of the observed target, making it less suitable for reporting deer location. While a volunteer might be able to use geolocation or a map user interface for her own location, it is more practical to report the location of deer through a map user interface. These and other types of uncertainty are described below. Uncertainty types are modeled with a probability of occurrence and magnitude based, where possible, on available literature. The uncertainty parameters described here may later be referenced as “standard” uncertainty.

Location Uncertainty

Two general approaches to acquire location information in terms of geographic coordinates are 1) geolocation using a device capable of reporting its location, and 2) a web-based map user interface that allows a volunteer to identify a location on a map. Although these two methods produce the same information, each method has very different uncertainty characteristics.

Geolocation

Geolocation technologies are complex and rapidly evolving and are one of the primary drivers of volunteered geographic information (Lu 2012, Haklay et al. 2008, Zandbergen 2009, Zandbergen 2011). Generally, there are a number of options available for determining location coordinates, including geocoding, GeoIP, smartphone-based sensors, web browser-based gelocation, and recreational GPS receivers. Geocoding, or resolving coordinates from some other piece of data like a street address, is not a viable option in the present context because of a lack of precision. If a volunteer and group of deer were both at the same address, they would both resolve to the same coordinates which would not work for any of the proposed reporting methods. GeoIP is a specific form of geocoding which uses an Internet addresses to resolve the location of the user (MaxMind,Skyhook). Most services may only resolve an IP address to the city or possibly area code in which it resides. Again, this would not provide sufficient precision for the reporting methods necessary here. Smartphones employ a sequence of approaches that use a combination of hardware, software, and network services in order to achieve the highest precision given the circumstances (Zandbergen 2009, Zandbergen 2011). A smartphone may first try to use its GPS receiver to determine its position. Failing that, it might try WiFi fingerprinting which geocodes based on WiFi networks that are in range. If that fails it may try using identifying information about the cellular network to determine a location. Direct use of the sensor requires platform specific software on each device, for example iPhone or iPad (iOS), Android, Windows phone or Blackberry. Incorporated into most modern web browsers, however, is the W3C Geolocation API (World Wide Web Consortium 2012), a uniform interface to geolocation technology available on supporting platforms. The API provides not only a location if available, but also an indication of the accuracy of the location. This allows the same web page to be rendered on any device and makes the best available geolocation technology supported on the platform available (given user permission) to code in the web page. This alternative works the same way across all compliant web browsers on smartphones, tablets, laptops, or desktop computers. While recreational GPS receivers are an option, because of the diversity of available models, lack of standardized ways of dealing with GPS data, the knowledge and experience required of the user to use the technology, and limited number of people who own recreational GPS receivers, they are not considered a feasible alternative in this study.

Reliance on a sensor for location coordinates implies a number of constraints; the sensor has to physically occupy the reported location, all of the elements of the geolocation technology, the hardware, software and network service elements, have to be functioning properly, and the application has to be running on the device. For circumstances where these conditions are not met, marking a position on a web map interface offers a reasonable alternative.

Map User Interface

Marking a position on a map is dependent on the care with which the volunteer takes in the process of marking the position and two issues are of concern; 1) accuracy of location, and 2) precision. Accuracy refers to the difference between the recorded location and the actual location while precision is influenced by the capability of the interface technology, specifically pixel to coordinate mapping at the rendered map scale. Bolstad et al. 90 examines error associated with manual digitization of paper maps. It indicates among operator differences in manual digitization error is significant, but within operator or by feature type is not. Errors in X or Y were independent, and mean distance deviation measured .054mm, sd .032mm and a max of .261mm. The errors in X and Y were approximately 2x the rated accuracy of the digitizing table. Meng et al. (1998) examines the distribution of error from manual digitization finding that the error distribution was between a normal and laplace distribution. A new distribution, NL distribution was further defined which the authors claim is more representative of error in GIS data. Of the slight deviations in X and Y from a normal distribution Bolstad et al. (1990, p.406) notes “these results indicate that the frequency distribution for signed positional uncertainty differs from a normal distribution in a statistical sense, although in a practical sense the difference appears to be quite small.”

A digitization table provides 1:1 scaling between the display -- the paper map, and interface device – the digitizing puck. Heads up digitizing, on the other hand, offers variable scaling because the source material can be displayed at different resolutions. Most computers are configured to move the cursor a greater screen distance than the distance the mouse is physically moved, yet people can still manage to get pixel accurate cursor positioning.

Modern display technology varies considerably in effective display resolution, increasing in pixel density as physical screen size decreases. For example, a representative 24" display that might be associated with a desktop computer offers 94ppi (pixels per inch) while the latest iPhone (4S model with 3.5" display) offers 326ppi. In addition to differences in display resolution, platforms also vary in pointing device technology. A mouse is more accurate than a touch pad, which is more accurate than a touchscreen.

Table 2.1 Display and Input Resolution

Platform	**Example**	**Pixels per Inch**	**Pointing Device**
Desktop	Dell Ultrasharp U2410 24" 1920x1200	94ppi (.27mm dot pitch)	Mouse
Laptop	Toshiba Satellite L745-S4310 14" 1366x768	112ppi (.23mm dot pitch)	Touch Pad
Tablet	Galaxy Tab 10.1 10.1" 1280x800	149ppi (.17mm dot pitch)	Touchscreen
Smartphone	iPhone4S 3.5" 960x640	326ppi (.077mm dot pitch)	Touchscreen

http://accessories.us.dell.com/sna/productdetail.aspx?cs=19&c=us&l=en&sku=320-8277

http://us.toshiba.com/computers/laptops/satellite/L740/L745-S4310

http://www.samsung.com/global/microsite/galaxytab/10.1/spec.html

http://www.apple.com/iphone/specs.html

Heads up digitization depends on the scale of the map as displayed on the screen or the map's "zoom level." In Central Texas, 1 pixel at Google Maps zoom level 18 represents a ground distance of approximately .52 m (unpublished calculations).

The max reported error of .261mm in Bolstad would reflect .97 px (.50m) error using the 24” display, or a 3.38 px (1.75m) error on the 3.5” display. These levels of accuracy are not likely given the technology, skill and attention of ordinary users. There appears to be a lack of literature that investigates digitization error in web-based map interfaces which would be particularly relevant for VGI. Here, digitization uncertainty is modeled based on a combination of display dot pitch and pointing device type at a single display scale (zoom level). In order to differentiate pointing devices, each pointing device type is assigned an error factor. For example, a Desktop computer with a mouse has an error factor of 4, a Laptop with a Touch Pad has an error factor of 6, and Tablets and Smartphones with Touchscreens have error factors of 8. The mean error for each device type is calculated as:

(.261mm x error factor / dot pitch) x ground resolution at zoom level 18

Standard deviation of error, for simplicity, is simply .33 of the mean error. Map uncertainty values are shown in the Map Uncertainty table.

Table 2.2 Map Uncertainty Table

	Mean Error (m)	**Error StdDev (m)**
Desktop	3.87	2.01
Laptop	6.81	3.54
Tablet	12.28	6.39
Smartphone	27.12	14.10

In addition to the precision/accuracy of the display, other sources of uncertainty include the accuracy of the base map, correct identification of the real world location on the map, and other map use problems which are not addressed explicitly here.

Location Uncertainty Implementation

Uncertainty related to location is implemented according to these rules:

There are two types of location points; observer points which are coincident with the position of the volunteer and can be established either by geolocation or by map interface, and map points which are established exclusively through a map interface. At simulation start, each volunteer is randomly associated with a specific device type, one of smartphone (P=.2), tablet (P=.1), laptop (P=.32), or desktop (P=.38) based on estimated device prevalence and assigned a location method, either map interface (P=.75) or geolocation (P=.25). If the requested point is an observer point and the location method is geolocation then the geolocation provider type is selected based on estimated probabilities in the provider type probabilities by device type table, table 2.3, below. Next, the location is displaced a random direction by a distance that is drawn at random from a triangular distribution whose parameters (min, max, and mean) are defined in the error by provider table, table 2.4, for the location provider type.

Table 2.3 Geolocation Provider Probabilities by Device Type

Geolocation provider type probabilities by device type:**P**	**Cell Network**	**WiFi**	**AGPS (Indoors)**	**AGPS (Outdoors)**
Smartphone	.05	.6	.3	.05
Tablet	.05	.6	.3	.05
Laptop	0	.9	.05	.05
Desktop	0	.97	.02999	.00001

Geolocation uncertainty by provider type

Table 2.4 Geolocation Error by Provider Type

Error (meters)	**Min**	**Max**	**Mean**
Cell	30	2731	599
WiFi	16	562	74
AGPS (Indoors)	.74	90.69	12.16
AGPS (Outdoors)	.53	58.35	10.14

If the requested point is an observer point and the location method is map interface or if the requested point is a map point the point is displaced in a random direction at a distance drawn at random from a distribution whose mean and standard deviation are presented in the Map Uncertainty table, table 2.2, based on the device type.

Distance Uncertainty

Several observation methods require that volunteers estimate the distance between themselves and one or a group of deer, or estimate the total distance that can be seen by the volunteer. Strauss and Carnahan (2010) reports that observers in an urban context tend to underestimate distance by 11.2%, 5.3% and 9.2% at short (22-30'), medium (148-211'), and long(330'-383') distances respectively, however, the distribution of error is not normal and there tends to be a large number of very large outliers (overestimates). Likewise, observers estimating distance to determine surveyed area for spotlight deer surveys tended to underestimate distance by 45% in habitats of less dense foliage and overestimate distance by 26% in habitats of more dense foliage (Whipple et al. 1994). In the present study, the urban landscape is more analogous to the less dense foliage in Whipple et al. further suggesting that distance would be underestimated in the present study.

Distance uncertainty is modeled according to these rules:

The probability of an outlier is (P=.05). If the observation is an outlier, the error distance is the product of the true distance and an error factor chosen at random from between 2 and 5 inclusive. If the observation is not an outlier, the error distance is the product of the true distance and an error factor. If the true distance is less than 9.14m, then the error factor is selected at random from a normal distribution with mean .888 and standard deviation .03. If the true distance is between 9.14m and 64.3m, the error factor is selected at random from a normal distribution with mean .947 and standard deviation of .03. If the true distance is greater than 64.3m, the error factor is selected at random from a normal distribution with mean .908 and standard deviation of .03.

Deer Detection Uncertainty

Deer detection is influenced by a number of factors including time of day, time of year and habitat type but not weather with the exception of snow (McCullough 1982). Spotlight surveys performed in darkness generally produce better results than daytime surveys (McCullough 1993). Roberts et al. (2006) reports a significant difference between spotlight surveys and infrared triggered camera surveys. ITC surveys captured nearly twice the number of deer observations as road surveys (including sunrise (n=90), sunset(n=93) and nighttime(n=70) surveys). Collier (2007) compares spotlight to thermal imaging surveys finding a spotlight detection probably of .31 to .66 with inter-observer variability up to 30%. Whipple et al. (1994) used simulation to evaluate spotlight detection probability over two habitat types finding probabilities of .67 in open habitats and .72 in closed habitats. In another study, spotlight surveys on average detected 53.3 deer whereas daytime surveys during the same period detected 38.6 deer suggesting that probability of detection during daytime surveys is .7242 or 72% of that of spotlight surveys (McCullough 1993). With sufficient evidence of a probability of detection for spotlight surveys around .66, then the probability of detection during the day can be estimated as .66*.7242 or .478.

For the present simulation, detection probabilities for each deer category are independent but all set to (P=.478).

Area Association Uncertainty

Area association uncertainty is specific to Method E where volunteers associate deer with predefined areas. Area association uncertainty is modeled as a probability of occurrence (P=.05) with the incorrect area chosen at random from the other defined areas.

Other Uncertainty

Uncertainty associated with the creation of predefined areas of Method E, uncertainty associated with specification of the direction and width of observation area in Method C, and uncertainty associated with malicious or intentionally erroneous observations have not yet been investigated.

Results

Simulation Set 1 was comprised of five simulation runs of 1000 iterations each. At each observer location, all five input methods were used to collect deer observations subject to standard uncertainty. The primary result of each simulation run was a collection of observed to actual deer density ratio values (OADR). Prior to simulation, acceptability criteria were established for input methods whereby an acceptable input method produces an OADR between .75 and 1.25 in 95 out of 100 simulation iterations. The results of each simulation run are shown in table 2.5, below. Shapiro-Wilk tests for normality on the distributions of OADRs for each input method suggest non-normal distributions, so for each method and simulation run, the number of "acceptable" OADRs (i.e. the number of ratios between .75 and 1.25), the median OADR and the interquartile range for the OADR is presented in table 2.5. The number of "acceptable" ratios provides a useful benchmark for the reliability of the input method. The median OADR provides an indication of the general accuracy of the method and the OADR interquartile range reveals the variability of the method. Figure 2.10 plots the distribution of OADR for each input method for the first run in Simulation Set 1. A summary of results for Simulation Set 1 showing the number of acceptable OADRs, median OADR and OADR interquartile range is presented in table 2.6.

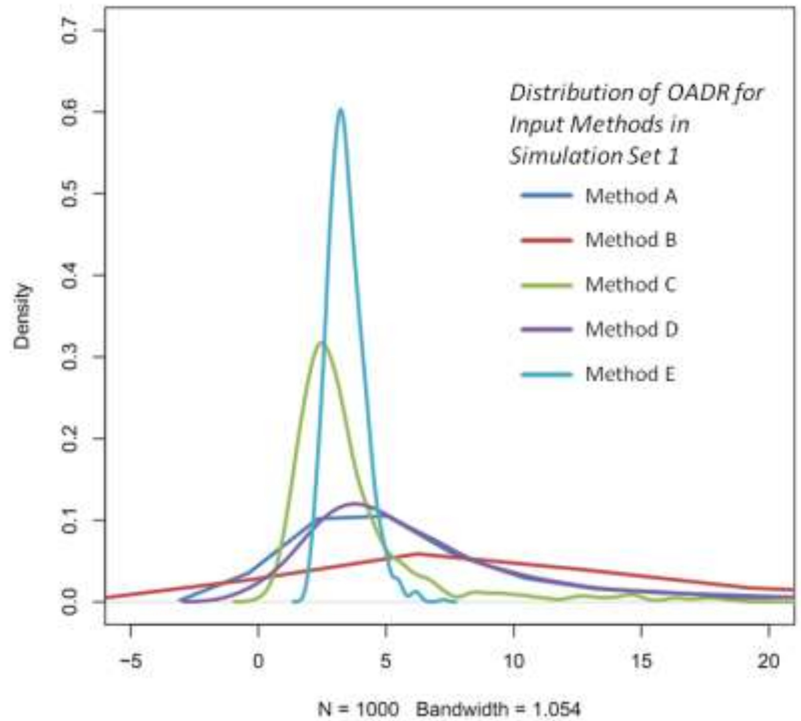


Figure 2.10 Distribution of OADR for Simulation Set 1

Table 2.5 Simulation Set 1 Results

Run 1			
Input Method	**Ratios Passed**	**Median**	**IQR**
A	24/1000	4.87	6.24
B	6/1000	8.79	13.39
C	42/1000	2.91	1.98
D	16/1000	5.39	5.68
E	0/1000	3.36	0.93
Run 2			
Input Method	**Ratios Passed**	**Median**	**IQR**
A	6/1000	9.54	14.12
B	5/1000	12.18	17.12
C	18/1000	3.74	2.58
D	14/1000	6.28	7.05
E	0/1000	3.32	0.95
Run 3			
Input Method	**Ratios Passed**	**Median**	**IQR**
A	20/1000	6.10	8.28
B	3/1000	9.30	11.65
C	19/1000	3.04	2.02
D	11/1000	5.15	4.78
E	0/1000	3.34	1.05
Run 4			
Input Method	**Ratios Passed**	**Median**	**IQR**
A	78/1000	2.14	4.82
B	40/1000	5.43	8.83
C	77/1000	2.48	2.85
D	14/1000	4.23	5.45
E	34/1000	3.06	1.99
Run 5			
Input Method	**Ratios Passed**	**Median**	**IQR**
A	24/1000	5.91	7.05
B	7/1000	8.06	11.51
C	42/1000	4.11	5.72
D	11/1000	5.34	5.97
E	0/1000	3.37	1.08

Table 2.6 Summary of Simulation Set 1 Results

Input Method	Ratios Passed *(out of 5000)*	Pct Passed	Median	IQR
A	152	3.04	4.87	6.24
B	61	1.22	8.79	13.39
C	198	3.96	2.91	1.98
D	66	1.32	5.39	5.68
E	34	0.68	3.36	0.93

No input method satisfied the acceptance criteria (ratio between .75 and 1.25 95% of the time) and in fact all methods performed poorly substantially overestimating deer density. Rank order of performance in terms of passing ratios was C, A, D, B, E but the best performing method, method C, only produced a passing ratio 217 out of 5,000 iterations (4.34%). With the exception of method E, all methods exhibited considerable variability, both within and between each trial. Method E, on the other hand, produced a passing ratio in less than 1% of the simulation iterations, but it exhibits very little variability within and between trials. Actual density remained constant for each simulation run and method E produced a stable median OADR for each run. Method E was consistent, if not correct.

Simulation Set 2: Sensitivity

Simulation Set 1 evaluated the performance of five input methods subject to possible uncertainty. A reliable input method should produce consistent results across a range of volunteer participation and deer density. Simulation Set 2 examined the performance of these same five input methods with various levels of volunteer participation and deer density. The primary metric for evaluation was median OADR.

Simulation runs were conducted using permutations of 6, 30 and 50 deer representing low, medium and high deer density, and 4, 20, 40 and 80 observers representing limited, low, medium, and high volunteer participation. Input methods were subject to standard uncertainty, that is, the default uncertainty profiles established in Simulation Set 1. Results from Simulation Set 2 are presented for levels of participation grouped by deer density.

Low Deer Density

Table 2.7 shows the median OADR for each input method at each participation level given low deer density. Figure 2.11 plots median OADR for each participation level given low deer density.

Table 2.7 Median OADR, Low Deer Density

	Number of Observers			
Input Method	**4**	**20**	**40**	**80**
A	14.78	10.94	16.16	17.99
B	13.19	19.87	20.23	29.10
C	6.25	10.75	9.75	12.48
D	12.58	18.49	16.60	18.21
E	13.44	11.71	12.17	12.30

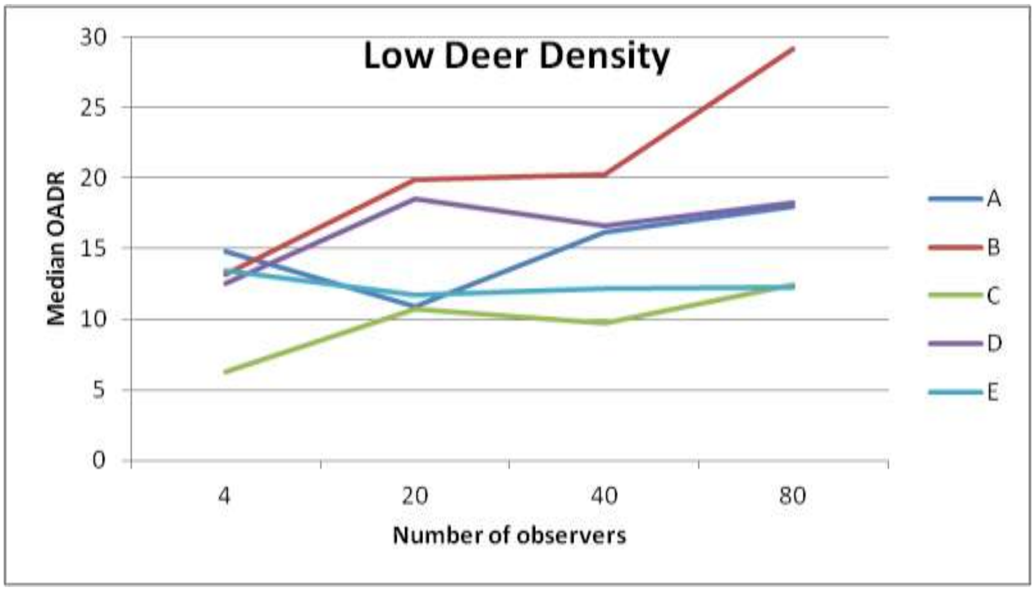


Figure 2.11 Plot of Median OADR, Low Deer Density

The median OADR varied across levels of participation for input methods A, B, C and D. Median OADR for input method E remained relatively stable across levels of participation. Absolute accuracy for all input methods was poor, for example input method E reported 13 times the actual density.

<u>Medium Deer Density</u>

Table 2.8 shows the median OADR for each method at each participation level given medium deer density. Figure 2.12 plots median OADR for each participation level given medium deer density.

Table 2.8 Median OADR, Medium Deer Density

	Number of Observers			
Input Method	**4**	**20**	**40**	**80**
A	6.08	5.539	5.89	8.39
B	6.15	9.66	7.97	10.95
C	3.68	3.52	4.30	3.72
D	6.88	5.78	5.29	6.90
E	2.88	3.26	3.45	3.44

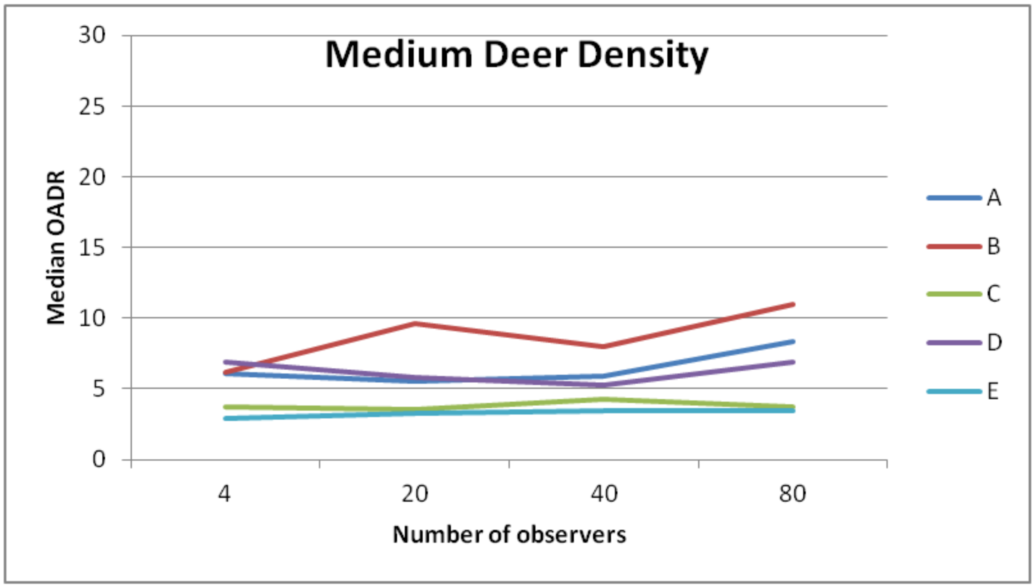


Figure 2.12 Plot of Median OADR, Medium Deer Density

Given medium deer density, median OADRs for methods C and E remain consistent across various levels of participation and for both methods there is an improvement in accuracy of median OADR.

High Deer Density

Table 2.9 shows median OADR for each input method at each participation level given high deer density. Figure 2.13 plots median OADR for each participation level given high deer density.

Table 2.9 Median OADR, High Deer Density

	Number of Observers			
Input Method	**4**	**20**	**40**	**80**
A	0.99	5.67	4.65	6.48
B	6.51	6.68	7.71	9.54
C	4.37	2.48	2.41	2.99
D	9.51	3.99	4.10	5.22
E	2.50	2.31	2.25	2.29

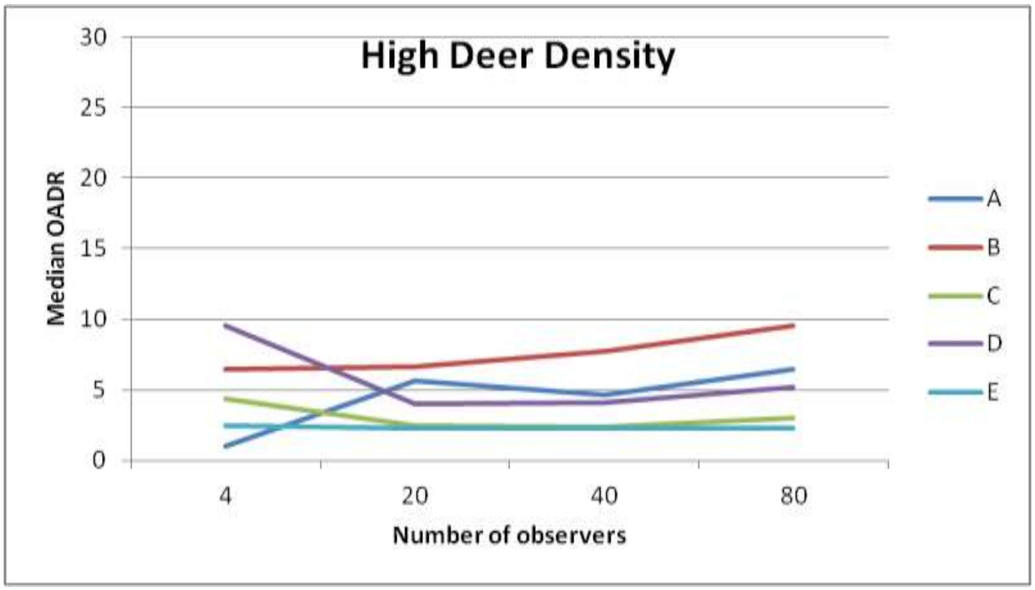


Figure 2.13 Plot of Median OADR, High Deer Density

Methods A, B and D continue to show considerable variability across various levels of participation whereas input methods C and E are generally stable across the range of volunteer participation and improve, once again, in terms of OADR accuracy.

<u>Results</u>

Generally, all methods showed improved accuracy with greater deer density, however, not all methods were stable across levels of participation for any given deer density. Methods C and E were the most stable across various levels of participation. Methods A, B and D are unstable across participation levels for any deer density and show an inverse relationship between participation and OADR accuracy. In practice this suggests these methods may degrade in performance with greater volunteer participation.

Method E and method C to a lesser extent, exhibited two important characteristics; 1) consistency across levels of participation at each deer density, and 2) improving accuracy with higher deer density.

Simulation Set 2 examined the performance of each input method relative to various levels of volunteer participation and deer density. Simulation Set 3 examines changes in input method performance with modifications to input methods to reduce uncertainty.

Simulation Set 3: Screening and Filtering

One strategy to improve the performance of an input method is to reduce potential uncertainty by modifying the input method. Some uncertainty types are more easily modified than others, for example, improving deer detection probability may be extremely difficult, but modifying location uncertainty by eliminating use of geolocation is easily accomplished and easily simulated. In Simulation Set 3A, input methods that allow the use of geolocation, input methods A, C and D, are modified to disallow the use of geolocation and to only allow map user interface location. In Simulation Set 3B, geolocation is re-enabled and distance estimation uncertainty is disabled simulating an improvement to input methods which rely on estimated distance, namely methods B, C and D. Simulation Set 3A and Set 3B included 40 observers and 30 deer. Simulation Set 3 results are compared to results from Simulation Set 1. An improvement in performance in Simulation Set 3A and Set 3B would mean either an increase in the number of acceptable OADRs, higher accuracy evidenced by a smaller OADR median, or a reduction in variability.

Simulation Set 3A: Geolocation Disabled

In Simulation Set 3A, the only available mechanism to determine a location is the map user interface. This simulates a limitation to input methods where volunteers may only enter observer locations through the map user interface. The simulation set is comprised of three simulation runs of 1000 iterations each. Table 2,10 shows the number of acceptable OADRs, median OADR and interquantile range for each input method. Note that methods B and E do not use observer location so these results are not shown.

Table 2.10 Comparison of Results from Set 1 and Set 3A

Input Method	Set 1 Pct Passed	Set 3A Pct Passed	Set 1 Median	Set 3A Median	Set 1 IQR	Set 3A IQR
A	3.04	0.23	4.87	9.99	6.24	13.02
C	3.96	2.63	2.91	3.20	1.98	2.10
D	1.32	1.70	5.39	5.89	5.68	6.06

Except for method D, the percentage of OADRs meeting acceptance criteria is lower for all input methods that use observer location when geolocation is disabled. Method D shows only a small 0.38% improvement. Median OADRs as well as OADR interquartile range for each method are higher with geolocation disabled.

Simulation Set 3B: Distance Estimation Uncertainty Disabled

In Simulation Set 3B, geolocation is re-enabled, but distance estimation uncertainty is disabled. This simulates an improvement to those methods that use distance estimation such that an accurate distance is always reported. Only methods B, C and D use distance estimation. The simulation set is comprised of three simulation runs of 1000 iterations each. Table 2.11 shows the number of acceptable OADRs, median OADR and interquantile range for each input method.

Table 2.11 Comparison of Results from Set 1 and Set 3B

Input Method	Set 1 Pct Passed	Set 3B Pct Passed	Set 1 Median	Set 3B Median	Set 1 IQR	Set 3B IQR
B	1.22	0.43	8.79	9.82	13.39	13.80
C	3.96	2.57	2.91	3.41	1.98	2.59
D	1.32	1.47	5.39	5.71	5.68	6.32

The percentage of acceptable OADRs was lower for each input method with distance estimation uncertainty disabled. In addition, each median OADR was higher and each interquartile range was larger with distance estimation uncertainty disabled.

Combined Results: Set 3A and Set 3B

Two measures of improvement in input methods are greater accuracy and lower variability. These can be tested statistically using nonparametric tests. The Ansari-Bradley test examines the dispersion (variability or scale) of the two result sets. The null hypothesis of the two-tailed version of the test assumes the dispersion is the same for each set. If dispersion is the same for both sets then it is assumed there is no improvement in variability. If there is no improvement in variability, a one-tailed Mann-Whitney test can test the null hypothesis that the location (median) of set 1 is less than or equal to set 3. A statistically significant result would suggest that the location of set 1 is to the right of set 3 (or that set 1 median is less than set 3 median), which, given the location of median, would be more accurate. If the two-tailed Ansari-Bradley test reveals a difference in dispersion, a second one-tailed Ansari-Bradley test is used to determine if the dispersion of set 1 is greater than or equal to the dispersion of set 3. Results from these tests are shown in table 2.12.

Table 2.12 Statistical Test Results, Sets 3A and 3B

		Ansari-Bradley		Mann-Witney		
Data Sets	**Input Method**	**(X=Y)**	**(X<=Y)**	**(X<=Y)**	**Less Variable**	**More Accurate**
1,3A	**A**	p < 2.2e-16	p = 1	NA	No	NA
1,3A	**C**	p < 2.2e-16	p = 1	NA	No	NA
1,3A	**D**	p = 0.8796	NA	p = 1	No	No
1,3B	**B**	p = 3.132e-4	p = 0.9998	NA	No	NA
1,3B	**C**	p = 3.616e-10	p = 1	NA	No	NA
1,3B	**D**	p = 0.1100	NA	p = 0.9926	No	No

Neither the removal of geolocation in Set 3A nor the removal of distance estimation uncertainty in 3B produced less variable results. Input method D showed no change in variability in sets 3A and 3B, but also no improvement in accuracy.

<u>Results</u>

In Simulation Set 3A, eliminating the use of geolocation failed to achieve a dramatic improvement in input methods A, C or D. If there is any evidence of improvement at all it is in method C and only marginal at best. Simulation Set 3B also failed to show dramatic improvement in input methods B, C or D with the elimination of distance estimation uncertainty.

Simulation Set 3 investigated possible improvement to input methods by reducing (or eliminating) specific types of uncertainty. The two approaches for reducing uncertainty failed to achieve useful improvements to input methods in terms of greater accuracy or reduced variability. Another possibility for improving the performance of input methods is to alter the method. This approach is examined in Simulation Set 4.

Simulation Set 4: Optimization

Simulation Set 1 compared five methods for volunteer deer reporting. Simulation Set 2 explored the sensitivity of those five methods to various levels of volunteer participation and deer density. Simulation Set 3 examined reducing or eliminating specific types of uncertainty in the five methods. Simulation Set 4 attempts to optimize the performance of the most promising methods from prior simulations. Methods C and E consistently demonstrate low variability, however, both methods suffer poor accuracy. Both methods tend to overestimate the number of deer in the simulation. One limitation of all input methods, including C and E, is that observations are reported only when deer are observed. This approach neglects to report areas searched that contain no deer which could lead to inflated observed deer density. A report of a searched area containing no deer is referred to as a zero-deer observation. In Simulation Set 4A, revised versions of methods C and E, named C2 and E2, are introduced that function like their respective namesakes but also report zero-deer observations.

An additional limitation to input methods E and E2 is that predefined observation areas are mutually exclusive. This restriction means that an area could only be associated with one vantage point, an unnecessarily restrictive limitation from the perspective of a volunteer. In Simulation Set 4B, two new predefined observation area input methods, methods G1 and G2, are tested. In method G1, observation areas are mutually exclusive. In method G2, observation areas may overlap.

Simulation Set 4A: Zero-deer Observations

Reporting a deer observation when none have been seen is counter-intuitive, but not reporting all searched areas could produce an inflated deer density estimate as too small of an area could be associated with the number of deer that are seen. In this simulation set, methods C and E which do not include zero-deer observations are adapted into methods C2 and E2 which do include zero-deer observations. Three simulation runs of 1000 iterations each were executed to compare the performance of these input methods. Each simulation included 30 deer, 40 observers and standard uncertainty.

Table 2.13 reports the results for each run and table 2.14 reports a summary of all three runs.

Table 2.13 Results Set 4A

Run 1			
Observation Type	**Ratios Passed**	**Median**	**IQR**
C	23/1000	3.99	5.16
C2	243/1000	0.57	0.57
E	0/1000	3.41	1.02
E2	680/1000	0.90	0.30
Run 2			
Observation Type	**Ratios Passed**	**Median**	**IQR**
C	25/1000	3.64	3.95
C2	112/1000	0.40	0.35
E	0/1000	3.31	0.92
E2	727/1000	0.93	0.28
Run 3			
Observation Type	**Ratios Passed**	**Median**	**IQR**
C	22/1000	3.51	2.71
C2	176/1000	0.51	0.53
E	0/1000	3.27	0.99
E2	763/1000	0.92	0.26

Table 2.14 Summary of Results for Set 4A

Input Method	Ratios Passed *(out of 3000)*	Pct Passed	Median	IQR
C	70	2.33	3.72	3.64
C2	531	17.70	0.48	0.50
E	0	0.00	3.33	0.98
E2	2170	72.33	0.92	0.28

In all three simulation runs, method C2 showed a considerable improvement over C in terms of reduced variability, increased accuracy and larger number of "acceptable" OADRs. Method E2 also showed a marked improvement over method E with smaller variability and higher accuracy. The result of higher accuracy and reduced variability is seen in the percentage of "acceptable" ratios. Method C2 using zero-deer observations shows 17.7% passing ratios up from 2.33% using Method C without zero-deer observations. Method E2 which includes zero-deer observations jumped to 72.33% passing ratios from none passing using Method E with no zero-deer observations. Given the improvement in performance of Method E2 from inclusion of zero-deer observations, sensitivity analysis using varying levels of participation and deer density similar to that performed in Simulation Set 2 was conducted with input method E2. Input method E2 was slightly modified to only report a zero-deer observation with a probability of (P=.75) to simulate observers not consistently reporting zero-deer observations.

Table 2.15 Acceptable OADRs and Correlation by Participation and Density

	Observers				
Deer Density	**4**	**20**	**40**	**80**	**Correlation**
Low (6 deer)	134	304	325	718	0.98
Medium (30 deer)	372	646	718	827	0.89
High (50 deer)	429	759	827	931	0.86

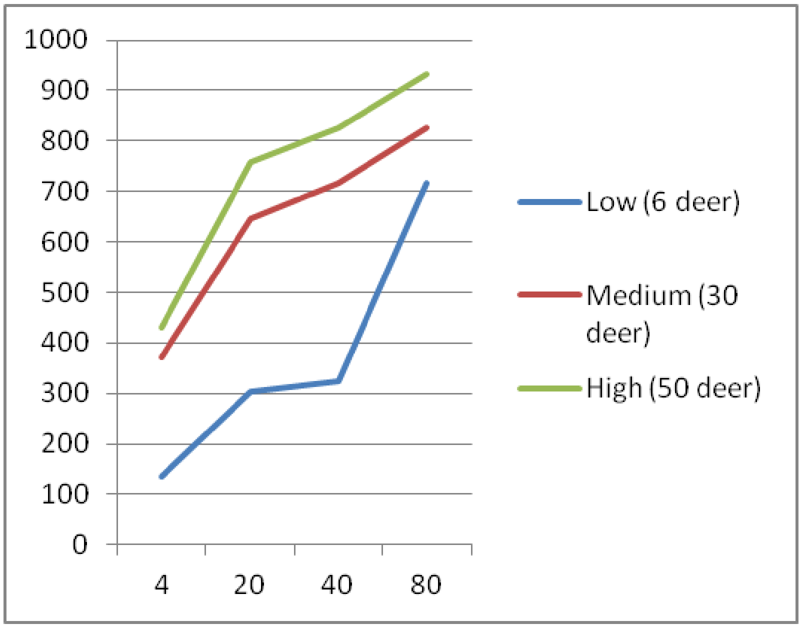


Figure 2.14 Acceptable OADRs by participation and density

Table 2.15 shows the number of passing OADRs out of 1000 simulation iterations for each combination of deer density and volunteer participation. These values are depicted graphically in figure 2.14. For each deer density, the number of observers was strongly and positively correlated with the number of acceptable OADRs. The number of observers in the simulation serves as a measure of participation and it is assumed that all volunteers participate equally. Another measure of participation is the percentage of the study area included within volunteer observation areas, in other words, the portion of the study area observed by all observers together. Table 2.16 shows the correlation between percent coverage of study area and number of passing ratios for each deer density.

Table 2.16 Correlation between Percent Coverage and Acceptable OADRs

Percent Coverage					
Average Pct Coverage	0.254	0.634	0.729	0.737	**Correlation**
Low (6 deer)	134	304	325	718	0.72
Medium (30 deer)	372	646	718	827	0.97
High (50 deer)	429	759	827	931	0.98

Again, correlation between passing OADRs and percent coverage is both positive and strong particularly at medium and high deer densities.

Results from method E2 show a substantial improvement over method E, but even with high deer density and a high level of volunteer participation (76.26% of the simulation frame observed), with only 931 out of 1000 passing ratios (93.1%), E2 fell shy of the 95% acceptance criteria.

<u>Simulation Set 4B: Overlapping Observation Areas</u>

Input methods E and E2 employ a very simple approach for creating sets of volunteer observation areas for each observer. Observation areas are created by buffering selected house locations and road segments. As each new observation area is added, existing areas are subtracted from the new area so that there is no overlap between existing areas and the new area. This observation area allocation procedure is limiting in two ways. First, it does not take into account multiple vantage points from within a single home, and second, it is unnecessarily limiting as it would be possible, and in fact likely for a volunteer to define observation areas that overlap. Simulation Set 4B introduces two new predefined observation area input methods, G1 and G2, which more closely simulate multiple vantage points within a single home. Input methods G1 and G2 create clusters of observation areas of random sizes at random distances from the observer location. In G1 these observation areas are mutually exclusive. In G2 these observation areas overlap and the observer selects at random any one of the containing observation areas to associate with each deer observation. The probability of an observer associating the wrong observation area with a report in both G1 and G2 is (P=.05). For both G1 and G2 input methods the number of observation areas is chosen at random between 3 and 5 inclusive. Each observation area is displaced from the observer location in a random direction at a distance chosen at random between 10 and 150 meters with a diameter chosen at random between 10 and 100 meters. In G1, the geometric union of previously allocated observation areas is subtracted from the new observation area so that the new area does not overlap any existing observation area. In G2 observation areas may overlap. Representative G1 and G2 observation areas are depicted in figure 2.15.

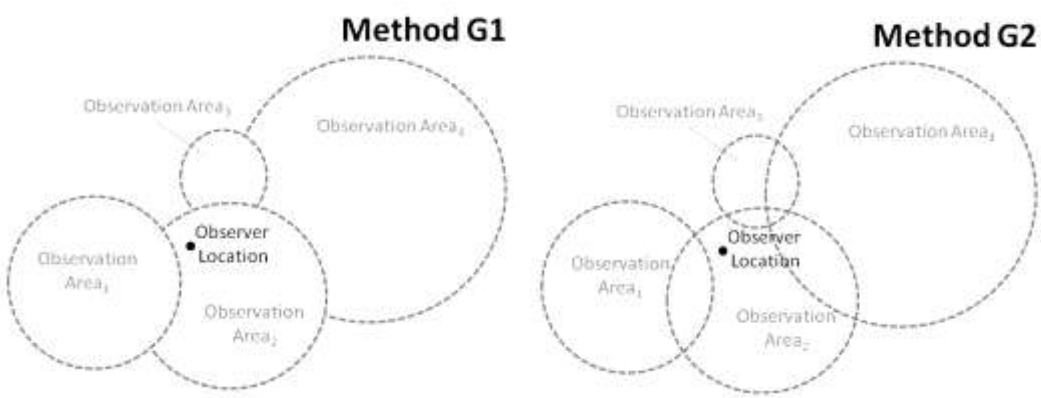


Figure 2.15 Input Methods G1 and G2

The purpose of Simulation Set 4B is not to improve a reporting method but rather to test the removal of an unnecessary constraint. Simulation Set 4B consisted of 5 simulation runs of 1000 iterations each, however, these results were grouped for analysis.

Table 2.17 shows the number and percentage of acceptable OADRs followed by the median and interquartile range for each input method.

Table 2.17 Set 4B Acceptable OADRs

Input Method	**Total Passed** *(out of 5000)*	**Pct Passed**	**Median**	**IQR**
G1	1975	39.5	1.30	0.63
G2	2241	44.82	1.24	0.60

The slightly larger percentage of passed OADRs, the smaller median value and smaller interquartile range of input method G2 relative to G1 provides some evidence that method G2 performs at least as well as G1. As in Simulation Set 3, the dispersion of OADR for the G1 and G2 simulation results was tested for equal scale. The Ansari-Bradley test failed to reject the null hypothesis of equal dispersion. The one-tailed Mann-Whitney test of the null hypothesis that the median of G1 is less than or equal to the median of G2 was rejected corroborating that the G2 input method with overlapping observation areas was no worse than G1, but in fact statistically superior to G1. The results of these tests are presented in table 2.18.

Table 2.18 Statistical Results, Set 4B

	Ansari-Bradley		**Mann-Witney**		
Input Method	**(X=Y)**	**(X<=Y)**	**(X<=Y)**	**Less Variable**	**More Accurate**
G1,G2	p = 0.41	NA	p = 1.24e-07	no	yes

Results

Simulation Set 4A shows that with predefined observation areas and zero-deer observations, method E2 offers comparatively little variability along with high accuracy across various levels of participation and deer density. These qualities make this type of input method the best choice from among the alternatives tested. Put another way, no tested input method fully meets the acceptance criteria, but input method E2 provides the best opportunity for a reliable result. The number of acceptable OADRs is strongly and positively correlated to percent coverage of study area.

The results of Simulation Set 4B show a slight but significant performance advantage for G2, the input method using overlapping observation areas.

Simulation Results and Discussion

The purpose of the simulation described in this chapter was to explore potential uncertainty associated with a planned VGI initiative in order to improve the likelihood of useful VGI results. Each simulation set had a specific focus and provided unique insight into the planned volunteer deer counting project. Lessons learned from simulation results guided development of the volunteer count project. There were a number of key conclusions drawn from simulation results. Using distance as a surrogate for area in deer density observation as in Methods A - D was generally ineffective. Method C which used an estimate of field-of-view distance was better than other distance-based input methods, but still it failed to perform as well as optimized methods using predefined observation areas. Filtering and screening techniques did little to improve the poor performance of any observation method. While it was possible to reduce uncertainty, the reduction made no material improvement on the data collection process. Of the five original observation methods, the best performing method given “standard” levels of uncertainty was Method E which associates observed deer with predefined observation areas. Its performance was consistent showing little variation between simulation runs, but not accurate. By adding zero-deer observations to Method E, performance increased dramatically, suggesting a similar method may make the best candidate for an actual VGI deer counting project. Overlapping (non-mutually exclusive) pre-defined observation areas work as well or better than mutually exclusive observation areas. Simulation results suggest a strong positive relationship between participation in terms of coverage area and population estimate accuracy for any given level of deer density. Participation is crucial for the VGI initiative, but it is highly unlikely that participation will be uniform across volunteers. A VGI campaign that uses a reporting method based on predefined observation areas in which there is a high level of participation (in an area with a large deer population) may provide very reliable results

During simulation, observers of any specific type were homogenous. While there was some variation in the size and arrangement of the observation areas for each observer of a specific type, the range of variability was fixed. It is reasonable to assume much more variation between individual observers during a VGI initiative, but the form of that variation is difficult to predict and harder to quantify. This presents a challenge addressed later in the volunteer incentive program.

Simulation was conducted using “snapshot” counts, that is, a count of deer by observers as they are arranged at a single point in time. While it is possible that a VGI project may use a single count, a more likely approach would use multiple counts, so handling multiple snapshot counts during the real-world VGI initiative must be resolved.

The relationship between participation and accuracy during simulation directly influenced the temporal structure of the volunteer count. The concept of “count block” and “pushes” were established specifically to promote participation at focused times. The incentive program reinforced focused participation during pushes. The importance of zero-deer observations was introduced in the volunteer count in the counting method by the suggestion to start each count block with a zero-deer observation and reinforced in the incentive program by scoring up to two observations per observation area per count block, one for a zero-deer observation and one for a total observation. Participation in terms of coverage area was implemented in the incentive program as part of the scoring process. Both the optimal size of any one observation area and total coverage area per count block were used in the incentive scoring process.

This chapter presented the simulation phase of the project. With the insight developed through simulation, the facilitated VGI initiative was designed, implemented and executed. The VGI phase is described in the next chapter.

3. VOLUNTEERED GEOGRAPHIC INFORMATION

During October of 2012 study area neighborhood residents were recruited to participate in a volunteer neighborhood white-tailed deer count. Volunteers interacted with a web-based data entry and reporting application to report observations and monitor volunteer participation. The design of this facilitated Volunteered Geographic Information initiative incorporated optimizations from the preceding simulation project presented in Chapter 2.

This chapter provides context and details surrounding the facilitated VGI initiative including a brief description of the study area, recruiting activities, the structure of the volunteer count and counting rules, details regarding the incentive program and the basic structure of the web site that facilitated the count. Next, volunteer count results and a number of observations regarding the results are discussed. This chapter concludes with several interpretations of the volunteer count results.

Study Area

The volunteer count study area included the Sierra Circle neighborhood and an adjacent portion of the Tanglewood neighborhood of San Marcos, Texas, USA. Factors contributing to the selection of the study area included the presence of nearby green space, anecdotal evidence of the presence of urban deer in the neighborhood confirmed through direct observation, and prior use of the Sierra Circle neighborhood in urban deer surveys conducted by the city of San Marcos (figure 3.1).

Figure 3.1 Study Area

Recruitment

VGI initiatives rely on the participation of volunteers and volunteer recruitment for this project took a number of forms. The first introduction of the project to the neighborhood took place at an outdoor neighborhood meeting with both a recruiting table with information about the project and a position on the meeting agenda to introduce the project.

During the project, recruiting channels included a recruiting table, email, bulk physical mail, and word of mouth including social media.

Neighborhood meeting

The project was formally introduced on October 2, 2012 at a Neighborhood Night Out event in support of the National Night Out campaign. The formal introduction through a recognized neighborhood organizer helped establish credibility for the project. While the explanation of the project was brief, sufficient time was allowed to address a number of questions from the neighbors particularly regarding use of the results of the project. Being able to address neighborhood concerns about the project up front and in a group setting was instrumental in establishing some volunteer inertia as some neighbors agreed to participate together.

Figure 3.2 Neighborhood Night Out Event

Recruiting Table

A recruiting table was set up and manned for several hours at a time in various locations throughout the study area periodically duration of the VGI initiative to promote and provide information about the project, as a reminder that the project was ongoing, to answer questions regarding project procedures and to generally maintain a physical presence in the neighborhood. Manned table hours generally followed evening key times, 5:30pm to 7:30pm, during volunteer pushes.

Figure 3.3 Recruiting table

Bulk Mail, Social Media and Word of Mouth

Bulk mailings to neighborhood addresses were distributed in weeks 2 and 3 of the project. The first mailing included 127 addresses. The second mailing included 125 addresses. A number of posts were made to the neighborhood Facebook group. Posts included invitations for participation, notices of upcoming pushes (concentrated counting periods) including participation goals, announcements of participation results and weekly incentive program winners. Finally, word of mouth promotion was highly encouraged.

Opposition

Urban deer are a sensitive and divisive topic and resistance towards the volunteer count was expected. The most commonly voiced concern was use of the volunteer count results by the city or by the State of Texas to initiate a deer control program. Neighborhood residents did not want anything to happen to their deer as a result of the study. Residents were eased by the fact that the city would only consider a deer control program based on their own deer survey data, and that the state, through the Texas Parks and Wildlife Department, would only offer technical assistance for a deer control project, not unilaterally initiate one, and only in a situation where there is an established neighborhood consensus that deer control is necessary. In virtually every case, demonstrating an appreciation for the concerns of neighborhood residents and providing assurance that project results would not be used for deer control purposes transformed anxiety into interest and often enthusiasm for the project.

Counting and Counting Rules

A facilitated VGI initiative is one in which there is a coordinator or facilitator (Seeger 2006) that provides the means to collect and store data, determines what data are to be collected and how and defines how data are to be analyzed and reported. This research uses preliminary simulation to guide the development of these procedures. Several important observations emerged from the simulations presented in Chapter 2. First, a reporting method which links observed deer to pre-defined observation areas provides more reliable results than a method which derives observation area from a representation of distance. Second, observations that report the absence of deer in an observation area, zero-deer observations, improve the performance of an observation method using pre-defined observation areas. Third, the reliability of an observation method using pre-defined observation areas may be improved with greater participation in the form of more observed area. These observations are integrated into the volunteer deer count method that follows.

Volunteers, mainly comprised of neighborhood residents, were recruited to count deer in the study area neighborhood during a 29-day volunteer deer count initiative concluding on October 30, 2012. After registering through the web-based volunteer deer count application, volunteers were prompted to create one or more pre-defined observation areas using the reporting application's interactive map interface. Preliminary simulation results favored the use of predefined observation areas over distance-based observation areas. Observation areas represent the entire visible area from a particular vantage point when looking for deer. Counting was allowed only during daylight hours and each hour was divided into 15 minute intervals called count blocks. Each count block comprised an independent deer survey, meaning a new count started every 15 minutes. Although volunteers could make observations during any daylight hour and did not have to follow any specific schedule, the project prioritized counts during certain times and on certain days. Volunteers were asked to report observations when they looked for deer, even if they did not see any, what was termed zero-deer observations. The importance of zero-deer observations emerged during simulation. Observations were recorded using the web application by selecting the appropriate count block and observation area and indentifying the number, sex and age of observed deer, if any. Volunteers were encouraged to start each count block by making a zero-deer observation in each observation area observed to indicate that the area had been observed during the count block. Additional observations for any observation area during the same count block were simply combined into an aggregate count. Volunteers were asked to count specific deer only once per observation area during a single count block.

Simulation revealed that observation results may benefit from greater participation, so a paramount goal of the VGI program was to focus counting into narrow windows of time. Toward this end, two mechanisms were used to convey and reinforce VGI program priorities; coordinated counting "pushes" and an observation scoring system which awarded points based on observed area and time of observation with observations during preferred times earning more points. "Pushes" were simply days and times scheduled for focused and coordinated volunteer deer counting. During the volunteer count there were three such pushes, aligning with the last three weekends of the project starting Friday evening and ending Sunday evening. Pushes focused counting effort on specific days, but the scoring system focused effort at specific times during the day. Associated with each count block was a count block factor which was used in combination with the area observed to determine an observation score. As a factor of the observation score its value denoted the relative value of an observation during the count block. The standard count block factor was 2. The key times to look for deer were during the eight count blocks starting one half hour before sunrise, generally from 7a.m. to 9a.m., and the eight count blocks ending 30 minutes after sunset, generally 5:30p.m. to 7:30p.m.. These key time count blocks had a count block factor of 10 to underscore the relative importance of observations made during these times. Key time count blocks during pushes had a count block factor of 50. The scoring system was used both as a feedback mechanism allowing volunteers to understand the relative value of their contribution to the project and also to determine the incentive program grand prize winner each week. It is important to note that count block score points were awarded for observations, specifically the activity of looking for deer, and were not influenced by the number of deer actually observed. This distinction was intended to prevent count inflation and to encourage zero-deer observations if appropriate.

In addition to enabling volunteer registration, observation area mapping, and observation collection, the web application also provided feedback for each volunteer on individual performance as well as the performance of the entire neighborhood. Feedback also came in the form of weekly incentive program winner emails and posts on social media.

Website

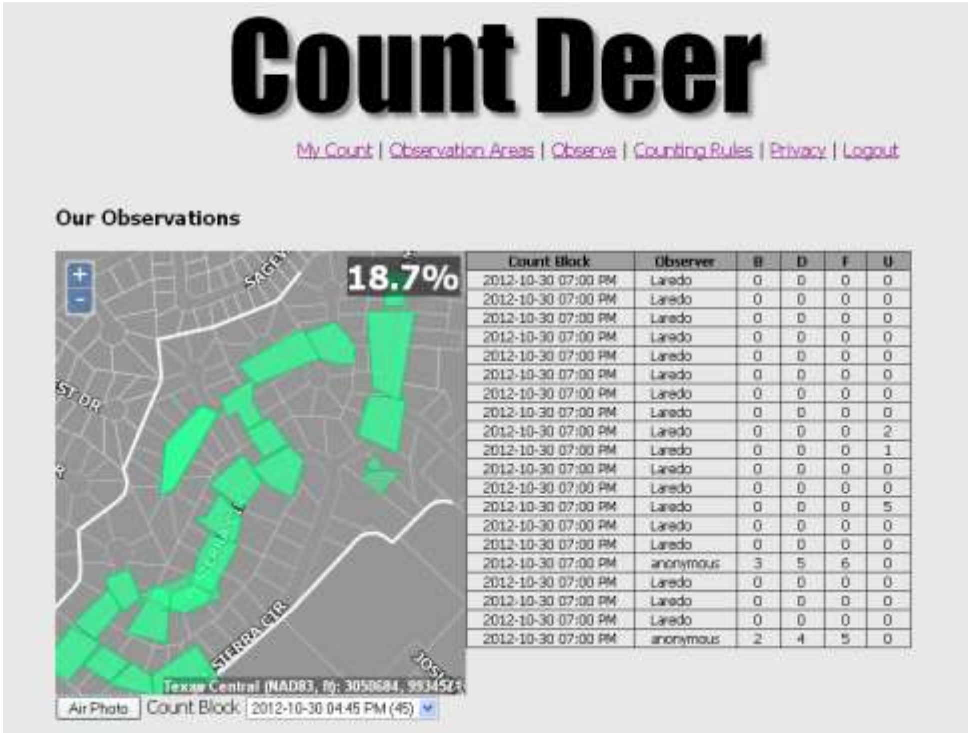
A web site was created to provide information about the project and facilitate the volunteer deer count. The project web site provided the following basic capabilities:

- User registration
- User authentication/authorization
- Add/delete user observation areas
- Record observations
- Review user and community participation

Volunteers were able to register with the site by providing a valid email address. An email with a confirmation link was sent to the registered email address to complete the two step registration process designed to limit unauthorized use of the system. Some information about the project was available without logging in so that site visitors could decide if they wanted to participate or not, but all of the data collection and data views required an authenticated user. After registration users were prompted to create one or more observation areas. Once the user had at least one observation area, users were allowed to submit observations. The observation form linked the count block, i.e. time of the observation, the observation area, and the number, age and sex of deer observed, if any. Users were encouraged to begin each count block with a zero-deer observation, if appropriate, in each observation area searched during the count block. Figure 3.4 shows the observation entry form.

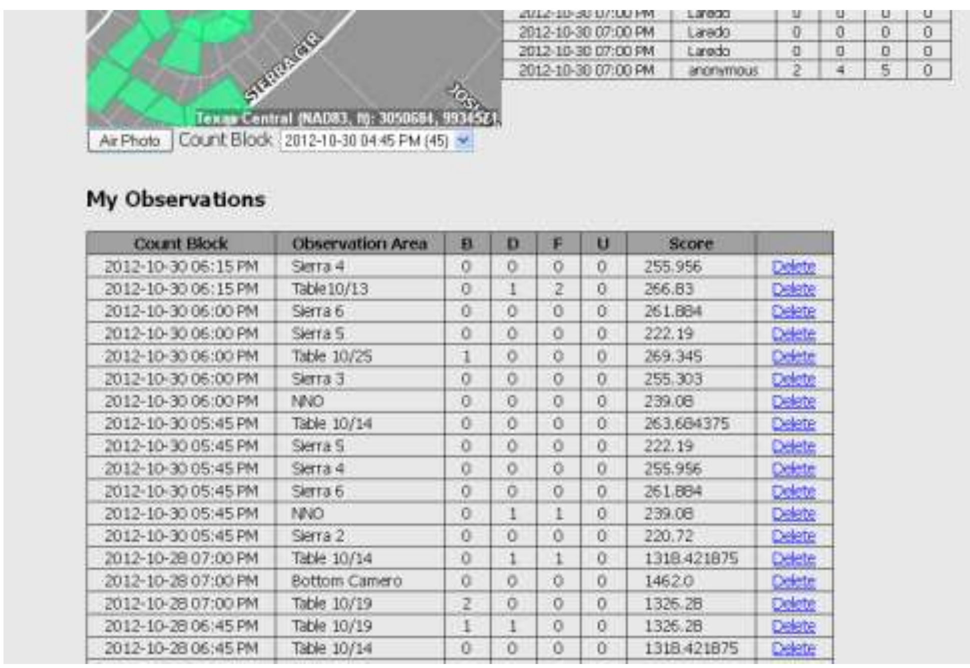
Figure 3.4 Observation entry form

User and community observations were displayed on a single status page. In the community section, labeled “Our Observations,” the collective area searched for the current count block was presented in map form next to a table showing the 20 most recent observations from all volunteers. The user results section, labeled “My Observations,” included a list of all individual user observations followed by a list of aggregated user observations for each count block. Figure 3.5 provides an example of the community results and figure 3.6 shows individual results.



Count Block	Observer	B	D	F	U
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	2
2012-10-30 07:00 PM	Laredo	0	0	0	1
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	5
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	anonymous	3	5	6	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	anonymous	2	4	5	0

Figure 3.5 Community results



2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	Laredo	0	0	0	0
2012-10-30 07:00 PM	anonymous	2	4	5	0

Count Block	Observation Area	B	D	F	U	Score	
2012-10-30 06:15 PM	Sierra 4	0	0	0	0	255.956	Delete
2012-10-30 06:15 PM	Table10/13	0	1	2	0	266.83	Delete
2012-10-30 06:00 PM	Sierra 6	0	0	0	0	261.884	Delete
2012-10-30 06:00 PM	Sierra 5	0	0	0	0	222.19	Delete
2012-10-30 06:00 PM	Table 10/25	1	0	0	0	269.345	Delete
2012-10-30 06:00 PM	Sierra 3	0	0	0	0	255.303	Delete
2012-10-30 06:00 PM	NNO	0	0	0	0	239.08	Delete
2012-10-30 05:45 PM	Table 10/14	0	0	0	0	263.684375	Delete
2012-10-30 05:45 PM	Sierra 5	0	0	0	0	222.19	Delete
2012-10-30 05:45 PM	Sierra 4	0	0	0	0	255.956	Delete
2012-10-30 05:45 PM	Sierra 6	0	0	0	0	261.884	Delete
2012-10-30 05:45 PM	NNO	0	1	1	0	239.08	Delete
2012-10-30 05:45 PM	Sierra 2	0	0	0	0	220.72	Delete
2012-10-28 07:00 PM	Table 10/14	0	1	1	0	1318.421875	Delete
2012-10-28 07:00 PM	Bottom Camero	0	0	0	0	1462.0	Delete
2012-10-28 07:00 PM	Table 10/19	2	0	0	0	1326.28	Delete
2012-10-28 06:45 PM	Table 10/19	1	1	0	0	1326.28	Delete
2012-10-28 06:45 PM	Table 10/14	0	0	0	0	1318.421875	Delete

Figure 3.6 Individual results

Incentive Program

Motivation is an important element of a volunteer project. A third-party donation to this project afforded an incentive program for volunteers. The goals of the incentive program were to 1) influence new volunteers to begin counting by making a first observation, 2) reward productive behavior without alienating productive volunteers, and 3) create a buzz and stimulate word of mouth promotion around the project. The incentive program ran for three weeks and each week started a new drawing. Three participating volunteers were drawn at random to receive a digital camera, a portable music player, and a streaming media player. Participating volunteers were volunteers that had made at least one observation during the week. The random drawing was intended to entice volunteers to get started making observations by becoming eligible to win a prize by simply making one observation. The grand prize each week was a tablet computer valued at approximately $500. Simply awarding the grand prize to the volunteer with the largest number of observations might have been perceived as a deterrent for runner-up volunteers, so an element of chance was included in the grand prize drawing. Like a lottery, the probability of winning the grand prize was made proportional to each volunteer's aggregate count block score which was determined by a formula based on a number of factors. Volunteers with the highest count block scores had the best chance of winning but any participating volunteer had the possibility of winning. One winning volunteer commented that the incentive program had persuaded him to participate where he might otherwise not have participated. An additional benefit of the incentive program was the introduction of constructive competition among volunteers and recognition for volunteer participation in weekly incentive award announcement emails.

Results

Participation

In total there were 4,914 observations across 764 count blocks. Nineteen of 26 registered users made at least one volunteer observation. As is seen in other VGI literature (Arsanjani, et al. 2013), a relatively small number of volunteers were responsible for a large percentage of total observations. In fact, the top two most productive observers generated over 50% of the total number of observations and the top 5 most productive produced nearly 90% of all observations. Table 3.1 presents the percentage of observations produced by the most productive observers.

Table 3.1 Cumulative Observations by Most Productive Volunteers

Volunteers	Cumulative Observations
Top 1 (5%)	33%
Top 2 (11%)	54%
Top 3 (16%)	71%
Top 4 (21%)	81%
Top 5 (26%)	88%

The project included a total 1,482 count blocks comprised of 1,338 non-push count blocks and 144 push count blocks. Slightly over half (51.5%, 764) of all count blocks had at least one observation. A total of 2,568 observations were recorded during the 1,338 non-push count blocks for an average of 1.91 observations per count block. A total of 2,346 observations were recorded during the 144 push count blocks for an average of 16.19 observations per count block. Volunteers were 8 times more likely to report an observation during a push count block. The largest number of observations in a single count block was 71 during count block 3341 which started at 7pm on Friday, 10/26/2012. This number represents a rate of one observation every 12.68 seconds. Figure 3.7 shows the total number of observations by each volunteer.

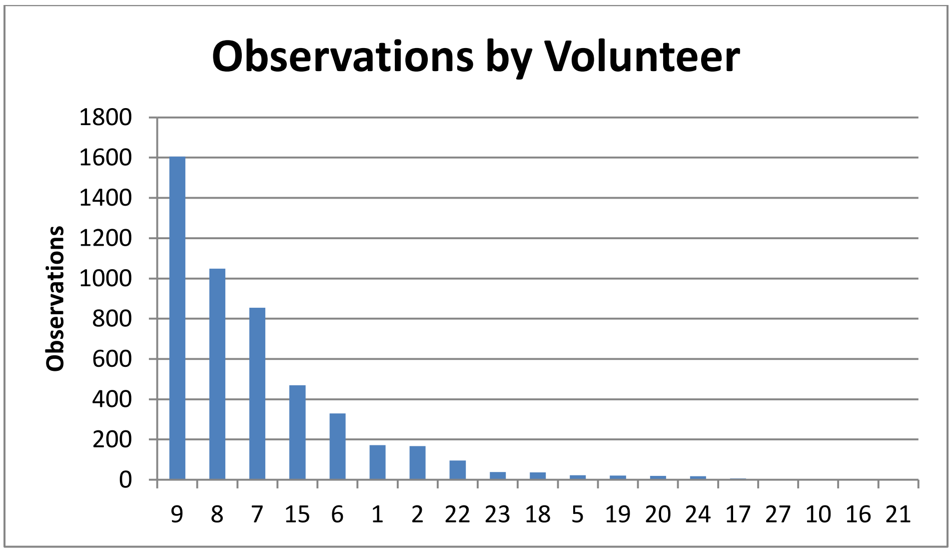


Figure 3.7 Total observations by volunteer

Simulation revealed study area coverage as an alternative metric for participation. Study area coverage, or simply coverage, is the fraction of the entire study area that was observed by any volunteer during a count block. More specifically, it is the ratio of the area of the geometric union of all observation areas during a count block to the total area of the study area. Figure 3.8 shows the distribution of coverage during non-push and push count blocks. Note that values of 0% coverage (no observations) have been omitted.

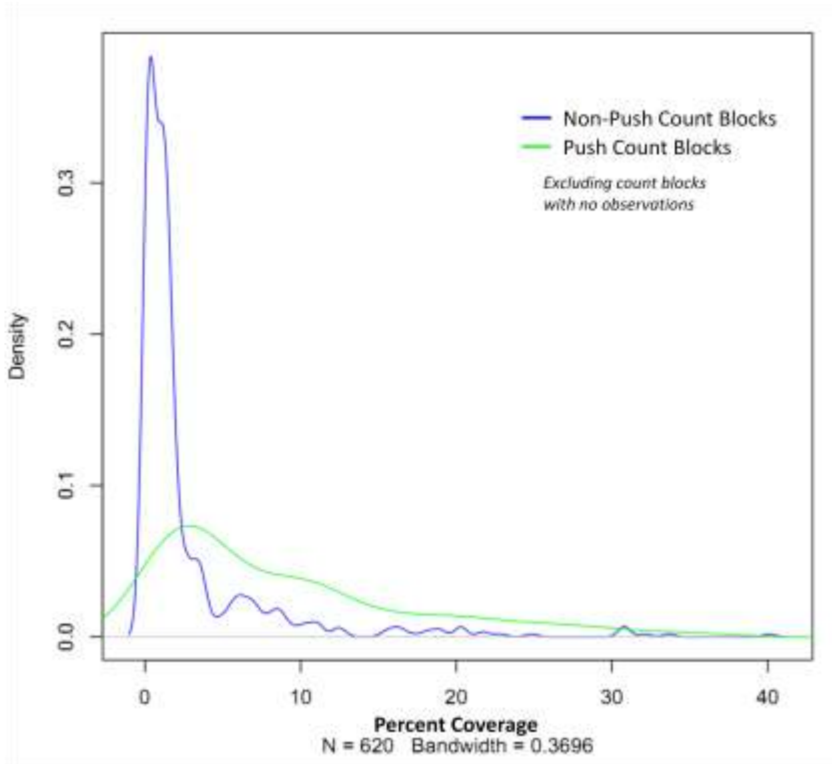


Figure 3.8 Distribution of observations per count block

Coverage median and interquartile range are shown in table 3.2 for non-push and push count blocks. Again, values of 0% coverage have been omitted.

Table 3.2 Aggregate Percent Coverage

Aggregate Percent Coverage		
	Median	**IQR**
Non-Push	1.10	1.99
Push	5.40	9.73

Figure 3.8 suggests that most non-push count blocks have relatively low coverage with very few count blocks having more than 5% coverage. Push count blocks, on the other hand, have coverage that is more widely distributed with a greater percentage of high coverage. Higher median coverage further suggests that coverage was higher during push count blocks than non-push count blocks which is consistent with greater

participation during push count blocks. Simulation also suggested that population estimate accuracy increases with coverage. Coverage is plotted against deer density in figure 3.9.

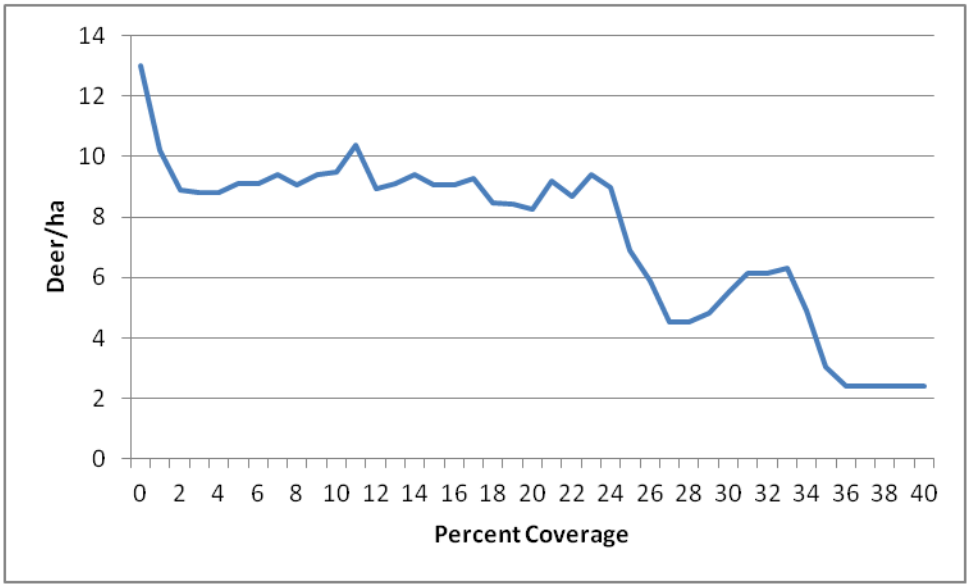


Figure 3.9 Deer density by percent coverage

As coverage increases, the number of count blocks with that level of coverage decreases meaning the density estimate is based on fewer count blocks. For example, 91 count blocks have at least 10% coverage and average density for these count blocks is 9.477 deer/ha, but the average density for the 11 count blocks that have 30% or greater coverage is 5.506 deer/ha. Estimated density generally decreases with an increase in coverage. This section has examined aggregate participation for the entire duration of the volunteer initiative. The next section examines volunteer participation over time.

Participation Over Time

Participation, measured in terms of number of observations or in terms of coverage generally increased from the beginning of the VGI initiative to the end. Figure

3.10 shows the number of observations by date and figure 3.11 shows maximum coverage by date.

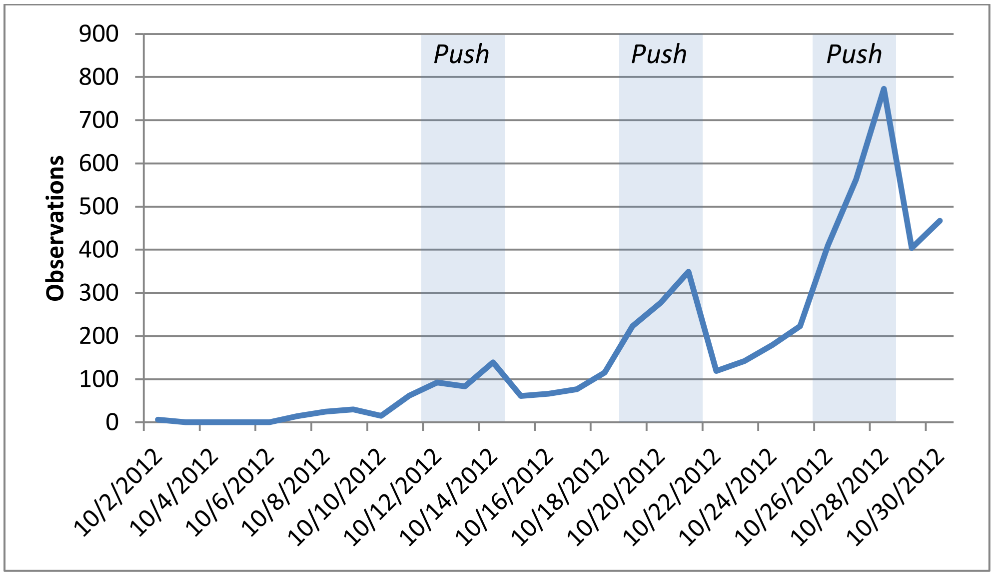


Figure 3.10 Observations by date

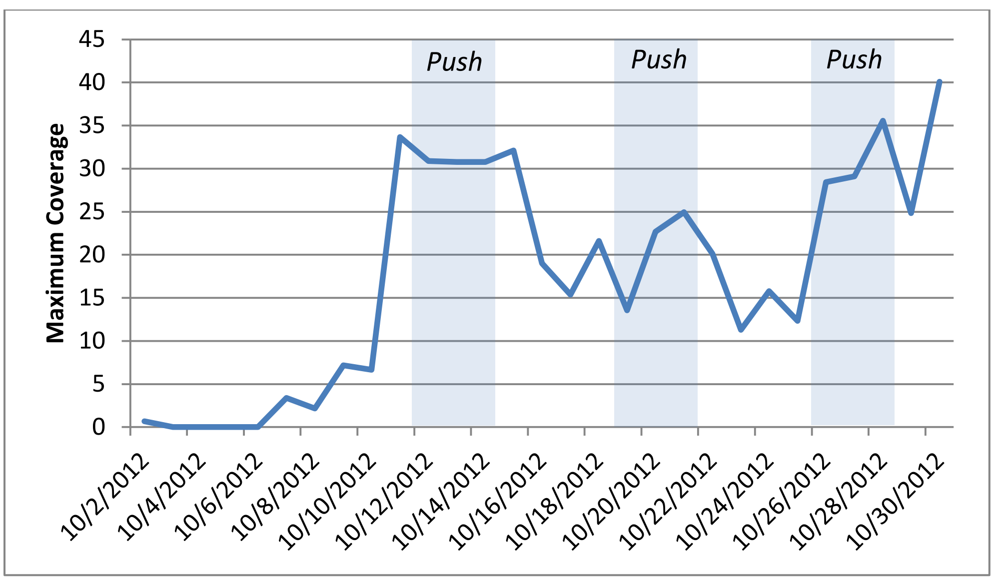


Figure 3.11 Coverage by date

Peaks in the number of observations are clear during pushes, particularly on the final day of pushes, and although less pronounced, slightly higher coverage is also observed during pushes providing further evidence of greater participation during pushes. The high level of coverage during the first push may be due more to a number of overstated observation areas than to actual observed area. During the first push, a few observations were associated with very large observation areas.

<u>Observation Areas</u>

Preliminary simulation used simple approaches to construct observation areas and each observer type used the same rules to allocate observation areas. This simplification was important for two reasons; to keep simulation models as simple as possible, but also creating observation areas another way would assume some knowledge about how people would allocate areas. With no precedent, any assumptions as to the definition of areas would be supposition. Figure 3.12 shows aggregate observation areas for four simulated Input Method E observers during a simulation run. Distinct home and road segment observation areas can be seen as well as the uniform construction of each individual observation area.

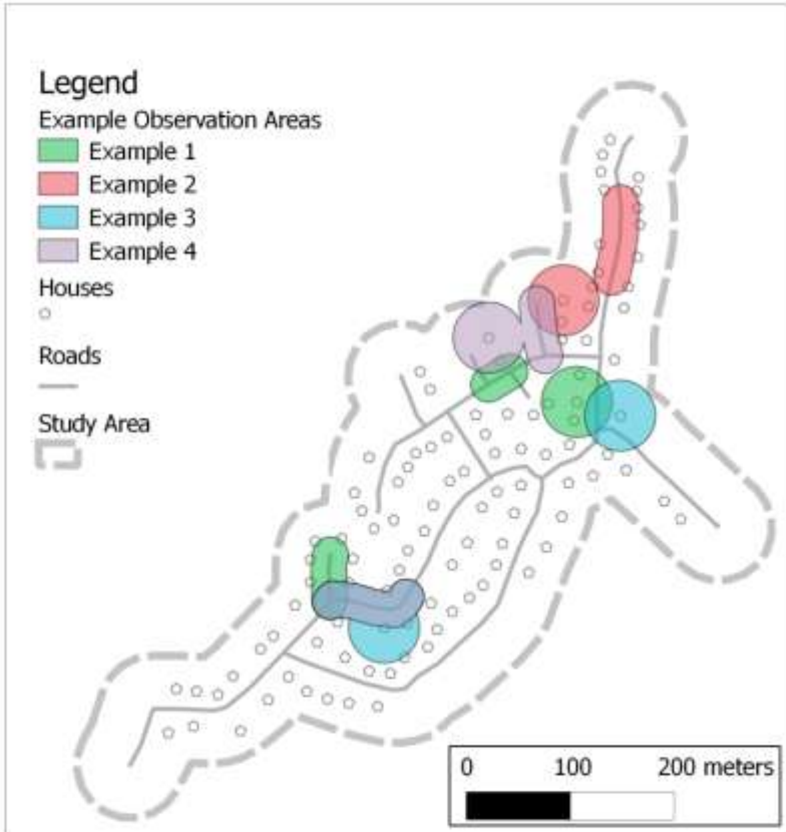


Figure 3.12 Aggregate observation areas for simulated method E observers

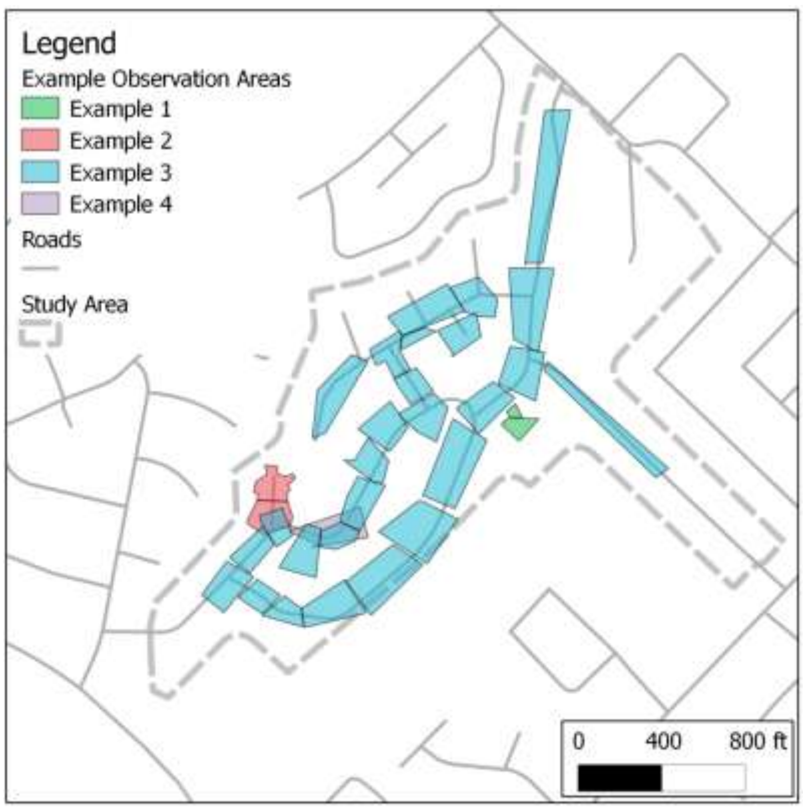


Figure 3.13 Aggregate observation areas for VGI observers

Figure 3.13 shows VGI observation areas reported with observations for four VGI volunteers during one count block. These observation areas are much less uniform reflecting the unique perspective each observer brought to the task of defining observation areas. Simulation observation areas were more distributed reflecting the stochastic nature of their origin whereas VGI observation areas reflect a more real-world arrangement. Simulated observers selected both house and road segment observation areas, whereas, VGI observers tended to use either house-oriented observation areas, like VGI Example 1 and Example 4, or road-oriented observation areas, like VGI Example 2 and Example 3. In fact, this behavior was pronounced enough to be the basis of one approach to VGI observer differentiation.

Observer Types

During simulation, all observers of a specific type behaved in exactly the same way, however, VGI observers demonstrated two patterns of behavior which can best be labeled “Sentinel” and “Scout.” Sentinels recorded observations associated with one or two house-oriented observation areas, whereas Scouts recorded observations associated with multiple road-oriented observation areas distributed along roads. VGI Examples 1 and 4 represent Sentinels and VGI Examples 2 and 3 are best described as Scouts. One indicator of the behavior type of any volunteer is the number of predefined observation areas as shown in figure 3.14. Volunteers with more than two predefined observation areas typically behave as Scouts. Volunteers with one or two observation areas typically behave as Sentinels. Individual volunteers tended to participate consistently according to one behavior pattern or the other although there were several examples where the same volunteer would behave as a Sentinel during some count blocks and as a Scout during other count blocks. The distinguishing feature of Scout behavior is the tour of a collection of adjacent observation areas along a roadway during a count block which closely mimics a traditional cruise survey.

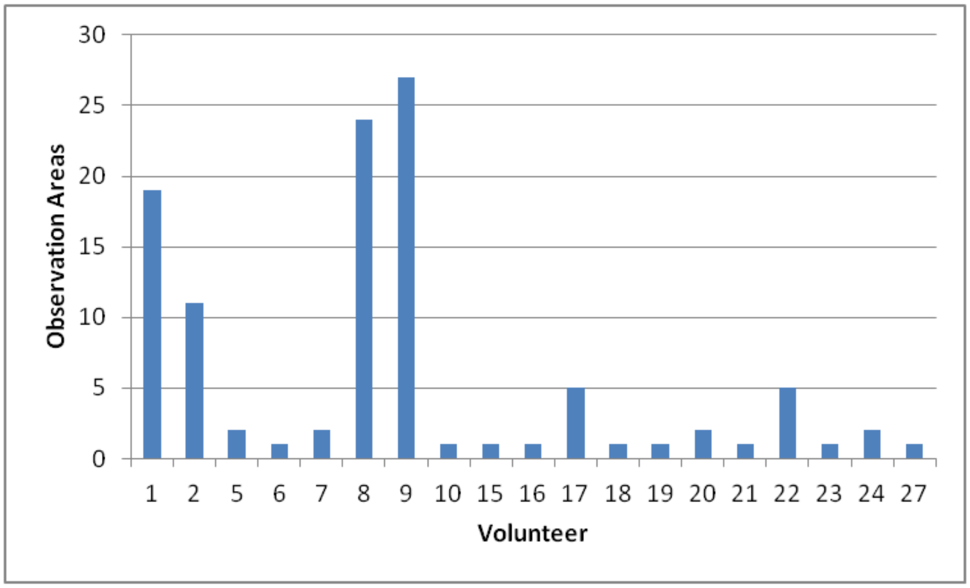


Figure 3.14 Number of observation areas by volunteer

Using the distinguishing characteristic of more than two observation areas, 6 of 19 (32%) volunteers could be labeled Scouts and 13 of 19 (68%) Sentinels.

Out of 2,346 push observations, 1,727 (74%) were recorded by Scouts and 619 (26%) were recorded by Sentinels. Scouts, on average, recorded 2 observations per push count block per observer versus 0.33 observations per push count block per Sentinel observer. Table 3.3 shows an ordered list of the number of observations by Scouts and by Sentinels along with the cumulative percentage of total observations of each type. Within each group, a relative small number of observers account for a large percentage of observations.

Table 3.3 Number and Cumulative Observations by Type

Scouts		
ID	**Push Observations**	**Cumulative %**
9	925	53.6%
8	607	88.7%
1	91	94.0%
2	65	97.7%
22	35	97.9%
17	4	100.0%
Sentinels		
ID	**Push Observations**	**Cumulative %**
7	235	38.0%
6	142	60.9%
15	141	83.7%
23	32	88.9%
19	20	92.1%
18	18	95.0%
24	18	97.9%
20	8	99.2%
5	2	99.5%
27	2	99.8%
21	1	100.0%

For both types of observer, a small number of observers contributed a large percentage of observations. One third of Scout observers contributed nearly 90% of Scout push observations and 3 out of 11 (27%) Sentinel observers contributed over 80% of Sentinel push observations. Number of observations is one measure of participation, coverage is another. Individual coverage can be defined as the ratio of the area of the geometric union of observation areas observed by a single observer during a count block to the area of the study area. It provides a measure of the amount of the study area observed by a single observer during a count block. Figure 3.15 shows the distribution of individual coverage for Scouts and Sentinels.

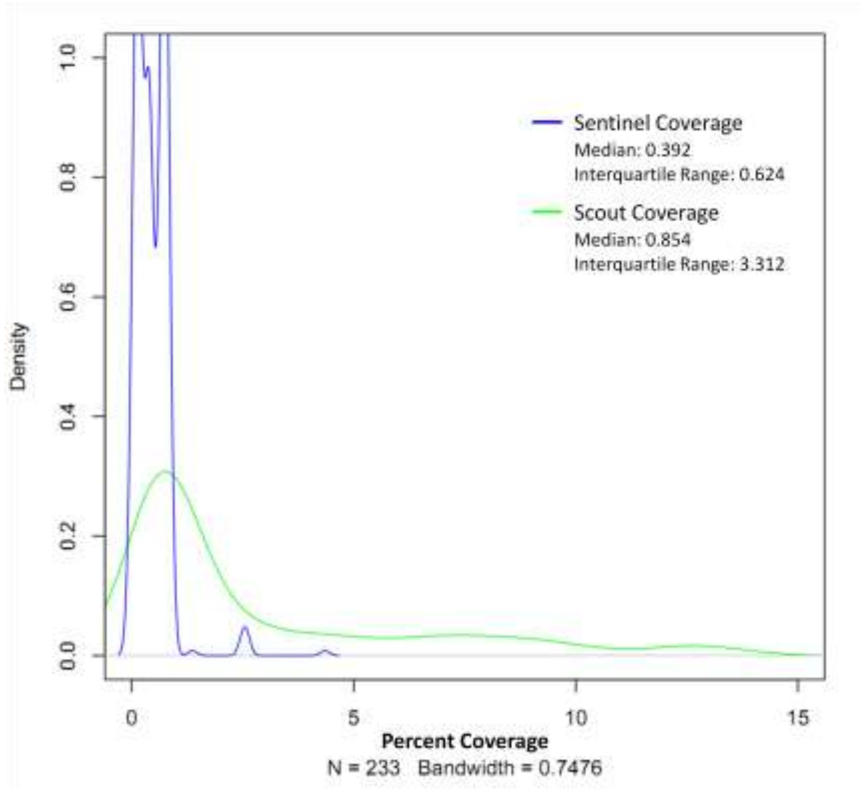


Figure 3.15 Distribution of coverage for Scouts and Sentinels

Sentinel coverage was limited to small areas whereas Scout coverage tended to be higher and more variable.

<u>Incentive Results and Count Block Score</u>

A well designed incentive program rewards behavior commensurate with its benefit to the project. In this project count block score takes into account the day, time and area of observations along with the number of observations. Table 3.4 presents an ordered list of the total count block score points earned by each of the top 10 scoring volunteers along with cumulative percentage of count block score points. The average observed deer density during push count blocks for each volunteer is shown in the last column.

Table 3.4 Count Block Score Points by Observer

Uid	Total Score	Cumulative % of total incentive points	Avg Push Density (deer/ha)
9	1,016,881	36.0%	2.306
8	866,580	66.7%	3.001
15	185,677	73.3%	2.341
1	118,509	77.5%	3.254
7	79,375	80.3%	9.273
2	61,710	82.5%	30.434
23	35,446	83.8%	3.077
20	34,231	85.0%	1.514
6	32,656	86.1%	13.637
18	23,921	87.0%	12.756

Observer 9, the top count block score point scorer alone earned 36% of all count block score points. Observer 8 adds another 30.7%, and the top 3 scorers combined account for 73.3% of all points awarded. While not as dramatic as other measures of participation, here again a relatively small number of volunteers account for a large amount of volunteer contribution.

Observed Area Distribution Map

The spatial arrangement of observations of Sentinels and Scouts vary dramatically. Sentinels recorded many observations within a limited number of observation areas whereas Scouts recorded observations over a larger number of spatially distributed observation areas. Figure 3.16 depicts the relative frequency of push observations by volunteers categorized as Sentinels and figure 3.17 shows the relative frequency of push observations by Scouts.

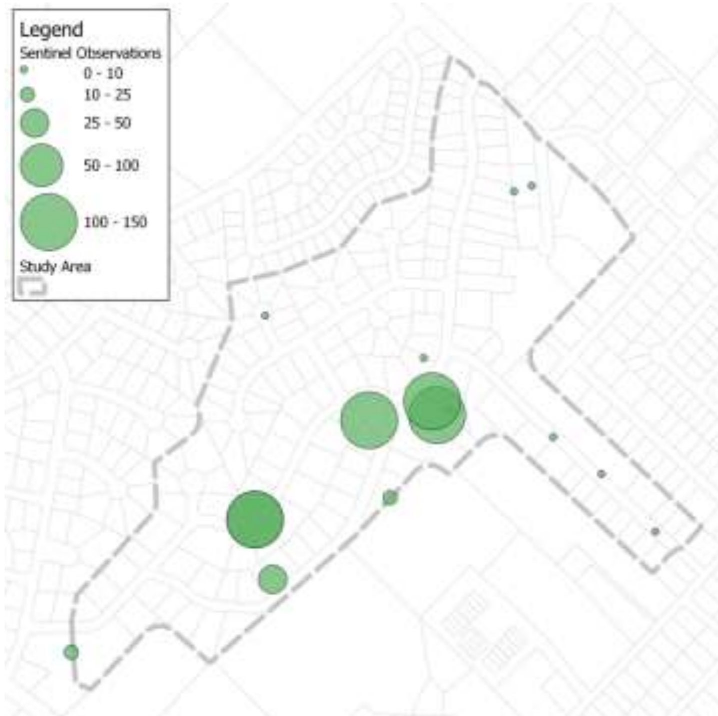


Figure 3.16 Relative frequency of push observations for Sentinels

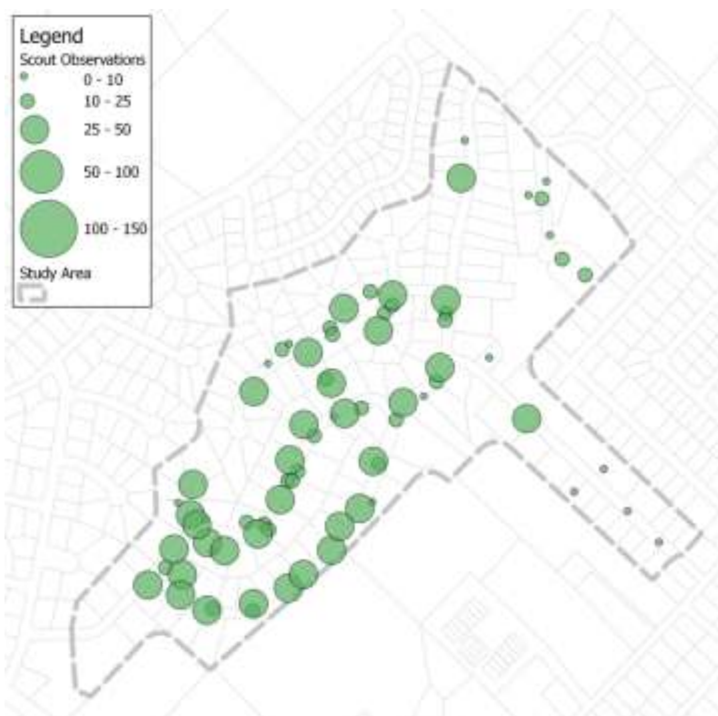


Figure 3.17 Relative frequency of push observations for Scouts

These maps show the areas that were searched for deer by Sentinels and Scouts during push count blocks. Figure 3.16 shows small, distinct, disconnected islands of observed area with low coverage whereas figure 3.17 shows distributed bands of coverage along neighborhood streets with comparatively greater coverage.

Deer Distribution Map

Figure 3.18 shows the total number of deer per observation reported during push count blocks by observation area. Many of the observation areas associated with higher total deer per observation are Sentinel observation areas. While it was not investigated further, this pattern may suggest a difference in behavior regarding zero-deer observation reporting between Scouts and Sentinels.

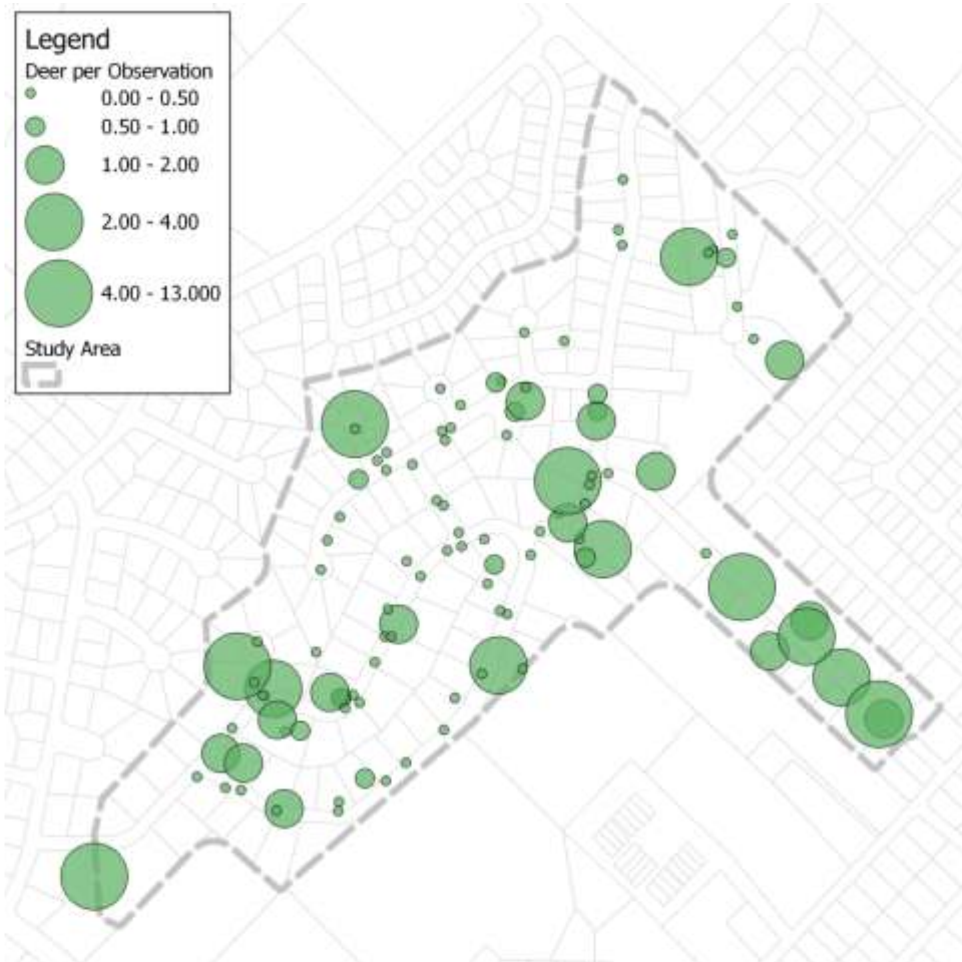


Figure 3.18 Observed deer per observation, push count blocks

<u>Aggregate Count</u>

The purpose of the VGI initiative was to produce a “useful” urban deer population estimate for the study neighborhood. Using the standard from simulation, a useful population estimate would be between 75% and 125% of the actual population. In order to arrive at an estimate some attention must be given to how to interpret the VGI data. Simulation revealed a number of factors that influenced the design and implementation of the VGI initiative, like the difficulty in using distance as a surrogate measure of observed area and the strong positive relationship between coverage area and population estimate accuracy. In a similar manner, simulation results also provide insight into VGI results interpretation. There are numerous approaches to interpret VGI data to establish an estimated deer population within the neighborhood and several of those approaches are applied here.

All In

The first approach establishes a mean deer density from all observations from all observers throughout the entire duration of the project, then extrapolates that density over the study area. This approach uses all volunteer effort, but it also includes (perhaps many) potential outliers and ignores information from simulation regarding coverage and participation. Using this approach, the estimated deer density was 13.019 deer/ha for a total of 421 deer in the study area.

All In, Push Only

Another approach establishes a mean deer density using observations from all observers but only during push count blocks, then extrapolates that density over the study area. This approach uses insight regarding coverage and participation at the expense of some volunteer effort. Using this approach, the estimated deer density was 9.666 deer/ha for total of 312 deer.

Coverage Threshold

Yet another approach capitalizes on the relationship between estimate accuracy and coverage observed during simulation to derive deer density. Simulation suggested a positive relationship between coverage and density estimate accuracy. As coverage increases, however, the number of count blocks and therefore the amount of volunteer effort used in the estimate decreases, that is, the estimate is based on a smaller sample of count blocks. The benefit of this approach is that it takes greater advantage of information learned from simulation, but a limitation of this approach is selection of an appropriate coverage threshold. At what level to set the coverage threshold is somewhat subjective. Table 3.5 below shows deer population estimates at several levels of coverage and corresponding number of count blocks.

Table 3.5 Population Estimates based on Count Block Percent Coverage

Percent Coverage	Number of Count Blocks	Percent of total Count Blocks	Density (deer/ha)	Population Estimate
10%	91	11.9%	9.48	307
20%	34	4.5%	8.25	267
25%	16	2.1%	6.88	223
30%	11	1.4%	5.51	178
35%	2	0.3%	3.04	98

By Type, All In

Analysis of results suggested two distinct types of observers, Scouts and Sentinels. Because participation was greater during push count blocks, only push count blocks are considered here. Average deer density and estimated deer population reported by Scouts and by Sentinels during push count blocks is shown in table 3.6. This approach considers the difference in observing style as well as the difference in participation between push and non-push count blocks.

Table 3.6 Deer Density by Type, All In

	Density (deer/ha)	Population Estimate
All Scouts	9.41	305
All Sentinels	9.79	317

By Type up to 80%

VGI results analysis also showed that for both Scouts and Sentinels, a relatively small number of volunteers were responsible for a large number of observations. Two Scouts produced over 80% of Scout observations and three Sentinels produced over 80% of Sentinel observations. Table 3.7 shows the average density and estimated population for these selected Scouts and Sentinels during push count blocks.

Table 3.7 Deer Density by Type, 80% of Observations by Type

	Density (deer/ha)	Population Estimate
80% Scouts	2.79	90
80% Sentinels	8.46	274

Incentive Score Weighted Average

If the incentive scoring system faithfully rewards desired behavior, then volunteer incentive score can be viewed as an alternative measure of contribution. The incentive scores of the three top scoring observers account for 73.7% of the total number of incentive points earned by all volunteers. The top three scoring volunteers included two Scouts and one Sentinel and are shown in table 3.8.

Table 3.8 Incentive Score Weighted Average Deer Density

UID	Total Score	Cumulative %	Avg Push Density (deer/ha)	Weight	Score
9	1,016,881	36.0%	2.31	0.49	1.13
8	866,580	66.7%	3.00	0.42	1.26
15	185,677	73.3%	2.34	0.09	0.21
				Sum	**2.60**

The incentive score-weighted average of average push count block deer densities of the top three scoring individuals is 2.6 deer/ha producing a deer population estimate of 83 deer. This estimate is representative of the average density reported by these three volunteers and takes advantage of over70% of volunteer contribution based on incentive score.

This chapter presented the VGI initiative and a number of interpretations of the VGI results. The list of interpretations presented here is not exhaustive but illustrates a variety of approaches which could be used to evaluate the VGI data. The next chapter provides details about an infrared-triggered camera deer survey that was conducted at the same time as the VGI count in order to estimate the neighborhood deer population using a scientifically accepted method.

4. INFRARED-TRIGGERED CAMERA DEER SURVEY

Chapter 2 described the use of computer simulation to inform the development and implementation of a Volunteered Geographic Information (VGI) neighborhood deer count initiative which is further described in Chapter 3. The primary result of the VGI initiative is a neighborhood deer population estimate. In order to produce an independent deer population estimate, a scientifically accepted Infrared-Triggered Camera (ITC) survey was conducted concurrently with the volunteer program. From September through December of 2012, a set of 8 infrared-triggered game cameras collected over 14,000 images, a subset of which were scored to produce an alternative neighborhood deer population estimate.

Study Area

The study area for the ITC survey was the same as for the volunteer deer count, the Sierra Circle neighborhood and a portion of the Tanglewood neighborhood in San Marcos, Texas. In total, eight cameras were deployed for the survey creating a camera density of 1 camera per 4 ha (10ac/camera), considerably higher than the suggested density of 1 camera per 65 ha demonstrated by Jacobson (Jacobson, et al. 1997) or 1 camera per 100 – 160 acres (40.5 – 64.8 ha) recommended by Texas Parks and Wildlife Department (Oetgen, Lambert and Whiteside 2008) or one camera per 46ha used in Roberts et al. (2006). While no cameras were located in adjacent green spaces, it is likely that these green spaces contributed to the effective habitat area for neighborhood deer. Mapping the specific locations of the cameras would reveal the identity of camera station hosts, so in the interest of privacy, the actual locations of cameras is not provided.

ITC Protocol

The camera survey protocol used in this project follows that established in Jacobson et al. (1997), further refined for public use by Texas Parks and Wildlife Department (Oetgen, Lambert and Whiteside 2008). The protocol includes image capture followed by image analysis. First, one or more infrared-triggered game cameras are placed in the study area for several days to several weeks to capture images of deer. At the end of the capture period, images are collected from the camera and analyzed. The total number of occurrences of bucks, does and fawns in the images are counted. Each unique buck is identified in the images providing a total number of bucks in the surveyed population. The ratio of the number of unique bucks to the total number of buck occurrences is multiplied by the number of doe occurrences to estimate the number of does in the population and multiplied by the number of fawn occurrences to estimate the number of fawns in the population. One departure from the recommended procedure was not baiting camera stations. There were two important reasons for this departure. First, baiting camera stations would have required the project to be conducted under an Institutional Animal Care and Use Committee (IACUC) protocol, guidelines established for the use of animals in research. By only capturing images and avoiding introducing anything into the habitat of the deer, the project remained IACUC exempt. A second and more influential reason for not baiting camera stations was neighborhood acceptance of the project. The consensus view towards feeding (or baiting) deer in the neighborhood was negative and neighborhood resistance to the project was a serious concern. Out of respect for the concerns of the neighborhood, this project did not bait camera stations.

Recruitment, Allocation and Establishment

Camera station hosts and locations were selected based on a number of factors including geographic distribution within the neighborhood, camera location characteristics and familiarity with the host. Most hosts were direct acquaintances while others were "friends-of-friends." Because camera station hosts needed to be able to perform periodic data collection activities, there were additional technical requirements for camera station hosts, for example, availability of a wireless network connected to the public Internet. Cameras were located in areas where deer were frequently seen but that also had a stable background to prevent false camera triggering. Camera locations were reasonably discrete in order to prevent tampering with the camera. The full list of camera station hosting requirements is shown below:

- Provide a reasonably secure location where camera is not likely to be stolen
- Free movement of deer (not fenced or low fence)
- Presence of things that attract deer, like landscape plants or decorative water features, feed, or that channel the movement of deer along a transportation corridor
- Shade -- 100% shade during October is better than sun/part sun
- Deer are already commonly seen at the location
- Location fits into distributed arrangement of camera stations
- With notice, provide access to camera by project personnel (to replace memory chips/batteries, etc.)
- High-speed Internet available for project use (Wifi is required, but could be arranged if necessary)
- Willing and able to remove data card for data upload every two to three days
- Accept that images may be scored by strangers, however, access to images will be strictly limited

Once hosts were selected, each camera location was tested for several days. In some cases cameras were moved or adjusted to prevent false triggering or improve capture results. Once the camera locations were set, they remained fixed for the duration of the camera survey. No cameras were moved, lost or stolen during the survey period.

Placing cameras on private property raised a number of privacy concerns, but steps were taken to address these concerns ahead of time. Perhaps the most important action was being proactive about privacy by raising the issue early and communicating a genuine concern for host privacy including discussing the measures taken to preserve privacy. These measures included not revealing the identity of any camera host even to each other, and being explicit about who would be able to view and score the images. Another important part of privacy was reinforcing the concept that the images captured by the host camera could be controlled by the host including the ability for the host to review images through a secure web site immediately after upload, and the right to request deletion of any images. Fortunately, there were no problems or concerns relative to privacy with the camera images.

Camera Station Operation

Camera Station

Each camera station was comprised of a game camera, memory card and upload assembly. The game camera used for the survey was a Wildgame Innovations model W8E (Micro Red 8) 8 Megapixel infrared game camera (Wildgame Innovations 2012). This model features a passive infrared sensor that triggers image capture. During the day (in the presence of adequate visible light), the camera captures normal color images. At night, in the absence of adequate visible light, the camera takes an infrared image using an infrared flash. Examples of each are shown in figure 4.1.

Figure 4.1 Normal color and infrared game camera images

Each camera was configured with the following settings:

- Resolution: high (8MP)
- Trigger: PIR (passive infrared sensor)
- PIR Mode: Still (capture still images rather than video)
- PIR Active: 24hr (capture images 24 hours/day)
- Delay: 30 seconds (wait 30 seconds before triggering next image)
- PIR Sensitivity: high (trigger on smallest movement)

Each camera was equipped with EyeFi model X2 4GB WiFi enabled SDHC card. A special feature of this memory card is the ability to transmit data captured on the card over an 802.11b/g wireless network (WiFi) to a storage server. In typical applications where there is consistent power to the card, images are transmitted within a few minutes, however, in the power-limited application of the game camera where the card is only powered on long enough for data to be written to the card, the card was not able to complete data transmission, so it simply operated as a standard SDHC memory card.

Outside of the camera using the upload assembly with a stable source of power and proper configuration, the WiFi enabled memory card was able to transmit data over the WiFi network to the data storage server. Each EyeFi memory card was programmed to use the camera host’s 802.11b/g wireless network to connect to the Internet and post new camera images to the storage server. Figure 4.2 shows an upload assembly and figure 4.3 depicts the upload process.

Figure 4.2 Upload assembly

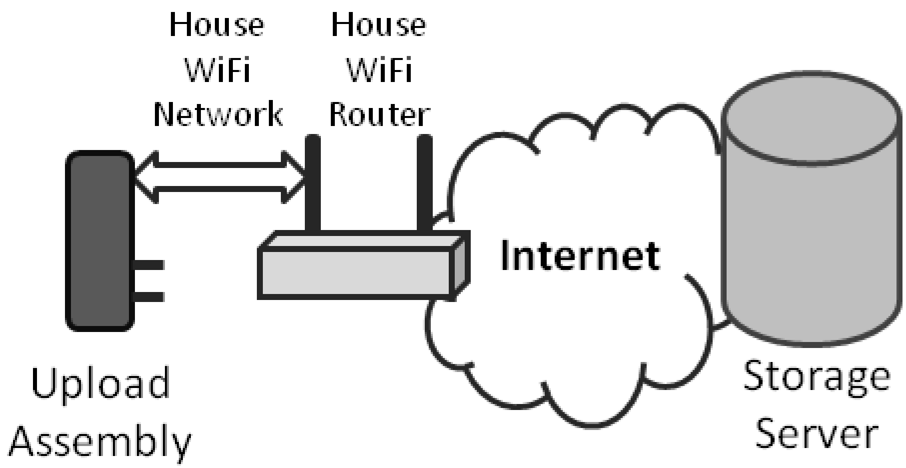


Figure 4.3 Upload process

<u>Image Upload</u>

Camera hosts were asked to perform an upload procedure 2 to 3 times per week to move images to the storage server. The upload procedure included removing the memory card from the camera, inserting the card into the upload assembly and plugging the upload assembly into wall power for about 30 minutes, then returning the memory card to the camera. The upload assembly was comprised of a USB SDHC card reader plugged into an A/C to USB power supply, a part commonly found in cell phone chargers.

Website

The web site for the ITC survey project provided three specific applications; 1) an image storage service, 2) an image index and screening application for camera hosts, and 3) an application for ITC survey image analysis and scoring.

Applications

Storage Service

The storage service processed and indexed images arriving from WiFi memory cards. On arrival of each new image, a process would catalog the new image recording image metadata into the project database and post the original resolution image, an intermediate resolution image (800x600) and a thumbnail image to a cloud-based object datastore and content delivery network.

Image Index and Screening

Through a secure logon, camera hosts were able to review images captured by their camera. The index was organized by date and provided a contact sheet view (thumbnails) of images as shown in figure 4.4. Clicking an image loaded the full size image. Each camera host was provided credentials (userid/password) authorizing access only to the host's images.

Figure 4.4 Image index web page

Image Scoring

Like the host screening capability, the image scoring facility required authenticated access to images. The scoring process included two passes through images. On pass one, image targets like does, fawns, deer of unknown type and other wildlife are marked with a point in each image and bucks are marked by a bounding box. On pass two, bucks marked in images were uniquely identified. As each buck image was presented the user could either add the buck as a new entry to the buck roster or associate the buck with an existing member of the buck roster. For this project a global buck roster was maintained for bucks identified in images from all cameras. A summary report displayed the number of images scored, the number each image target identified, the number of bucks unique to the camera and a population estimate for the camera.

Results

The ITC deer survey officially ran from September 22, 2012 through December 20, 2012 collecting 14,234 images from eight cameras, however, for this project only a subset of 6,242 images captured between October 1, 2012 and October 30, 2012, the date range corresponding to the volunteer deer count, are included in the analysis. Table 4.1 shows the number of images captured and scored by camera.

Table 4.1 Number of Images Captured by Camera

Camera	Scored Images
1	390
2	1823
3	489
4	329
5	832
6	305
7	1521
8	553
Total	6242

The scoring process begins with counting all occurrences of bucks, does and fawns in the images, then identifying the number of unique bucks in the images. The ratio of unique bucks to total buck occurrences provides a population estimate multiplier that can be used to scale the number of does and fawns to come up with a total population estimate. It is important to note that deer images were scored by the author who does not have any more than ordinary familiarity with deer. In other words, the images where not scored by a deer expert. Also, deer of unknown age/sex are ignored. Population estimates for each camera are presented in table 4.2.

Table 4.2 Population Estimates for Individual Cameras

	Targets			Population Estimates			
	Bucks	**Does**	**Fawns**	**Bucks (unique)**	**Does**	**Fawns**	**Total**
camera01	129	306	391	19	45	57	121
camera02	47	180	96	10	38	20	69
camera03	30	133	68	10	44	22	77
camera04	11	141	48	5	64	21	90
camera05	25	109	442	6	26	106	138
camera06	25	41	32	5	8	6	19
camera07	69	1184	1175	15	257	255	527
camera08	31	235	283	13	98	118	230

Population composition ratios for each camera are shown in table 4.3.

Table 4.3 Population Composition Ratios for Individual Cameras

	B:D	**D:F**
camera01	0.42	0.78
camera02	0.26	1.88
camera03	0.23	1.96
camera04	0.08	2.94
camera05	0.23	0.25
camera06	0.61	1.28
camera07	0.06	1.01
camera08	0.13	0.83

Population estimates in table 4.2 are for individual cameras and vary considerably from camera to camera. Some variation is expected between cameras. Population composition ratios also vary widely between cameras. While some variation in population composition might be expected between cameras, the dramatic variation present here is more likely caused by the position of the camera within the deer habitat. For example, camera07 captured many images of does and fawns resting at night, whereas camera03 captured deer moving from one location to another along a trail.

Table 4.2 presents population estimates based on individual cameras. What is needed is an aggregation of these results for the entire neighborhood. Several alternative approaches to aggregation follow.

All In

The TPWD Infrared-Triggered Camera survey instructions specify grouping images from all cameras rather than using individual camera images to determine unique bucks and number of occurrences of bucks, does and fawns. Composite results using this approach are presented in table 4.4.

Table 4.4 Aggregate ITC Results, All In

Targets			Population Est			
Bucks	**Does**	**Fawns**	**Bucks (unique)**	**Does**	**Fawns**	**Total**
367	2329	2535	30	190	207	428

The total population estimate using this interpretation is 428 deer.

Population Composition

An alternative approach to aggregating results across multiple cameras is to assume stable population composition ratios across the study area and use these ratios to scale the number of does and fawns. It is assumed that composition ratios are more stable over space. This is consistent with the simulation model created in chapter 2. Median buck to doe ratio and doe to fawn ratio across all cameras are .2275 and 1.1445 respectively. Given 30 uniquely identified bucks, population estimates for does and fawns we would be 132 does and 115 fawns for a total population estimate of 277 deer.

Redundant Images

An issue with several cameras was repeated images of the same one or more deer during a short period of time. The first step in evaluating whether or not these redundant images had a negative effect on population estimates was to quantify the occurrence of redundant images. Rather than using a subjective and time-consuming manual review of images, a simple heuristic was used to classify images as redundant. If an image included the same number of bucks, does, fawns and unknown deer as the previous image within a specified time window, the image was considered redundant. While it is possible some images were classified as redundant when they were not, this approach allowed efficient repeat analysis using various time windows. Table 4.5 contains the percentage of redundant images for each camera using several time windows.

Table 4.5 Percent Redundant Images by Time Window

	2 min	**5 min**	**10 min**	**15 min**	**20 min**
Camera	**120 sec**	**300 sec**	**600 sec**	**900 sec**	**1200 sec**
camera01	8.5%	9.6%	9.6%	11.4%	11.8%
camera02	12.0%	19.2%	22.5%	25.4%	25.7%
camera03	14.0%	17.1%	18.7%	20.2%	20.2%
camera04	15.1%	18.7%	21.7%	22.3%	22.3%
camera05	11.8%	16.8%	19.8%	21.4%	22.7%
camera06	7.8%	11.1%	13.1%	14.4%	17.0%
camera07	27.0%	26.3%	43.8%	46.3%	48.1%
camera08	7.7%	0.0%	12.2%	12.7%	13.7%

Generally, as the time window increased from five minutes to ten minutes and beyond, the percentage of redundant images became stable. The number of redundant images across cameras ranged from about 10% to almost 50% suggesting a rather high level of redundancy, but the real question was the influence of redundant images on population estimates. Table 4.6 shows population estimates for all images, images filtered using a 10 minute time window and images filtered using a 15 minute time window.

Table 4.6 Population Estimates, No Filter, 10, and 15 Minute Filter

No Filter	Targets			Population Est				
	Bucks	Does	Fawns	Bucks (unique)	Does	Fawns	Total	Pct Change
camera01	129	306	391	19	45	57	121	
camera02	47	180	96	10	38	20	69	
camera03	30	133	68	10	44	22	77	
camera04	11	141	48	5	64	21	90	
camera05	25	109	442	6	26	106	138	
camera06	25	41	32	5	8	6	19	
camera07	69	1184	1175	15	257	255	527	
camera08	31	235	283	13	98	118	230	
10 minute window	Targets			Population Est				
	Bucks	Does	Fawns	Bucks (unique)	Does	Fawns	Total	Pct Change
camera01	122	288	378	19	45	59	123	101.4%
camera02	39	143	76	10	37	19	66	95.9%
camera03	24	100	56	10	42	23	75	97.4%
camera04	11	96	42	5	44	19	68	75.3%
camera05	22	87	341	6	24	93	123	88.9%
camera06	21	37	29	5	9	7	21	109.0%
camera07	47	560	686	14	167	204	385	73.1%
camera08	28	204	264	12	87	113	213	92.4%
15 minute window	Targets			Population Est				
	Bucks	Does	Fawns	Bucks (unique)	Does	Fawns	Total	Pct Change
camera01	121	284	377	19	45	59	123	101.5%
camera02	38	141	73	10	37	19	66	96.1%
camera03	24	97	56	10	40	23	74	95.8%
camera04	11	95	42	5	43	19	67	74.7%
camera05	22	87	333	6	24	91	121	87.4%
camera06	21	37	29	5	9	7	21	109.0%
camera07	46	526	655	14	160	199	373	70.9%
camera08	28	203	262	12	87	112	211	91.9%

Filtering for redundant images increases individual camera population estimates as much as 9% and decreases estimates as much as 29%. Using the TPWD aggregation method on the images filtered using a ten minute redundant image filter the neighborhood population estimate is 342. With data filtered using the fifteen minute redundant image filter the estimate is 336 deer. Using median population composition approach, population estimates for the 10 and 15 minute filtered data are 259 and 256 respectively.

Representative Cameras

Large variations in population composition figures may suggest cameras were capturing areas of differing habitat function, for example shelter area, transportation corridor or feeding area. Cameras 2, 3 and 5 have relatively consistent population composition ratios that correspond to median ratios for all cameras, however, camera 5 includes a large number of redundant images. Rejecting camera 5 leaves cameras 2 and 3 as representative. Cameras 2 and 3 contain a combined total of 16 unique bucks. Combined information for these two cameras is shown in table 4.7.

Table 4.7 Population Estimate for Representative Cameras

Targets			Population Est			
Bucks	**Does**	**Fawns**	**Bucks (unique)**	**Does**	**Fawns**	**Total**
77	313	164	16	65	34	115

These cameras combined suggest a population estimate of 115 deer.

This chapter described the ITC survey that was conducted concurrently with the VGI initiative described in chapter 3 to serve as an independent and scientifically accepted method for determining a deer population estimate. The next chapter concludes by reviewing the results of this chapter and of chapters 2 and 3 in light of the research questions and outlining a number of other observations.

5. DISCUSSION, INTERPRETATION AND CONCLUSION

In the preceding chapters, computer simulation was used to inform the design and implementation of a facilitated VGI initiative, a neighborhood white-tailed deer count. The neighborhood count was conducted concurrently with an Infrared-Triggered Camera (ITC) survey for use as alternative scientifically accepted deer survey technique. The focus of this research is on four general questions: Can simulation guide what geographic information is collected and how? Can simulation reveal the influence of potential uncertainty on usefulness and provide a method to reduce its impact? Can simulation show the effect of participation on VGI usefulness? Do simulation results correspond to actual VGI results? The first three research questions deal specifically with the use of simulation to inform or improve the design and implementation of a facilitated VGI initiative, including a go/no-go decision. The fourth research question looks at similarities and differences between the simulation and the real VGI project. These questions are somewhat subjective, so support for conclusions is drawn from the experience of completing the project. Each question is reviewed below.

Research Questions

<u>Can simulation guide what geographic information is collected and how?</u>

Simulation set 1 employed five alternative methods for collecting deer observations. Four of the five alternatives methods, methods A through D, included variations on the use of distance as an indicator of search area. The fifth method, method E, associated observations with pre-defined search areas. While none of the methods met acceptance criteria, method E showed the most consistent result with the least variability.

The results of Simulation Set 1, therefore, guided the use of pre-defined search areas rather than the use of distance as a surrogate measure of area during the VGI initiative. Rather than collecting locations of volunteers or deer or the distance between the two, pre-defined observation areas were collected.

Simulation Set 2 examined the performance of each observation method relative to various levels of volunteer participation and deer density. Methods C and E showed greater stability across ranges of deer density and volunteer participation providing guidance for the selection of the observation method used during the VGI initiative.

Simulation Set 4A introduced zero-deer observations, reporting of searched observation areas that contained no deer. Simulation Set 4B tested the use of overlapping observation areas. The results of Simulation Set 4A showed a dramatic performance improvement in both Method C and Method E when zero-deer observations are reported. In this way the results of Simulation 4A guided how volunteers reported observations, by including zero-deer observations. The results of Simulation Set 4B, on the other hand, showed no worse performance in a reporting method using overlapping observation areas compared to non-overlapping observations areas. The results of this test guided how volunteers were allowed to create observation areas during the VGI initiative, specifically the use of more user-friendly overlapping observation areas.

These examples suggest that simulation is useful in selecting what geographic information is collected during a VGI initiative and how it is collected. Simulation provides a way to test and compare alternative approaches to collecting volunteer geographic information as well as subtle variations in those approaches.

Can simulation reveal the influence of potential uncertainty on usefulness and provide a method to reduce its impact?

Simulation Sets 3A and 3B focused on location uncertainty and distance estimation uncertainty. In Simulation Set 3A location uncertainty was limited to that associated with a map-based user interface, that is, geolocation was disabled, but this modification failed to produce material improvement in any input method. In Simulation Set 3B distance estimation uncertainty was disabled, but again, it had no influence on the performance of any input method. These simulations revealed the limited influence of these two types of uncertainty on result usefulness. This is a helpful finding because it allows the VGI facilitator to focus attention elsewhere as efforts to improve geolocation or distance estimation accuracy are unlikely to improve VGI results. Because of the limited influence of uncertainty in these simulations there was not an opportunity to use simulation to reduce it. Had uncertainty been more influential it is likely simulation could have been used to reduce its impact. Within this project the impact of uncertainty was managed by using all of the simulation results in combination to establish a best performing input method, one that used pre-defined, overlapping observation areas and that included zero-deer observations.

Can simulation show the effect of participation on VGI usefulness?

Volunteer participation in a new VGI initiative is difficult to predict, yet the level of participation could have a profound impact on the usefulness of the initiative results. Simulation Set 2 looked specifically at the performance of various input methods in terms of consistency at several levels of participation. Two input methods, C and E, emerged as being relatively robust to variations in participation across 3 levels of deer density.

Consistency is a desirable trait of an input method, but accuracy is important, as well. Sensitivity analysis using input method E2 including zero-deer observations in Simulation Set 4A showed a dramatic improvement in useful results with greater participation across all deer densities. These insights, taken in combination, not only provided confidence in the choice of input method for the VGI initiative but also highlighted the need for focused counting times to improve participation during certain time intervals.

Do simulation results correspond to actual VGI results?

Simulation provided a facility for experimentation with different aspects of VGI data collection including input method, various types of uncertainty and levels of participation. Simulation differentiated alternatives, for example, the relative performance of each input method in Simulation Set 1 or the effectiveness of uncertainty screening and filtering in Simulation Sets 3A and 3B. Simulation showed performance differences under alternative conditions, like in Simulation Set 2 where input methods were compared under various levels of deer density and participation. Simulation revealed general relationships and trends, for example in Simulation Set 4A the strong positive relationship between participation, represented as percent observed area, and population estimate accuracy, represented as the number of acceptable OADRs. Rather than arriving at a single representation of the actual VGI initiative, simulation encompassed a number of very generalized, limited, and purpose-specific representations of specific alternatives for a VGI initiative. There were a number of differences between the computer model used during simulation and the way in which volunteers participated in the VGI project. In other words, there was no single simulation that accurately corresponded to the actual VGI initiative, but all of the simulation steps in combination provided a more refined and informed starting point for the VGI initiative.

VGI vs. ITC

This section compares the results of the VGI initiative with the results of the ITC survey. Chapter 3 discusses the VGI initiative and several interpretations of the results. Chapter 4 does the same for the ITC survey. The difference between the All In population estimate of the VGI project and the All In population estimate of the ITC project is less than 2%, but the VGI estimate of 421 deer and the ITC estimate of 428 deer are unlikely. Deer density estimates for the Edwards Plateau region of Texas range from 15 acres per deer (0.165 deer/ha) to 3 acres per deer (0.824 deer/ha) (Armstrong and Young 2002) whereas a population of 421 deer in the 32 ha study area represents a density of 13.156 deer/ha. As discussed in both the VGI chapter and the ITC chapter, the results of each method can be interpreted in several ways.

The Incentive Score Weighted Average approach to analyzing VGI result allocates credibility to volunteers based on the number of incentive points earned. If the incentive program properly aligns with desired behavior, then the highest scorers are the ones that contributed the most to the project. The top three scorers account for over 70% of the total number of incentive points awarded meaning a very large amount of volunteer contribution is taken into consideration. In addition, this interpretation takes advantage of insight from simulation suggesting that population estimate accuracy improved with coverage area. In fact, VGI coverage area during push count blocks was considerably higher than non-push count blocks, so the use of only push count blocks in this interpretation is well founded. For these reasons the Incentive Score Weighted Average is selected as the best interpretation of the VGI results producing a population estimate of 83 deer.

Variability in both population estimates and population composition ratios among cameras combined with large numbers of redundant images on some cameras complicates ITC result interpretation. Using the Representative Cameras approach for cameras 2 and 3 addresses both the population composition issue and the redundant image issue. Population ratios for cameras 2 and 3 are consistent and each camera has relatively few redundant images. In addition, population estimates for each camera individually varied by less than 12%. Using only 2 cameras in the 32 ha study area is still well above other reported camera densities and the geographic distribution of cameras 2 and 3 is relatively balanced. For these reasons the Representative Cameras interpretation using cameras 2 and 3 is selected as the best interpretation of the ITC results producing a population estimate of 115 deer.

At 72% of the ITC population estimate, the VGI population estimate falls short of the 75% to 125% acceptability criteria meaning under the strictest interpretation of the stated standard the VGI initiative failed to produce useful results. The acceptability criteria for the VGI population estimate is specified relative to the "actual" population which is unknown. The ITC population estimate is intended to be representative of the real population but it, too, is an estimate subject to interpretation. The range of variability of population composition ratios and population estimates from individual cameras in the ITC survey was surprising and begs further investigation.

Although the VGI population estimate did not meet the defined standard for accuracy, simulation played an important role in the development of the VGI project. The next section presents additional observations regarding the VGI project.

Additional Observations

Part of the value of this research lies in what was discovered through the process of using simulation to inform the design and implementation of a VGI initiative. Each simulation step revealed insight that to a certain degree influenced subsequent simulation steps. For example, the results of Simulation Set 1 suggested that input method E was very inaccurate, but the choice was made to continue investigating the use of method E because it also demonstrated very low variability. Subsequent simulation sets, in particular numbers 3 and 4 tested for ways to improve the accuracy of method E while preserving consistency. Rather than being a faithful representation of a VGI initiative, simulation provided a tool for exploration, a mechanism to test alternatives for individual elements, like overlapping observation areas or zero-deer observations, for integration into the VGI initiative. In this capacity, simulation was extremely valuable. Without simulation, it is probable that a distance-based input method would have been used and considerable more effort invested in screening distance estimate uncertainty.

In the analysis for coverage area over time a large peak in study area during the first push was primarily the result of one volunteer's very large observation area. While the observation area did not represent the area that could be seen from a single vantage point, there was not a mechanism in place other than the incentive program to prevent this type of observation area from being drawn so the facilitator takes partial credit for this problem. One goal for this research was to try make every volunteer's contribution meaningful, but it may be that there will always be some volunteer effort that is seen as contributing noise to VGI. In a facilitated VGI project the facilitator and volunteer share in the responsibility for this.

Contributions from relatively few volunteers made up the majority of the VGI dataset. Volunteers do not contribute in equal amounts.

Simulation did not predict the two distinct types of observers that emerged during the VGI initiative, Sentinels and Scouts. Sentinels intently watched a few areas whereas Scouts repeatedly toured many areas. The Scout method of observation is very similar to a traditional cruise deer survey. Some research suggests that cruise surveys may underestimate deer populations (Roberts et al. 2006).

The pattern of aggregate observation was quite similar between simulation and VGI initiative in that in both cases areas around some homes and areas along roads were searched for deer, however, who searched these areas was different. During simulation, method E observers searched both house search areas and road search areas, but during the VGI initiative, house search areas were used by Sentinel type volunteers and Scout type volunteers primarily searched road areas.

With the new found knowledge of Sentinels and Scouts, the input method used by each should be revisited as there is no particular need that they use the same one. Perhaps there are different input methods better suited for each one.

The relationship between estimate accuracy and coverage highlighted through simulation influenced the use of concentrated observation times, or pushes, during the VGI initiative. The incentive program was carefully designed to reinforce coordinated volunteer behavior. VGI results indicate volunteers were 8 times more likely to participate during a push count block suggesting the effectiveness of the incentive program.

While simulation did not address it at all, VGI participation built gradually over several weeks and through each push. This phenomenon was not examined in this project but may be worth future investigation.

Screening, filtering, analyzing and interpreting data are important tasks in the shift towards VGI (Kuhn 2007, Coleman 2009), but effective screening and filtering is difficult without a baseline dataset. Using simulation to screen for location position uncertainty or distance estimation uncertainty was not particularly effective in this study but may be more so under different circumstances.

In preparation for simulation, only area-based survey methods and relatively simple methods that were considered for volunteers, but an opportunity exists to use more sophisticated ecological survey methods in cases where volunteers are able to provide observations of sufficient quality to support the method. What is asked of volunteers should be simple, but the underlying method for aggregating and analyzing the data can be as complex as necessary. For example, if a method was developed that allowed volunteers to reliably detect and measure distances to deer, then distance sampling might be a very good strategy.

Addressing neighborhood residents' concerns about the use of deer count results up front helped with recruitment.

The map user interface that was part of the VGI web application was difficult for some users to use and allowed a number of observation area “outliers.” Uncertainty related to drawing observation areas in the map user interface was not included in preliminary simulation but is a good candidate for experimental research.

Simulation is useful, but it is easy to get bogged down in the complexity of simulation. Simple models and simple tests worked the best. Simulation may be best used as an exploratory tool to arrive at a better answer rather than relied upon for a “best” answer.

There were three unexpected results from the ITC survey. First, there was considerable variability in population estimates and population composition ratios among individual cameras. Second, some cameras included a large number of redundant images. Third, results from all cameras aggregated according to standard procedures produced an unlikely population estimate. These results suggest the need for further investigation into possible differences between urban ITC surveys and rural ITC surveys.

Next Steps

During this project two vast datasets were collected, the VGI deer observation dataset and the ITC survey dataset. There are many ways in which one or the other or both could be further analyzed.

Perhaps the highest priority next step is to better understand the results of the ITC survey, for example the variation in population composition estimates as well as population estimates among cameras. Another avenue for research with the ITC dataset is to examine the use of volunteers to score ITC images. A third use of the ITC images takes a look at temporal patterns of deer occurrences in images, not only daily patterns, but also patterns over the 6 week total duration of the camera survey.

Using the VGI deer observation dataset, one could examine the value proposition of this VGI project by estimating the cost of the volunteer effort in terms of dollars and compare that cost and quality with the cost and quality of a scientific or professional survey. Another project could more closely examine the temporal and spatial patterns of participation throughout the project. Another project might compare deer detection patterns and zero-deer observation frequency between Sentinels and Scouts. Although only two observer types were detected in the volunteer count, others may emerge with more investigation. In addition, studies of other VGI datasets like OpenStreetMap could be replicated with this VGI dataset.

The volunteer deer counting method could be refined, particularly in light of the two emergent types of volunteers, finding input methods that are most effective for each type. Simulation is a good strategy for conducting this further work.

Other projects might look at experimentally quantifying potential uncertainty types, for example, web map user interface digitization uncertainty, particularly across devices, and real-world geolocation uncertainty, that is, geolocation uncertainty as it might be encountered by volunteers.

Summary

This project used computer simulation to investigate potential uncertainty and levels of participation in order to inform the design and implementation of a facilitated VGI initiative, a neighborhood white-tailed deer count. The project demonstrated that simulation can inform the choice of data to collect and how to collect it, that simulation can reveal the influence of potential uncertainty and may provide a method to reduce its impact, and that simulation can show the effect of participation on VGI usefulness. The project did not demonstrate that simulation results correspond to actual VGI results because of the difficulty in predicting volunteer behavior in a new VGI initiative. Rather, it showed that simulation is a useful tool for exploring alternatives for specific elements of a VGI initiative or the influence of certain conditions on the VGI initiative which may lead to incremental improvement in the design of the VGI initiative.

This project is meaningful to GIScience literature in three ways. First, it creates a new, original VGI dataset from the volunteer deer count. VGI research to date relies heavily on the venerable OpenStreetMap dataset. This project provides an alternative dataset for comparison and further examination. Second, this project highlights an opportunity to improve VGI through research on uncertainty and error associated with methods commonly used in VGI, like the use of web map user interfaces or geolocation technologies. Third, this project demonstrates the use of simulation to improve the design and implementation of a facilitated VGI initiative which is useful not only for new projects, but also for improving existing projects.

APPENDIX SECTION

APPENDIX A: IRB EXEMPTION REQUEST APPROVAL

Exemption Request EXP2011K5173 - Approval

AVPR IRB [ospirb@txstate.edu]

You replied on 4/1/2011 4:30 PM.

Sent: Friday, April 01, 2011 1:23 PM

To: Nicosia, David A

DO NOT REPLY TO THIS MESSAGE. This email message is generated by the IRB online application program.

Based on the information in IRB Exemption Request EXP2011K5173 which you submitted on 03/22/11 16:39:41, your project is exempt from full or expedited review by the Texas State Institutional Review Board.

If you have questions, please submit an IRB Inquiry form:

http://www.txstate.edu/research/irb/irb_inquiry.html

Comments:
No comments.
=====================================
Institutional Review Board
Office of Research Compliance
Texas State University-San Marcos
(ph) 512/245-2314 / (fax) 512/245-3847 / ospirb@txstate.edu / JCK 489
601 University Drive, San Marcos, TX 78666

WORKS CITED

Ahola, Terhi, Kirsi Virrantaus, Krisp Matthias, and Gary J. Hunter. "A spatio-temporal population model to support risk assessment and damage analysis for decision-making." *International Journal of Geographical Information Science* 21, no. 8 (2007): 935-953.

Armstrong, W. E., and E. L. Young. *White-tailed Deer Management in the Texas Hill Country.* Texas Parks and Wildlife, 2002.

Arsanjani, Jamal Jokar, Christopher Barron, Mohammed Bakillah, and Marco Helbich. "Assessing the Quality of OpenStreetMap Contributors along with their Contributions." *16th AGILE Conference on Geographic Information Science.* Leuven, Belgium, 2013.

Batty, Michael, Andrew Hudson-Smith, Richard Milton, and Andrew Crooks. "Map mashups, Web 2.0 and the GIS revolution." *Annals of GIS* 16, no. 1 (March 2010): 1-13.

Birkin, Mark, Nick Malleson, Andy Hudson-Smith, Steven Gray, and Richard Milton. "Calibration of a spatial simulation model with volunteered geographical information." *International Journal of Geographical Information Science* 24, no. 8 (2011): 1221-1239.

Bishr, Mohamed, and Lefteris Mantelas. "A trust and repudiation model for filtering and classifying knowledge about urban growth." *GeoJournal* 72 (2008): 229-237.

Bolstad, Paul V., Paul Gessler, and Thomas M. Lillesand. "Positional uncertainty in manually digitized map data." *International Journal of Geogrphical Information Systems* 4, no. 4 (1990): 399-412.

Budhathoki, Nama Raj, Bertram Bruce, and Zorica Nedovic-Budic. "Reconceptualizing the role of the user of spatial data infrastructure." *GeoJournal* 72 (2008): 149-160.

Budhathoki, Nama Raj, Zorica Nedović-Budić, and Bertram (Chip) Bruce. "An Interdisciplinary Frame for Understanding Volunteered Geographical Information." *Geomatica* 64, no. 1 (2010): 11-26.

Chen, X, and FB Zhan. "Agent-based modelling and simulation of uban evacuation: relative effectiveness of simultaneous and staged evacuation strategies." *Journal of the Oprational Research Society* 59 (2008): 25-33.

Coleman, David J., Yola Georgiadou, and Jeff Labonte. "Volunteered Geographic Information: The Nature and Motivation of Produsers." *International Journal of Spatial Data Infrastructures Research* 4 (2009): 332-358.

Collier, Bret A., Stephen S. Ditchkoff, Joshua B. Raglin, and Jorgan M. Smith. "Detection Probability and Sources of Variation in White-tailed Deer Spotlight Surveys." *The Journal of Wildlife Management* 71, no. 1 (2007): 277-281.

Connors, John Patrick, Shufei Lei, and Maggi Kelly. "Citizen Science in the Age of NeoGeography: Utilizing Volunteered Geographic Information for Environmental Monitoring." *Annals of the Association of American Geographers* 102 (2011): 1-23.

Cooke, J. G. "Improvement of fishery-managment advice through simulation testing of harvest algorithms." *ICES Journal of Marine Science* 56 (1999): 797-810.

Coolidge, Jena. "City takes on rising deer population." *University Star*, February 24, 2011: 1.

Crampton, Jeremy W. "Cartography: maps 2.0." *Progress in Human Geography* 33, no. 1 (2009): 91-100.

Elwood, Sarah. "Geographic information science: emerging research on the societal implications of the geospatial web." *Progress in Human Geography* 34, no. 3 (2010): 349-357.

Elwood, Sarah. "Geographical Information Science: new geovisualization technologies -- emerging questions and linkages with GIScience research." *Progress in Human Geography* 33, no. 2 (2009): 256-263.

Elwood, Sarah. "Grassroots groups as stakeholders in spatial data infrastructures: challenges and opportunities for local data development and sharing." *International Journal of Geographical Information Science* 22, no. 1 (January 2008): 71-90.

Environmental Protection Agency. *Monitoring and Assessing Water Quality - Volunteer Monitoring.* 2012. http://water.epa.gov/type/rsl/monitoring/index.cfm (accessed 9 15, 2012).

Flanagin, Andrew J., and Miriam J. Metzger. "The credibility of volunteered geographic information." *GeoJournal* 72 (July 2008): 137-148.

Girres, Jean-François, and Guillaume Touya. "Quality Assessment of the French OpenStreetMap Dataset." *Transactions in GIS* 14, no. 4 (2010): 435-459.

Goodchild, Michael F. "Citizens as sensors: the world of volunteered geographic information." *GeoJouran* 69 (November 2007): 211-221.

Goodchild, Michael F. "Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0." *International Journal of Spatial Data Infrastructures Research* 2 (2007): 24-32.

Goodchild, Michael F. "Commentary: whither VGI?" *GeoJournal* 72 (2008): 239-244.

Goodchild, Michael F. "Geographic Information Science and Systems for Environmental Management." *Annual Review of Environmental Resources* 28 (2003): 493-519.

—. "Spatial Accuracy 2.0." *Proceedings of the 8th international symposium on spatial accuracy assessment in natural resources and environmental sciences.* Shanghai, 2008.

Goodchild, Michael F., and J. Alan Glennon. "Crowdsourcing geographic information for disaster response: a research frontier." *International Journal of Digital Earth* 3, no. 3 (September 2010): 231-241.

Goodchild, Michael. "NeoGeography and the nature of geographic expertise." *Journal of Location Based Services* 3, no. 2 (June 2009): 82-96.

Gouveia, Cristina, Alexandra Fonseca, António Câmara, and Francisco Ferreira. "Promoting the use of environmental data collected by concerned citizens through information and communication technologies." *Journal of Environmental Management* 71 (2004): 135-154.

Gouveia, Cristina, and Alexandra Fonseca. "New approaches to environmental monitoring: the use of ICT to explore volunteered geographic information." *GeoJournal* 72 (2008): 185-197.

Grira, Joel, Yvan Bédard, and Roche Stéphane. "Spatial Data Uncertainty in the VGI World: Going from Consumer to Producer." *Geomatica* 64, no. 1 (2009): 61-71.

Haklay, M. "Geographical Citizen Science - clash of cultures and new opportunities." *Position paper for GIScience workshop on the role of VGI in advancing science.* 2010.

Haklay, Muki, Alex Singleton, and Chris Parker. "Web Mapping 2.0: The Neogeography of the GeoWeb." *Geography Compass* 2, no. 6 (2008): 2011-2039.

Harvey, Francis. "Commentary." *Environment and Planning B: Planning and Design* 34 (2007): 761-764.

Helton, J. C., J. D. Johnson, C. J. Sallaberry, and C. B. Storlie. "Survey of sampling-based methods for uncertainty and sensitivity analysis." *Reliability Engineering and System Safety* 91 (2006): 1175-1209.

Hudson-Smith, Andrew, Andrew Crooks, Maurizio Gibin, Richard Milton, and Michael Batty. "NeoGeography and Web 2.0: concepts tools and applications." *Journal of Location Bases Services* 3, no. 2 (2009): 118-145.

Jacobson, Harry A., James C. Kroll, Randy W. Browning, Ben H. Koerth, and Mark H. Conway. "Infrared-Triggered Cameras for Censusing White-Tailed Deer." *Wildlife Society Bulletin* 25, no. 2 (1997): 547-556.

Jester, Steve, and Jim Dillard. *Conducting White-Tailed Deer Spotlight Surveys in the Cross-Timbers and Prairies Regions of North and Central Texas.* Austin: Texas Parks and Wildlife Department, n.d.

Koenen, Kiana K. G., Stephen DeStefano, and Paul R. Krausman. "Using Distance Sampling to Estimate Seasonal Densities of Desert Mule Deer in a Semidesert Grassland." *Wildlife Society Bulletin* 30, no. 1 (2002): 53-63.

Kuhn, Werner. "Volunteered Geographic Information and GIScience." Position Paper for the NCGIA and Vespucci Workshop on VGI; Santa Barbara, CA Dec. 13-14, 2007, 2007.

Lu, Yongmei. "Pervasive location acquisition technologies: Opportunities and challenges for geospatial studies." *Computers, Environment and Urban Systems* 36, no. 2 (2012): 105-108.

MacEachren, Alan M., et al. "Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know." *Cartography and Geographic Information Science* 32, no. 3 (2005): 139-160.

McCullough, Dale R. "Evaluation of Night Spotlighting as a Deer Study Technique." *The Journal of Wildlife Management* 46, no. 4 (1982): 963-973.

McCullough, Dale R. "Variation in Black-Tailed Deer Herd Composition Counts." *The Journal of Wildlife Management* 57, no. 4 (1993): 890-897.

McKinley, William T., Stephen Demarais, Kenneth L. Gee, and Harry A. Jacobson. "Accuracy of the Camera Technique for Estimating White-tailed Deer Population Characteristics." *Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies.* 2006. 83-88.

Meng, Xiaolin, Wenzhong Shi, and Dajie Liu. "Statistical Tests of the Distribution of Errors in Manually Digitized Cartographic Lines." *Annals of GIS* 4 (1998): 52-58.

Metropolis, Nicholas, and S. Ulam. "The Monte Carlo Method." *Journal of the American Statistical Association* 44, no. 247 (September 1949): 335-341.

National Audubon Society, Inc. *Christmas Bird Count.* 2012. http://birds.audubon.org/christmas-bird-count (accessed 9 15, 2012).

NOAA. *NOAA Fisheries Toolbox.* 2011. http://nft.nefsc.noaa.gov/index.html (accessed 3 19, 2012).

Oetgen, Jesse G., Billy C., Jr. Lambert, and Jay D. Whiteside. *Surveying White-Tailed Deer Populations Using Infrared-Triggered Cameras.* Austin, TX: Texas Parks and Wildlife Department, 2008.

OpenStreetMap contributors. *OpenStreetMap.* 2012. http://www.openstreetmap.org/ (accessed 2012).

O'Reilly, Tim. "What is Web 2.0: Design patterns and business models for the next generation of software." *Communications and Strategies* 65 (2007): 17-37.

Pearce, Jennie L., and Mark S. Boyce. "Modelling distribution and abundance with presence-only data." *Journal of Applied Ecology* 43 (2006): 405-12.

Pfeffer, Max J., and Linda P. Wagenet. "Volunteer Environmental Monitoring, Knowledge Creation and Citizen-Scientist Interaction." In *The SAGE Handbook of Environment and Society*, 235-249. 2007.

Rinner, Claus, Carsten Keßler, and Stephen Andrulis. "The use of Web 2.0 concepts to support deliberation in spatial decision-making." *Computers, Environment and Urban Systems* 32 (2008): 386-395.

Roberts, Clay W., et al. "Comparison of Camera and Road Survey Estimates for White-tailed Deer." *Journal of Wildlife Management* 70, no. 1 (2006): 263-267.

Seeger, Christopher J. "The role of facilitated volunteered geographic information in the landscape planning and site design process." *GeoJournal* 72 (2008): 199-213.

Silvertown, Jonathan. "A new dawn for citizen science." *Trends in Ecology and Evolution* 24, no. 9 (2009): 467-471.

Strauss, M., and J. Carnahan. *Observed Errors in Distance Estimation.* SAE Technical Paper, SAE International, 2010.

Sui, Daniel Z. "The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS." *Computers, Environment and Urban Systems* 32 (2008): 1-5.

Tulloch, David L. "Is VGI participation." *GeoJournal* 72 (2008): 161-171.

Turner, Andrew J. *Introduction to NeoGeography*. O'Reilly Media Inc., 2006.

van Oort, P.A.J. "Spatial data quality: from description to application." January 13, 2006.

Whipple, J. David, Dale Rollins, and Walter H. Schacht. "A Field Simulation for Assessing Accuracy of Spotlight Deer Surveys." *Wildlife Society Bulletin* 22, no. 4 (1994): 667-673.

Whitelaw, Graham, Hague Vaughn, Brian Craig, and David Atkinson. "Estabilishing the Canadian Community Monitoring Network." *Environmental Monitoring and Assessment* 88 (2003): 409-18.

Wildgame Innovations. *Wildgame Innovations.* 2012. http://www.wildgameinnovations.com/ (accessed 2012).

Yahoo! Inc. *Flickr.* 2012. http://www.flickr.com/ (accessed 2012).

Zandbergen, Paul A, and Sean J Barbeau. "Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones." *The Journal of Navigation* 64 (2011): 381-399.

Zandbergen, Paul A. "Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National Standard for Spatial Data Accuracy." *Transactions in GIS* 12, no. 1 (2008): 103-130.