DECISION SUPPORT FOR CRITICAL INFRASTRUCTURE RECOVERY

by

Okan Pala

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2014

Approved by:

____
Dr. David C. Wilson

____
Dr. William Tolone

____
Dr. Heather Lipford

____
Dr. Russell Bent

____
Dr. Ertunga Ozelkan

ABSTRACT

OKAN PALA. Decision support for critical infrastructure recovery.
(Under direction of DR. DAVID C. WILSON)

Protecting critical infrastructure systems, such as electrical power grids, has become a primary concern for many governments and organizations across a variety of stakeholder perspectives. Critical infrastructures involve multi-dimensional, highly complex collections of technologies, processes, and people, and as such, are vulnerable to potentially catastrophic failures and cascading effects with escalating impact across multiple infrastructures. Understanding the impact of service outages in, for example, utility services such as electric power, water, and natural gas, is a key part of decision-making in response and recovery efforts. In this dissertation, I present research that investigates the design and development of more effective decision support tools as part of critical infrastructure analysis. Thus my overall research question is: "How can spatial decision support systems be improved to facilitate more accurate and efficient decision-making processes in critical infrastructure analysis for critical infrastructure recovery?" To address this question, I have conducted three primary research studies. In the first, I develop a recommender framework approach and prototype interactive geovisualization environment for static decision-making tasks in critical infrastructure reconstitution. User study experimental results indicate that the decision makers can make better decisions

with less time and lower cognitive load using a software tool based on this decision recommender framework.

In developing the decision support environment, it became clear that an accurate knowledge base to support impact analysis within component critical infrastructure networks is essential for improving the decision support tools. Often, however, data that encapsulate the source-sink relationships between utility service points and customers are confidential or proprietary, and, therefore, unavailable to external sources due to their sensitive nature. As a result, during emergencies, external decision-makers often rely on estimations of service areas produced by various methods. During the decision support tools user study, critical infrastructure and geographic information analysis experts expressed a need for highly accurate estimates. Several types of estimation methods are prevalent in practice, but little information is available on their comparative effectiveness.

Therefore, in the second research study I tested and compared traditional distance-based and cell-based methods for service area approximation on an electric power network for a mid-size Midwestern US city. Experimental results showed substantial differences in accuracy between the methods, indicating significant tradeoffs in method selection. For example, methods that take capacity and demand into account outperform standard methods, and cell-based methods can be more accurate when demand is closer to sources, indicating that cell-based methods may work best for large areas, such as states, while distance-based methods may work best for locations with uniform demand.

In the third research study, given that I had found substantial accuracy differences among methods in practice, I investigated whether new estimation approaches could be more effective overall. In order to understand the potential, I applied insights gained from the second study in order to develop and test several new estimation approaches. I developed two novel service area estimation methods based on road network optimization techniques. To better understand the relative merits of each method, I also devised a novel adaptation of accuracy assessment methods from land cover classification to service area estimation. Experimental results from the third research study, estimating water service areas for the state of Kentucky, indicate that service area estimation methods based on road network optimization produce more accurate results compared to distance or cell-based estimation methods. In order to cross-check the results, I set up a comparative experiment applying the new methods to the electric power data set for the mid-sized US city from the second study. Experimental results showed that the traditional service area estimation methods outperform transportation-based methods for this specific spatial location and dataset. This is likely due to the absence of zoning or industrial demand area definitions in the reference dataset. This variation in results highlights the importance of the distribution of sources and sinks in critical infrastructures, and the tradeoffs in modeling for decision support. Overall, these studies show that particular care is needed to ensure that approximation methods are chosen to align with the properties of the service network, the population distribution, and the available source and demand data.

This dissertation details the research I have conducted in order to investigate how spatial decision support systems can be improved to facilitate more accurate and efficient decision-making processes in critical infrastructure analysis for critical infrastructure recovery. It provides insight and design guidance on decision support environments for cross-infrastructure analysis, as well as modeling tradeoffs in service area approximation methods for increased recommendation accuracy. Overall, the results of this dissertation research will help support better design and implementation of decision support systems for geovisual decision-making, particularly in critical infrastructure analysis.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. David Wilson for his continuous support of my Ph.D. study and research, for his patience, motivation and enthusiasm. I also would like to thank the rest of my dissertation committee, in no particular order, Dr. Russell Bent, Dr. William Tolone, Dr. Heather Lipford and Dr. Ertunga Ozelkan for their insightful comments, support and encouragement. I also would like to thank the Los Alamos National Laboratory Energy and Infrastructure Analysis group for allowing me to spend close to a year on site collaborating with the staff to perform valuable research.

I am especially grateful for my wife Dr. Tiffany Barnes for her continued support to make this dissertation possible. She has been a great partner giving me enough leeway that allowed me to find my way out. I learn from her every day both academically and personally. She is my North Star and role model. She is also my partner in crime raising a son, now almost 5 year-old Ender Pala, who has been an amazing blessing and fun distraction. I cling to every moment that I spend with him as I know it will not last. My parents Ismail Pala and Munibe Pala has been a great support with their encouragement and prayers always showing their love and trust in me. Not but least, I would like to put a note here for my son to read when he grows up: Nothing is easy, but persistence and hard work can beat almost anything. The moment that you are closest to giving up is the moment that you are closest to the success. Remember this and try learning from past experiences so you don’t repeat them. Trust yourself, your intellect and your charming personality. I will end this with a quote from my undergraduate advisor Erdal Unal: “You can do it. You will do it. If you cannot, nobody can. Good luck.”

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

In recent years, the protection of critical infrastructures (CI), such as electrical power grids or communication networks, has become an increasingly significant concern for many governments and organizations across a variety of stakeholder perspectives. Critical infrastructures have evolved into multi-dimensional, highly complex collections of technologies, processes, and people, and as such, are vulnerable to catastrophic failures (intentional or unintentional) on many levels. Aside from the increased complexity, cross-infrastructure dependencies have been shown to give rise to cascading effects, with escalating impact across multiple infrastructures (Tolone et al., 2009; Tolone et al., 2004). A well-documented example can be seen in the August 2003 blackout in the northeastern U.S. and eastern Canada (Andersson et al., 2005; Hauer et al., 2004). The extent of the disaster was unforeseen. A series of unintentional events led to cascading failures across 263 power plants, with a loss of power for approximately 50 million people, including businesses and homes. Moreover, failure in the electrical power infrastructure had serious impacts on other critical infrastructures. For example, the loss of power also led to a loss of water for approximately four million people across many communities, as water systems depend heavily on power to operate the pumping systems that deliver water for consumption. Further, Amtrak rail services in the northeastern corridor were stopped, passenger screening was not possible at various airports, and flights were canceled due to e-ticket systems being out of service. Even pumping gas into vehicles was not possible in many places, since gas pumps typically rely on power. Moreover, there were also disruptions in communication networks for cellular telephone, cable, and radio broadcast services. The tight couplings within and across these infrastructures and the brittleness that can result were evident in the length of time it took to restore both power and the related infrastructures to normalcy in the affected region. It also became evident that failure isolation is a difficult task within complex infrastructures, let alone across infrastructures.

There have since been many more examples of similar catastrophes, of varying degrees of severity, which have occurred in the U.S. and elsewhere. The prevailing opinion held by most citizens is that the government was, and still is, unprepared to respond to such large-scale disasters (Scavo et al., 2008). In truth, simply understanding the interrelations among different infrastructures, a necessity for initiating adequate response, is a difficult task. As an example, utility services, including electric power (EP), water, and natural gas (NG), serve their customers through various functional sources, such as electric power substations (EP), pumps and pipes (water, NG). Understanding the impact of service outages in such services is a key part of decision-making in response and recovery efforts.

Critical infrastructure protection involves both safeguarding against potential disaster scenarios and effective response in the aftermath of infrastructure failure. There have been many discussions across the globe regarding the best approaches to improving government response to disaster (Fazel Zarandi et al., 2011; Nigim et al., 2006; Ouyang, 2014). Both to guard against and respond to critical infrastructure failures, multi-dimensional infrastructure modeling and simulation has been proposed as a way to support analysis and decision-making (Belton and Stewart, 2002; Johansson and Hassel, 2014; Masucci et al., 2009). An interactive geovisualization interface provides a natural context for this infrastructure analysis support (Mac Aoidh et al., 2008; Pala and Wilson, 2013; Tolone, 2009; D. C. Wilson et al., 2008; D. C. Wilson et al., 2009).

For my research in critical infrastructure protection, I wanted to answer the question of “How can spatial decision support systems be improved to facilitate more accurate and efficient decision-making processes in critical infrastructure analysis for critical infrastructure recovery.” To address this question, I have conducted three primary research studies. In the first, I develop a recommender framework approach and prototype interactive geovisualization environment for static decision-making tasks in critical infrastructure reconstitution (D. C. Wilson et al., 2009). In user study evaluations, critical infrastructure and geographic information analysis experts expressed a need for highly accurate estimates (Pala and Wilson, 2013). Accuracy in the framework and analysis environment depends on good underlying service area models. While several types of service area estimation methods are prevalent in practice, little information is available on their comparative effectiveness Therefore, in the second research study I tested traditional distance-based and cell-based methods for service area approximation on an electric power network (Pala et al., 2014). Given substantial accuracy differences among methods in practice, I also wanted to investigate whether new estimation approaches could be more effective overall. Thus, in my third research study, I applied insights from the second study in order to develop and test several new estimation approaches. Overall, my research outcomes contribute to the development, implementation, and understanding of innovative approaches to modeling and assessing critical infrastructure systems in support of decision-making.

1.1 Organization

The overall research question of this work is: How can spatial decision support systems be improved to facilitate more accurate and efficient decision-making processes in critical infrastructure analysis for critical infrastructure recovery? In order to address this question, the remainder of the dissertation is organized as follows.

Chapter 2 details foundational background on the Critical Infrastructure Protection (CIP) problem, including the challenges of data inaccessibility and resulting needs for approximation, and the US-government-proposed methods for CIP. Difficulties that arise in the decision-making process, and decision support software, in general and specifically created for CIP, are explored. Decision-making for reconstitution after emergencies is challenging because of the cascading nature of disablements. Relationships that allow for cross-infrastructure cascades further complicate decision-making.

Chapter 3 describes related work on decision support in general and decision support specifically for CIP. As a foundation for developing and evaluating the novel service area approximation methods proposed here, an overview is provided on methodologies for understanding interdependence among critical infrastructures. This includes spatial approximation methods (computational geometry, Thiessen polygons, and Cellular automata), location allocation and accuracy assessment approaches.

Chapters 4 and 5 describe the first research study, investigating the following specific research question:

> Specific Research Question RQ1: Is a spatial recommender system focused on critical infrastructure cross-infrastructure effects more efficient and effective than using commonly-used, industry-standard GIS tools for Critical Infrastructure recovery decision-making tasks when multiple networks are interrelating?

To answer this question, I developed (1) a framework for a decision recommendation tool for CI, (2) developed a Decision Recommendation Tool (DRT) called the Critical Infrastructure Explorer (CIE) based on the framework, and (3) performed a user study to compare CIE with standard GIS tools. Chapter 4 presents the framework and describes the CIE implementation, while Chapter 5 details the user study evaluation. In the user study, performance data were collected for system experts and GIS analysts using CIE and standard GIS tools to reenact several disablement analysis scenarios.

Results for user task efficiency, task completion, and cognitive load show that, given the same scenario, expert CIE users successfully completed more scenarios more accurately, in less time, and with lower cognitive load. Overall, the results indicate that adoption of such integrated approaches would be useful to support better decision-making and complex, multi-dimensional analysis for critical infrastructure.

Chapter 6 describes the second research study, investigating service area modeling accuracy in support of approaches like CIE. During an outage, cross infrastructure cascade effects are determined by the extent of the service areas for service distribution sources. This is one of the foundational modeling assumptions in the first study. Grounding cross-infrastructure impact analysis relies on accurate service area estimation for the critical infrastructure networks. Thus, the specific research question for the second study is:

> Specific Research Question RQ2: What are the differences in effectiveness among various service area estimation techniques that are used for CI enablement scenarios?

In order to answer this question, I conducted experiments to assess the relative accuracies of several service area estimation approaches, commonly used in practice, in comparison with a reference data set provided by the electric utility company for a mid-size Midwestern city (~1 million). I employed two different evaluation approaches to perform the accuracy assessment of the results: aggregate statistical accuracy analyses and spatial accuracy analyses. Moreover, as part of the evaluation I adapted some existing accuracy assessment processes (Congalton and Green, 1999, 2008) to the CI domain, and also implemented two new analyses specific to the CI domain – polygon relaxation analysis and proximity confidence analysis.

Results from this study shed a light on the relative accuracies of common service area approximation methods, which can help system builders to choose the most appropriate service area estimation method for the type and context of impact analyses on electric power networks. In particular, our study showed that cellular automata (CA) algorithms perform better near source stations, suggesting the idea that CA methods may work better when utility usage is highest near source stations and tapers off as the distance grows. Thiessen polygon (TP) methods seem to perform better in uniform usage situations.

Chapters 7 – 9 detail the third research study, investigating new methods for service area estimation, which are based on transportation optimization metrics. Given the substantial differences observed in the accuracies of commonly employed methods, the natural next step was to investigate the potential for new methods. Thus the specific research question for the third study is:

> Specific Research Question RQ3: Will applying metrics for transport optimization to service area estimation improve accuracy in comparison to common techniques?

In order to address this question, I adapted and implemented transportation network-based service area estimation techniques for water systems service area approximation. The intuition is that many utilities install their distribution networks along streets and roads, which may provide additional context to improve estimation accuracy. Chapter 7 describes and evaluates two service area estimation methods based on travel distance optimization in order to create service areas for CI elements. To support evaluation, this chapter also details the adaptation of established land cover classification accuracy assessment methods for service area estimation accuracy(Congalton and Green, 1999, 2008). The proposed service area estimation methods are compared to the cell- and distance-based methods using aggregate impact and point impact accuracy. Results show that using source capacity and demand amount for network optimization produces significantly better results in approximating the service areas than using the network optimization without the source and demand values. Based on these results, CI simulation systems should consider transportation networks, source capacities, and demand locations into account to refine service area approximation.

Chapter 8 investigates these new methods applied to the electric power network previously explored in Chapter 6, and Chapter 9 compares the accuracies for distance-, cell-, and transport-based service area estimation across the power and water network studies. Results show that while it is possible to improve accuracy with new approaches, there are tradeoffs to consider in the context of the available data. Thus, Chapter 9 presents an analysis of how the various methods considered differ across network types and study area sizes, discussing the potential trade-offs in method applicability. For example, differences may arise due to the absence of industrial zoning data for electric power demands when their usage is still included in source substation capacities. Whereas, water usage from water treatment plants is not impacted as strongly by this kind of zoning difference, because these plants are likely not providing large volumes of water for industrial use, as this kind of water does not need the same treatment as drinking water.

1.2 Contributions

This dissertation research shows that spatial decision support systems can be improved to facilitate more accurate and efficient decision-making processes in critical infrastructure analysis for critical infrastructure recovery. The research contributions are summarized as follows:

Research Study 1

1. Development of a decision recommendation framework including target model, user model, recommender engine and a simulation engine, as well as the development of a prototype tool implementing the framework approach.

2. User study experimental analysis showing the advantage of the approach for cross-infrastructure analysis scenarios – indicating decision makers may be able to better decisions with less time and less cognitive load using this kind of approach.

Research Study 2

1. Adaptation of area and point accuracy assessment processes from land cover classification (Congalton and Green, 1999, 2008) to apply to the CI domain for city-scale electric power network service area estimation.

2. Implementation of point and area accuracy assessment for power networks.

3. Experimental evaluation showing that weighted point and area service area estimation approaches are more accurate in approximating source-sink relationships, and weighted distance based methods have higher point impact accuracy, in an electric power network context.

4. Development of confidence interval approaches for more detailed analysis and interpretation of accuracy assessment results in the CI domain – polygon relaxation analysis and proximity confidence analysis.

5. Confidence interval analysis of service area estimation accuracy results on power networks, showing that as the distance increases between critical points to sources, the rate of decrease in accuracy is higher in polygons produced by non-weighted methods.

Research Study 3

1. Adaptation of formal significance (Kappa) analyses to apply to the CI domain, in the context of water networks.

2. Implementation of previously-described point impact accuracy assessment for water networks.

3. Experimental evaluation showing that weighted service area estimation approaches are more accurate in approximating source-sink relationships in a water network context.

4. Development and implementation of two new transportation network-based service area estimation methods for a state-scale water network.

5. Experimental evaluation showing that the proposed transportation-network-based methods are significantly more accurate than standard service area estimation methods for the state-scale water network.

7. Implementation of the proposed transportation network-based service area estimation methods in the previously tested context of a city-scale electric power system.

8. Experimental evaluation showing that transportation network based methods on a city scale electric power network does not perform as well as standard distance or cell-based methods.

Overall, the outcomes of this research provide insight for researchers and practitioners in geographic information systems, and particularly for developers and users of applications for critical infrastructure modeling, analysis, and decision-making. Enabling new kinds interaction through tools and environments can support improved analysis and thereby decision-making. And since a key enabling element is accurate knowledge about the infrastructures themselves, understanding the tradeoffs involved in modeling infrastructures and infrastructure interactions can lead to more accurate modeling, enabling improved analysis and decision-making. Results of the accuracy analysis here indicate that decision makers and system designers should weight tradeoffs and select carefully when applying different service area approximation methods to CI source-sink analysis. Methods using source output amounts as weights to approximate the service areas provide better results. Networks at larger scale, such as state or regional, do well using transportation-based optimization methods and if demand information and road network is available it is possible to improve the accuracy using those as a part of the method. Dense networks at smaller scale, such as city scale, are not as susceptible to transportation network optimization and using weighted version of distance based methods provide best accuracy.

CHAPTER 2: BACKGROUND – THE CRITICAL INFRASTRUCTURE PROTECTION PROBLEM

Findings that I present in this dissertation can be useful for system design or analyses for different aspects or stages of preparation, response and recovery. To provide a detailed picture of current state of research in this field, in this chapter I focus on characterizing the Critical Infrastructure Protection (CIP) problem and provide examples of previous research done to cope with various aspects of this problem at various phases of the disaster timeline.

In general, critical infrastructures can be thought of as the fundamental enabling systems or networks upon which the smooth functioning of society is particularly dependent, with the implication that the failure of such a network has lasting effects on the public good (J. P. Cohen, 2010; Reinermann and Joachim, 2003). Infrastructure systems widely regarded as "critical" include fundamental enabling assets, such as transportation and utilities (e.g., energy, water, telecommunications, wireless sensor networks), along with assets that depend on them, such as medical care and government institutions or services (e.g., law enforcement, schools, and libraries) (Franco et al., 2012; Mbowe and Oreku, 2014; Mendonça et al., 2014; Oreku and Mbowe, 2014). I adopt as a working definition the U.S. government's characterization of critical infrastructures (USA Patriot Act, 2001; Franco et al., 2012) as "Systems and assets, whether physical or virtual, so vital that the incapacity or destruction of such may have a debilitating impact on the security, economy, public health or safety, environment, or any combination of these matters, across any Federal, State, regional, territorial, or local jurisdiction."

Many of these critical infrastructure systems, such as the power grid, are effectively physical processes that are operated by electronic devices. Over the last decade, these electronic devices have begun to be operated remotely, often through wireless networks (Buttyán et al., 2010; Cetinkaya et al., 2010; Sousa et al., 2009). This has increased system complexity and created cross-infrastructure dependencies between many base infrastructures and communications infrastructure. Failures of distinct components within the power system, for example, can cause a general failure of the delivery of power to specific load points (Volkanovski et al., 2009). In some cases, a full blackout may occur.

In general, critical infrastructures have evolved into multi-dimensional, highly complex collections of technologies, processes, and people, and as such, are vulnerable to catastrophic failures (intentional or unintentional) on many levels. Aside from the increased complexity, cross-infrastructure dependencies have been shown to give rise to cascading effects, with escalating impact across multiple infrastructures. Thus, the protection of critical infrastructures, such as electrical power grids or communication networks, has become an increasingly significant concern for many governments and organizations across a variety of stakeholder perspectives.

As a primary example, I consider the problem of critical infrastructure protection in the context of disaster response and recovery. The field of disaster response and recovery is one of several disciplines that have focused attention on the challenge of properly understanding critical infrastructure behavior and interdependence. From the perspective of the disaster response task domain, an extreme event colliding with a vulnerable situation creates a disaster (Kreimer, 1991; Mbowe and Oreku, 2014). The World Bank characterizes a disaster as an extraordinary event with limited duration that seriously dislocates a country's economy (World Bank 1995).

For example, a chemical, biological or radiological (CBR) incident is considered to be a wide-area disaster that is anticipated to disrupt a large number and variety of critical infrastructure assets. In addition, natural disasters, such as hurricanes, tornadoes, floods, tsunamis, and drought can easily be just as disruptive. When infrastructure breaks down, results can be catastrophic and may have long-term regional consequences (Coffrin et al., 2011; Franco et al., 2012).

In the most industrialized countries, some of the major effects of disasters have included blackouts, with loss of water and power, disruption in transportation, closing of public facilities, and many other impacts (Coffrin et al., 2011). Businesses, too, would be negatively impacted to the extent that the overall economy is affected. Even in less-developed nations, these types of incidents have frequently occurred, leaving large segments of the population without supplies and communication (Memmott and Hanks, 1992).

When critical infrastructures are disrupted, the ability for a community to function is reduced and may dissolve altogether (Franco et al., 2012). In particular, from the U.S. national perspective, critical infrastructure protection (CIP) is part of a national program to ensure that critical infrastructure failures should have a minimal impact on

- The national government, performing essential national security missions and ensuring general public health and safety;

• Regional and local governments, maintaining order and to delivering minimum essential public services; and

• The private sector ensuring orderly functioning of the economy and delivery of necessary telecommunications, energy, financial and transportation services (Clinton, 1998).

Any disruptions of critical infrastructures supporting these functions must be “brief, infrequent, manageable, geographically isolated and minimally detrimental” (Clinton, 1998). While this characterization is derived from a national government point of view, the general notion applies equally to almost any organization with critical infrastructure dependencies, though the profile of focus infrastructures may well vary.

The definition and securitization of these infrastructures creates various hierarchies, along with exclusions, among interrelated structures (Aradau, 2010). The decisions made regarding which community assets and systems are critical and how they are prioritized often depend on which technical systems are most important for the smooth running of daily life (Mbowe and Oreku, 2014). As discussed later, this kind of prioritization activity, including infrastructure modeling and stakeholder perspectives on impact, is central to CIP decision-making and an essential consideration in developing CIP decision support tools.

To determine which systems are critical, the impacts and consequences of a disaster are assessed, along with the interdependencies among assets and the ability to find workarounds should a specific system fail (Franco et al., 2012). The prioritization may vary depending on the specific community. Determining the impacts and consequences, however, is based on the distribution of source-consumer relationships, while interdependencies among assets are determined based on the source-sink relationships among the infrastructure elements.

Thus, effectively addressing the complex problem of critical infrastructure protection, requires (1) understanding or modeling critical infrastructures and dependencies between infrastructures (e.g., source-sink analysis, geographic scope), (2) understanding the impact of service outages within and across infrastructures (e.g., community services, economic), (3) in order to make decisions about priorities from different stakeholder perspectives. The staggering complexity of decision-making for critical infrastructure protection at any significant scale, gives rise to the need for decision-support tools.

2.1 The Problem of Decision Support in Critical Infrastructure Protection

Decision makers face many challenges in prioritizing resources for protection and response during and after the occurrence of a disaster. In this section I review some of the main challenges and context considerations that impact CI decision makers, which serves as design context for developing CIP decision support tools.

2.1.1. Difficulties in Decision-Making for Disaster Recovery

The decision-making environment both during and after an infrastructure disaster can be extremely complex and dynamic. One method used for decision-making in various types of complex situations is referred to as the analytic hierarchy process (Saaty, 1994). Treated as a hierarchy, the problem can be decomposed into sub-problems, and the solutions of the sub-problems can then be aggregated into a decision. In practice, the hierarchy referred to by Saaty (1994) can be difficult to define during infrastructure disasters. The nature of a disaster and its aftermath is constantly shifting. Decision makers may need to re-evaluate plans rapidly in response to those changing conditions. For example, the environment during a terrorist attack can be considered as opposing sides wanting to control the states of the critical infrastructure systems to achieve their goals (Haimes, 2006; Saaty, 1994).

Another major issue with decision-making during disasters is inadequate planning, despite any efforts made (Guikema, 2009). For example, in the pre-planning stage, homeland security and emergency management planners discover and isolate service areas within a power grid that are likely to be impacted by a disaster. However, this process is often hindered by limited information about the utilities. This is due, in part, to the fact that approximately 85 percent of critical infrastructures in the United States, including power grids, are held by the private sector. These organizations are held responsible for the security of information. Unauthorized access or destruction of an organization's information assets arising from malicious acts, errors, or disasters could result in compromised information, along with numerous other consequences (Holgate et al., 2012).

On the other hand, the growth of information technology has created scenarios in which there is actually an overabundance of data. This scenario creates the necessity for a "needle in the haystack" approach to finding answers in the data. Data quality also factors into the lack of usability of the data risk assessments (Guikema, 2009). As a result, the pre-planning is difficult, may be based on inaccurate assumptions, and may ultimately not be sufficient for quick enactment and remediation during and after a disaster takes place (Havlin et al., 2012).

2.1.2 Cross/Multi-Infrastructure

Magnifying the challenges for analysts and decision-makers are the numerous inherent interdependencies that exist among critical infrastructures (Havlin et al., 2012; Usov et al., 2010). Modern infrastructures consist of complex cyclic interdependencies. As a result, any restoration process must take a holistic approach to be successful (Coffrin et al., 2012). For example, electric power systems depend upon transportation networks to deliver fuel to generation facilities. These same generation facilities often depend upon water systems for cooling purposes. In addition, electric power systems depend heavily upon telecommunication networks to support the Supervisory, Control and Data Acquisition (SCADA) systems that manage power transmission and distribution (Ebrahimy, 2014; Mahmood et al., 2010).

The list of interdependencies among the critical infrastructure sectors is long and in many cases, interdependencies are poorly understood. Furthermore, infrastructure interdependencies are often very strong, time-sensitive, and essential to overarching system operation. The result is a brittle "system of systems" that contributes to the potential for catastrophic occurrences as a failure cascades and escalates across related infrastructures (Havlin et al., 2012). The problem of understanding the behavior of critical infrastructures and their interdependence as part of reconstitution efforts remains difficult and open. The limitations of single-dimensional approaches are by no means trivial. Multi-dimensional approaches, while theoretically promising, have produced few results. Analysts and decision-makers face extremely complex issues in understanding and responding to multi-dimensional CIP problems. They must account for a variety of contextual elements, including task goals (e.g., remediation vs. response), geographic scope, infrastructure and cross-infrastructure knowledge, resource allocation, and outcome measurement (Usov et al., 2010). Particularly given constraints on response time and dynamic contextual alterations, the complexity can rapidly lead to information overload, which can significantly impact the efficiency and quality of response. In such a shifting decision-making environment, it can be just as important to reduce the decision-making attempt as it is to increase decision quality (Todd and Benbasat, 1992). Specifically, simplifying the efforts to prioritize which assets to dedicate resources is vital to recovery following a disaster. This balance is also an important design consideration in CI decision support environments.

2.2 Requirements for Critical Infrastructure Analyses

To meet these CIP goals, frameworks for critical infrastructure assurance have been developed, such as the six phases of the Critical Infrastructure Protection Life Cycle (England, 2005), as shown in FIGURE 1. Each of these phases gives rise to distinctive priorities and goals as a context for decision-making.

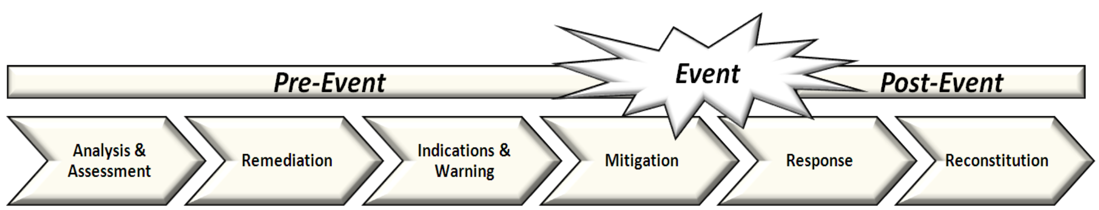


FIGURE 1: Phases of the critical infrastructure protection life cycle

Prioritization must occur at each stage of this process. Factors affecting prioritization of components of critical infrastructure include (Franco et al., 2012):

- Disaster consequences and impacts
- Recovery objectives
- Status and relationships of services to recovery objectives

- Status and contribution of services to Infrastructure assets
- Interdependencies
- Workaround potential
- Milestone requirements

2.2.1 Pre-Event Preparation

In the pre-event phase, decision-makers and planners analyze infrastructures to determine potential vulnerabilities and focus points for remediation and monitoring. Specifically, in the pre-event phase analysis, critical infrastructure systems must be considered in their entirety to identify and prioritize potential weaknesses, with the goal of minimizing potential impacts from large-scale disasters on services provided by critical infrastructure systems (Mendonça et al., 2014).

There are several important actions that should be considered by recovery planners and enacted by managers of emergency personnel and processes, which would support a quicker economic recovery, for example, after a wide-area CBR (Franco et al., 2012). These are summarized below

- Baseline knowledge of critical infrastructure characteristics and assets must be developed, including such information as service types, dependencies and interdependencies, and workaround potential.
- Prioritization approaches need to be developed and evolved for execution within a multi-disciplinary group of stakeholders.
- Strategies for disaster economic recovery must be developed, along with enactment plans, and aligned between the government (federal and local) and the private sector.

- Adequate capital must be ensured for recovery (include agreements, funding mechanisms, and insurance policies).
- Economic resistance and resiliency must be built into the regional economy.

Pre-disaster preparation may also involve training exercises that have been developed with anticipated hazards in mind (Franco et al., 2012; Mendonça et al., 2014). Ultimately, the goal of these exercises is to develop tools, response and training techniques that aid in achieving organizational resilience, or the capacity to retain control, continue operating, and rebuild where necessary (Mendonça et al., 2014).

During the pre-disaster phase, criticalities of network elements are studied as a part of vulnerability analysis. At this point, it is important to define criticality and explain how vulnerability analyses play an important role to help decision makers as a part of pre-event analyses.

Criticality

One quality of infrastructure system components considered crucial to discovery and understanding is criticality – the highest levels of importance. Each critical infrastructure system has a distinct mission, generally defined by a decision-maker, and can be altered as needed (Quirk and Fernandez, 2005). Energy infrastructure, for example, possesses the critical mission to reliably supply power to businesses or residents. As society relies more on the internet, to the degree that it is now considered indispensable, criticality of that infrastructure has been realized (Yan et al., 2010).

Pre-planning should incorporate potential effects of damage to a particular network component that would result in the loss of critical functionality. Topologic characteristics are often used to convey criticality in a network system. The assessment of criticality has taken many forms, using various tools such as graph-theoretical analysis, route-based analysis, traffic-based analysis, and consequence-based analysis (Yan et al., 2010).

Vulnerability

Definitions of vulnerability and methodologies to capture them vary in their context and specificity (Collins et al., 2011). Criticality and vulnerability are intrinsically distinct from each other. Criticality may, in fact, be a constituent of vulnerability (Quirk and Fernandez, 2005). Determination of the degree of vulnerability for a component of a critical infrastructure can be affected by numerous factors such as population or a single point of failure within or across networks. Particularly vulnerable components should be prioritized in critical infrastructure analysis. Any CI analysis should consider the specific population served by a component, which would be particularly negatively affected in the event of a disaster (Matisziw and Murray, 2009; Schintler, Gorman, et al., 2007; Schintler, Kulkarni, et al., 2007). For example, when the size of the population served by a particular component is large, that population is particularly vulnerable (Bush, 2005).

The type of population is also a factor, for example, an inpatient hospital houses a population of more susceptible people, and the failure of critical infrastructure could substantially increase mortality in that particular population. It is also important to note that the population served by a specific component of the critical infrastructure may vary throughout the day. Subsequently, the vulnerability of that component may also vary. If the service area served by a component consists of businesses and the critical functionality of those businesses is at risk, then that will influence the determination of vulnerability(Schintler, Gorman, et al., 2007).

The U.S. Department of Homeland Security (DHS) recently developed a methodology for evaluating protection defenses and vulnerability of key resources and critical infrastructure. This methodology for detecting and measuring vulnerability is part of a greater DHS effort (referred to as the Enhanced Critical Infrastructure Protection Program) to mitigate any vulnerabilities, enhance relationships between separate entities, and improve the flow of information among public and private entities (Collins et al., 2011). The complexity and recent nature of this methodology highlight the need for research into effective ways to manage critical infrastructures.

According to Koger and Landry (2010), several factors directly impact the vulnerability of critical infrastructure. These factors are grouped as being technological, societal, natural, system-related, or institutional. Societal factors include the appeal of a particular service area for attack, demographics, and the ability to quickly communicate with the public in that area. System-related factors center around a network's sophistication and interdependence. Natural factors involve the availability of vital resources and the incidence of natural hazards. Technological factors consist of likelihood of failure. For example, inconsistencies in a network design exposes that network to potentially severe outages, thereby increasing overall system vulnerability (Zio and Golea, 2012). Institutional aspects comprise the existence of historic structures, specifically related legislation, and market organization (Cetinkaya et al., 2010; Koger and Landry, 2010).

Vulnerability analyses are also used to determine the resilience of CI networks. The concept of resilience has been introduced as the ability to recover from a disaster and to reduce impacts from future disasters. Territorial resilience can be determined by the supporting elements for organizations when they confront crisis situations (Garbolino et al., 2013). Resilience is the ability of a network to provide a desired service even when challenged by large-scale disasters or other types of failures (Sterbenz et al., 2013).

2.2.2. Event and Post-Event Phase

During an event, first responders focus on moderating the overall effects of the disaster. There is a focus on containment of the effects, which could lead to disabling currently functional infrastructure assets to reduce the probability of a cascading failure. Such response capabilities are characterized as a “system for responding to a significant infrastructure attack while it is underway, with the goal of isolating and minimizing damage” (Clinton, 1998).

In the post-event phase, recovery planners and decision-makers assess infrastructures to determine the most effective allocation of resources to rebuild and restore damaged infrastructures. A process for prioritization of critical infrastructure is necessary (see FIGURE 2). Analysis is more precisely focused on known areas of impact, which may be radically different from areas of focus identified in pre-event analysis. Reconstitution capability can be characterized as a “system to reconstitute minimum required capabilities rapidly” (Clinton, 1998; Franco et al., 2012).

FIGURE 2: Process for prioritization of critical infrastructure (Franco et al., 2012)

2.2.3. Recovery Phase

Managers of civil infrastructure systems face the problem of restoring essential public services after failures of the system (Cavdaroglu et al., 2013). Recovery after a disaster, natural or man-made, consists of a variety of components. The planning developed during the pre-disaster phase is expected to be enacted immediately following the disaster event. Where possible, data are collected and compared to pre-event baseline data for regional critical infrastructure assets. Pre-determined prioritization is reassessed and utilized. Private and public sectors coordinate efforts and share information. Interim provisions are made to protect and preserve life (Franco et al., 2012). If effective, the preplanning should speed the response, smooth the transitional periods, and mitigate damage to people and the economy.

There are numerous considerations related to supporting a region's critical infrastructure recovery. Due to the limitations of available resources, recovery team members who are responsible for planning recovery efforts must employ a transparent, analysis-based process for prioritizing infrastructure restoration to determine the components and tasks needed for temporary installment or repair, along with the optimal assignment of the tasks to work groups and schedules for the completion of the tasks. This involves extensive assessments and data collection and modeling to ascertain the current state. This information will assist the decision-making and prioritization, and relieve the general uncertainty regarding the need and availability of critical services and functions (i.e., drinking water, food, mobility, medical care) that can immediately follow a disaster (Cavdaroglu et al., 2013).

Transportation of people, goods, and services is critical. These resources sustain the operators of infrastructure assets, and, if the resources are unavailable, operators may not be able to perform their duties (Cavdaroglu et al., 2013). Transportation is considered a focal point. Impacted industries considered to be “cornerstone,” and the critical infrastructures they rely on, need to be identified and prioritized accordingly for restoration. Agricultural infrastructure, as a source of food and as a fundamental component of economy, should be prioritized (Franco et al., 2012).

CHAPTER 3: PREVIOUS WORK ON DECISION SUPPORT APPROACHES

In the previous chapter I defined the CIP problem, examined different aspects of the problem and provided examples of difficulties facing those aspects at various stages of the disaster timeline. The complexity of the CIP problem motivates the need for decision support tools and environments to help CI analysts and decision makers. In this section I will provide background information on decision support approaches and their applications on various aspects of critical infrastructure analyses. In addition, since I analyze CI service area estimation methods enabling knowledge components for decision support, I also include an overview of foundational concepts for service area estimation.

3.1 Spatial Decision Support

A Decision Support System (DSS) can be defined as "a computer based system that aids the decision-making" (Finley and Sanders, 1994). Because most computerized systems support many critical and non-critical decisions the area is very open for interpretations from different fields. Keen (1980) suggested that providing a precise definition of DSS that includes all of its facets is impossible, demonstrating the complexity of the interplay of factors important to such systems.

The methods used in DSSs involve designing intelligent components capable of delivering informed recommendations, which can greatly assist decision-makers as they make choices (D. R. Wilson and Martinez, 2000).

3.1.1 Spatial Decision Support Systems (SDSS)

Spatial Decision Support Systems are defined as flexible systems that enable analysis of geographical information in a decision-making environment (Densham, 1991). It is possible to think of SDSS as a flexible wrapper around set of geographic information analyses processes and tools. According to Densham, a SDSS framework is used for integrating database management systems with analytical models, graphical and tabular reporting capabilities, and expertise of decision-makers (Densham, 1991).

SDSS evolved from traditional DSS designed for business-oriented applications, and then followed a parallel development pattern. DSS and decision-support science essentially form the foundation of SDSS. DSS has traditionally been defined as having six core components (Densham, 1991; Geoffrion, 1974):

1. They are specifically designed to solve ill structured, ill-defined problems.
2. They have an easy to use, powerful user interface.
3. They can combine complex analytical data.
4. They utilize built in models to generate alternatives for the user to select from.
5. They support different decision-making styles.
6. They support interactive, recursive problem solving.

Densham (1991) further extends this definition to describe SDSS by including a few additional capabilities:

1. They allow for input of spatial data.
2. They allow for the representation of complex spatial relationships.
3. They can be used to perform spatial analysis.
4. They provide output in some sort of spatial format (i.e., maps)

Examples of the use of SDSS in critical infrastructure assessments include the application of a web-based SDSS for flood risk management (Jumadi, 2013), management of storm water and assessment of water quality (Kaunda-Bukenya et al., 2012), and determination of the optimal sites to use for physical development (Baloye et al., 2010).

3.1.2 Multi Criteria Decision Analysis

Multi Criteria Decision Analysis (MCDA) was first designed to provide solutions for problems with coexisting conflicting interests in Operations Research field. In such cases there exists a need for identification of priorities according to multiple criteria to support analysts and decision-makers (E. Gomes and Lins, 2008; E. G. Gomes and Lins, 2002). It also is possible to describe MCDA as "an umbrella term to describe a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter" (Belton and Stewart, 2002).

MCDA allows for weighting, or ranking, of specific criteria important in decision-making. For example, MCDA has been utilized to aid in the response to an electrical system failure for the determination of the optimal approach for restoration (Wang, 2001), in generation of a suitability map for biodiversity conservation within a region in Italy (Bottero et al., 2013) and in ranking the potential failure of equipment within certain energy substations (Moreira et al., 2009). Achieving MCDA is a long-term goal that can be facilitated by findings of this dissertation. In the decision recommender system discussed in in chapters 4 and 5, I propose and implement simple decision analysis with ordinary weighting. Expanding this to perform an experiment testing different MCDA techniques requires detailed infrastructure data for a large area and multiple infrastructures, so that simulations can be meaningful enough to make comparisons to human judgment.

3.2 Decision Support in Critical Infrastructure Recovery

According to Kreimer (1991), the actions that immediately follow a natural disaster generally include activities designed to

- Provide emergency assistance
- Reduce the possibility of secondary damages, and
- Speed up recovery operations.

The key to meeting these criteria is gaining greater knowledge and understanding of the affected entities and interrelations among them. Post-disaster restoration of infrastructure typically requires consistently close collaboration across many sectors. Unfortunately, the technological capacity to support this type of collaborative training lags behind current training needs. Additionally, the rarity of such large-scale disruptive events reduces the opportunity for observation, data collection, and comparisons of impacts, responses, and recovery efforts (Mendonça et al., 2014).

There are several reasons for the increasing research emphasis on more effective disaster response support at a national level, specifically in the United States (Tolone et al., 2004). Events such as the 2003 northeast blackout and Hurricane Katrina have created broader awareness of the complex problems involved in disaster response and insufficient planning and resources. Historically, critical infrastructures were not as integrated and complex, and man-made chemicals, which may potentially be used as weapons, were not available in great abundance (Santella et al., 2009). With advances in technology and automation, as well as increased coupling of infrastructure management with communication systems for managing CI, more recent events, including natural disasters or terrorist attacks, have had more far-reaching consequences than ever before (Usov et al., 2010; Volkanovski et al., 2009). Natural disasters, in particular, are anticipated to continue to increase in severity due to climate change (Andersson et al., 2005). In addition, technological advancement has also led to increased knowledge availability. For example, the National Hurricane Center has developed new models that can predict the direction, timing, and magnitude of tropical storms and hurricanes (Coffrin et al., 2011). Unfortunately, this knowledge is not easily incorporated into disaster response planning (Coffrin et al., 2011).

Aside from disaster recovery, the problem of understanding critical infrastructure behavior and interdependence is a significant issue across many well-established disciplines, such as urban and regional planning, civil and environmental engineering, operations research, landscape architecture, and emergency management (Tolone et al., 2004). In pursuit of knowledge and understanding of this behavior and interdependence, some researchers emphasize brokered integration of data and information within software systems, while others focus on how to prevent such occurrences by vulnerability analysis (Haimes, 2006), whereas other researchers focus on responding to events in the most effective way (Coffrin et al., 2011).

The field of Geographic Information Science and Technology (GIS&T) is receiving increasing attention for this purpose (D. C. Wilson et al., 2008). GIS&T researchers have approached the study of critical infrastructure behavior and spatial interdependence from three main vantage points (Sinton, 1992). The first examines critical infrastructure interdependence with tools from spatial statistics and econometrics, an approach typically identified as spatial data analysis (Cressie, 1991; Haining, 1990). The second depicts geographic correlations among critical infrastructure elements via traditional map overlay methods for spatial data aggregation within a geographic information systems (GIS) environment (Burrough, 1990; Goodchild, 1990; Greene, 2002). The third employs rule-based inference engines, usually grounded by human expert knowledge, in the delineation and manipulation of interdependence (Gronlund et al., 1994). Static models, similar to descriptive, reflect an unchanging status, whereas dynamic models give the user a view of changing processes over time (Maguire and Longley, 2005).

The goal of GIS-based modeling is to reflect, simplified but accurate view of the real world under varying specific conditions, in order to gain an understanding of, and predict the impacts of, certain events and conditions. To meet real world needs, software with geospatial analysis capabilities must be developed with the end user in mind. It must be easy to learn, function efficiently, and provide straightforward output. Researchers have made substantial efforts to improve geospatial software by analyzing the end user's behavior, in addition to better visualization tools and enhanced exploratory data analysis capabilities (Mac Aoidh, 2008; Mac Aoidh et al., 2008; D. C. Wilson et al., 2008).

GIS-based modeling can be used in the application of CIP, as an infrastructure modeling tool (Tolone et al., 2004). The Urban Flood Model for ArcGIS is an example of an external model and it integrates the urban flood model output directly into ArcGIS (Kang et al., 2010). Embedded models, such as those created using a GIS model builder, are directly implemented in GIS. These models require more time for the actual integration, but are much quicker to use (Goodchild, 1993; Goodchild et al., 1993;

Maguire and Longley, 2005; Westervelt, 2002). The Interdependent Energy Infrastructure Simulation System (IEISS), which was developed by the Los Alamos National Laboratory for the National Infrastructure Simulation and Analysis Center (NISAC), simulates the interactions both within and among critical infrastructure systems. Specifically, the IEISS seeks out potential cascading failures that impact more than one critical infrastructure network (Bush, 2005). Within any urban system, a common GIS platform gathering all networks, city infrastructures, populations and issues could be of high value in planning for a more resilient city (Toubin et al., 2012).

Tolone and colleagues (2004) have identified three important characteristics that are necessary to modeling architecture. These include scalability, flexibility, and extensibility. The flexibility component would allow for varying models to be constructed for a given infrastructure and region. Scalability allows for multiple models of similar infrastructure types to be simultaneously incorporated into a simulation. Extensibility is simply the expansion of the simulation by adding new infrastructure model types (Tolone et al., 2004). Advancements in user-friendliness are crucial to CIP planners, who may not have expert knowledge of geospatial science and technology, yet still have to work across disciplines incorporating GIS-based models to conduct effective planning (D. C. Wilson et al., 2008).

3.3 Previous Research on Critical Infrastructure Analyses

Most research on critical infrastructures has been conducted in industrialized countries, and has focused on transportation and energy components of the overall infrastructure. Less-developed countries have not been as extensively investigated, primarily due to scarcity of data in those regions (Usov et al., 2010).

CI research has been limited due to such things as differences in the data that are available across fields of study (Usov et al., 2010). Climate and climate-impact data tend to be collected using large-scale resolution, whereas critical infrastructure data are typically at a much finer resolution. More recently, climate data are being scaled to more closely match power grid data. This area of study is still emerging, and new problems have arisen with regard to how the climate and power grid data can be integrated and analyzed (Bhaduri and Kloos, 2013).

A recent example of emerging studies includes research on the relationship between climate change, or natural disasters, and energy infrastructure, such as energy grids (the network of transmission lines, substations, and power plants) (Santella et al., 2009). Within power systems, Volkanovski et al., (2009) developed a method that incorporates both the fault tree analysis and the analysis of power flows within the system to conduct reliability assessments. Their method allows for identification of the main contributors to reliability within the power system (Volkanovski et al., 2009). The Critical Infrastructure Protection Decision Support System (CIPDSS) modeling tool, developed for the Department of Homeland Security, was tested by a group of researchers trying to determine the potential capabilities available through modeling. Specifically, they modeled road and telecommunication disruptions, and studied cascading effects on related infrastructure sectors (Santella et al., 2009). More recent efforts using different models have aimed to incorporate disaster-specific information into optimization techniques, whereby total watts of power outages are minimized (Coffrin et al., 2011).

Yet, as sophisticated as mathematical modeling techniques are, those applied to critical infrastructure data have not greatly enhanced the understanding vulnerabilities of critical infrastructure networks. There is still a great need to build support for analysis and decision-making in these networks. Therefore, this dissertation explores an integrated tool for decision support and compares service area estimation techniques to a city-wide power network and a statewide water networks, to better understand the components needed for successful analysis and decision-making for critical infrastructures.

A number of researchers, such as Getis (1994) and Sinton (1992), have advocated a multi-dimensional approach to the study of behavior and spatial interdependence of critical infrastructures. Instead of divide and conquer, they suggest a strategy that combines the strengths of a multi-dimensional approach and investigates interdependence from all three viewpoints. However, though there have been some genuine efforts in this direction (Flowerdew and Green, 1994; Getis, 1994), progress along this route has yet to meet the advocates' expectations.

In 1995, Jankowski identified two main perspectives on the use of GIS in a decision support system (Jankowski, 1995). The first perspective, the SDSS perspective, incorporates geographical information in the generation, evaluation, testing, and production of recommendations to spatial decision-making problems. In this perspective, GIS is identified as the core of the SDSS. The second perspective, the integration perspective, involves incorporating existing models into the framework of a GIS.

In MCDA, GIS is widely used and has been studied in fields such as land suitability or planning scenarios. Methods used in MCDA include outranking methods and weighted summation. However, due to the ease of employing map algebra, weighted summation is used most often, as reflected in publications by Malczewski (2006). Very little attention has been given to attribute-based GIS models in the assessment of vulnerability and criticality in energy networks. Most of the research examining vulnerability and criticality within energy networks focuses primarily on the use of graph theory. The flow of energy throughout the system is not incorporated into the model (Arianos et al., 2009).

DHS has supplied a listing and description of the software tools currently used by the federal government for the purpose of decision support (Franco et al., 2012). Here, I provide an overview of these tools, showing that each one serves a particular purpose but there is no single tool that integrates the information that could support the diverse purposes of analysis, simulation, or prediction of critical infrastructures that could be critical in decision-making.

The PATH/AWARE tool assists emergency planners with infrastructure prioritization by using a quantitative methodology. Planners can use this software to scrutinize interactive GIS maps of a given restoration area, based on real-time data. Weights can be input into the model, which should be based on specific recovery objectives. By using a quantitative algorithm, the software yields a prioritized listing of infrastructure components in need of restoration. A user can make changes to the priority list as needed. The difference between this tool and the tool that I developed for this dissertation is in the cross-infrastructure cascading effects. My approach cascades disablements in networks within and across networks based on defined dependencies to recommend prioritization.

Another software suite, **FASTMaps,** provides GIS-based results and statistical data on essential economic sectors, in addition to the physical position of critical infrastructure and assets considered to be at risk. It can either provide answers to questions directly or serve as pre-modeling data for more detailed analyses.

Economic impact can be estimated using **REAcct**, which yields county-level estimates of economic impact as indicated by changes in Gross Domestic Product and status of employment. This model can be applied any region within the United States, via the integration of geo-spatial tools and economic data.

Another software tool, **N-ABLE**, simulates environments and larger-scale supply chain models for estimation of economic impacts over time. Long-run structural alterations to the economy can be modeled using **REMI.** This software is publicly available and is used to produce annual models retrospectively. This is an economic impact modeling tool specifically used for supply-chain modeling whereas the tool that I explain in Chapter 4 of this dissertation focuses on recovery after the disaster, and the priorities can be modeled on any information layer such as population effects, effects on other stakeholders and critical point impacts instead of just economic data.

Each of these software tools addresses several factors that could be important for critical infrastructure analysis. However, none of these tools is tailored specifically to the task of decision support, and each has parts missing that would be needed in emergency management.

3.4 Research to Develop Methodology for Understanding Interdependence of Critical Infrastructures

In general, studies in the reliability engineering area focusing on interdependence have primarily emphasized topological properties like betweenness and disruption of connectivity (Dueñas-Osorio and Vemuru, 2009; Ouyang and Dueñas-Osorio, 2011). In the sphere of artificial intelligence, more studies focusing on the restoration of multiple interdependent infrastructures are needed. Power system restoration scenarios have been considered in studies promoting good methodology in the application of planning, configuration, and diagnostic techniques. However, many researchers use connectivity as a foundation of their models. Inasmuch, their reliability is insufficient in situations in which complex interdependencies exist (Coffrin et al., 2012).

Apostolakis and Lemon (2005) conducted a study in which they examined three critical infrastructures on the MIT Campus: electric, water and natural gas, as well as the interactions between them (Apostolakis and Lemon, 2005). They focused on developing a process for identifying critical locations in infrastructures, given a specific threat as a component of their vulnerability analysis. Following on that work, Michaud and Apostolakis (2006) developed an approach using geographical information that takes water supply network capacities and repair into account to create a vulnerability screening methodology in the context of potential terrorism scenarios (Michaud and Apostolakis, 2006). Patterson and Apostolakis (Patterson and Apostolakis, 2007) pooled the prior two methods and tailored them to produce a more effective system for mapping the geographic-valued worth, which was exhibited in a color scheme representative of the numerical ranking of distinct geographical areas (Koonce et al., 2008; Michaud and Apostolakis, 2006; Patterson and Apostolakis, 2007).

Tolone and Chu (2010) have created a multi-infrastructure modeling system that can simulate interdependencies and vulnerabilities among numerous entities. Through simulation, the impacts of infrastructure failure components on other infrastructures can be predicted. However, this system requires substantial collaboration and data sharing.

Coffrin, et al (2012) devised a “last-mile restoration approach” to be applied to multiple complex interdependent infrastructures. Their approach uses mixed-integer programs (Argany et al., 2011) to model interdependent networks (specifically, power and gas), through the combination of a linearized Direct Current model for the first network (power) and a flow model for the other (Ehlen et al., 2010).

Usov (2010) also recommends simulation for critical infrastructure dependency analysis to test methods of risk reduction, and evaluation of historical failures. Moreover, coupling the simulations with external threat models, such as a river flood model, can help with decision-making processes in more complex situations. More research is necessary, specifically on the reusability of coupling solutions and on ontologies, development of tools for scenario management, models, and critical infrastructure data, and development of other potential uses of the simulation coupling methods (Usov et al., 2010).

Each of these studies has contributed to the general understanding of interdependencies among infrastructures, but progress hinges upon the resolution of certain issues. Sharing of data and information among distinct infrastructures and access control is a prominent issue (Tolone et al., 2005). Expansion of system environments to acquire and support novel, discovery-based analyses is necessary for system evolution, hence increasing their value. Incorporating probabilistic representations of dependencies and failures among and within infrastructures, including the use of more complex analyses when dealing with fuzzy effects of probabilistic events is also important (Mendonça et al., 2014; Tolone et al., 2004) . A better understanding of cascading and/or escalating failures within and among critical infrastructures can be attained through the study of alternate scaling of common cause failures in which multiple infrastructures become disabled due to a single cause.

The general interface between simulation and decision-making models needs to be further studied, requiring a wider spectrum of expertise and skills. Most essentially, validating data, information, and knowledge across infrastructures is fundamental to enhancing protection of critical infrastructures (Tolone et al., 2004). For example, the accuracy of certain topological metrics used for modeling infrastructure systems has recently come into question (Hines et al., 2010). This is a primary issue that I address as one of my main research questions here. Chapters 6 through 9 focuses on the accuracy of modeling critical infrastructure service areas.

Mendonça, et al. (2014) published a description of a group of prototype tools that were designed for analysis in post-disaster environments and supporting the restoration of infrastructure systems through training exercises. The system consists of large-scale displays, novel interaction abilities, and realistic data, all run through discrete event simulations. The parameters (i.e. time available for task execution, the complexity of the networks) of each simulation could be modified as needed. This work shows promise for aiding in understanding complex interactions within and among infrastructures, and continues to be developed and tested (Mendonça et al., 2014).

Kulawiak, in a 2013 paper, presented a system offering tools for target analysis, simulations, and spatial analysis for use in analyzing municipal Critical Infrastructures with a remote, web-based geographic information system. This system was applied to research in the city of Gdansk, Poland, including blast attack, chemical contamination, and flood hazard scenarios. The system also used a spatial density algorithm, that identifies events where the proximity of certain infrastructures can influence their susceptibility to attack (Kulawiak and Lubniewski, 2013).

Liu (2014) used a computational model of failures (incorporating field knowledge, records, results from inspection, and sensory data) within the infrastructure of water transmission and distribution systems, with the goal of facilitating the decision-making process in water main renewal. The model used fuzzy theory-based techniques (fuzzy synthetic evaluation, and a fuzzy Markov process) (Liu et al., 2009).

Recently a paper by Ouyang (Ouyang, 2014) reviewed studies across the critical infrastructure field and broadly groups the existing modeling and simulation approaches into six types: empirical approaches, agent based approaches, system dynamics based approaches, economic theory based approaches, network based approaches, and others. The author also offers future research directions and identifies critical challenges in the field. Future directions that are identified are a) “data access and collection,” b) “comprehensive modeling and analysis,” c) “Integration and co-simulation” and d) “validation and applications.” This dissertation addresses Ouyang’s areas of comprehensive modeling and analysis, validation, and integration. In Chapters 4 and

Chapter 5 of this dissertation, I delve into comprehensive modeling critical infrastructure systems and also propose a decision support framework that would help integrate different approaches. In Chapters 6-8 I evaluate existing service area estimation techniques and proposes new ones showing that some of these methods perform satisfactorily would lead to decrease in dependency to detailed distribution system data.

3.5 Previous Work on Service Area Modeling

Service area estimation methods are used to determine source-sink relationships between critical infrastructure service distribution nodes and demand nodes. Chapters 6 to 9 describe my research on service area estimation as an enabling knowledge model component for decision support. Here I provide an overview of concepts and related work in service area estimation as a foundation for the development and testing of my approach.

3.5.1 Spatial Approximation Methods for Service Areas

Utility services, including electric power (EP), water, natural gas (NG), and telecommunication (Telecom) serve their customers through various functional service source facilities, such as power substations (EP), pumps and pipes (water, NG), and switch controls and cell towers (Telecom). Each of these sources is related to a geographical area that represents the customers in their service area.

Defining service areas accurately has long been a problem, but estimating these boundaries accurately is very important in disaster recovery situations. During emergencies, external decision-makers often rely on estimates of service areas produced by various methods. Typically, a geographic boundary for each serving point is defined to estimate the source-sink relationships between the serving network entities (“sources”) and the entities using those services (“sinks” or “demands”). Data that encapsulate the source-sink relationships between these serving points and the customers are often unavailable to external sources. However, the notion of source-sink analysis as the foundation of the network models underneath critical infrastructure analysis is extremely important. This kind of analysis is used in Chapters 5-7 to weight and improve the accuracy of all investigated service area estimation methods for power and water data. Increased estimate accuracy could lead to a more efficient recovery. It is important to understand the comparative merits of the estimating approaches to support decision-makers as they develop mitigation and remediation strategies after a damaging event (Castongia, 2006; L. J. Fenwick and Lyne, 1999; P. Newton, 1997; Sulewski, 2013).

Before a useful process can be designed, spatial approximation methods must be fine-tuned for better accuracy within critical infrastructure systems (Keen, 1980). The delineation of specific service areas can be done using software (such as ArcGIS). However, much of this software requires unobtainable data, and is designed especially for utility companies. Furthermore, even though utilities have detailed information about specific distribution source-sink relationships between their assets, this information is not designed or organized to facilitate large-scale analyses, nor is it documented by public regulatory agencies. In fact, these data are often considered sensitive or proprietary. As a result, CIP planners have started using other spatial approximation methods for delineating service areas within distinct power grids (Meyers, 2001). The two main methods used are spatial proximity and rule-of-thumb methods. These two methods are of limited value because they do not incorporate factors such as variation in supply and demand.

3.5.2 Computational Geometry

Computational geometry is based in geometric modeling. Geometric modeling, in turn, consists of the mathematical depiction of geometric shapes, such as curves and surfaces. Computational geometry branches from basic modeling, with emphasis on developing, analyzing, and implementing computer algorithms that were derived in geometric modeling. The application of computational geometry is broad, with substantial use in engineering, biology/ecology, geology, and other core sciences (Patrikalakis and Maekawa, 2009). De Berg (2000) defines computational geometry as "the systematic study of algorithms and data structures for geometric objects, with a focus on exact algorithms that are asymptotically fast." The field has produced a large collection of efficient and useful geometric algorithms that are used today (De Berg et al., 2000).

GIS is rife with geometric problems, because geographical data consists of shapes of land masses, including heights and depths, vegetation types and density, population data, and weather patterns. These systems have a lot of relevance within the study of critical infrastructures, as they can also store data for human-made structures (i.e. roads, cities, electricity lines). Voronoi diagrams are one method used to solve a geometric problem. For example, they can be used for mapping out the optimal route from one location to another. These diagrams have also been paramount in the study of critical infrastructures and interdependence among them (De Berg et al., 2000).

3.5.3 Thiessen polygons (Voronoi diagrams)

Research in the definition of service areas for critical infrastructure has utilized the Voronoi approach (Akabane et al., 2002; P. Newton, 1997; A. Okabe and Sugihara, 2000). Thiessen polygons, also known as Voronoi diagrams, are useful for data presentation and analysis data since they uniformly and systematically partition an area. They are simple geometrical structures that are easy to create, computationally less demanding than other methods, and the output is a uniform discrete dataset composed of polygons separating the space into equal pieces based on the center points. Thiessen polygons define abstract boundaries, which is critical when approximating substation service areas, which tend to be constrained by geographic traits such as water bodies (rivers, lakes, and wetlands) or mountain ridges.

Given some number of points in a Euclidean plane, a Thiessen diagram divides the plane according to a nearest-neighbor rule: each point is associated with the region of the plane closest to it (Aurenhammer, 1991). To create the region boundaries, this method draws a straight line between all pairs of points; on each line's mid-point, a perpendicular line is drawn at equal Euclidean distance to each endpoint. Thiessen polygons take shape when perpendicular lines are trimmed at intersections with other lines (Boots, 1986) (FIGURE 3). Work by Okabe (2000) and Okabe et al. (1992) provides detailed discussions on the concept of Thiessen diagrams from both historical and geometric viewpoints.

From a critical infrastructure perspective, the literature includes papers (Akabane et al., 2002; Held, 2004; P. Newton, 1997; H. Okabe et al., 1992) that detail efforts for using this approach to create critical infrastructure service boundaries. For example, some researchers have applied Thiessen diagrams to the approximation of service areas when looking to determine the optimal sites for power quality control centers (Akabane et al., 2002).

One drawback of this approach is that it assumes each point is homogeneous (as seen in FIGURE 3). This is generally not the case because each source point provides varying degrees of service. For example, electric power substations have different load outputs, and natural gas gates have different pressures and output capacities. For weighted Thiessen polygons, the critical infrastructure elements with smaller outputs are assigned smaller service areas calculated by using weighted Euclidean distances (Dong, 2008). In practice, this approach is potentially more realistic than Thiessen polygons with equal weighting. For example, an EP substation with a 2-megawatt power output is estimated to serve a smaller area than its neighbors with larger load outputs, as shown in FIGURE 4.

Research by (Gahegan, 2000, 2000; Gahegan and Ehlers, 2000; Gahegan and Lee, 2000; Gahegan et al., 2000) and (Dong, 2008) represents some of the primary attempts at creating software applications to produce accurate weighted Thiessen polygons.

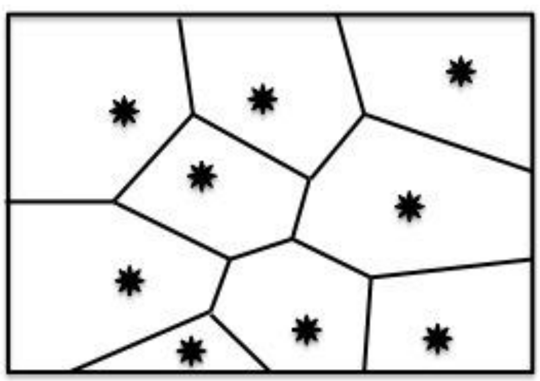


FIGURE 3: Example thiessen polygons

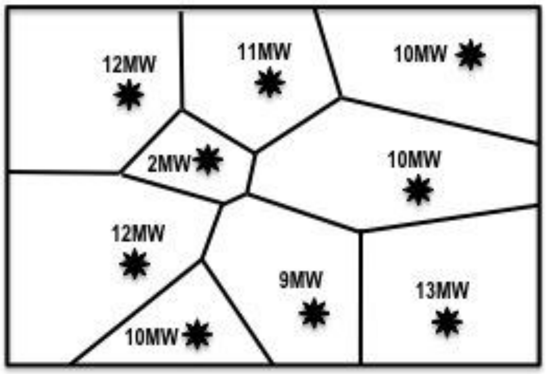


FIGURE 4: Example thiessen polygons with weights

Geosensor network optimizations often employ Voronoi diagrams, along with Delaunay triangulation (Argany et al., 2011). Voronoi models have also been used to estimate the volume of reef-top sediment bodies in Hawai'i (Bochicchio et al., 2009). Interest in quantifying the emission of methane from landfills has led to the use of Voronoi diagrams and importance sampling for robust prediction of gas and other volatile organic compound emissions (Mackie and Cooper, 2009).

3.5.4 Cellular Automata

Discrete computational systems that are composed of a finite or enumerable set of homogeneous, simple cells as a part of a spatially and temporally discrete grid structure are called cellular automata (CA) (Berto and Tagliabue, 2012). Often, CA are explained as mathematical models for complex natural systems that contain large numbers of simple identical components with local interactions (Wolfram, 1994). I adopt the following as a formal definition:

> "[A] Cellular automata system [is] composed of adjacent cells or sites (usually organized as a regular lattice) which evolves in discrete time step. Each cell is characterized by an internal state whose value belongs to a finite set. The updating

> of these states is made in parallel according to a local role involving a neighborhood of each cell." (Chopard, 2012)

There are three fundamental features defining a cellular automaton:

- Uniformity in which cell states are updated using one set of rules;
- Synchronicity in which cell states are all updated simultaneously;
- Locality in which the local rules are used (Schiff, 2011).

The CA approach emerged with the onset of digital computing in the 1940s (Von Neumann, 1953, 1993). However its initial usage in geographic science occurred in the 1970s (Barto, 1975; Tobler, 1979). A growth in geographic information technologies in the 1990s led to an increase in the use of CA within a geographic context (Dowell and Maheshwari, 2000; Torrens, 2000, 2006, 2009; Torrens and Benenson, 2005; Torrens and O'Sullivan, 2001). In retrospect, the adoption of CA by geographic science was natural as both fields intrinsically rely on proximity, adjacency, distance, spatial configuration, spatial composition, and diffusion. In addition, remote sensing, geographic information systems, relational databases, object-oriented programming and CA share mathematical and algorithmic structures (Torrens, 2009). It is also possible to estimate service areas using CA (L. J. Fenwick and Lyne, 1999), and I test these methods in Chapters 5-7 for accuracy. Although CA is applied to a wide variety of fields, CA techniques were not used for service area calculations until the late nineties and early 2000s (L. J. Fenwick and Lyne, 1999; H. Linger and Burstein, 2001).

The CA approach is an iterative approach, whereby each substation cell claims one neighboring cell at a time. The claimed area continues to grow iteratively until the boundary of a neighboring substation service area is met, or until other constraints that have been added to the model, such as the total capacity of a substation, have been met (L. J. Fenwick and Lyne). Similar to the Thiessen polygons, CA algorithms are also run with equal weights or weights based on the actual load values of the substations.

Tools that use CA-based approaches to estimate service and outage areas include the IEISS (Bush, 2005; G. Loren Toole et al., 2008), TranSims, and Water Infrastructure Simulation Environment (TN McPherson and SJ Burian, 2005; D Visarraga et al., 2005).

Los Alamos National Laboratory has developed a modified CA method, called the Constrained Cellular Colonization (C3) method. This method can also be exploited for defining service and/or outage areas for electric power grids (Bush, 2005). Oak Ridge National Laboratory formulated another modified CA approach, which is deemed to be more robust in its ability to determine service areas. Their method incorporates population data derived from Oak Ridge National Labs' LandScan datasets for population distribution, derived using GIS data, remote imagery, and census data, combined with load and location data for the substations. The Moore (square-shaped) neighborhood approach is used in this method, and neighbors are acquired iteratively based on data on supply sources within the given radius. The algorithm assimilates neighbor cells until all cells (non-zero) are accounted for (Omitaomu and Badiru, 2007). CA can also be easily incorporated into GIS software, as a GIS shapefile (Bush, 2005).

Examples of the use of CA in research include a wide range of fields. Chu (2009), upon studying facility design for emergency evacuation including a pedestrian model along with a selection model, used a CA model of the movement of pedestrians (Chu, 2009). Another interesting study uses the CA approach to simulate galaxy formation. The researchers used simple percolation model, with a polar grid having rings that rotate at different rates (FIGURE 5)(Schulman and Seiden, 1986).

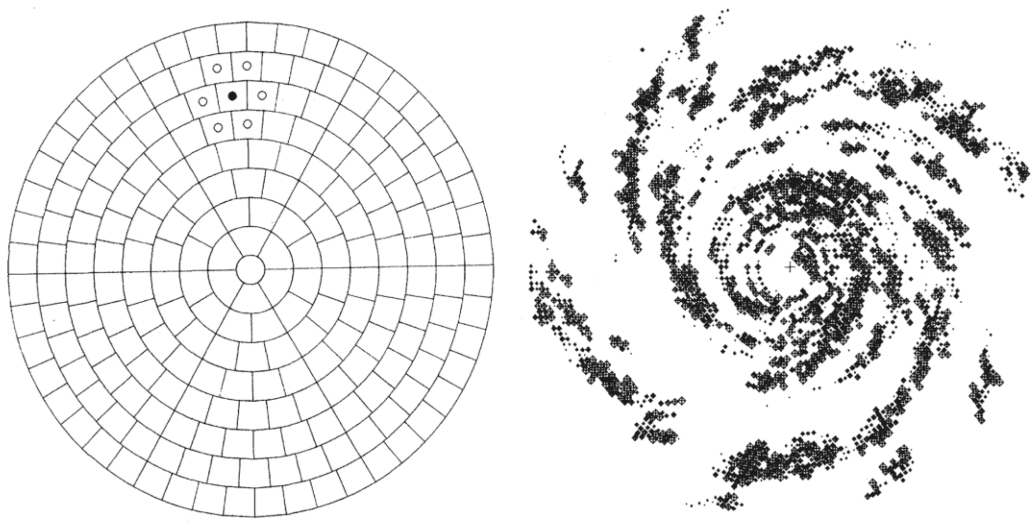


FIGURE 5: Percolation model with polar grid. Source: Shulman and Seiden [1986]

3.5.5 Location Allocation

Chapter 6 and 7 of this dissertation deal with service area approximation for critical infrastructures. In Chapter 6 I introduce two approximation methods based on road network optimization. One of these methods uses location allocation, and here I provide an overview of the basic location allocation approach.

Location and the allocation of specific components of an organization to certain locations are important factors in cost, quality of services and successful service delivery, and accessibility of resources, among other measures of successful functionality of the organization. Location allocation is described as an optimization problem with the task of simultaneously locating facilities and allocating known demand points to those facilities (ESRI, 2014).

General examples include the location of a small café or hairstylist shop that will serve specific local clientele. If placed in a location that is difficult to access or out of the way, that alone can cause the business to suffer and potentially fail. In addition, there are international organizations with factories and distribution centers located across multiple countries. In this case, costs of doing business in a given country, proximity to raw materials, and shipping costs, among many other factors, can contribute substantially to overhead costs. Other examples of general location-allocation problems include optimizing locations for public-sector service facilities including schools, hospitals, fire stations, emergency response services, and many more. In all of these cases, ease of access, and speed of response is critical, and may drastically impact the ability to save lives (2014).

Jenelius et al. (2010) considered the problem of location allocation for critical infrastructure assets, with specific consideration given to the ability to protect specific assets from malicious attacks. They present a modeling framework that assumes attackers have imperfect knowledge of the system, and show that optimal resource allocation can differ substantially from what would be expected given an attacker with perfect knowledge (Jenelius, 2010; Jenelius et al., 2010).

Location-allocation modeling was utilized by Al-Rasheed and El-Gamily (2013) to study how planning and decision-making can be improved for educational resources available to children in Kuwait. Their results showed many shortcomings, which, if remedied, would have an enormous impact on average knowledge and skill levels within Kuwait society (Al-Rasheed and El-Gamily, 2013).

In other applications of location-allocation models, researchers examined chronic health effects due to exposure to certain toxins or otherwise harmful substances in the environment (Liu et al., 2009), planning road transport (Jenelius, 2010), creating land-use maps (Rozenstein and Karnieli, 2011), and siting waste landfills (Ferretti, 2011).

Disaster responders have used location-allocation tools extensively for situational awareness in disasters, which has played a critical role in timeliness and effectiveness of responses to emergencies by allowing for expeditiously delivered data of high volumes and reasonable accuracy. A wireless architecture, DistressNet, was introduced and applied to support disaster response. The architecture is designed to include collaborative sensing mechanisms, topologically-aware routing mechanisms, and reliably accurate resource localization. In addition to DistressNet, a set of protocols and other applications for disaster response are recommended (George et al., 2010). Even though the potential benefit from this type of architecture is high, it has not been widely applied for disaster management situations, due to the relative scarcity and lack of predictability of large-scale disasters (George et al., 2010).

3.5.6 Accuracy Assessment

Decisions about resources typically require a map. Inaccuracies in the map lead to ill-informed, and possibly harmful, decisions, particularly if the decision-maker is unaware of the inaccuracies. As a result, it is important to measure the accuracy, and thus, the reliability, of a map prior to using it for decision-making purposes (Congalton and Green, 2008).

As a part of this dissertation, I survey and test commonly used service area approximation algorithms and introduce two approaches that have not previously been used for this purpose. In order to evaluate and compare the accuracies of these service area approximation methods, I apply accuracy assessment procedures (Congalton, 1991; Congalton and Green, 1999, 2008). These procedures have been commonly applied to test the accuracy of land cover type classification in remote sensing field, but not for service area estimation, and I adapt them for assessment in the context of critical infrastructure analysis.

An accuracy assessment allows for the determination of the accuracy of a map created using remotely-sensed data. There are two types of map accuracy assessment, thematic and positional. Positional accuracy emphasizes the accuracy of the position of features on the map, and measures the distance of these features from their true locations. A thematic measure of accuracy focuses on the accuracies of labels on the map. For example, a wetland area that is characterized as an open field would be a labeling inaccuracy (Congalton and Green, 2008).

Thematic accuracy cannot be depended on to characterize geometrical properties of classification maps with very high-resolution images. Persello and Bruzzone (2010) recommend a protocol incorporating two indices from separate sources. The first is a traditional thematic accuracy index. The second is a novel set of geometric indices derived from separate geometric properties of objects within the images. They recognize five types of geometric errors commonly occurring in maps. These include over-segmentation, under-segmentation, edge location, shape distortion, and fragmentation. They applied their methodology to digital environmental models derived from the QuickBird land observing satellite with sub-meter accuracy, and found greater effectiveness compared to standard protocols (Persello and Bruzzone, 2010).

Foody (2010) studied the accuracy of land cover change detection, wherein the ground reference data contained errors. Through simulation, the impacts of various ground data imperfections on accuracy of land cover change were studied, and it was discovered that even relatively small amounts of known error in the ground reference data had the potential to produce large error in the results. The author suggested methods for removing the ground reference data error, using simple algebraic techniques to estimate the actual accuracy and extent of change (given that the imperfections were known), and using a more complex latent class analysis to assess the classification accuracy and to estimate change extent without using ground reference data (Foody, 2010). These types of methods are similar to those presented here to estimate accuracy of service area estimation methods and demonstrate how such methods could be integrated to make predictions for decision support planning.

In this dissertation I am exploring ways to harness existing data to support critical infrastructure decision-making processes and analysis. I show that both tools for exploration through geovisualization and assessments of accuracy are needed to make decisions and analyses of CI data more efficient and correct. In Chapters 4 and 5 I detail the decision recommender tool framework, its implementation into the Critical Infrastructure Explorer prototype, and a user study comparing DRT with common GIS tools.

Correctness of the results of the simulation, and therefore the decisions made by the user, is directly related to the accuracy of the information related to cross dependencies. This accuracy is defined by all of the source-sink relationships in a service area. And this led me to investigate the accuracy of existing service area estimation methods and then introduce two new ones based on transportation network optimization. Chapter 6 investigates the accuracy of common estimation methods to a power network. Chapter 7 investigates the accuracy of two new methods, as well as common methods, to a statewide water network. Chapter 8 applies the new methods to the electric power network. Chapter 9 provides a clear comparison of the service area approximation accuracy results, and the dissertation concludes in Chapter 10.

CHAPTER 4: RECOMMENDATION-BASED GEOVISUALIZATION SUPPORT FOR RECONSTITUTION IN CRITICAL INFRASTRUCTURE PROTECTION

In order to address the general research question of how spatial decision support systems can be improved to facilitate more accurate and efficient decision-making in CI analysis and recovery, I conducted three main research studies. This chapter introduces the first research study (D. C. Wilson et al., 2009), which investigates the following specific research question:

> Specific Research Question RQ1: Is a spatial recommender system focused on critical infrastructure cross-infrastructure effects more efficient and effective than using commonly-used, industry-standard GIS tools for Critical Infrastructure recovery decision-making tasks when multiple networks are interrelating?

To answer this question, I developed and evaluated an approach to CI decision support, which was then implemented within an interactive geovisualization environment. This chapter describes (1) the framework for CI decision recommendation tools, and (2) the implementation of a particular Decision Recommendation Tool (DRT) called the Critical Infrastructure Explorer (CIE) based on the framework. In the following chapter, I go on to describe the user study evaluation of the approach.

My approach to cross-infrastructure modeling and simulation for reconstitution leverages the strengths of a multi-dimensional approach for a variety of stakeholder contexts. I believe this approach provides an appropriate foundation for multi-dimensional analyses of critical infrastructure interdependencies, and I include some initial scenario-based results to demonstrate the kinds of analyses and subsequent understandings to be gained from this research. The remaining sections of this chapter describe the approach, illustrated with components of an example scenario.

4.1 Recommendation Approach

My approach to providing active support for decision-making during infrastructure reconstitution efforts is grounded in a framework for prioritizing infrastructure elements based on potential impacts. The framework provides recommendations that highlight high-impact infrastructure elements to consider in planning resource allocation to reconstitution efforts. This framework is comprised of four primary components, listed below and as illustrated in FIGURE 6.

- **User model**: Defines general task goal categories, specific weightings of those categories to reflect the user's individual viewpoint, and the geographic scope of interest to the user
- **Target model**: Defines individual infrastructure knowledge, cross-infrastructure dependency knowledge, and the knowledge necessary to enable metric assessment of desired outcomes
- **Simulation engine:** Calculates effects within and across infrastructures based on initial conditions and perturbations
- **Recommender engine**: Maps user goals to metric assessment in the context of modeling and simulation to prioritize the impact of infrastructure elements

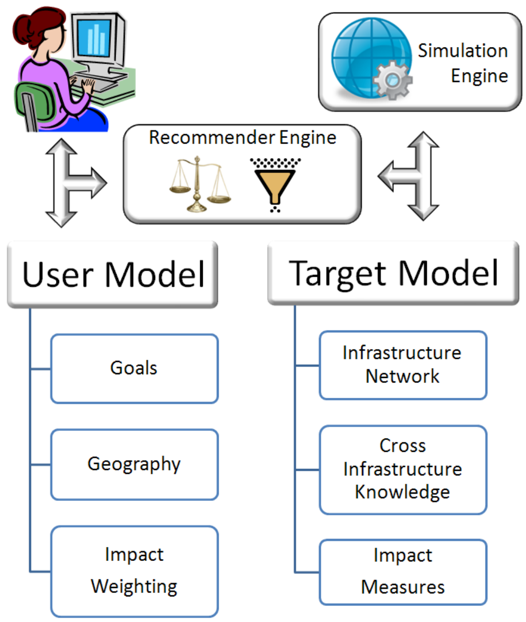


FIGURE 6: Recommender framework

The following sections provide further detail on the component aspects of my approach, illustrated with a scenario example of the implementation. In the initial implementation, my goal has been to develop a proof-of-concept for the overall approach. Thus I have chosen simple component techniques (e.g., for determining cross-infrastructure interactions). Future research would test a variety of finer-grained variations on component modeling techniques.

4.1.1. User Model

Reconstitution decisions are highly contextual and dependent upon stakeholder viewpoints. The user model component of my framework captures this characteristic through three primary aspects. For a given stakeholder perspective, I consider (1) the geographic scope of the stakeholder mandate; (2) the main stakeholder goals in terms of reconstitution; and (3) the relationship of these goals to the individual infrastructures being modeled. To this end, I have established an approach to specify individual stakeholder perspectives (e.g., emergency services, national guard, and citizenry) in terms of different recovery levels. For example, emergency services may have an initial focus on prioritizing transportation network recovery, as they may have their own generation equipment and primarily need ease of access to canvas the region for search and rescue. The corresponding degree may also be characterized in different ways. For example, a minimal baseline level of transportation network functionality may suffice for heavy-duty rescue vehicles (e.g., primary roads accessible, or certain proportion of network), as opposed to a full reconstitution of all roadway feeders and arteries.

This user model is embodied in a Recovery Evaluation Matrix (Vatcha et al., 2009) structure. Currently, the REM is manually specified and the geographic scope is fixed for each stakeholder, but more flexibility could be added as tool support for user model specification develops. Moreover, the current REM settings do not reflect stakeholder expert domain knowledge. The benefit of the REM is that it enables domain experts to specify and prioritize their goals as part of the analysis modeling and simulation.

The user model embodied in the REM is then used to evaluate recovery options within the modeling and simulation environment. Thus, a decision maker can be recommended a candidate prioritization for recovery resources, given (1) a network (e.g., electric power); (2) a layer that represents an impact criterion (e.g., population impact by service area); and (3) a list of unavailable network elements. This recommendation would display the network element that provides the most benefit to a pre-defined target criterion on a digital map, and therefore should be restored first.

4.1.2 Target Model

Reconstitution decisions are highly dependent upon the associated infrastructures and their functional and spatial properties. The target model captures this understanding for a given region. Each target model instance is comprised of individual infrastructure knowledge, cross-infrastructure knowledge, and grounding metrics for impact assessments.

4.1.2.1 Individual Infrastructure Knowledge

Modeling and simulation for infrastructure analysis, of course, relies on the fidelity of the underlying infrastructure data. In my work here, I do not address the enabling and complementary problem of acquiring such knowledge for individual infrastructures. Rather, I presume that such knowledge already exists to support baseline, unassisted analysis activity. I know that significant critical infrastructure data is often controlled by the private sector, which has a vested interest in keeping such data secret. My subsequent research studies address the issue of modeling the supporting knowledge. In this study I employ infrastructure models that were developed as part of a related project on techniques for building just such models from available data sources (Tolone et al., 2007).

FIGURE 7 shows an illustrative electric power network model that was used as part of my research. In this model, points represent power plants (letters) and substations (green/orange squares), while lines represent the connecting power lines (green lines). This illustrative model supports the identification of neighboring network elements, as well as the determination of basic network flow; and thus intra-infrastructure dependencies.

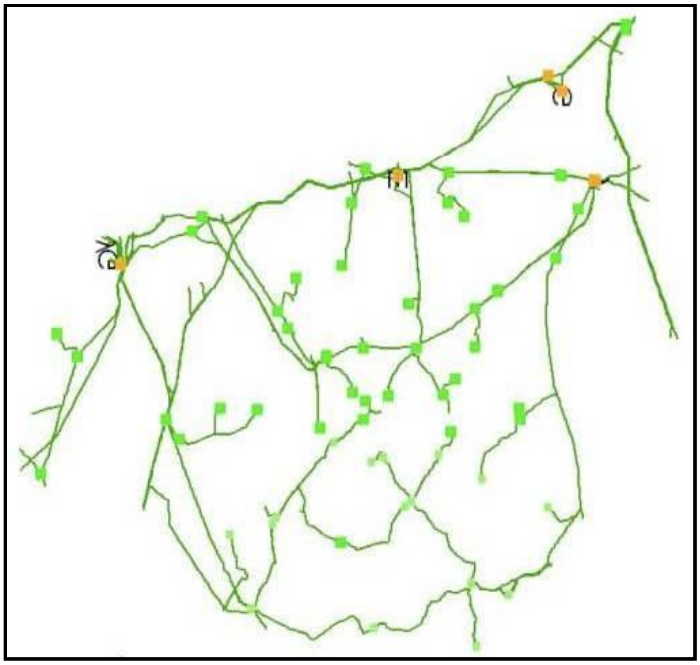


FIGURE 7: Example electric power network

4.1.2.2 Cross-Infrastructure Knowledge

To model cross-infrastructure dependencies, I employ a proximity-based approach guided by domain knowledge. For example, I divide the power network into a set of spatially distinct polygonal “service areas” delineating each infrastructure element. When an element of a related and dependent infrastructure, such as a cell tower in the communication network, falls within the service area of a power network element (e.g., substation), the related element (cell tower) is assigned a dependency relationship with the provider element serving that area (substation). FIGURE 8 shows an example of the service areas associated with power network substations approximated using Thiessen polygons.

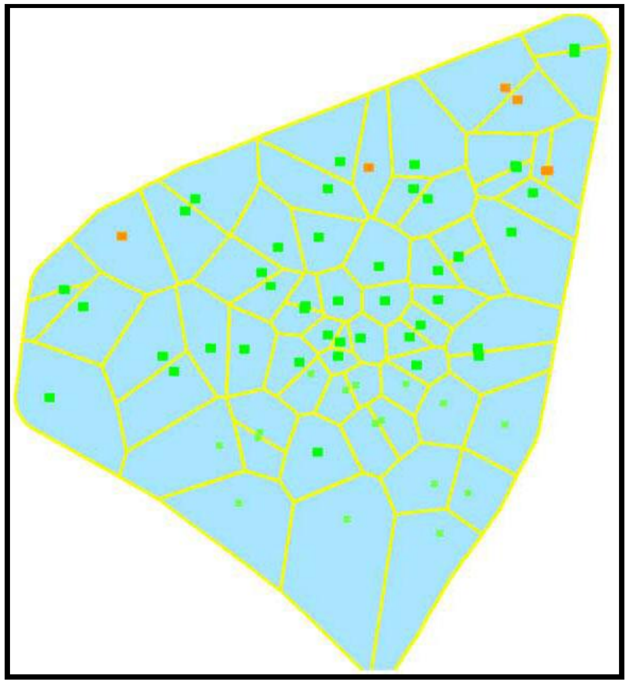


FIGURE 8: Service area determination

4.1.2.3 Impact Metrics

I employ service impact metrics to provide concrete measures of an infrastructure’s relevance to task goals. For example, if the local government stakeholder wants to ensure that services are restored most quickly to the greatest number of constituents, they could measure the impact by prioritizing based on areas of greatest population density. Other measures might include prioritizing healthcare (e.g., hospitals) or educational (e.g., schools) infrastructure elements.

To illustrate, I have implemented a population density impact metric in the form of a process flow that relates power substations to the number of people for which they provide power.

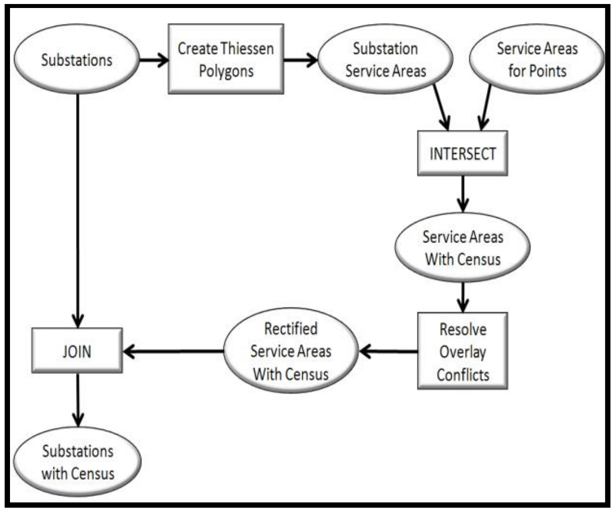


FIGURE 9: Determining population impact measure

The second step in the process involves combining the service area layer with population layers. I employ Block data from the US Census Bureau for population, as shown in FIGURE 10 for an example region. I combine service areas and Census population data to carry population data over to the service areas. However, the process requires additional steps, as some service area polygon boundaries intersect census block data polygon boundaries. So, one census block may belong to two or more service area polygons. I included an additional stage in the process to resolve such conflicts automatically and assign an appropriate proportion of the population to each service area, based on the percentage of the census block polygon overlapping the service area.

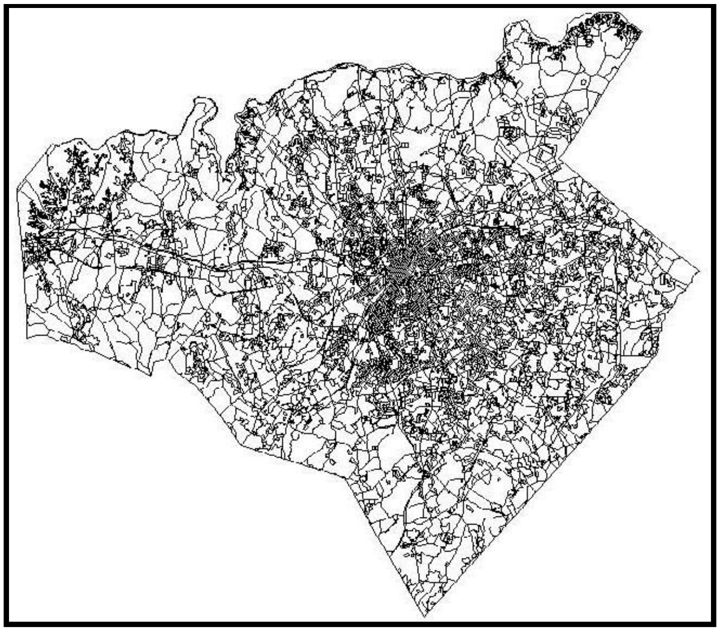


FIGURE 10: Census block data for example region.

I first recalculate the areas of each polygon after the union process, and then calculate the percentage of the new intersected polygon based on new area and old area. Next I recalculate the population information based on this percentage. As an example, assume I have a census block with 1000 people in it and the service area boundary intersects this polygon so that 60% of the polygon by area now belongs to service area A and the remaining 40% now belongs to service area B. I disperse the population accordingly, so that 60% of the original census block population now should belong to service area A, and 40% of the original census block population now should belong to service area B.

The final stage of this process is to carry the population information that is now inherent to service area polygons over to the power substation points via a spatial join operation. The overall processes both for relating point (as in my communication cross-infrastructure example) and polygon (as in my population grounding example) is implemented in ESRI Model Maker, as well as provisioned as a generic tool interface in ArcToolbox.

4.2 Recommender & Simulation Engines

Given representations for infrastructure and cross-infrastructure connectivity, as well as related grounding measures, I can model the impact of various failed infrastructure elements, in order to measure the importance of an individual element to the user's task goals.

4.2.1 Simulation Dependency Knowledge

Simulating disablement and enablement scenarios requires utilizing not only connectivity of the network elements but also the dependencies of the network elements. I developed an independent application that utilizes ESRI libraries and data structures. There are two possible approaches to follow. The first is to employ ESRI programming structures to simulate network outages. The second is to export infrastructure network data to an external junction dependency matrix, simulate the outage using this matrix and send the result back to ESRI tools for display. Given simulation performance considerations and potential complexity in cross-infrastructure interactions, I chose to use the latter approach. I export network data at the beginning of a session and create a dependency matrix used to perform the simulation.

The first phase of the process to create the dependency matrix extracts the network connectivity graph and creates a matrix data structure where rows and columns represent network elements, with matrix entries representing connectivity. This provides a matrix of network connectivity, but network element dependencies (e.g., flow direction) are also needed for simulating outage and reconstitution effects. I then perform a dependency analysis on the network data structure to create a dependency matrix. FIGURE 11 and FIGURE 12 show a simplified example matrix and a network that it represents. A similar process is employed to create a set of dependency matrices for cross-infrastructure dependencies, with each matrix representing the dependencies present between two distinct infrastructures. There are a variety of ways to determine cross-infrastructure dependencies, depending on the fidelity of the source network data. In the absence of direct cross-infrastructure dependency information, geographic proximity, such as service sheds approximated with Thiessen polygons, may be employed, as shown in FIGURE 13. In the next section I explain how I perform basic simulations based on the dependency matrix.

		TO								
		0	**1**	**2**	**3**	**4**	**5**	**6**	**…**	***n***
FROM	**0**	**1**	**1**	**1**	**0**	**0**				
	1		**1**	**0**	**0**	**0**				
	2			**1**	**1**	**0**				
	3				**1**	**1**				
	4					**1**				
	5						**1**			
	6							**1**		
	…								**1**	
	m									**1**

FIGURE 11: Example dependency matrix

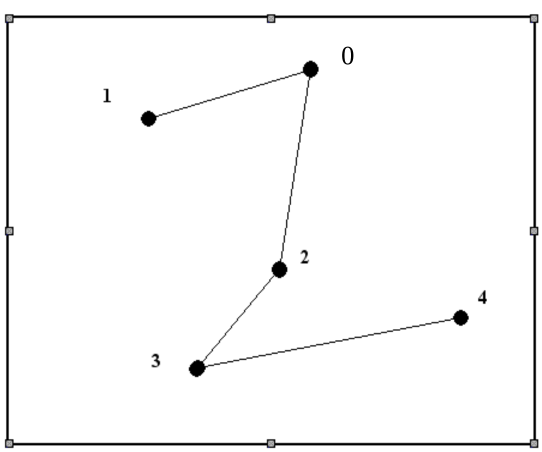


FIGURE 12: Example network for dependency analysis

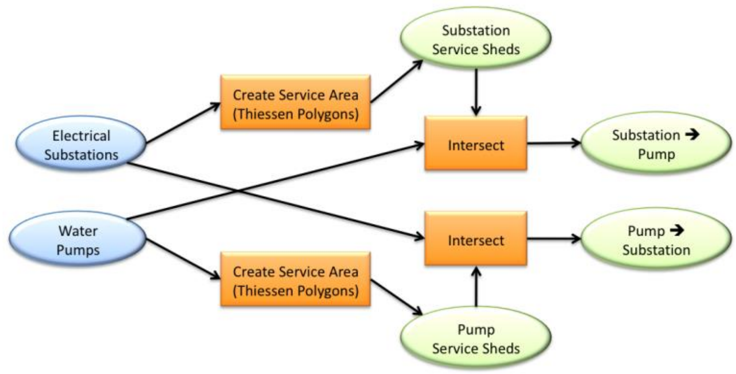


FIGURE 13: Cross infrastructure service area modeling

4.2.2 Simulation Metrics

To perform a simulation to determine potential impacts of infrastructure elements, I begin with a representation of the state of each infrastructure network. I employ a state matrix for each network, in which each element is represented as functional or non-functional. At the start of a simulation, the user sets the initial state of the system, in order to represent the failed infrastructure elements in the current task context, as shown in FIGURE 14.

A dynamic placeholder matrix is used to hold currently disabled nodes. For each element that is disabled, the simulation environment follows the dependency matrix and obtains a list of elements that are directly dependent on the current element. It then searches through the network and obtains a list of dependent objects to each of the nodes that are being held in the temporary matrix. This process is repeated until it goes through all the dependencies. When a step in the simulation process would re-enable a failed infrastructure element, either directly or indirectly, available grounding measures are evaluated and added to the overall impact measure. By cycling through failed infrastructure elements across each infrastructure of interest to evaluate overall impacts, I compute an individual relevance measure for each element.

Using this information it is straightforward to test the effect of each node on the population, given that population information has been associated with each substation in the network. Based on this information I provide a recommendation to the decision maker about the candidate junctions to be enabled that will enable service to the greatest number of people. FIGURE 15 shows an example screenshot. The power substation highlighted is the node recommended for restoration first.

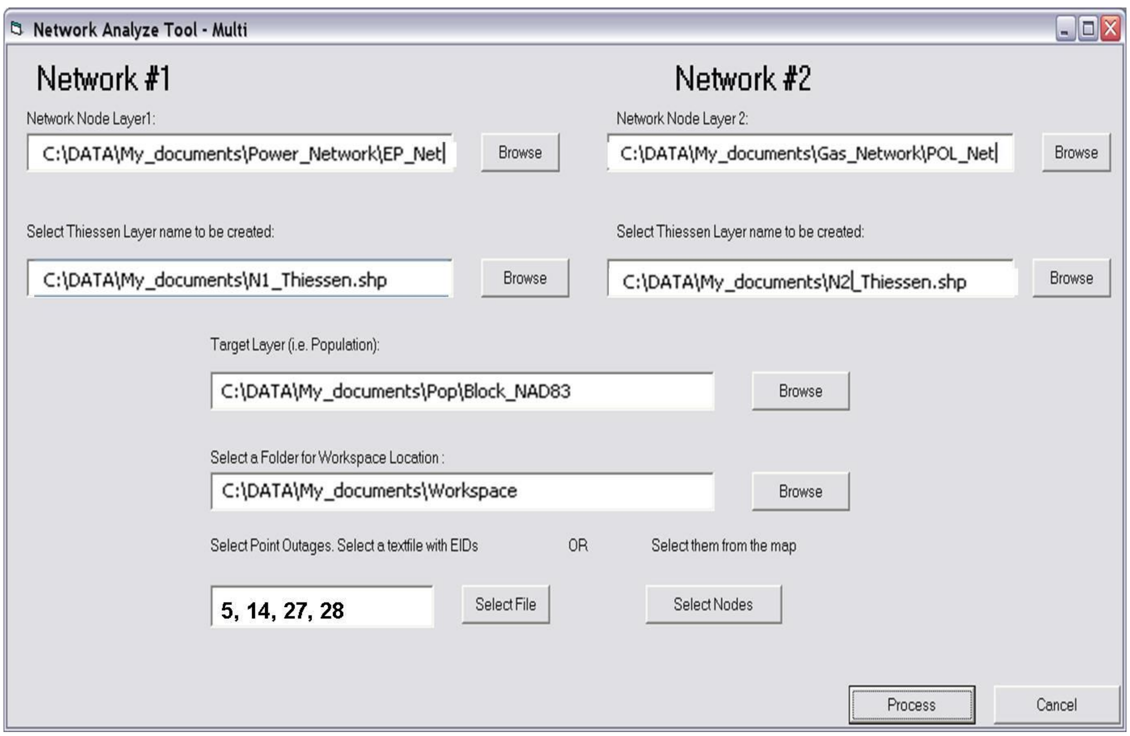


FIGURE 14: Scenario initialization

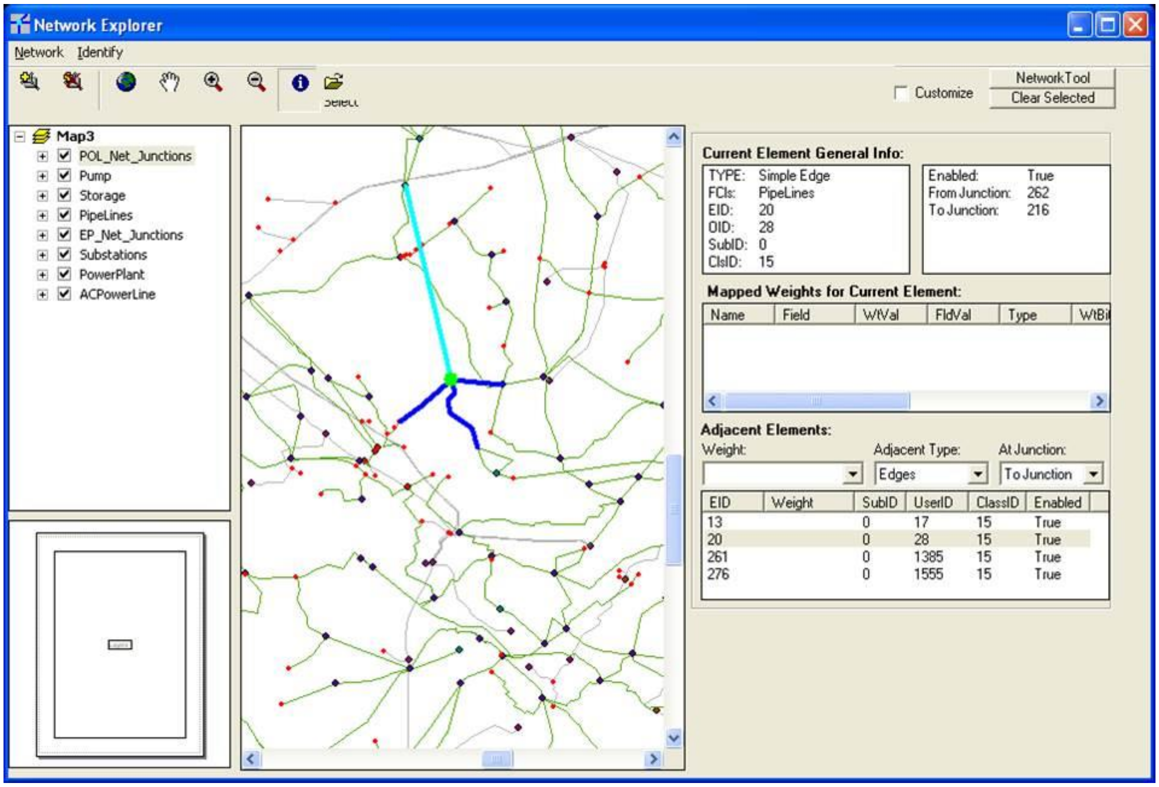


FIGURE 15: Recommended infrastructure elements for prioritization.

4.3 Decision Support Recommender Framework Conclusion

This chapter presented my recommendation-based approach to support analysis and decision-making for cross-infrastructure reconstitution tasks. The contributions of this chapter are: (1) development of the user model that accounts for varying stakeholder perspectives and priorities, as implemented in the recovery evaluation matrix; (2) development of the target model, as implemented in the cross-infrastructure dependency and grounding metric analyses; (3) development of the baseline modeling and simulation engine coupled with the recommendation engine for prioritizing infrastructure elements for consideration; and (4) providing this analysis functionality within an interactive geovisualization interface. These elements are integrated within my prototype GIS environment, as a proof-of-concept for the approach. Since decision-making environments can be extremely complex and dynamic, it is as important to reduce the cognitive load and overhead as to improve decision quality (Todd and Benbasat, 1992). My approach and prototype system accomplishes this by approaching the decision maker with a recommendation based on specified criteria.

The work I have reported here is a first step in creating a complex system that supports analysis and decision-making for critical infrastructure networks. An effective critical infrastructure analysis and decision recommendation tool is critically dependent on deriving cross-infrastructure dependencies. This ability depends on accurate service area estimation. To address the needs for supporting decision-making and analysis for critical infrastructures, it is important to understand the impacts of a decision recommendation tool on users' task performance, while also understanding the potential tradeoffs in accuracy underlying the methods used to estimate service area coverage. In the next chapter, I present a user study evaluation of the critical infrastructure explorer tool (CIE) that I built using the DRT framework presented in this chapter. In later chapters, I explore the accuracy of service area estimation techniques.

CHAPTER 5: USER STUDY ANALYSIS OF A GEOVISUALIZATION DECISION SUPPORT ENVIRONMENT FOR CRITICAL INFRASTRUCTURE RECOVERY

Given the DRT framework and CIE analysis tool developed in the previous chapter, I conducted a user study evaluation in order to investigate my first specific research question (Pala and Wilson, 2013) – (RQ1): Is a spatial recommender system focused on critical infrastructure cross-infrastructure effects is more efficient and effective than using commonly-used, industry-standard GIS tools for Critical Infrastructure recovery decision-making tasks when multiple networks are interrelating? This chapter describes a user study evaluation of the Critical Infrastructure Explorer (CIE) system conducted in order to address RQ1.

The user study focused on GIS system experts and GIS analysts as participants. It collected performance data as users ran through several disablement analysis scenarios separately with the CIE and ArcGIS tools. I chose ArcGIS with the Utility Network Analyst extension (referred to as STDGIS for standard, or commonly used, GIS) since the study's system experts employ this ArcGIS extension regularly for analysis of critical infrastructures. I performed quantitative analyses comparing results for the two platforms in terms of effectiveness measures for end users. It is important to note that the scenarios used were static, in order to investigate how two different systems could support a single decision for critical reconstitution. It was assumed that the emergency event has completed, and instantaneous updates were not made. Real-time systems with instantaneous updates could impact the results of such a study.

5.1 CI Decision Support Approach

My approach for CI decision support is grounded in the decision recommender framework, presented in the previous chapter, for prioritizing infrastructure elements based on potential impacts – recommending high-impact elements to consider in planning resource allocation to CI recovery efforts.

5.1.1 Decision Recommendation Tool: CIE

At the start of a simulation, the user sets the initial state of the system to represent the failed infrastructure elements in the current scenario. The CIE models resulting disablements through the initial CI network and linked CI networks that depend on service from the initial network. It also relates each network node to target layer components in order to determine the effect of disablement of each network node (e.g., population affected). Coupling this with cascading disablement simulation provides an indication of the overall effect of each initial disabled point's effect on the ground.

The CIE affords an interactive interface enabling the user to explore various cross-infrastructure scenarios, visualize the effects of disablements, and thereby explore the best options for CI reconstitution. This includes simulation animations, tabular data, network table of contents, and three information tabs (FIGURE 16):

- "Disablements" shows network disablements in a tree-view structure,
- "Options" shows disablements and details in tabular form. Details can be clicked to initiate related animations on the map (see FIGURE 17). A summary is also provided for each initial disabled network element, showing the effect of its enablement on target service layers. For example, if node #15 is enabled then

electrical service would be restored for approximately 15,000 people and communications (dependent on electric service) for approximately 39,000 people.

- "Network Detail" enables the user to click for detailed information on individual nodes in the networks.

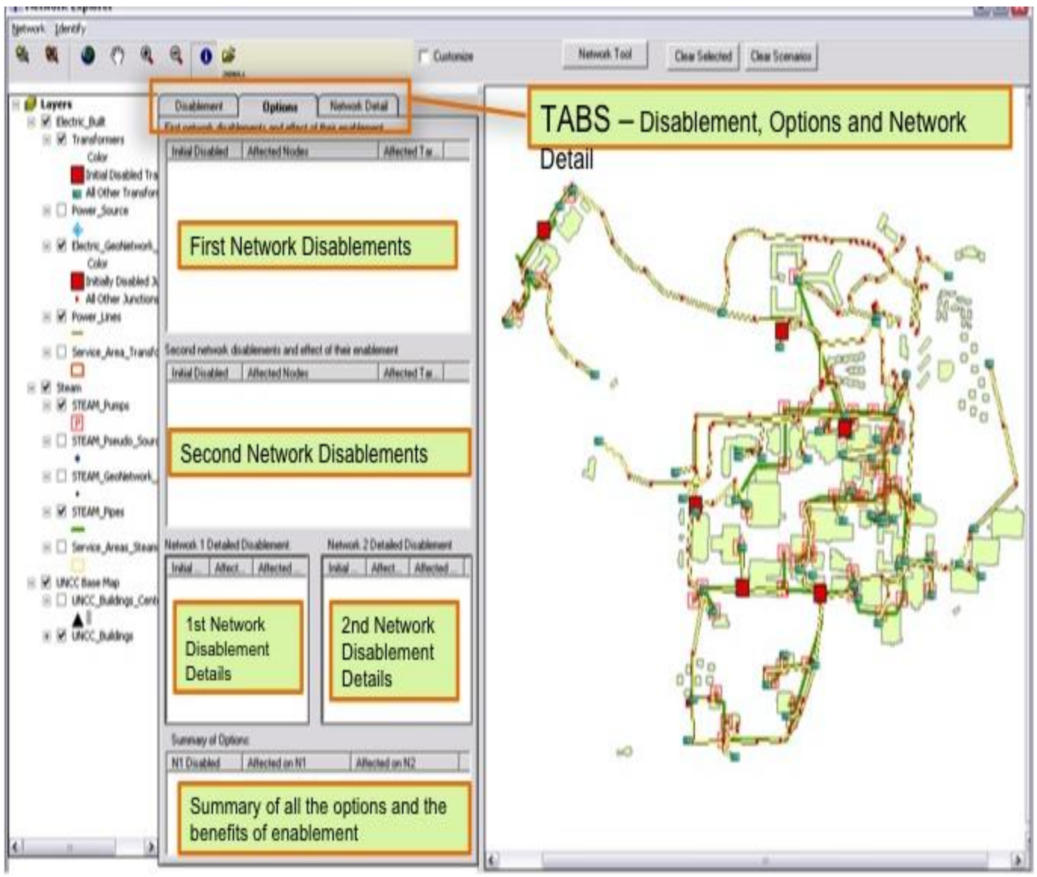


FIGURE 16: CIE interface details

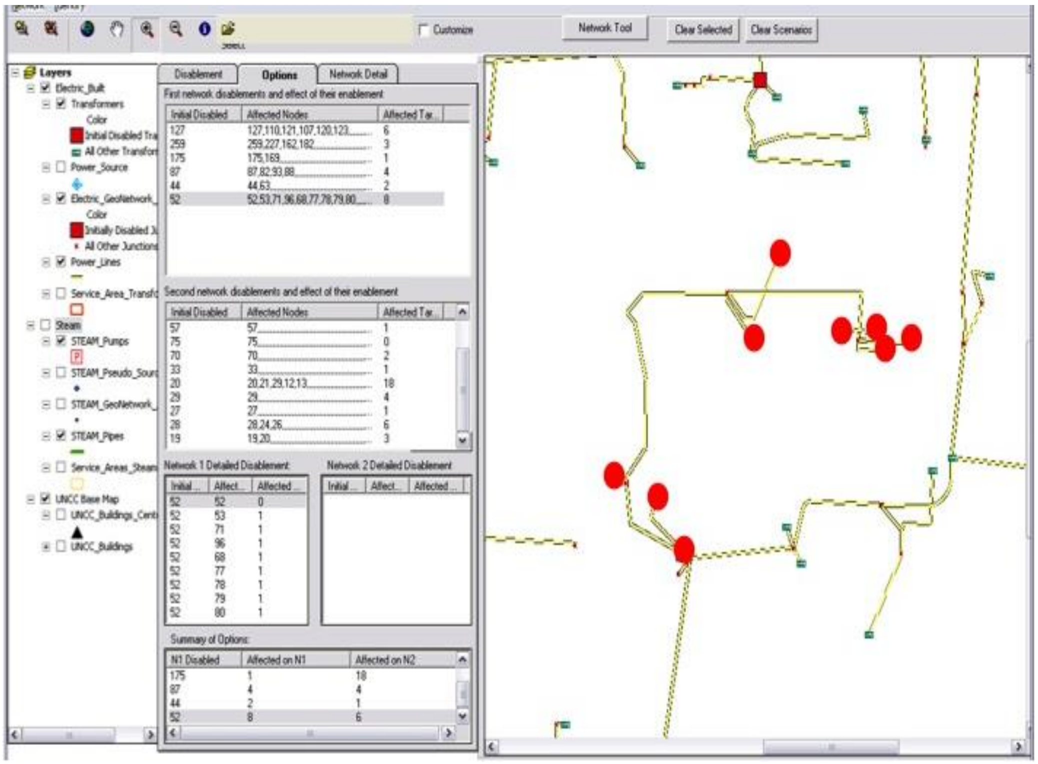


FIGURE 17: On screen outage simulation with CIE

5.2 Study Design

To evaluate my approach, I have conducted a user study with system experts and GIS analysts. The study compares my implemented DRT approach (CIE) with off-the-shelf industry-standard GIS tools (ArcGIS with Utility Network Analyst extension, denoted STDGIS), which are often employed by analysts for CI support, as indicated by my study experts. Such controlled studies typically compare novel tools or techniques to state of the art (Koua et al., 2006; Network, 2014). For example Plaisant et al. (2002) compared three tree visualization tools, space tree, hyperbolic and window explorer. ESRI ArcGIS Desktop software products are commonly used tools for GIS analysis.

Hypotheses:

A Decision Recommendation Tool (DRT) developed for critical infrastructure recovery based on my proposed framework is more efficient than using existing GIS tools for Critical Infrastructure reconstitution when multiple networks are interrelating. More specifically,

H1-1. Given the same scenario, decision makers can make decisions with less time using DRT than using the most common GIS suite of tools.

H1-2. Given the same scenario, decision makers can make decisions with lower cognitive load using DRT than using the most common GIS suite of tools.

H1-3. Given the same scenario, decision makers prefer using DRT than using the most common GIS suite of tools.

For each participant I recorded the screen, user voice and user video with the Morae usability software. Users were asked to complete an exit survey on tool use, cognitive load, user preference, and general feedback. Users were also asked to fill out the NASA Task Load Index (NASA-TLX) Mental Demand evaluation questionnaire twice, once for each tool in the same context. The NASA-TLX is a survey that rates perceived workload to assess a system’s effectiveness (Hart, 2006; Hart and Staveland, 1988). All the materials including the user surveys and tasks can be found in APPENDIX A: Supporting Documents for User Study.

5.2.1 Participants: US National Research Labs employees and UNCC GIS analysts

My user base consists of professional critical infrastructure system experts and GIS analysts from UNC Charlotte. The CI system experts were employees of one of the US National Research Laboratories who have been working on CI-related projects for the last five years or more. GIS Analysts were recruited from UNC Charlotte staff and student GIS users who are proficient in GIS theory and usage of GIS tools. Participants were recruited to gain insight both into the decision makers' approach and the approach that GIS analysts would take to make recommendations to decision makers.

I recruited 5 system experts from national labs and 10 GIS users from UNC Charlotte. Out of 15 participants 13 had at least three or more years of GIS experience. More than half of all the users defined themselves as experts in GIS. For my study, domain experts are considered to be individuals who have been working professionally with critical infrastructure analysis (my national lab subjects).

System experts with at least five years of experience in Critical Infrastructure decision support at Los Alamos National Labs were recruited by email invitation to participate in the study during their regular workdays. As employees of a federal national lab, these system experts have been providing support to real decision makers throughout the U.S. in emergency situations involving critical infrastructures.

GIS users were recruited by sending emails to the UNC Charlotte Geography department and GIS email listserves, as well as posting flyers posted in the UNC Charlotte Geography department. GIS users are the students, faculty and staff professionals who are proficient in GIS and therefore are qualified to play a "GIS Analyst" role with training in CI analysis that I provided. Demographic and expertise info can be seen in FIGURE 18, FIGURE 19, FIGURE 20, and FIGURE 21.

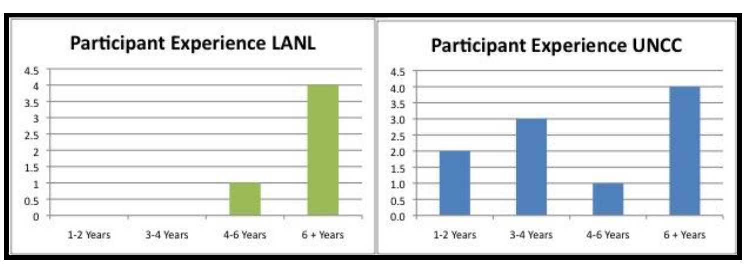


FIGURE 18: Expert user participants' GIS experience at LANL and UNC Charlotte

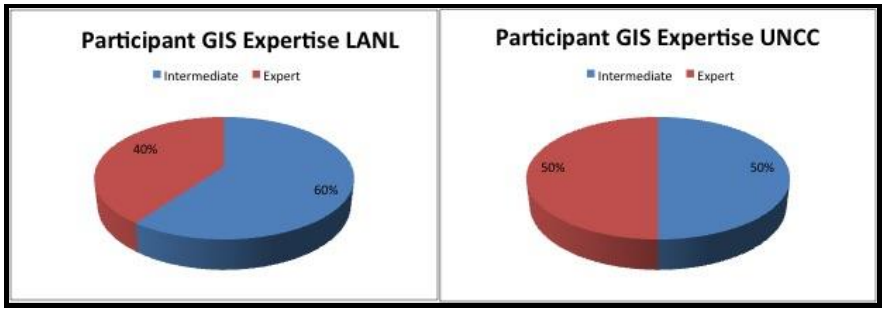


FIGURE 19: Participant’s GIS expertise at LANL and UNC Charlotte

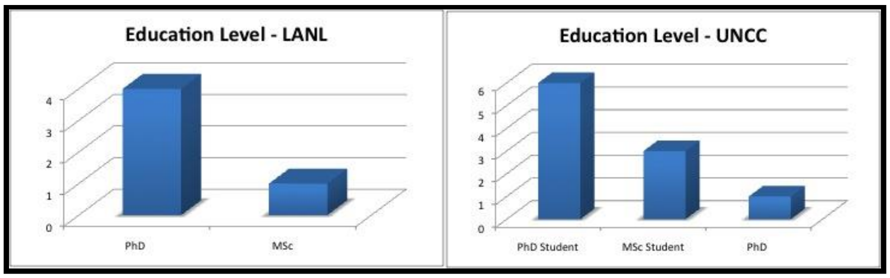


FIGURE 20: Participants’ education levels

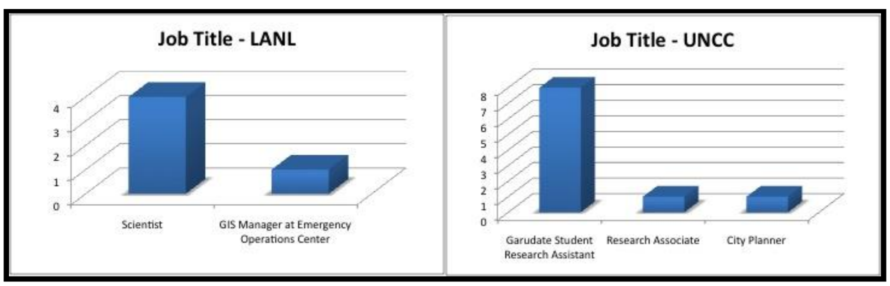


FIGURE 21: Participants' job titles

I selected a within-subject study design because I am comparing the performance and experience of the same group of users in different scenarios with two different tools, and because my targeted user population is small. I counter-balanced the presentation of the interfaces allowing half to be presented with STDGIS first (Group A, 8 participants) and the other half with CIE first (Group B, 7 participants). I have equally assigned system experts to my counter-balanced participant sample.

5.2.2 Measures

To evaluate system effectiveness, I considered task efficiency (time) and outcome quality. I also considered cognitive load, which is important for improving decision quality for decision makers (Todd & Benbasat, 1992). Quantitatively I measured: time spent on each task, rate of outcome correctness, and rate of analysis correctness. These were measured and validated through analysis of recorded participant sessions. To measure cognitive load, I utilized the NASA Task Load Index (TLX) tool (Hart, 2006; Hart and Staveland, 1988).

5.2.3 Experiment Setup

To familiarize users with the software tools in the study, each participant was provided an initial training session on a sample CI outage scenario covering cascades and cross infrastructure effects. I first drew an example scenario on paper – one network with two initial disabled points. Then I showed the participants how the service areas are utilized to determine the buildings that would be affected by these outages. Then I drew the second network elements overlaying the first outage and showed participants how to determine the second network outages based on the first network’s service areas. Next I cascaded down the second network outage and showed how those would be related to the number of buildings in the service areas of the disabled second network elements. Based on this I created a table that lists the first network disabled points and overall effect of each of those in the buildings with respect to type of service being disrupted. After the paper disablement scenario demonstration I ran through an outage scenario once with STDGIS and once with CIE. These example scenarios had one initial disabled point on the first network and two interacting infrastructure networks.

Users were then asked to work through four outage scenarios at increasing levels of complexity. Complexity was set to be similar at each level but with different initial disablements. I applied four levels of complexities, so users worked through eight different scenarios. Users were instructed to act as the GIS analyst in an outage-emergency situation where they are required to provide a report to the Decision Maker (DM) on the priority of the initial disabled points in importance of their effect in each specific situation. This way the DM could allocate appropriate resources to the CI elements with greatest effect on target layer elements for optimum recovery. The network data employed is adapted from UNC Charlotte Critical Infrastructure network data. For this experiment I used CI network data, buildings and building center point layers (FIGURE 22).

The first scenario included one network (water) with two network elements disabled initially. The second scenario included one network (electric power) with six network elements disabled initially. The third scenario included two networks (water, gas) and two elements initially disabled on each network, with the water network interacting with gas in a source-sink relationship (e.g., pump cooling). The last scenario had the electrical power network interacting with the steam network in a source-sink relationship with six initial disabled elements. Users were asked to determine which of the initial disabled network elements should be restored first to provide most benefit. All the initial disablements were on the first network, which was presumed to be providing services essential for elements of the second network to function. For the purposes of this study, I presumed a one-way source sink relationship between networks.

An exit survey was given to all the participants and their verbal comments were recorded using Morae Recorder software. Then all the recordings including exit survey recordings and participant task recordings were imported into the Morae Manager software tool. Through this tool I was able to mark the task beginning and ending times for all the tasks and users. I also was able to mark the task completion rates. Then I exported all the videos and imported those into a software tool called InqScribe to create transcripts of the participant's speech and record timings.

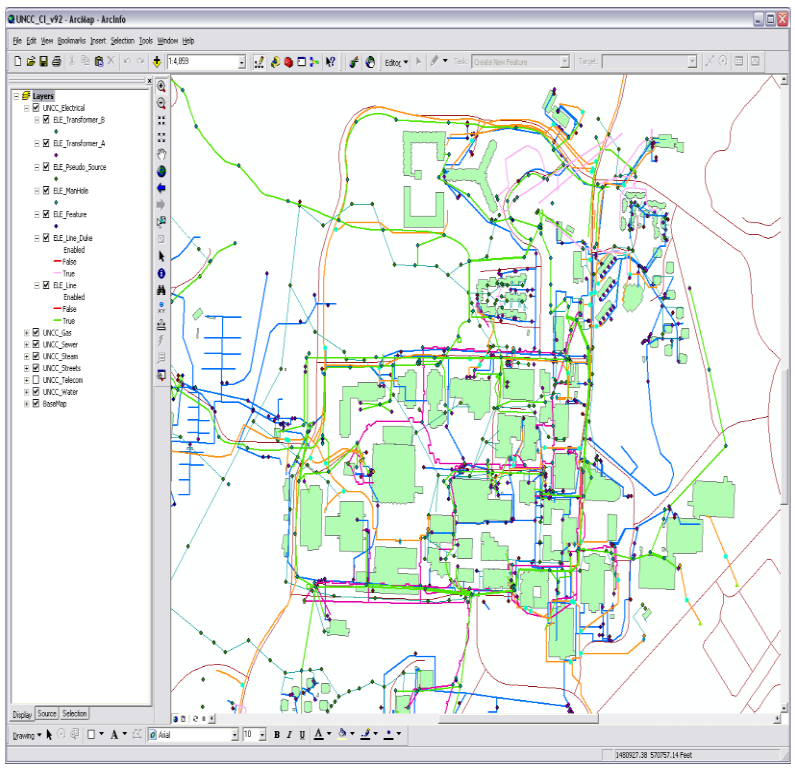


FIGURE 22: Scenario critical infrastructure overview

5.3 Results Comparing DRT to Standard GIS tools for CI analysis

Results of the study were evaluated and analysed in terms of task completion and correctness, time, cognitive load, and preferences for users using CIE and STDGIS.

5.3.1 Correctness: DRT results in more correct analyses

Results for successful task completion and correctness of supporting analysis are shown in FIGURE 23a and in FIGURE 23b. As shown in FIGURE 23a, participants who worked through the scenarios in CIE completed the tasks with correct conclusions based on correct analysis in almost all cases. However, as shown in FIGURE 23b, participants working with STDGIS reached correct answers significantly less frequently. Only 70% and 60% of participants completed Tasks 1 and 2, respectively, successfully achieving the correct analysis using STDGIS. Only 30% of users performing Tasks 3 and 4 with STDGIS had correct analyses. For Tasks 2 and 4 in FIGURE 23b, where the participants are prioritizing among six alternatives, 10% and 50% of the participants respectively did not reach the correct conclusion. Some participants carried over the disablement to the second network, but neglected to cascade the disablement through the second network, whereas some simply lost track of details in the process. Even for straightforward Tasks 1 and 3 (2 initial disablements), when using the STDGIS tools 30% of the participants on Task 1 and 70% of the participants on Task 3 did not have the correct numbers even though they reached the correct overall conclusion (FIGURE 23b).

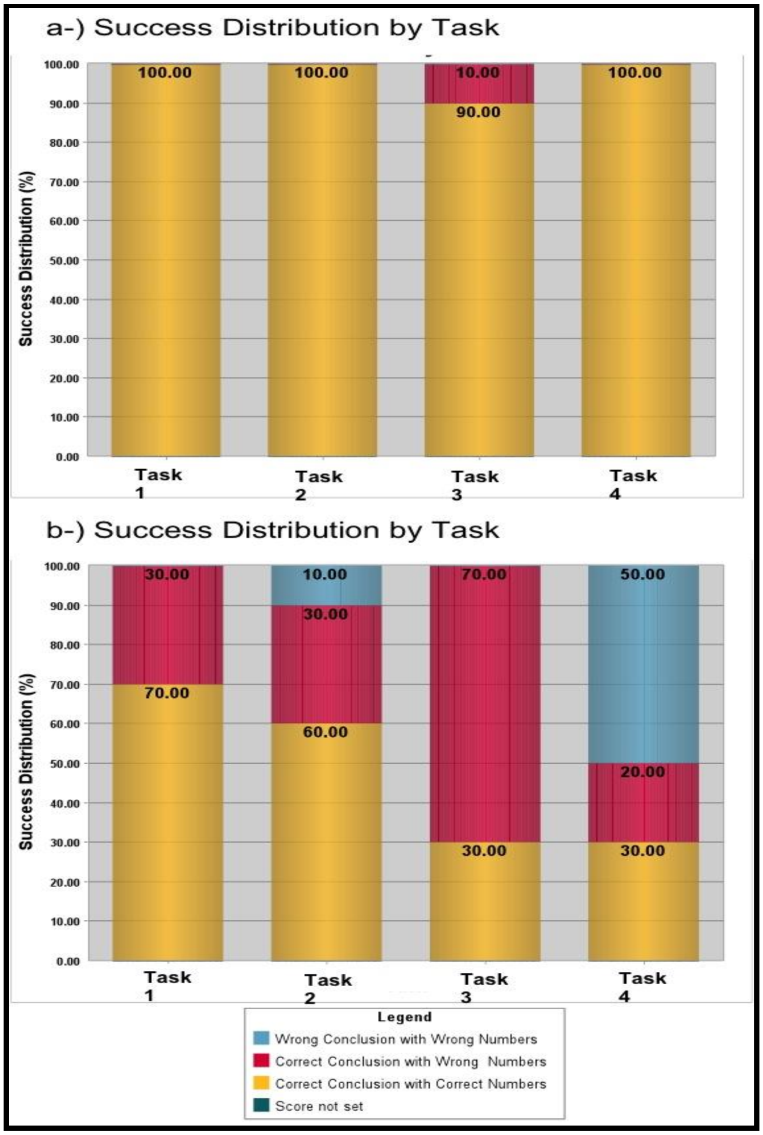


FIGURE 23: Distribution of successful task completion among participants using (a) CIE, (b) STDGIS

5.3.2 Time: DRT takes less time

Hypothesis H1-1 was accepted, that with DRT participants would be able to make decisions in a shorter time compared to standard GIS tools. Results for average task completion time are shown in FIGURE 24 (overall) and FIGURE 25 (only correct conclusions with correct analysis). Tasks were set up with increasing complexity and therefore difficulty. Overall, as the participants progressed through the tasks it took them longer to make the connections and come up with a conclusion. Completion time in Task 2 is lower for CIE. My observations indicate that users spent additional time exploring and familiarizing themselves with the CIE tool upon first use, accounting for additional time spent on a simpler task.

Task completion time increases in direct proportion to complexity for STDGIS tools, while completion time for CIE remains relatively flat. In the most complex scenario, STDGIS takes almost three times as long. This is even more apparent if I use the data from only those who reached the correct conclusion using correct numbers (FIGURE 25). Solutions with STDGIS remain a more manually-driven process where the participants have to pay a lot of attention to the task at hand to produce the correct numbers so that they can base their prioritization decision on correct numbers.

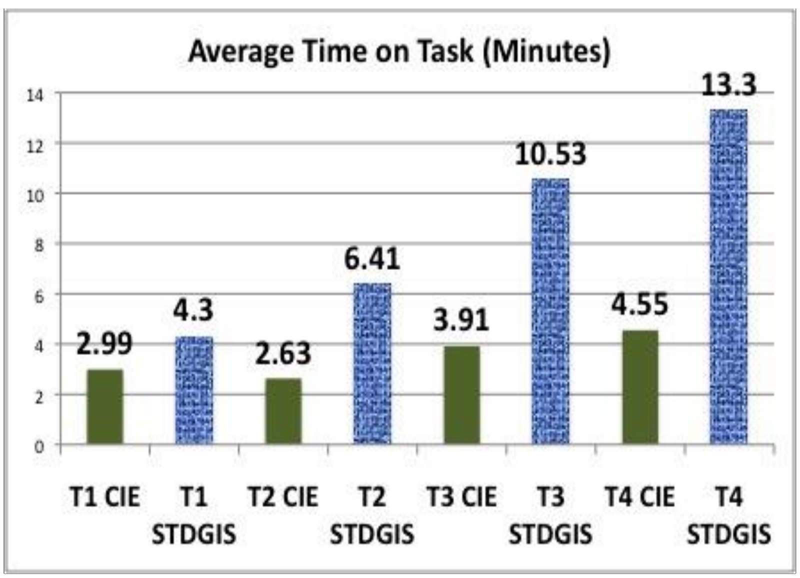


FIGURE 24: Average time spent on each task in minutes including data from all GIS user participants

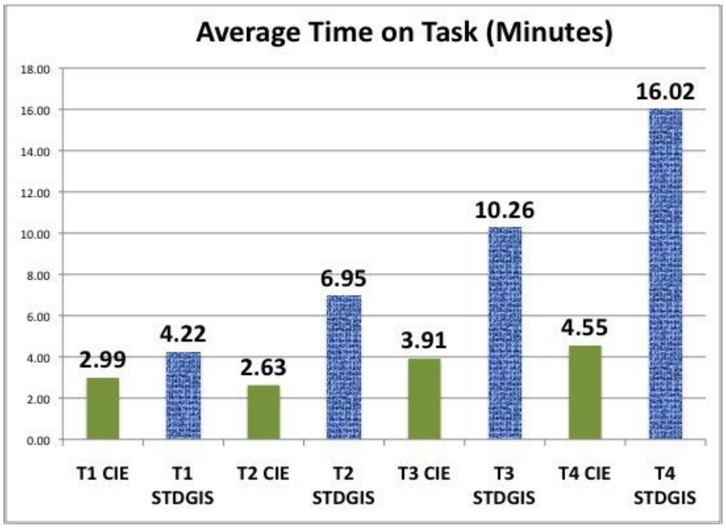


FIGURE 25: Average time on task: Only the participants who reached the correct result using correct numbers

Based on the correctness results participants can be considered to have made better, more well founded decisions using DRT than standard GIS tools. Based on timing results, DRT users complete scenarios accurately in shorter time frames, hence improving efficiency. The combination of shorter time and higher correctness indicate that DRT is more effective than STDGIS for these kinds of scenarios.

5.3.3 Mental demand: DRT takes less cognitive load

Another measure that I employed for evaluation was in terms of mental demand. Hypothesis H1-2 posited that decision makers would make decisions with less cognitive load while making decisions using DRT versus standard GIS tools. Results on the NASA TLX mental task demand evaluation support the hypothesis H1-2 about reduced cognitive load, as shown in TABLE 1. I found a significant difference between the average TLX score using CIE (M=21.26; SD=12.22) and STDGIS (M=65.26; SD=13.21); t (14)= -9.032, p= 0.00. The TLX score of 21.26 for CIE is significantly smaller than the TLX score for STDGIS with a TLX score of 65.26 where they both have similar standard deviations. I can conclude that participants' mental demand was significantly lower with CIE than STDGIS tools. This interpretation is also supported by comments from users' exit interviews where they indicated that they had to concentrate much harder to get to a conclusion using STDGIS tools compared to CIE, and there was much greater room for mistakes and confusion even if they keep their attention at the highest level.

I also found a significant difference between the average TLX scores for STDGIS tools, based on experiment type. Group A has lower TLX score than Group B for STDGIS tools: Group A: M=58.54, SD=13.56, Group B: 72.95, SD=7.94; t(11.49)=-2.547, p=.026. In other words, while evaluating STDGIS tool for mental demand, participants didn't find performing the tasks as equally demanding if they performed the task with STDGIS tools first and then a similar one with CIE. I can interpret this as performing the task first with CIE they experienced a tool that provides easier interaction and better visualization. And therefore doing a similar task with a tool that requires more manual interaction appeared to be taking a greater mental toll, hence the higher TLX score. Moreover, if the participants first performed the tasks with STDGIS tools and then with CIE, they indicated higher mental demand required for CIE. Thus the users seem to be mentally fatigued upon starting to use CIE if they performed the task with STDGIS first.

TABLE 1: Group TLX statistics based on experiment type

	Group	N	Mean	Standard Deviation	Standard Error Mean
TLX: CIE	A	8	23.08	14.69	5.19
	B	7	19.19	9.34	3.53
TLX: STDGIS	A	8	58.54	13.57	4.80
	B	7	72.95	7.94	3.00

5.3.4 Survey results

As a part of the exit survey I asked participants four main questions. One of these questions was related to mental demand. I asked participants to compare the mental demand for each tool and also to comment about the level of mental demand as the task complexity increased. Most participants made comments that support the quantitative analysis discussed previously for the comparison of the tools. This also supports hypothesis H1-2 regarding mental demand. They also pointed out that as the complexity increased with each task, the mental power that they had to exert did not increase linearly. Participants pointed out that mental demand for Task 1 and 2 was almost the same. They did not think that Task 2 was more mentally challenging than Task 1 even though they were required to prioritize six alternatives instead of two. Task 1 and Task 2 scenarios had only one network. Similarly, participants indicated that mental demand for Task 3 and 4 was not significantly different where they had to work with two networks and cross infrastructure effects. It is likely that having to perform the task with two networks and two initial disablements first helped them learn the process and therefore it was not significantly harder to prioritize six alternatives later.

The exit survey included questions comparing the effectiveness and the ease of use of the tools. All participants felt that CIE was more effective for these kinds of tasks and was easier to use. When asked if they had any problems completing any of the tasks, several participants pointed out the difficulty of cascading down the disablements and especially cascading cross infrastructure disablements using ArcGIS tools. All the participants (100%) stated that they prefer the CIE compared to standard GIS tools; this supports hypothesis H1-3, which says given the same scenario, decision makers prefer using DRT over commonly-used GIS tools.

5.4. User study conclusions

Results for user task efficiency, task completion, and cognitive load consistently support my hypothesis: given the same scenario, decision makers can make better decisions (H1-4) with less time (H1-1) and less cognitive load (H1-2), and prefer (H1-4) using CIE. Applying my approach for CI reconstitution, users successfully completed more scenarios more accurately (FIGURE 23 a and b), in less time **(**FIGURE 24**,** FIGURE 25)**,** and with lower cognitive load (TABLE 1). Overall, I believe that such approaches are essential to address the information overload problem in complex, multi-dimensional analysis for CI in general and reconstitution efforts in particular. Results from this user study provide a baseline for my investigation of recommender-based geovisualization tools for CI decision support.

Important limitations to this study are that the CIE takes into account only a static emergency situation in a single type of CI network. Tools with instantaneous reconstitution data would be much more accurate, but would also be much more complex. It is possible that the results of this study would be different in these situations. However, I believe that the complexity of changing scenarios makes it even more important to build targeted tools to support emergency reconstitution decision-making.

The results from this study show that interactive geovisualization tools to support CI analysis, particularly in cross-infrastructure scenarios, can play an important role in making more accurate and timely decisions for deploying resources in disaster recovery. However, this is only one aspect needed to support decision makers.

During an outage, cross infrastructure cascade effects are determined by the extent of the service areas for service distribution sources. While working on the first study, we saw that even small changes in the extent of the approximated service area can make a big difference in the outcome. Adding one critically placed secondary network infrastructure element to the outage can introduce a cascade in a secondary network that might cover a large area. Hence, grounding cross-infrastructure impact analysis relies on accurate service area estimation for the critical infrastructure networks.

Because of the great responsibility that emergency managers have, these users have particular interest in using support tools that employ the most appropriate and accurate service area estimation algorithms. To that end, the following two studies were conducted to understand the nature of performance of algorithms for service area estimation for power and water.

CHAPTER 6: ACCURACY ASSESSMENT OF SERVICE AREA APPROXIMATION ALGORITHMS FOR CRITICAL INFRASTRUCTURE RECOVERY

The critical infrastructure and geographic information analysis experts in the CIE user study emphasized the need for highly accurate estimates on infrastructure knowledge. And decision-makers in general have a need to understand the accuracy of existing methodologies for critical infrastructure service area estimation and developing new approaches to improve existing techniques (J. W. Fenwick and Dowell, 1999; K. Newton and Schirmer, 1997; Sulewski, 2013). Several types of estimation methods are prevalent in practice, but surprisingly little information is available on their comparative effectiveness. This led to the second research study (Pala et al., 2014), investigating my second specific research question – (RQ2) What are the differences in effectiveness among various service area estimation techniques during an emergency operation for CI enablement scenarios?

It is important to understand the comparative merits of the estimating approaches used to support decision-makers as they develop mitigation and remediation strategies after a damaging event. Without knowledge of the accuracies of service area estimates, decision makers will not be able to trust decision-support systems, and will therefore rely on commonly used systems that may not support efficient or correct decisions.

In the previous chapter, one of the main reasons that some study participants did not come to the correct conclusion was because they missed the connection of a cross infrastructure cascading outage to one specific node. This node ends up cascading down the network to affect a large area, which is very close to the service area of the one of the disabled nodes in the first network. Thus a small change in the approximated service area can have a profound effect on the outcome of the scenario. The previous study employed Thiessen polygons to approximate the service areas as the enabling knowledge model for infrastructure interactions, and I wanted to investigate how the outcomes of the scenarios might change if another approximation method were used. My personal experience and judgment combined with information from system experts led me to test the accuracies of existing algorithms that are commonly employed in CI analysis.

In order to compare estimation approaches, it is necessary to employ well understood methods for accuracy assessment. Here, I have adapted accuracy assessments from land cover classification (Congalton and Green, 2008) to the domain of CI analysis. These metrics are applied to compare a representative set of service area approximation algorithms used in critical infrastructure recovery analysis. In this chapter, I focus on service area approximation using distance-based Thiessen polygon estimation, described in section 3.5.3 Thiessen polygons, and cell-based approaches, described in section 3.5.4 Cellular Automata.

In this study, I assess the accuracy of four different methods commonly used for estimating infrastructure impacts after a disruptive event. The methods I evaluate include distance-based Thiessen (Voronoi) polygons with weights (WTP) and without (TP), and cell-based cellular automata with weights (WCA) and without (CA). The term “impact” refers to the inability of a utility to provide a service, such as power or gas, due to infrastructure system damage. I focus on two types of impacts: (1) aggregate impacts, such as economic activity and population contained by the outage, and (2) point data impacts, such as whether a specific asset is included in an outage.

I compare the methods for overall accuracy with a reference model of a power distribution network for a Midwestern, mid-size U.S. city that does not have unusual geographic elements that might disrupt analysis of service area estimation. This work was performed in consultation with an electric power (EP) subject matter expert with a high level of familiarity with the nuances of the system developed reference service areas.

This chapter is organized as follows: In the next section, I discuss the literature related to service area estimation methods. Then, I describe the methods used to create distance- and cell-based service area estimates and assess their accuracy. Finally, I discuss my results, showing that weighted methods outperform their standard counterparts, confirming hypothesis H2-2, but also showing that cell-based methods do not always outperform distance-based ones (so H2-1 cannot be accepted). However, I discuss some possible limitations of the study that suggest that tuning may be important for cell-based methods may be particularly important for accuracy.

6.1 Background

Power, gas, water, and other infrastructure system assets (e.g., power substations) serve customers in a geographical area. These regions are termed service areas. Even though utilities have detailed information about specific distribution source-sink relationships between their assets, this information is not designed or organized to facilitate large-scale analyses, nor is it documented by public regulatory agencies. In addition, these data are often considered sensitive or proprietary. In the absence of data, defining the service areas accurately has long been a problem, but estimating these boundaries accurately is very important in disaster recovery situations (Castongia, 2006; J. W. Fenwick and Dowell, 1999; K. Newton and Schirmer, 1997; Sulewski, 2013). During emergencies, external decision-makers often rely on estimates of service areas produced by various methods. Typically, a geographic boundary for each serving point is defined to estimate the source-sink relationships between the serving network entities ("sources") and the entities using those services ("sinks"). Increased estimate accuracy could lead to a more efficient recovery. It is important to understand the comparative merits of the estimating approaches to support decision-makers as they develop mitigation and remediation strategies after a damaging event. In this research, I focus on approaches commonly employed as part of CI analysis – Voronoi estimation and CA approaches.

6.1.1 Thiessen Polygons (Voronoi Diagrams)

Thiessen polygons, also known as Voronoi diagrams, have been used in different ways to present and analyze data. The success of this method originates from its ability to uniformly and systematically partition an area. A Voronoi diagram divides the plane according to a nearest-neighbor rule when a number of points are provided and each point is associated with the region of the plane closest to it (Aurenhammer, 1991). This method draws a straight line between all of the points; on each line's mid-point, a perpendicular line is drawn to create the boundaries representing the point. Thiessen polygons take shape when perpendicular lines are trimmed at intersections with other lines (FIGURE 3). Work by Okabe (2000) and Okabe et al. (1992) provides detailed discussions on the concept of Thiessen diagrams from both historical and geometric viewpoints. From a critical infrastructure perspective, the literature includes papers (Akabane et al., 2002; A.

Okabe, 2000; Tolone et al., 2009; G Loren Toole and McCown, 2008) that detail efforts for using this approach to create critical infrastructure service boundaries.

Standard Thiessen method assumes that the dataset is homogeneous. This is generally not the case because each source point provides varying degrees of service. For example, different water treatment plants have different daily outputs. Using a weighting approach based on source points might enhance Voronoi-based approaches. This approach creates Thiessen polygons by calculating weighted Euclidean distances (Dong, 2008). For weighted Thiessen polygons, the critical infrastructure elements with smaller outputs are assigned smaller service areas. In practice, this approach is potentially more realistic than Thiessen polygons with equal weighting. For more information on Thiessen polygons please see the background section "3.5.3 Thiessen polygons (Voronoi diagrams)." FIGURE 3 and FIGURE 4 provide graphical representation of the Thiessen polygons.

6.1.2 Cellular Automata

Discrete computational systems that are composed of a finite or enumerable set of homogeneous, simple cells as a part of a spatially and temporally discrete grid structure are called cellular automata (Berto and Tagliabue, 2012, 2012). Often, CA are explained as mathematical models for complex natural systems that contain large numbers of simple identical components with local interactions (Wolfram, 1994).

It is also possible to estimate service areas using CA (J. W. Fenwick and Dowell, 1999). Although CA is applied to a wide variety of fields, CA techniques were not used for service area calculations until the last decade (J. W. Fenwick and Dowell, 1999; S. P. Linger and Wolinsky, 2001; Werley, 2002). Similar to the Thiessen polygons, CA algorithms are also run with equal weights or weights based on the actual load values of the substations. Tools that use CA-based approaches to estimate service and outage areas include the Interdependency Environment for Infrastructure Simulation Systems (IEISS) or TranSims (Bush, 2005; J. W. Fenwick and Dowell, 1999; G. L. Toole et al., 2001; G. Loren Toole et al., 2008; Werley, 2002), and Water Infrastructure Simulation Environment (T. McPherson and S. Burian, 2005; D. Visarraga et al., 2005).

Detailed background information about the cellular automata technique and examples of its applications can be found in the background section “3.5.4 Cellular Automata.”

6.2 Methodology

Creating the polygon datasets using each approximation method and evaluation of the accuracy assessment of these polygon layers are the main tasks of this study. I hypothesize that cell-based methods will outperform distance-based, and that weighted methods will outperform non-weighted ones.

Hypotheses for Research Question RQ2:

H2-1. Cell-based Service Area (SA) estimation techniques produce more accurate results compared to distance-based ones.

H2-2. Weighted SA estimation techniques produce more accurate results compared to their non-weighted counterparts.

In this chapter I compare four existing service area calculation methods for power network: (1) Thiessen Polygons (TP), (2) TP with weights from electric power substation load (WTP), (3) Cellular Automata (CA), and (4) CA with weights (WCA).

The experimental setting uses the CI network data of an Electric Power (EP) network for a regularly laid-out, Midwestern mid-size city in the US with around 150 substations. The data set includes the distribution network, the substations, and demand for power. It also includes the service areas for the substations. The service areas are polygons delineated by an EP system expert to reflect the service area for each substation. Finally, the data set also includes economic and population information. The population and economic impact data is driven from the 2010 LandScan data which was developed by Oak Ridge National Laboratory with funding by the Department of Defense (Bright et al., 2012). I also utilized LANL Daytime/Nighttime population information (ESRI, 2014; Tolone et al., 2007).

I used standard ArcGIS tools to implement the ordinary Thiessen polygons (TP). I also created the WTP using a publicly available ArcGIS extension, as described by Dong (2008). To create the CA and WCA polygons, I used functionality embedded in a desktop software tool, the Interdependent Energy Infrastructure Simulation System (IEISS) created by LANL (Bush, 2005; G Loren Toole and McCown, 2008). To create CA and WCA polygons using IEISS, I created two versions of XML data based on the power infrastructure layers. This XML file contains locations of all the substations, lines connecting them, study area boundary coordinates and settings related to various functionality in IEISS, including the service area estimation. I left the original power output numbers in the XML file. However, in order to create the non-weighted CA, I inserted a high power output number that is the same in all the power substations. This ensured that the power output numbers would not affect the algorithm and create regular (non-weighted) CA service areas. The algorithm starts growing cells using a raster format starting from each source point, in this case, EP substations, and grows them until it runs out of space or EP resource. Next, during the export process, the data set is converted to vector format. Than the accuracy of each service area calculation is assessed through comparison with the reference data set described earlier.

I use two different approaches to perform the accuracy assessment of the results: aggregate statistical accuracy analyses and spatial accuracy analyses. One contribution of this research is the adaptation of impact analysis from remote imagery classification accuracy assessment for use in the CI domain.

6.2.1 Aggregated Impacts

Aggregated impacts are used in situations where coarse information about the service areas is required. Examples include total population, total economic activity, or total area. In these situations, error in the spatial extent of the service area is acceptable, as long as the extent of the area generates the correct values. To generate these comparisons, I first calculate the daytime population for each polygon associated with substations, using each of the four methods. I then compare the results from each method to the actual population associated with the substation in the reference model. In these results, the method that has the smallest error is considered the better performing approach. This process is repeated for nighttime population, economic activity indicators, and total area. The economic activity indicators include direct, indirect, and induced economic impacts, as well as economic impact on business and employment. Similarly, the smallest difference between the numbers produced by the reference dataset and the method is considered the best approach. Direct economic impacts are based on the types of businesses in each area. Indirect economic impacts are derived from suppliers of commodities in a service area. Induced economic impacts are caused by the reductions in factor incomes in each service area. Economic impact on businesses and employment considers the overall effect on known businesses and approximated employment (Ewers, 2008).

6.2.2 Point Data Impacts

For other types of analysis, the spatial accuracy of a service area is important. It is often important to determine if an infrastructure outage impacts other infrastructure assets. For example, an asset that depends on electric power from a certain substation may be unable to function if the substation is out of service. For this analysis, I calculate the spatial agreement between the reference service areas and the calculated service areas. I evaluated spatial accuracy using a point accuracy test. For this metric, I create 10,000 points, randomly located within the study area. I overlay the service areas created by each method on the random points (FIGURE 26, FIGURE 27, FIGURE 28, and FIGURE 29). The point analysis assesses the accuracy of matching critical facilities with their corresponding service source point through service areas. This type of analysis is widely used in land cover classification accuracy assessments studies (Congalton, 1991). For each randomly placed point, I determine which service area the point belongs to in the reference model, as well as for the TP, WTP, CA, and WCA approaches. Using this information, I can calculate overall accuracy for evaluating the approaches, as shown in the process flowchart of FIGURE 30.

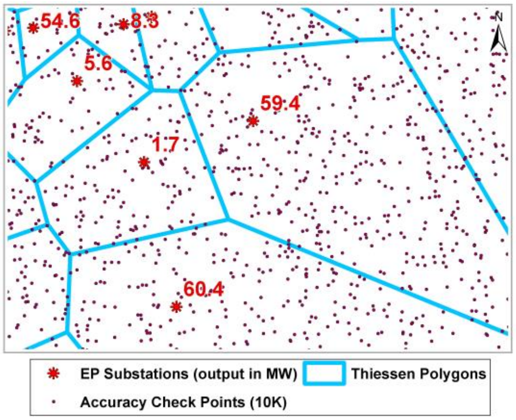


FIGURE 26: Point layer (10K) overlaid with standard thiessen polygon layer

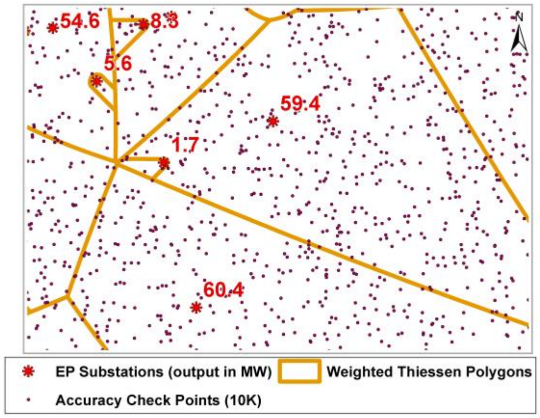


FIGURE 27: Point layer (10K) overlaid with weighted thiessen polygon layer

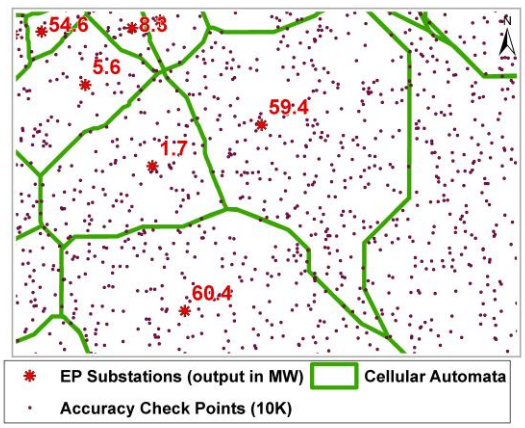


FIGURE 28: Point layer (10K) overlaid with standard CA polygon layer

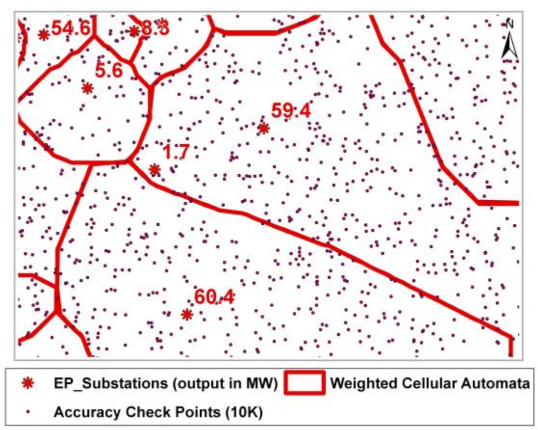


FIGURE 29: Point layer (10K) overlaid with weighted CA polygon layer

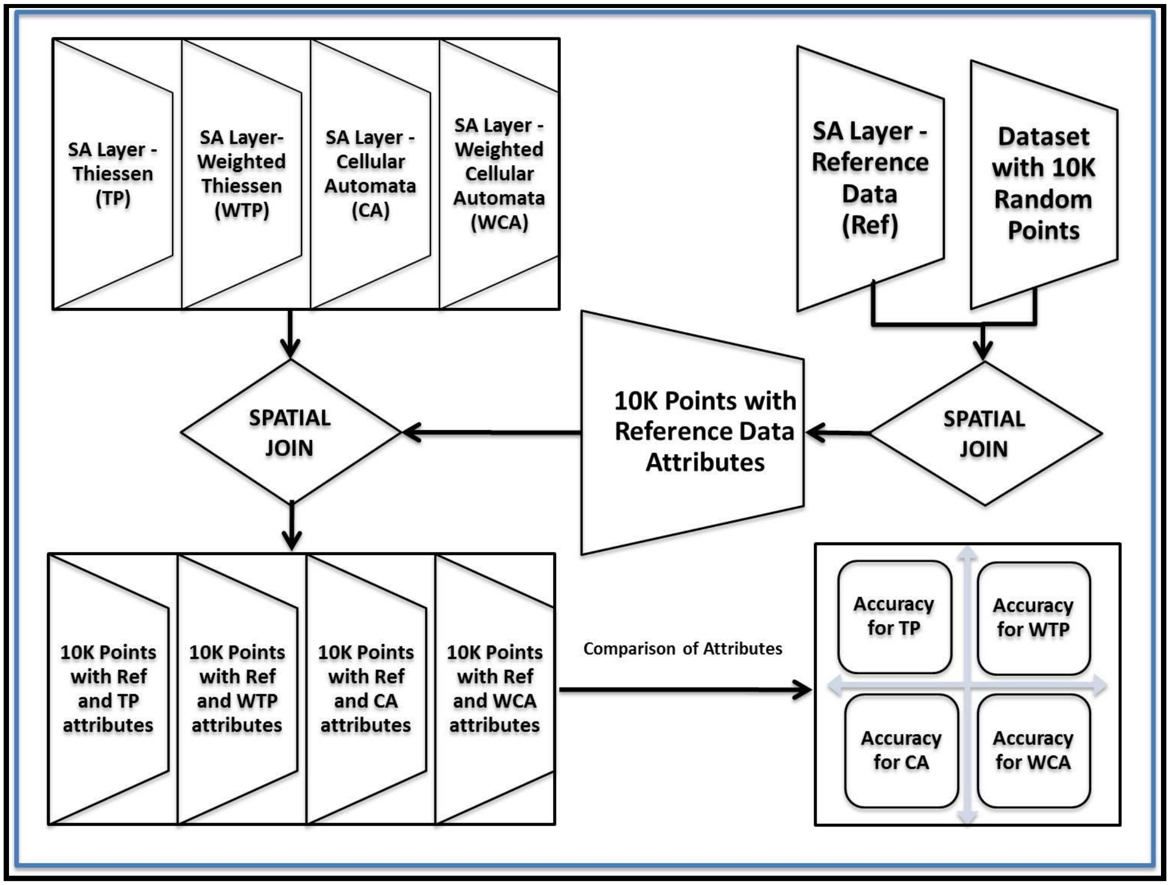


FIGURE 30: Point accuracy assessment data preparation flowchart

Decision makers need to trust decision support tools to provide them with accurate information. However, as much of the information for critical infrastructure is estimated, it is important to understand the strengths and limitations of the methods used. In particular, here I also performed proximity confidence analyses to estimate if the proximity to the source point in a polygon affects the accuracy of the estimation. To do so, I measure the distance between each of these points and the substation, defined as the serving point by the reference dataset. To classify the points uniformly based on their proximity to the serving source point, I normalize these distances based on the size of the service area polygon that overlays the point for each method. This approach allows a decision-maker to quantify the quality of the results based on where a point is located within a service area. Specific facilities that are closer to the service source facility (i.e., EP substation) naturally have higher confidence values then those that are further from the service source facility.

As an example, consider two hospitals as point data. The first one is 500 yards from EP substation A. The second hospital is 2 miles away from a substation B. If the service area sizes are the same for both substations, it is reasonable to compare the hospital to EP substation distances and to calculate my confidence that the hospitals are correctly associated with substations. However, if substation A's service area is much smaller than substation B's, distances must be normalized. Below I formulate the normalization and point classification based on distance to the source:

Let *P(s)* be the service area polygon of serving point *s*

Let *A* be the area of *P*

Let *r* be the radius of a circle that has the same area *A* as the service area polygon *P(s)*

Let *i* be a randomly placed point within the agreement zone (the agreement zone is where the reference data polygon and the polygon produced by the service area estimation method overlap)

Let *d* be the distance between *i* and *s*. *d* is normalized and classified as follows:

If $d < (r/4)$ then *i* is classified in Proximity Class #1 (closest 25%)

If $(r/4) < d < (r/2)$ then *i* is classified in Proximity Class #2 (25%-50%)

If $(r/2) < d < (3r/4)$ then *i* is classified in Proximity Class #3 (50%-75%)

If $d > (3r/4)$ then *i* is classified in Proximity Class #4 (furthest 25%)

Based on the normalized distances, I can classify the points and use these classifications to measure the effect proximity has on the accuracy of point data and quantify confidence in a method when reference data is unavailable. I also investigate the methods' ability to estimate the accuracy of determining the source sink-relationships if the neighboring polygons are also considered as a part of the equation. To achieve this, I created a lookup table for each method. This lookup table lists all the existing service area polygons for all the methods as well as the neighboring polygons for each of those. Using this lookup table, I was able to recalculate the point accuracy values. For each point that was misplaced on the reference dataset, I determine if the point was correctly associated with a neighboring polygon. Thus, the accuracy of determining a correct source-sink relationship for critical point locations can be tested by considering the neighboring service area polygons.

6.3 Experimental Results

In this section, I describe the performance of the four approaches. For the weighted methods, I use peak electric power consumption in megawatts (MW) for the weights. FIGURE 31 shows examples of some of the results; it contains four screenshots. Each example displays one service area creation method together with the reference dataset. The Figure also shows comparisons between the methods and the reference set. The top left Figure (A) is comparison of reference set and the TP approach. The top right image (B) in FIGURE 31 shows a comparison of WTP with the reference dataset. The bottom left Figure (C) shows the polygons produced by CA method and bottom right image (D) shows the WCA overlaid with the reference dataset.

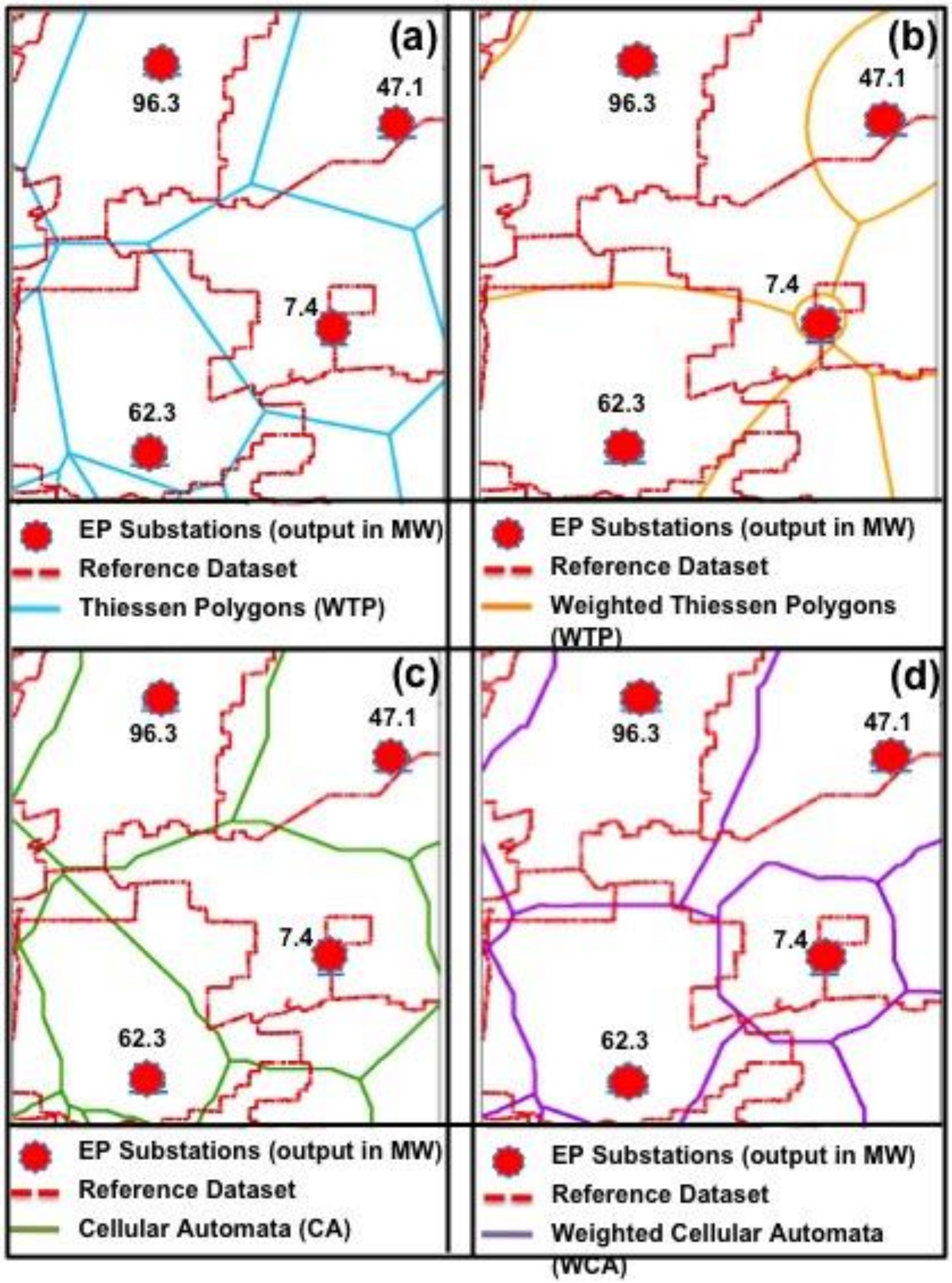


FIGURE 31: Service area polygon examples: (A) TP, (B) WTP, (C) CA, (D) WCA.

Using aggregate statistical accuracy analyses, I compare the area, population, and various economic indicators with results of the reference service areas. FIGURE 32 shows the mean difference in daytime and nighttime population between the calculated and reference service areas. A smaller value indicates a better result because it shows that the population number of the method is closer to the population number produced for the reference service area. In both population types, WCA produced the best result with a minimum difference when compared to the reference data. CA produced results with the highest difference. The cumulative sum of differences in population analysis yields similar results (FIGURE 33). As the difference between the results from WCA and reference set was the smallest, I conclude that the WCA method performed best out of the four methods. The WTP method showed slightly better results than the regular Thiessen and CA methods.

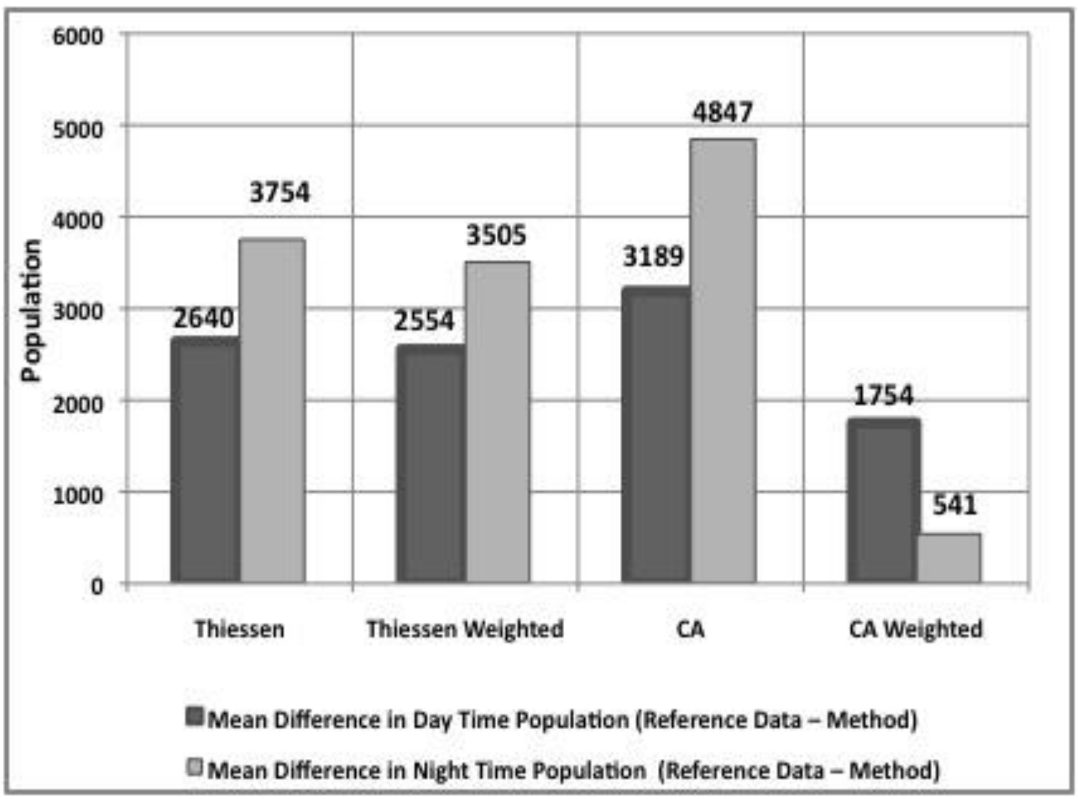


FIGURE 32: Mean difference in population

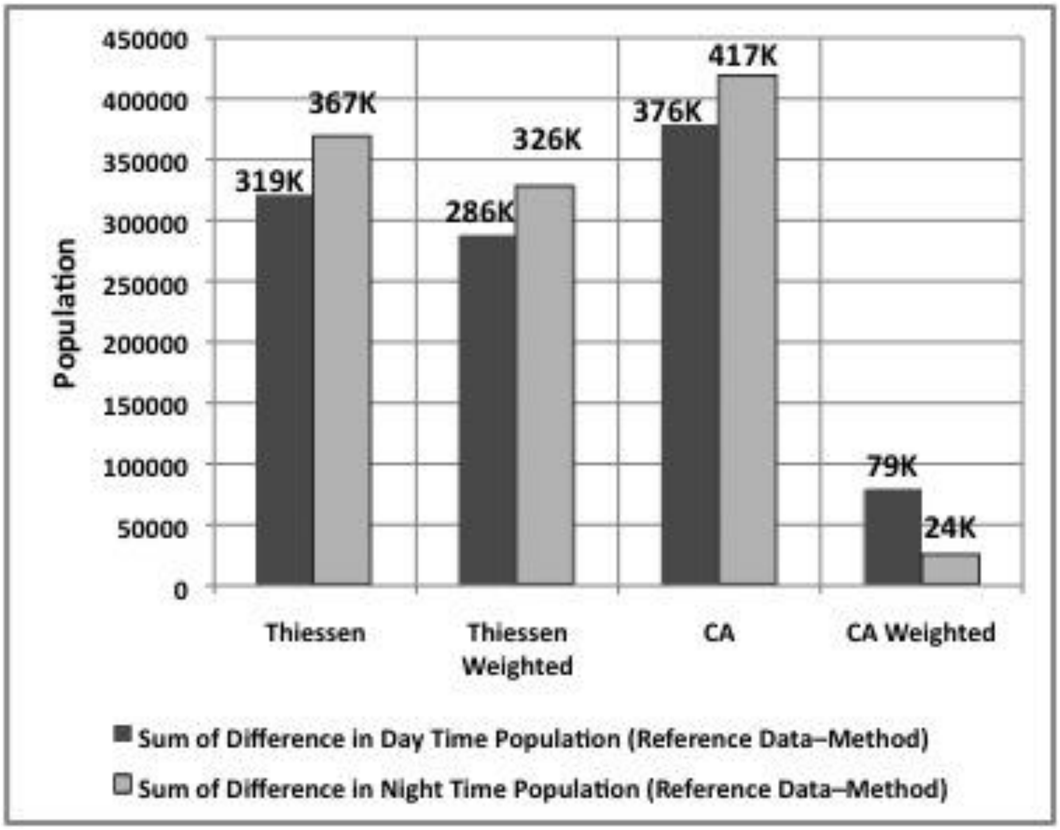


FIGURE 33: Sum of differences in population

FIGURE 34 and FIGURE 35 show the mean of the difference in economic impact for various types of economic impact metrics, such as direct, indirect, induced, employment, and business. In all of these economic impacts, the difference is smallest in WCA, second smallest is CA, and third smallest is the WTP. The only exception to this is economic impact on business (Figure 36), where WTP and CA swap places. The largest mean difference occurred in TP. Although the mean difference in WTP is larger than CA (with the exception of economic impact on business), the differences are not as notable as the differences in the other categories. FIGURE 36 and FIGURE 37 show the summation of the differences; these results follow a trend similar to mean differences. These results indicate that decision support systems should not uniformly adopt one estimation method over another, but must consider the spatial qualities of the data and how these relate to the chosen methods. In this instance, it is clear that weighted methods are preferred, but that the nature of the most important impact has bearing on the method to choose.

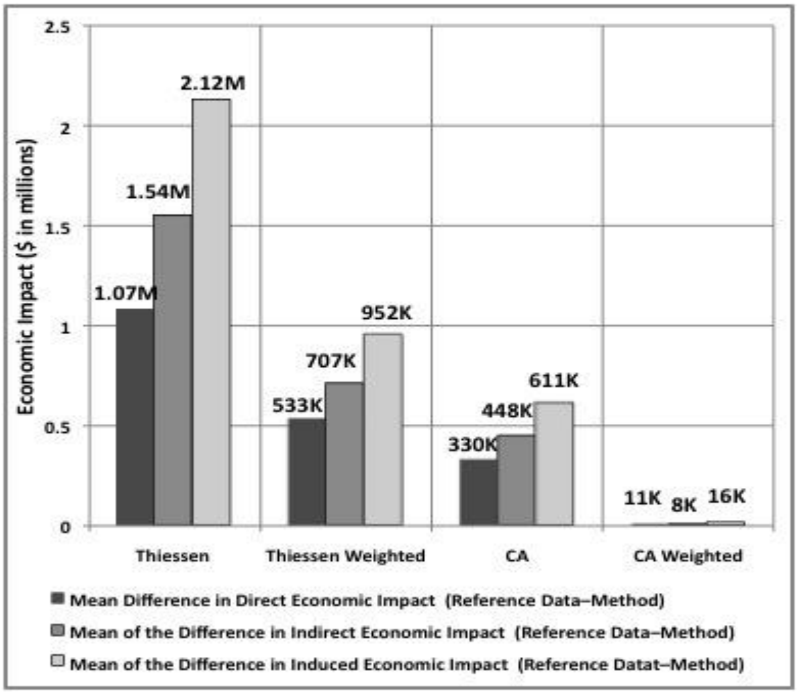


FIGURE 34: Mean of the difference in economic impact (direct, indirect and induced)

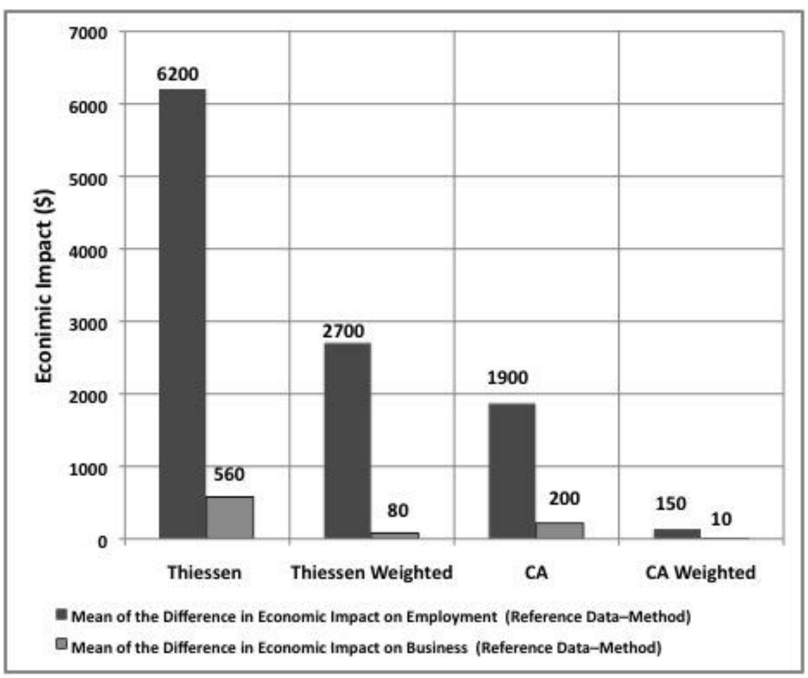


FIGURE 35: Mean of the difference in economic impact on employment and business

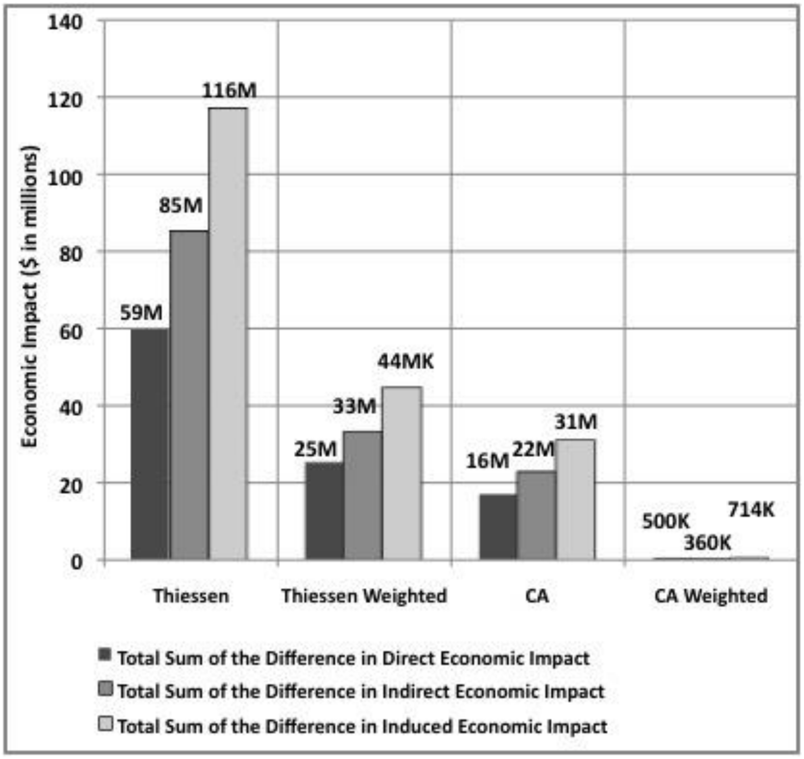


FIGURE 36: Total sum of difference in economic impact (direct, indirect and induced)

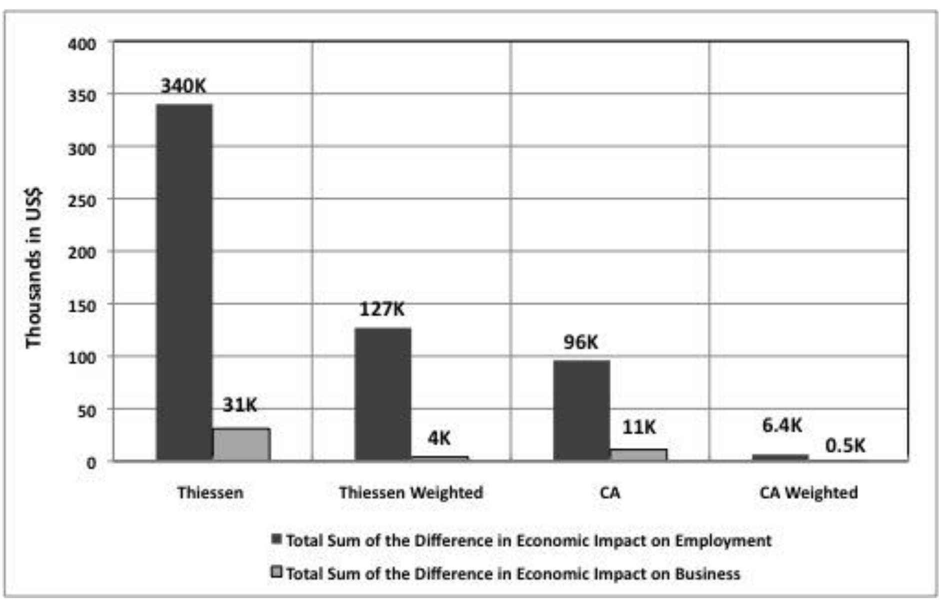


FIGURE 37: Total sum of difference in economic impact on employment and business

My final aggregate statistical comparison considers the total surface area of the polygons. The results for surface area comparisons indicate that the average reference polygons' area is 1,033 square acres. As shown in FIGURE 38, WCA and TP are closest in size, on average, to the average of the reference polygon sizes. The CA algorithm is the least accurate approximation for this metric.

For the point accuracy analysis, I place 10,000 random points across the study area and I calculate the overlay agreement accuracy by applying a series of spatial overlay processes (FIGURE 30). FIGURE 39 shows that WTP performed the best overall with 68.9%, followed by WCA at 59.5%, TP at 54.1%, and trailed by CA at 52.3% overall accuracy. These results are nuanced, as seen in FIGURE 40. Here, WCA has the highest point accuracy (91%) when I consider the points in closest 25% area of each polygon, followed by WTP with 86%, CA with 85%, and TP with 81%. The further from the source point, the lower the accuracy of the non-weighted methods (TP and CA). Accuracy of the weighted methods (WTP and WCA) decreases considerably with distance, but they are considerably higher than the non-weighted methods.

It is important to note that the accuracy of all approaches improves dramatically when neighboring polygons are included. Instead of assigning a point to a single polygon, I instead assign a point to a single polygon and any neighboring polygon. This relaxes the analysis to indicate that a point is associated with one from a set of source facilities. These results are shown in FIGURE 41, where points are correctly assigned to a set of source facilities over 95% of the time for all methods.

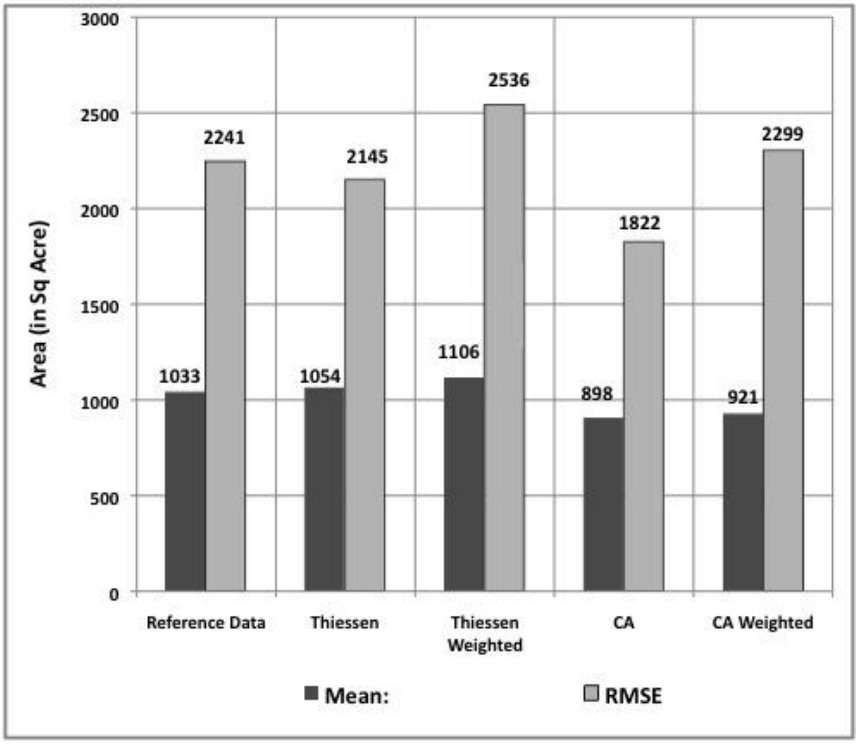


FIGURE 38: Average service area polygon size

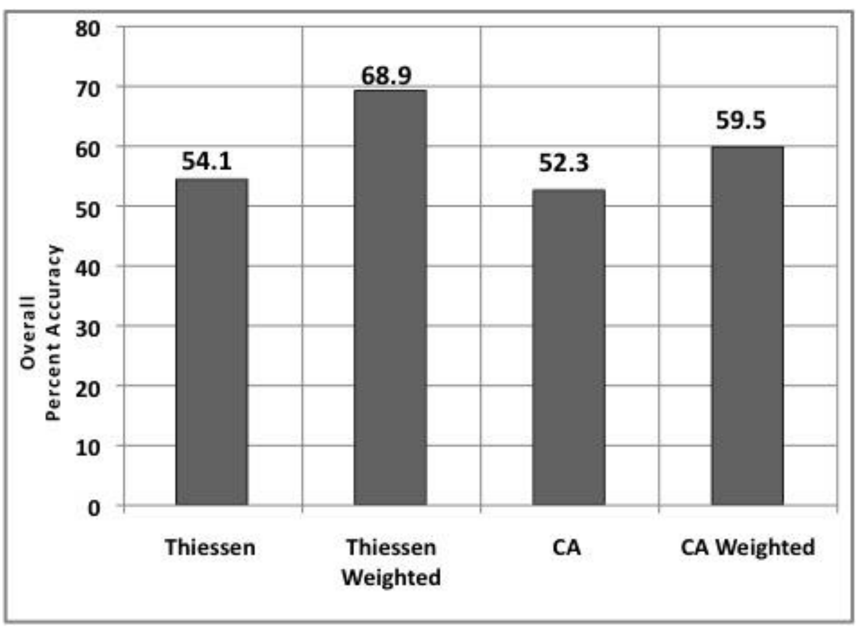


FIGURE 39: Overall accuracy through point analysis (% accuracy)

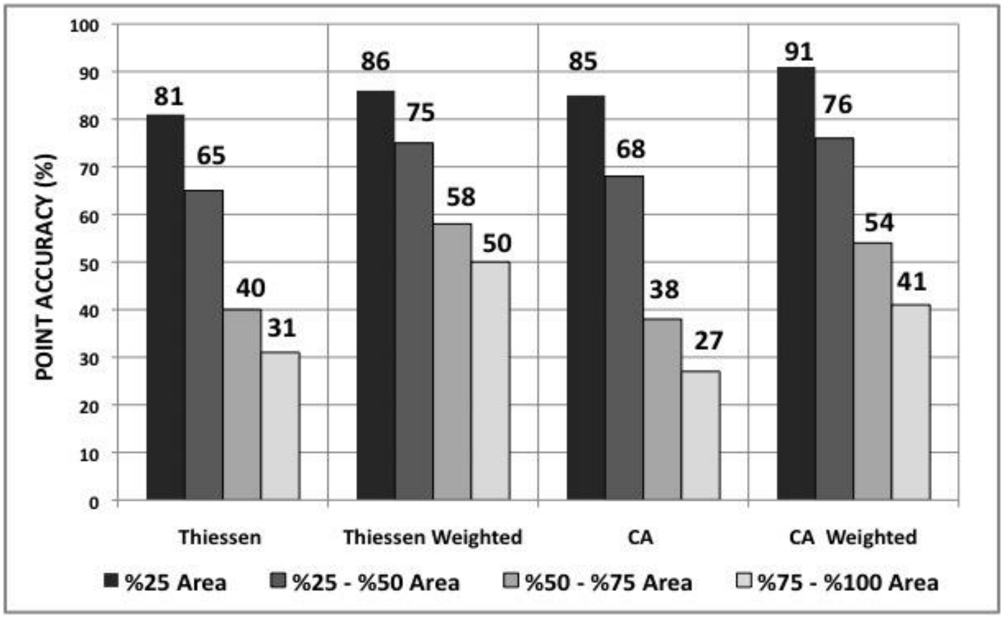


FIGURE 40: Proximity confidence analysis results (% accuracy)

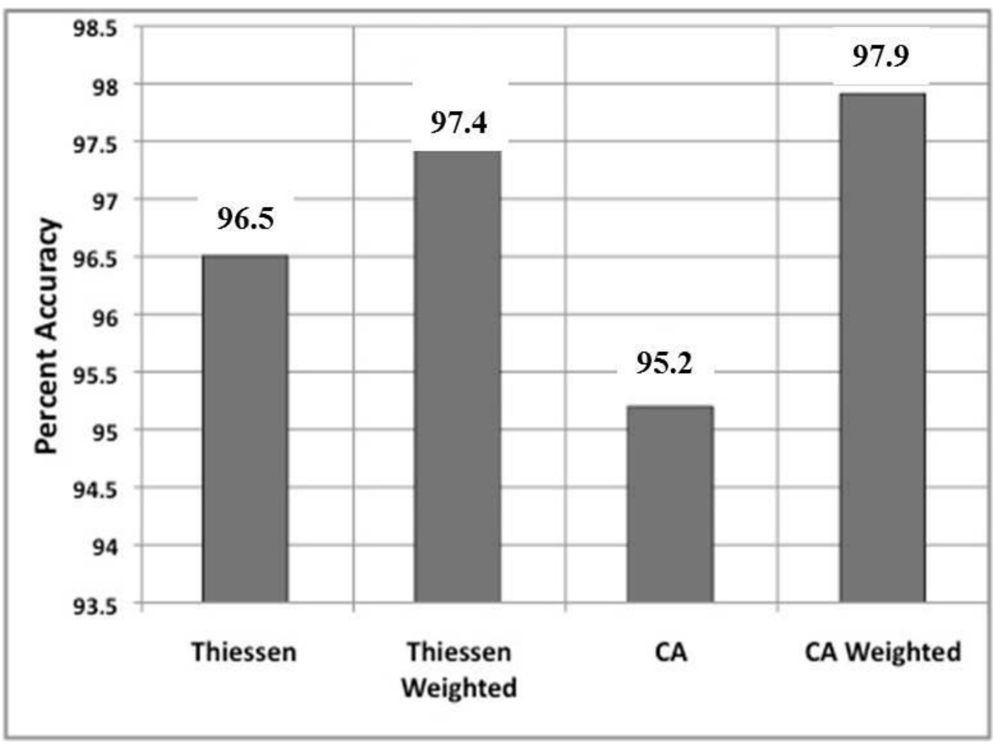


FIGURE 41: Point accuracy analysis based on polygon neighborhood relaxation

6.4 Power Network Accuracy Discussion and Conclusion

In this chapter, I have studied the accuracy of commonly used techniques for service area estimation in CI (Thiessen polygons and Cellular Automata) on a mid-size Midwestern US city with no unusual geographic qualities to ensure transfer to other cities. I tested two hypotheses. The first hypothesis H2-1 suggests that the cell-based methods work better than distance-based methods and the second hypothesis H2-2 suggests that the weighted version of these methods would work better than their non-weighted counterparts. To test these I implemented two types of accuracy assessment; namely Aggregate Statistic Accuracy Analyses and Spatial Accuracy Analyses. Through aggregate statistic accuracy analyses I have tested the differences in the population and economic indicator numbers caused by the difference between four estimation methods (TP, WTP, CA, WCA) and the reference data set. Mean difference and total sum of difference in daytime and nighttime population graphs show that WCA numbers are closest to the reference dataset numbers, with minimal difference. It is followed by the other weighted method WTP and the non-weighted methods CA and TP in accuracy. This result is similar in economic impact numbers, except that WTP and CA change places for some of the measurements. This also holds true for the rest of my analyses as WCA and WTP perform better than CA and TP in load population – Landscan population difference, average service area polygon size estimate and point accuracy analysis. This result supports hypothesis H2-2 – that weighted methods perform better than their non-weighted counterparts. However results are more complicated when testing H2-1 on the relative accuracy of cell-based approaches compared to distance-based methods.

My original hypothesis H2-1 held that cell-based approaches would perform more accurately than their distance-based counterparts. Results from mean and total sum of difference in day and night time populations, point accuracy analysis and some of the economic impact analysis show that this is not always true. Therefore I reject hypothesis H2-1 regarding the cell-based approaches performing unilaterally better. There are two main reasons for the performance differences between CA and TP methods. First, TP methods seem to be well-suited for uniform distribution of consumption, while CA methods seem better suited for denser areas of consumption around source stations with less usage between stations. This intuitively makes sense, since TP methods are partitioning the space under consideration into a grid, while CA approaches are building ever-increasing disks around sources. Second, CA approaches seem to suffer more from anomalies in resource allocation on the border of the region, as discussed in more detail below.

The results showing that cell-based approaches are not always more accurate than Thiessen polygons is somewhat surprising as the cell-based approaches are more sophisticated; this is especially surprising for the point accuracy analysis. I investigated possible reasons for this unexpected result. Visual inspection for the WCA polygons compared to polygons from the reference (ground truth) set provided some insight into the point accuracy result.

There are two visible occurrences that lead to the lower accuracy of WCA polygons in point accuracy analyses. The first occurrence happens mostly on the WCA polygons at the outer edge of the study area. As all the source points start growing cells and the space starts the run out, the cells from inner sources start taking up all the available space. Outer sources with larger power outputs run out of space and start channeling their growing cells further outwards. This phenomenon results in a few unexpected and unrealistic cell growths that contribute to the lower accuracy results on CA methods. This suggests that performing CA estimation on a larger area then clipping down to a focus area for study might be a better process for CA approaches than the constraint that all source output be absorbed only in the region being considered.

The second anomaly that I discovered in the dataset is that there are a few cases where the ratio of power output number for a specific substation to the total service area in the ground truth dataset is too large. This indicates that some of the power is provided to an industrial complex. The dataset that I employed does not indicate the type of power output so it is possible that the dataset includes some substations that are dedicated for industrial purposes. Inclusion of these substations with large outputs and small area coverage on the reference dataset also contribute to the unexpected results. As with all of the methods, it may also be possible to tune the CA parameters for better performance on the specific dataset.

The results of the confidence proximity analysis suggest that, as the distance between the point sources increases, the rate of decrease in point-impact accuracy is higher in polygons produced by non-weighted methods. I also found that, for cell-based methods, the closer to the source point, the better the point-impact accuracy. This suggests that, although the WTP method was more accurate overall for point-impact accuracy, WCA is more accurate in close proximity to the source. Therefore, when choosing among service area estimation methods and accuracy measures to use, it is important to consider where the highest accuracy is needed. For example, if it is important to have the highest point-impact accuracy in high-density locations, near source points, then Weighted Cellular Automata would be the best choice. On the other hand, in situations where the aggregate point-impact accuracy is more important, then weighted distance-based methods should be more accurate.

Through this study, I provide an empirical evaluation of service area estimation techniques, showing weighted methods are preferred, but that a more uniform distribution of source and demand points limits applicability of cell-based methods. There are several limitations to the generalizability of these results. One limitation is that the weighted methods only took into account the capacity of power substations, but not differentiated demand among the population – there could be impacts based on residential, industrial, and commercial zoning, such as a data center or manufacturing.

The city considered was a mid-size Midwestern city with no irregular geographic features, making the results generalizable to power networks in similar cities. However, this study only studied service area estimation for power, so more studies are needed to understand potential differences among CI types. The next chapter investigates a water network, but not in the same city, because of data access limitations.

Another limitation of this study for its generalization is that many cities may not have reference data to compare with the estimations. Therefore, it may be necessary to develop other metrics to assess accuracy. For example, each substation has MW load associated with it. This load could be compared with the expected consumption of population and businesses to assess the accuracy of a calculated polygon.

A major limitation in this study is that each service area estimation algorithm usually needs considerable tuning by system experts to address each algorithm's particular limitations. To provide a fair comparison, this tuning was not performed for each algorithm. In addition, since the inherent error and uncertainty in each service area algorithms varies, formal probability-based methods need to be developed to better assess the methods.

In the next chapter, I introduce, adapt, and implement two new methods for service area approximation based on transportation network optimization, study them in comparison to distance- and cell-based methods for a large water network, and expand methods for accuracy assessment to include error matrices and Kappa analysis.

CHAPTER 7: WATER UTILITY SERVICE AREA APPROXIMATION

Given that the substantial accuracy differences found among methods in practice from the previous study, my third research study was designed to investigate whether new estimation approaches could be more effective overall. In order to understand the potential, I applied insights gained from the second study in order to develop and test several new estimation approaches. I developed two novel service area estimation methods based on road network optimization techniques, in order to investigate my third specific research question:

> Specific Research Question RQ3: Will applying metrics for transport optimization to service area estimation improve accuracy in comparison to common techniques?

In this chapter, I perform a novel adaptation of transportation network analysis to create two new service area approximation algorithms, and compare them to cellular automata and Thiessen polygons for accuracy. Since utility networks are often aligned with transportation networks, this investigates the intuition that using transportation network layout as a weighting parameter can help to refine service area estimation methods and improve accuracy. The study also serves to demonstrate the more general applicability of the methods for accuracy assessment to water networks. The two figures below (FIGURE 42) illustrate visual inspection confirmation that this alignment holds true for water pipelines in Kentucky. It is likely that other CI networks will also have similar alignments with transportation networks, since they are often constructed together. For example, when new neighborhoods are built, utilities and roads are planned together to service them. Therefore, using road networks to estimate service areas would be expected to perform well since these are often planned in conjunction with CI networks.

FIGURE 42: Visual comparison of water (A) and road (B) networks in urban setting

For this work, I use the Kentucky infrastructure authority water and waste water datasets available for free download (Authority, 2013).

In the previous chapter, I compared four service area approximation methods that are heavily used in the disaster/emergency response community. In this chapter, I introduce two new methods, service area optimization (SAO) and location allocation optimization (LAO) and apply them to service area estimation for the Kentucky water network. Both SAO and LAO are based on road network travel distance optimization. However, LAO also incorporates into the model the amount of output the service source provides and matches that to the demand locations. I compare the results of these two methods to the previous four approximation methods discussed in Chapter 6.

It is reasonable to think that service areas are related to ease of connectivity to a source point. In this chapter, I estimate ease of connectivity using the transportation network. Service areas are estimated by using the road distance from source points. In GIS software, I can define a service area as a region that encompasses all accessible streets within specified "impedance." For example, a 10-minute service area for a source point includes all the streets that can be reached within 10 minutes from that point.

Location allocation optimization, also known as the facility location problem, has been utilized in scenarios such as the building of a new manufacturing plant. This optimization method uses the baseline knowledge of the capacity of the manufacturing plant, all the buyers that it has access to, and all other manufacturing plants. Using this information, the minimally expansive transportation routes based on the transportation network (e.g. roads) are determined, and an optimal location for the new plant is found based on the capacity, demand points and the amount of demand. The LAO can be considered the weighted version of SAO, similar to those tested in Chapter 6.

Many major critical infrastructure distribution network service delivery components (water pipes, power lines, etc.) are often planned and located along the transportation networks. Therefore, modeling critical infrastructure based on road networks might be of value. To this end, I have adapted transportation service area approximation processes (SAO and LAO) for critical infrastructure (or utility) service area approximation. I compare two versions of the new methods to each other, as well as to the four methods discussed in the previous chapter.

Hypotheses

I hypothesize that estimating service areas based on transportation networks will be more accurate than common methods, since many CI distribution networks are closely aligned with transportation networks. To understand and quantify the differences among the estimation methods, I test the following four hypotheses on the water network in the state of Kentucky:

H3-1. Road-network-based Service Area Optimization (SAO) will produce more accurate point impact results compared to cell-based (CA) or distance-based (TP) estimations.

H3-2. Road-network-based Service Area Optimization (SAO) will produce more accurate aggregate impact (area estimation) results when compared to cell-based CA or distance-based TP estimations.

H3-3. Road-network-based Location Allocation Optimization (LAO), weighted using capacities and demands, will produce more accurate point and area results than all other methods.

H3-4. Weighted SA estimation techniques (WTP, WCA, LAO) produce more accurate results compared to their standard counterparts (TP, CA, SAO).

The accuracy of each service area calculation is assessed through comparison with the ground truth service areas. I use two different approaches to perform the accuracy assessment of the results: aggregate (area) impact accuracy analyses and spatial point impact accuracy analyses. In the next section, I will discuss the data used for the study.

7.1 Data and Location

Kentucky has developed a system called Water Resource Information System Portal (WRIS) through the cooperative efforts of water systems and local, regional and state agencies (2013). It is one of a very few web-based mapping and analysis tools that also allows data download. The WRIS portal not only contains valuable system information, but also serves as the statewide registry for water and wastewater projects in the Commonwealth of Kentucky. Through this system, comprehensive water and wastewater datasets are made available for free download to aid water resource planning, such as watershed protection and infrastructure development. The dataset covers the Commonwealth of Kentucky and includes detailed information about each component.

Layers that are provided are water lines (FIGURE 43), water treatment plants, water tanks, surface and well sources, purchase sources, water meters, pump stations, and water pumps. Detailed technical information about these layers can be found in APPENDIX B: KENTUCKY GEOSPATIAL and on the WRIS website (2013) in the "Geospatial Data" section. In addition, I used data from Kentucky Geography Network's KY Geoportal (Authority, 2013). From this source, I obtained the state and local road line layers to create the Kentucky road network. I used USGS's 10-meter-resolution digital elevation model (DEM), in addition to miscellaneous layers like the state boundary layer and the state counties layer to clip the polygons to consider only the state of Kentucky. One last data source that I used for my analysis was the US Census bureau website (Bureau, 2012) to be used in demand estimation. This dataset included 2012 Census Tracts, Census Groups, and Census Block points. The layers were in TIGER format with GCS_North_American_1983 as their Geographic Coordinate System and D_North_American_1983 as their datum. All data layers were imported into an ESRI ArcGIS Geodatabase structure as feature datasets and grouped under feature classes. The Kentucky water system data is provided in North American Datum, Kentucky State Plane

FIPS 1600, and Lambert Conformal Conic projection system. To ensure accurate overlays, I transformed all other data layers into this projection.

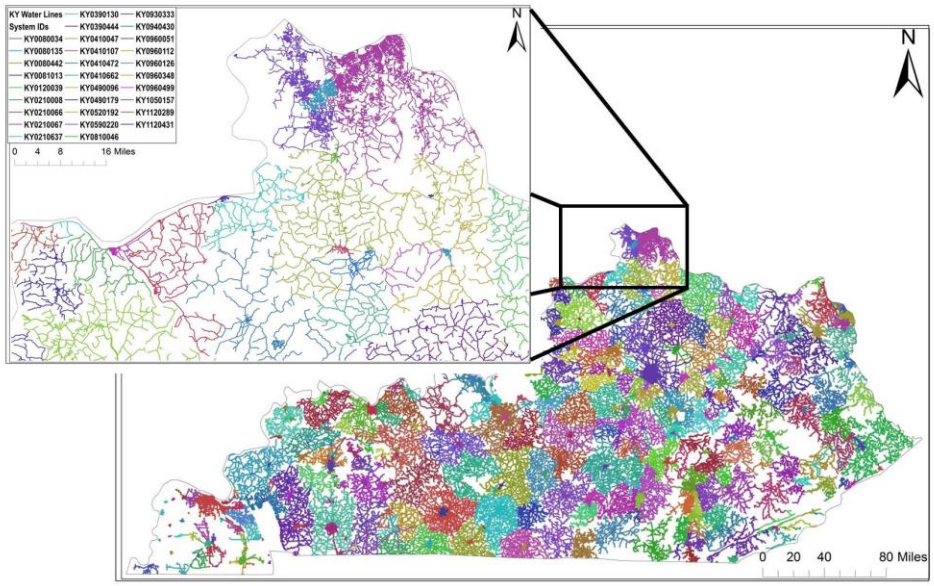


FIGURE 43: Kentucy state water lines

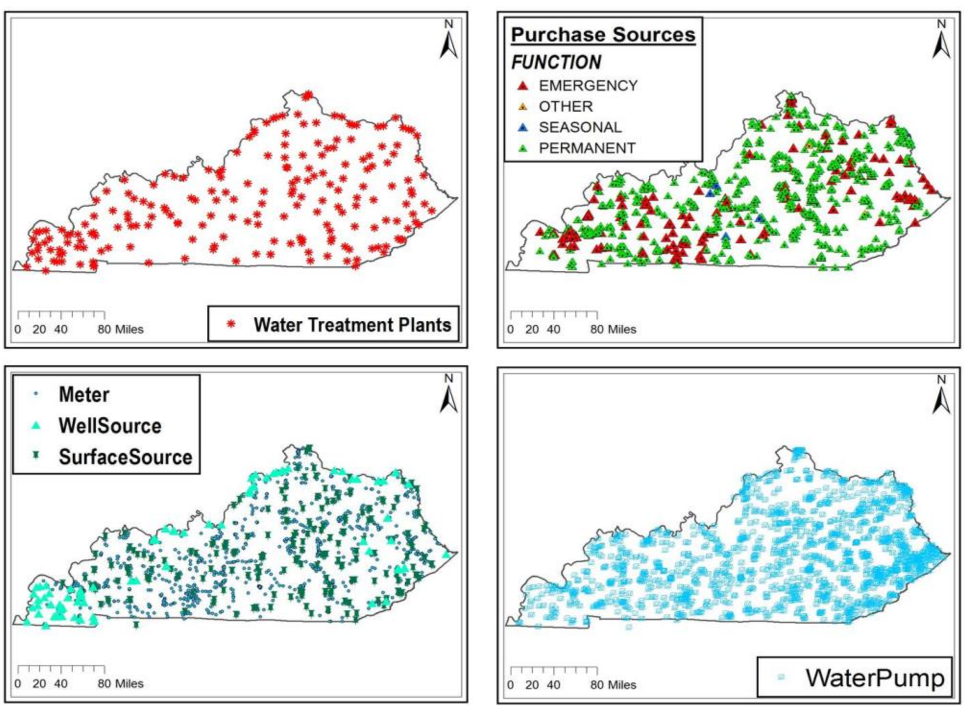


FIGURE 44: Elements of the Kentucky water network

The census database included the variables: track, block group (polygon) and block (point). Census blocks are designed as subsets of zip code zones, and block groups are designed as subsets of census tracks with up to 6000 people in each block group. Census block point data is a subset of block group data spread out in each census block group. Each census group point represents blocks with up to 3000 people; however, the mean of block point population values is 47.7. In the census data, I deleted block points with zero population values. This helped reduce the confusion and unnecessary data processing in the optimization system.

FIGURE 45 shows the census block points with population and census block data.

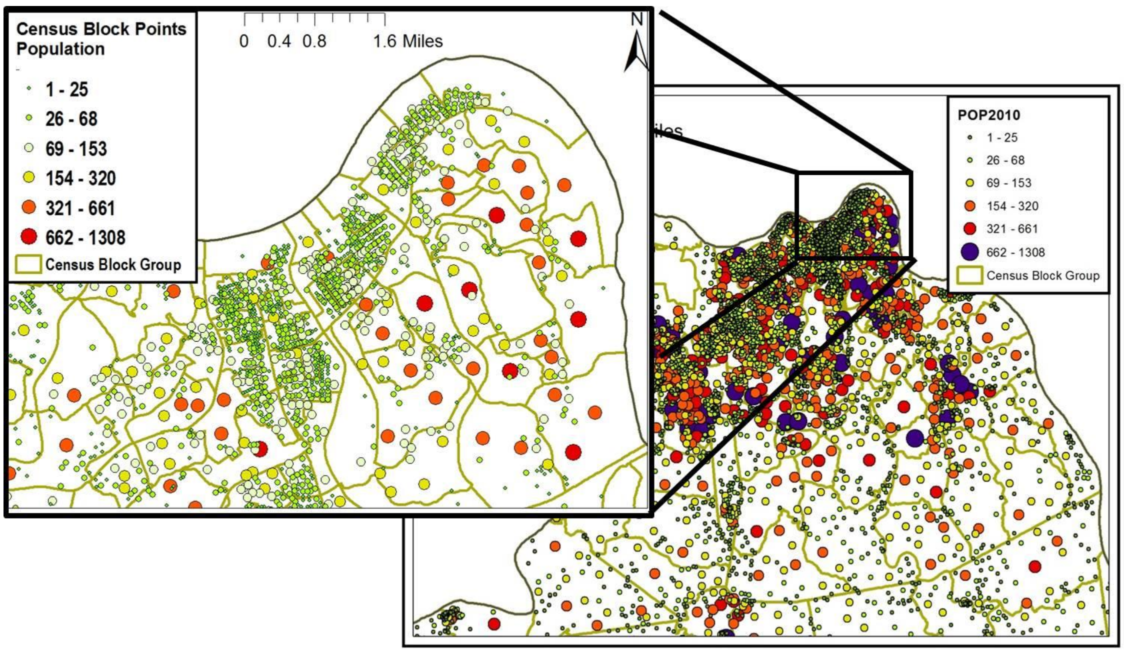


FIGURE 45: Census block population data

Water systems are typically organized around water treatment plants (WATP) as the source of water. To compute the service areas for the Kentucky, I began with the WATP dataset, and then performed cleanup. Five WATPS with 0 thousand gallons average daily production (AVGDP) WATPs with less than 10 thousand gallons average daily production (AVGDP) were removed from the dataset, because they are significantly smaller than most other WATPs, and are only used to serve areas in very close proximity to the WATP. This reduced the total number of WATPs from 217 to 203. I combined the two largest WATPs into one serving Louisville, since they were 7 miles apart and produce 102.36 and 92.36 million gallons for the city. The next largest Kentucky-American Water Company WATP is 75 miles away, producing 30.57 million gallons. I used a combination of system IDs and WATP names to consolidate duplicate WATPs.

7.2 Service area methods

Using Kentucky's water network, I re-tested four methods that were previously tested for an EP network (Chapter 6) and tested the two new methods for critical infrastructure service area approximation. In section 7.3.1 I review CA, WCA, TP, WTP methods and show examples of the resulting polygon layers for each. In section 7.3.2 I discuss the two new methods, namely service area optimization (SAO) and location allocation optimization (LAO), which I employed to create service areas. In section 7.3.3 I discuss the reference set.

7.2.1 Thiessen Polygons and Cellular Automata

In the previous chapter, I explained the creation of polygons using cellular automata (CA), weighted cellular automata (WCA), Thiessen polygons (TP) and weighted Thiessen polygons (WTP) for the electric power network of a mid-size US city. For the Kentucky water dataset, I followed a similar procedure to create water treatment plant service area polygons. I used ArcGIS's built-in Thiessen polygon creation tool to create the service area polygons using TP approximation. I then clipped this layer with the state border layer. FIGURE 46 depicts Thiessen Polygons created as an approximation of water treatment plant service areas.

As in Chapter 6, I created the WTPs using the publicly available ArcGIS extension (Dong, 2008). This method creates a raster grid file based on the cell unit setting defined by the user. The higher the cell unit setting is, the smaller the pixels of the grid are, and higher the resolution of the grid. I set this cell unit to 7000, which was the highest value available based on my computational power. The computer used and the setting are the same as what I used for EP network service area estimation in Chapter 6. I then clipped the output raster and polygon layers with the state border layer. The grid file and the weighted Thiessen polygons are shown in FIGURE 47. The lines on FIGURE 47 are anomalies that arose during the raster to vector conversion for WTP. This seems to be an implementation defect that introduces errors on donut-shaped polygons while performing the grid-to-polygon transformation.

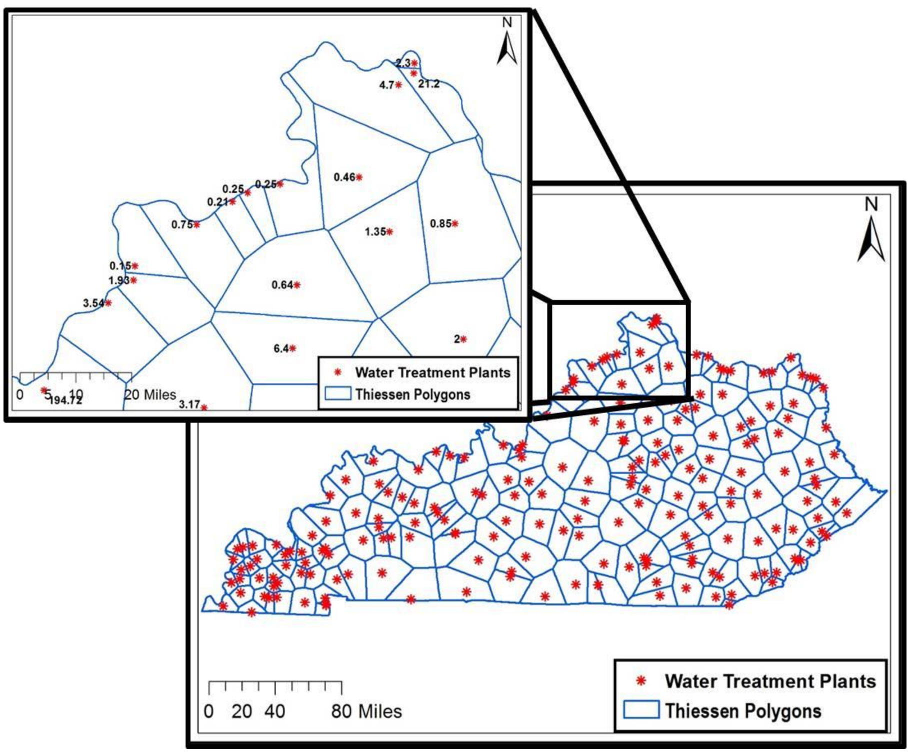


FIGURE 46: Thiessen polygons approximating service areas for water treatment plants.

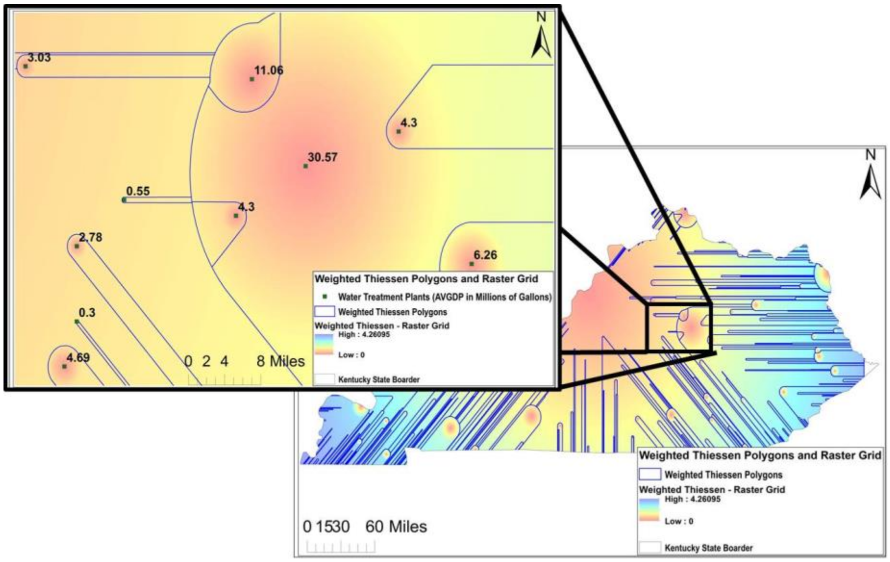


FIGURE 47: Weighted thiessen raster grid and polygons for approximating service areas

Cellular automata layers were created using the functionality built into the Interdependency Environment for Infrastructure Simulation Systems IEISS (Bush, 2005; G Loren Toole and McCown, 2008). For this I first re-projected the water treatment plant layer to the geographic coordinate system (WGS 84) and inserted two new columns for latitude and longitude coordinate values. I used WATP Name, X Coordinate, Y Coordinate, AVGDP, and ID and filed values into the appropriate XML tags by automating the process through MS Excel. For CA, I did not include the AVGDP value but instead I placed a sufficiently high number in its place. This number (1000) was the same for all the WATPs. For each water treatment plant, I also had to have a junction defined at the same exact location. WATPs were linked to these junctions through the "connections" tag linking to junctions "ID" tag. Please see "APPENDIX C: IEISS XML INPUT EXAMPLE: KENTUCKY WATER SYSTEM" for an example XML file for one WATP.

After creating the CA and WCA polygons based on water treatment plant layer information, I ran a "repair geometry" tool and manually cleaned the geometries of some donut shaped polygons. Next I clipped the layers with the Kentucky state border layer to have uniform study area boundaries. FIGURE 48 shows the process I applied to create the CA and WCA layers, and FIGURE 49 and FIGURE 50 show the resulting layers.

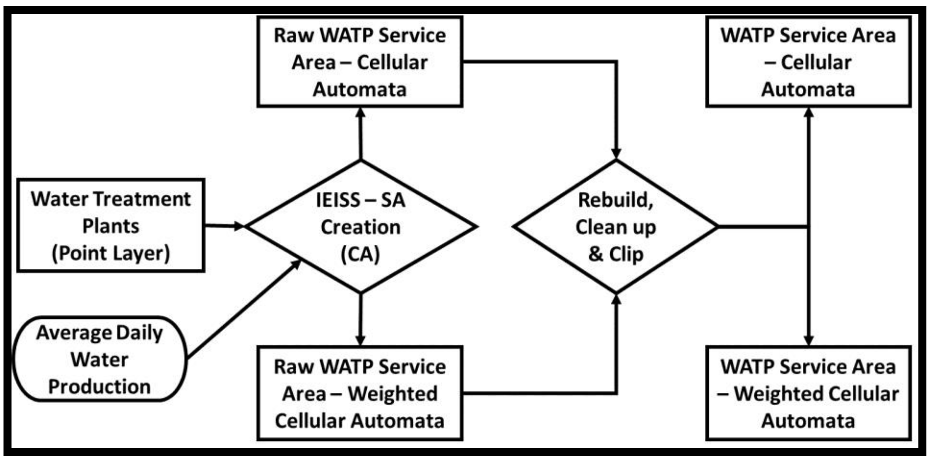


FIGURE 48: Process for creating polygons for standard and weighted cellular automata

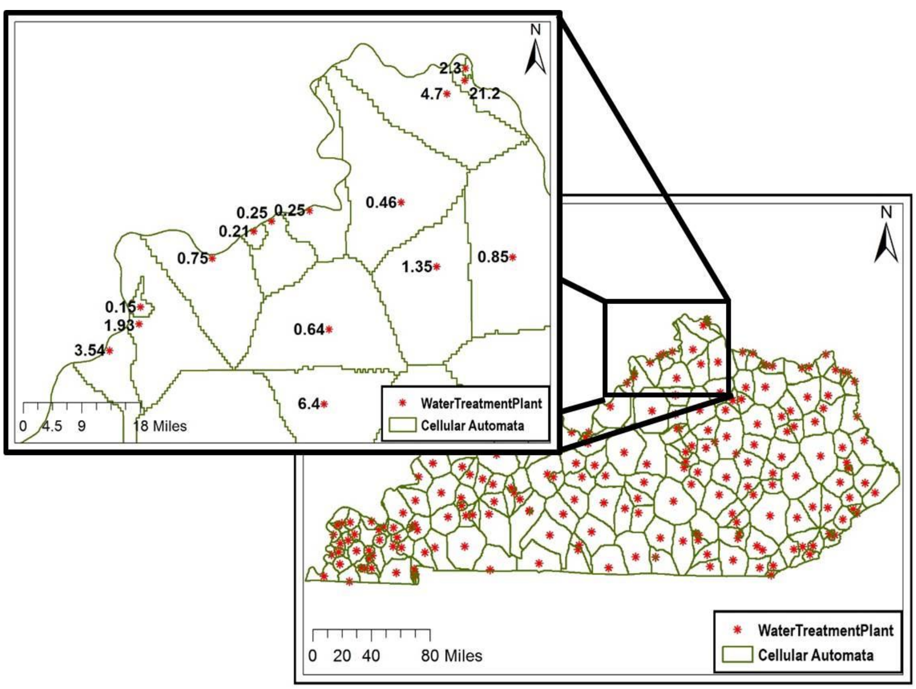


FIGURE 49: Service areas for water treatment plants using cellular automata

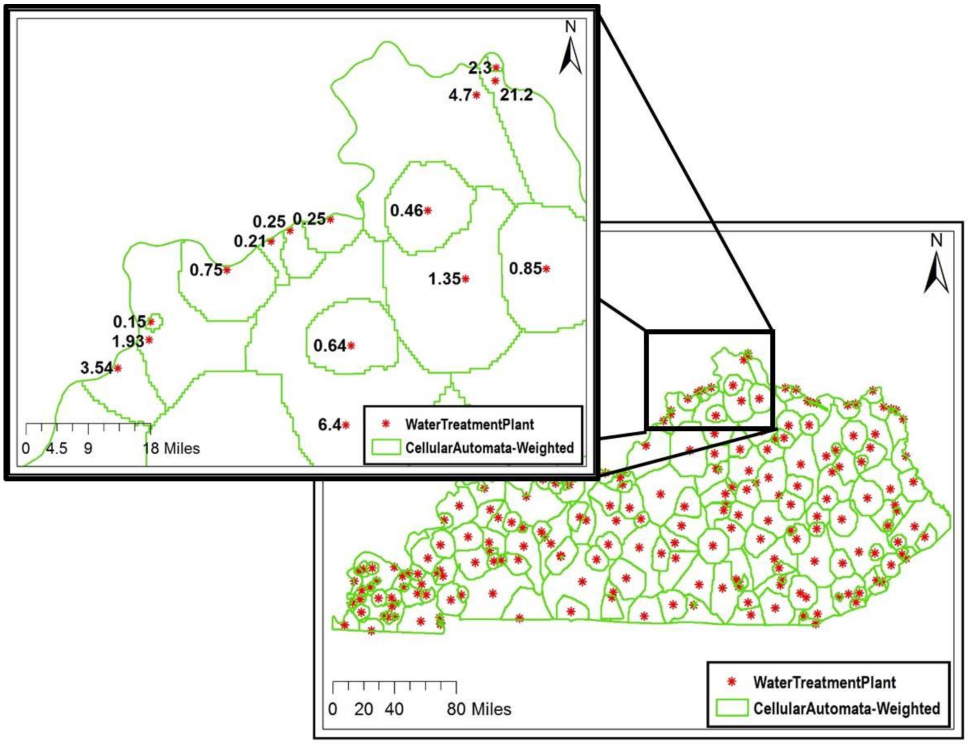


FIGURE 50: Service areas for water treatment plants using weighted cellular automata

7.2.2 Service Area Optimization and Location Allocation Optimization

Service area optimization and location allocation optimization processes both are based on the transportation (road) network data. I used the ArcGIS Network Analyst tool to generate a travel-distance based road network from GIS data layers for local and state roads for Kentucky. To do this, I set the ArcGIS Network Analyst tool impedance to the road length attribute, the merge field to "No overlap," and the overlap type to disks. I set the default break values as 264K feet (50 miles) to ensure coverage of the whole state (FIGURE 51). Next, I loaded the water treatment plant dataset as the facilities source point data. The SAO algorithm then computes the shortest path from each WATP to each road segment, and assigns each road to its nearest WATP. The final SAO polygons for each WATP are the hull of the WATP's assigned road segments.

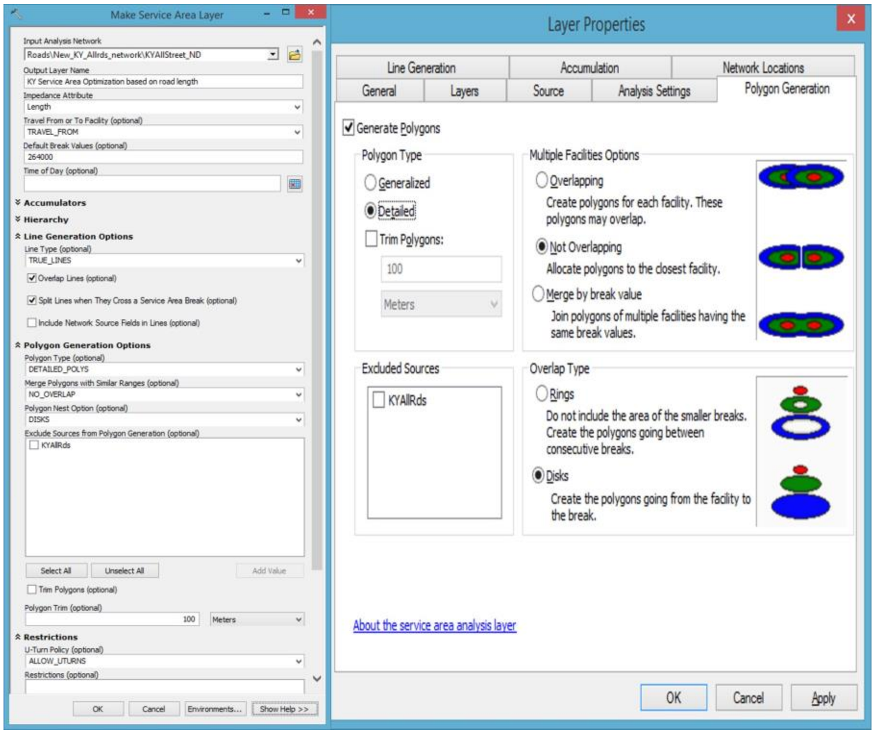


FIGURE 51: Service area optimization (SAO) layer set up

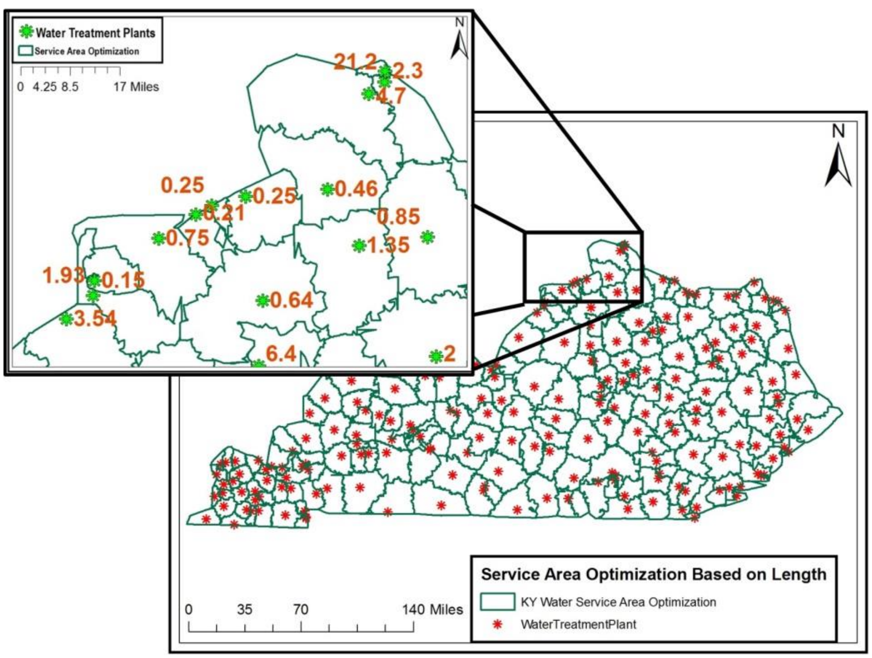


FIGURE 52: Overview of the polygons created using SAO

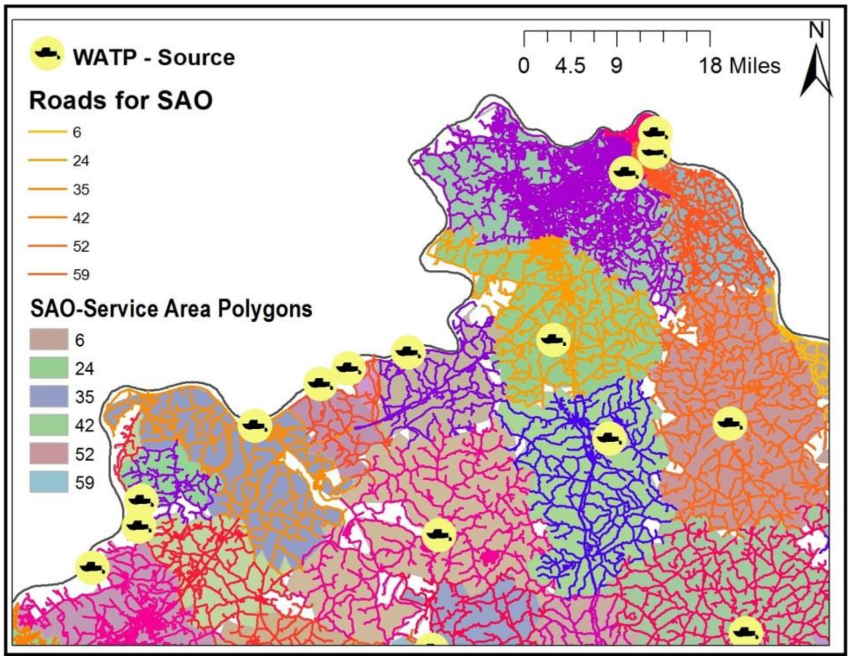


FIGURE 53: Road extent defining the service area polygons created using SAO

In a similar way, I used ArcGIS Network Analyst to create LAO service areas based on the road network. This is an application related to the facility location problem that seeks to optimally locate facilities to serve customers. The distances are set using the shortest path on the road network. A customer is characterized by a demand, and a source facility WATP by its daily capacity (AVGDP). In this case, a customer is a census block, and its demand is computed as the average block population (47) multiplied by the average daily demand per person (67 gallons (Kenny et al., 2009)). A table that shows average daily demand is shown in “APPENDIX D: .”

I created the LAO location allocation layer in ArcGIS Network Analyst, using WATPs for source facilities, and the study area separated into census block-size cells using the Thiessen polygon tool as point data layer for demand. For the problem type, I choose “maximized capacitated coverage,” that chooses WATPs so that all or the greatest amount of demand can be served without exceeding the capacity of any facility. In addition to satisfying capacity limit, it selects facilities such that the total sum of weighted impedance (demand allocated times distance) for all WATPs is minimized.

As a last step, I join the location allocation demand points and their assigned source points with the block area layer, according to WATP name. See FIGURE 54 for the flowchart of LAO layer creation process. See FIGURE 55 for the resulting service area polygons, as well as the source and demand points connected by relationship lines.

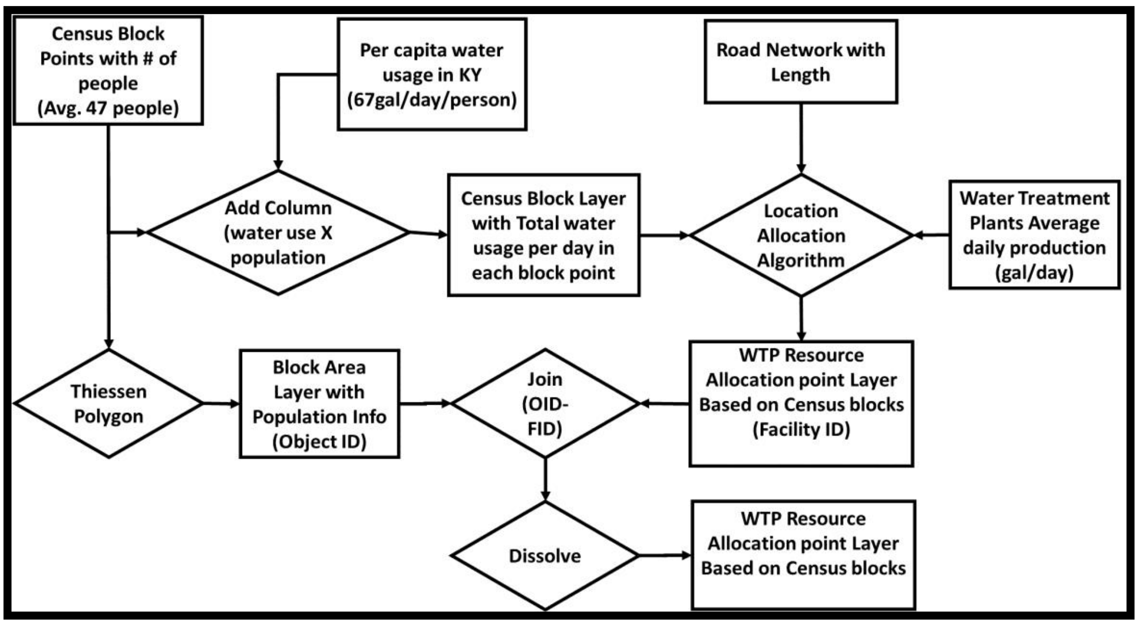


FIGURE 54: Process for LAO based on travel distance and population

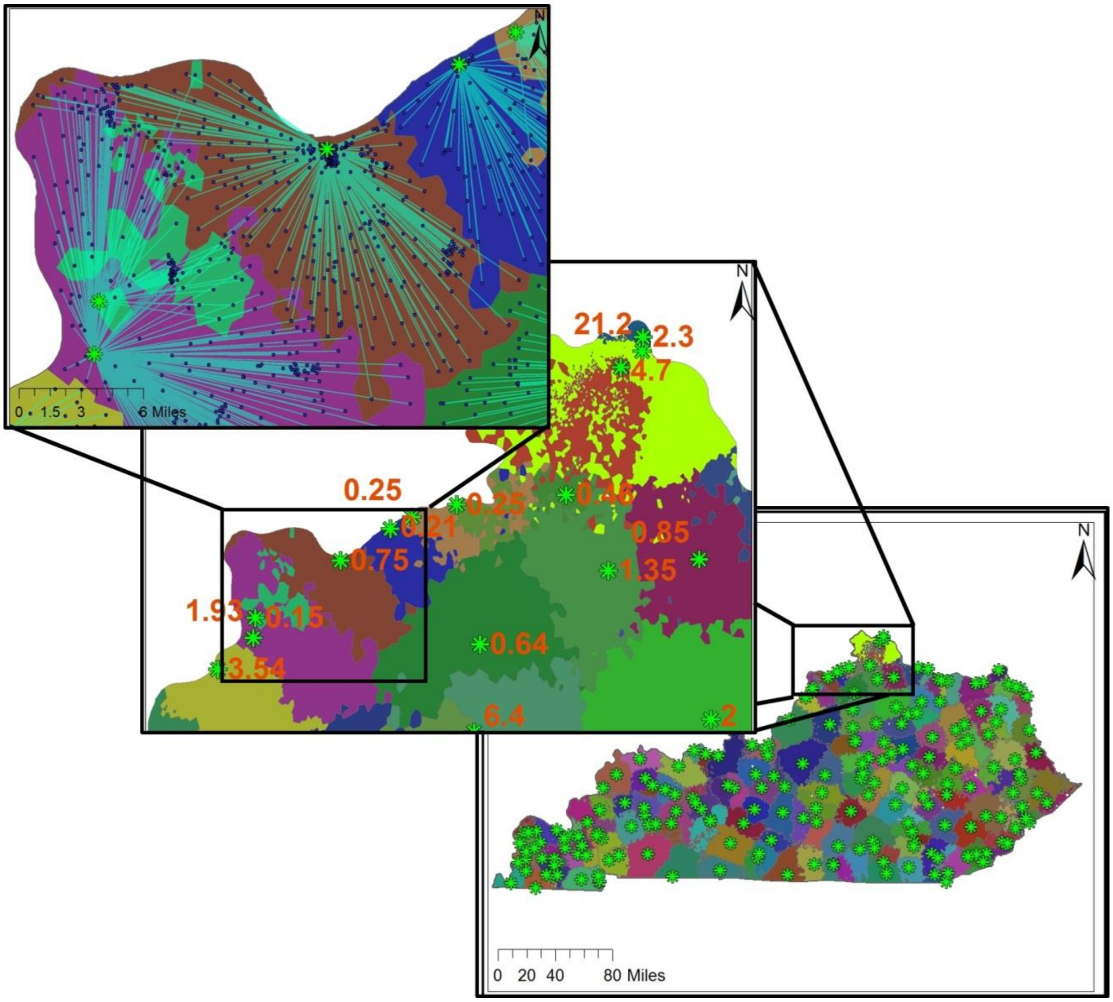


FIGURE 55: Service area polygons created by location allocation optimization

7.2.3 Reference Dataset

For an accurate comparison of the methods, it is crucial to have an accurate reference dataset. Data provided by the Kentucky infrastructure authority does not include the information about which area is getting service from which water treatment plant (WATP). However it includes the information about the systems that owned the infrastructure components like water pipes and treatment plants. Using this with additional data sources and collaboration with a principal Kentucky Infrastructure Authority GIS expert, I was able to finalize the reference dataset defining the WATP service areas.

The reference dataset polygons are based on distribution system water lines using the locations of these components associated with a water system (utility):

- Water lines (system name, their reach and location, connection to purchase points and water treatment plants)(FIGURE 43)
- Purchase points (amount of daily average sale, selling system name, buying system name and function as permanent, seasonal or emergency)(FIGURE 44).
- Water treatment plants (locations and amount of average daily output and system name) (FIGURE 44)

The majority of the systems were self-sufficient. Delineating the service area polygons for these self-sufficient systems uses the general shape of the water lines and does not cover areas where there are no pipelines (FIGURE 56). In some cases, self-sufficient systems or system groups might be grouped together with very minimal water exchanged between them (FIGURE 57). During this process I consulted the principal Kentucky WRIS GIS domain expert, who corroborated my analysis as useful for making accuracy comparisons. The resulting final reference dataset can be seen in FIGURE 58.

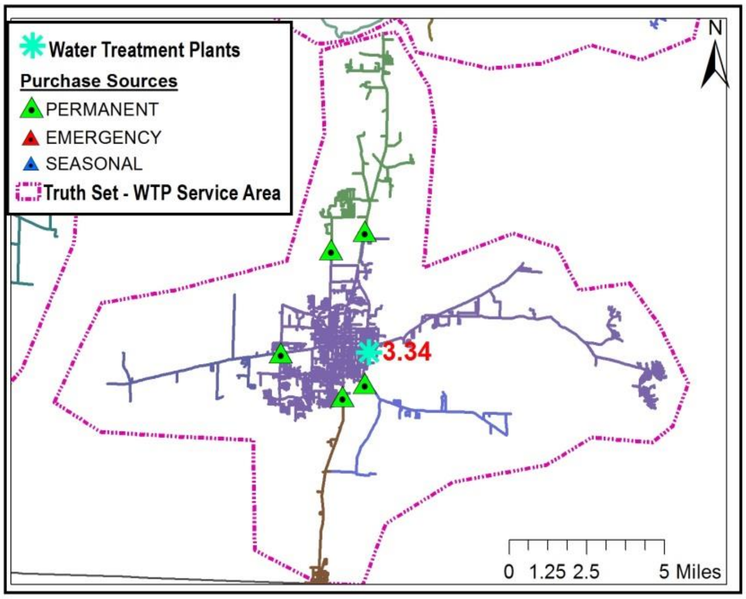


FIGURE 56: Service area reference dataset delineation for a self-sufficient system

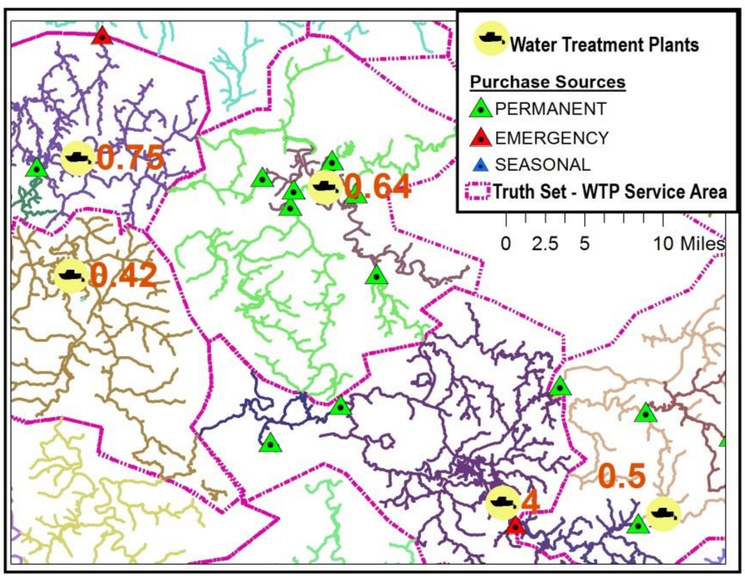


FIGURE 57: Service area reference dataset delineation - multiple systems

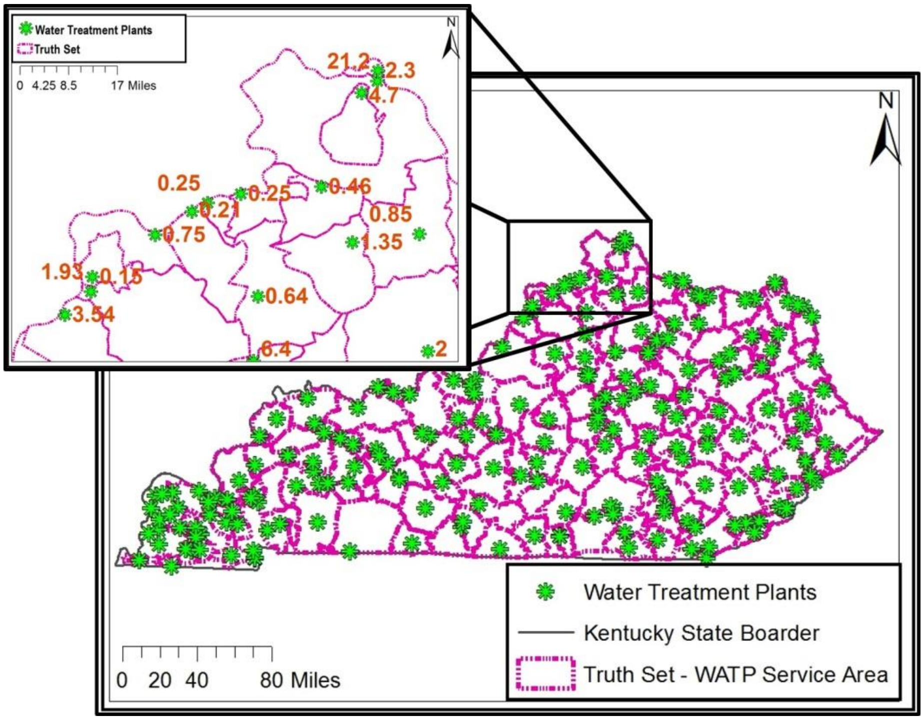


FIGURE 58: Reference dataset for water treatment plant service areas

7.2.4 Accuracy Checkpoints: Point Data Impact Analysis

Point data impact analysis requires random points placed across the study area. For each polygon, the service source is determined and agreement between the actual source and source assigned by the service area approximation method is checked. For each accuracy checkpoint, I determine from which water treatment plant they would get their service at their current location. Then I check if the calculated service area polygon agrees with that or not. Repeating this for all the points gives a measure of overall accuracy. To achieve this, I used two reference sets with different fidelities. The first reference dataset was a point layer with close to 1.1 million points quasi-randomly spread across the state using ArcGIS tools. The distribution was not totally random because I wanted to avoid smaller polygons not being well represented. The second reference dataset was a point layer, but these were specifically placed on top of the water lines.

Sample size and distribution are one of the most discussed and documented areas of accuracy assessment. A number of researchers have published guidelines for choosing appropriate sample sizes (Ginevan, 1979; Hord and Brooner, 1976; Rosenfield et al., 1982) in remote sensing accuracy assessment. Usage of an equation based on a normal approximation to binomial distribution or just the binomial distribution is appropriate if all you want is an accuracy number.

In this chapter, I introduce the computation of an error matrix to better understand the analysis (Congalton and Green, 2008). An error matrix represents a situation where for each point there are *n* possible source points, there is only one correct answer and (n-1) incorrect answers. To create a statistically valid error matrix, the sample size decided by multinomial distribution is recommended (Tortora, 1978). Sample size is selected to best represent the area and the classes; however, ideally, it would be most accurate if I had as many samples as my minimum mapping units (Congalton, 1991).

My minimum mapping unit in this case is the census blocks with an average of 47 persons (12-15 households). Using the random point distribution, with the total numbers equaling the number of households in each census block group, results in a total number of 1,097,512 accuracy points. I then performed an overlay with the reference dataset and eliminated the points that do not overlay the reference dataset (FIGURE 59). The number of points that do not overlay the reference dataset is inherently small because naturally, population numbers are considerably lower at locations where there is no public water service. After the overlay I eliminated 24,798 points, and overall accuracy was computed using 1,072,714 points (FIGURE 60).

Accuracy checkpoints were also created by placing the points directly on the lines representing the water pipes, because of the concern for potential errors produced by the reference data service area delineation process. Each line had 4 equidistant points placed on top of them. With over 280 thousand water lines and four points in each line, there were over 1.1 million total accuracy checkpoints. I consider a dataset the high-fidelity point accuracy reference set (FIGURE 61).

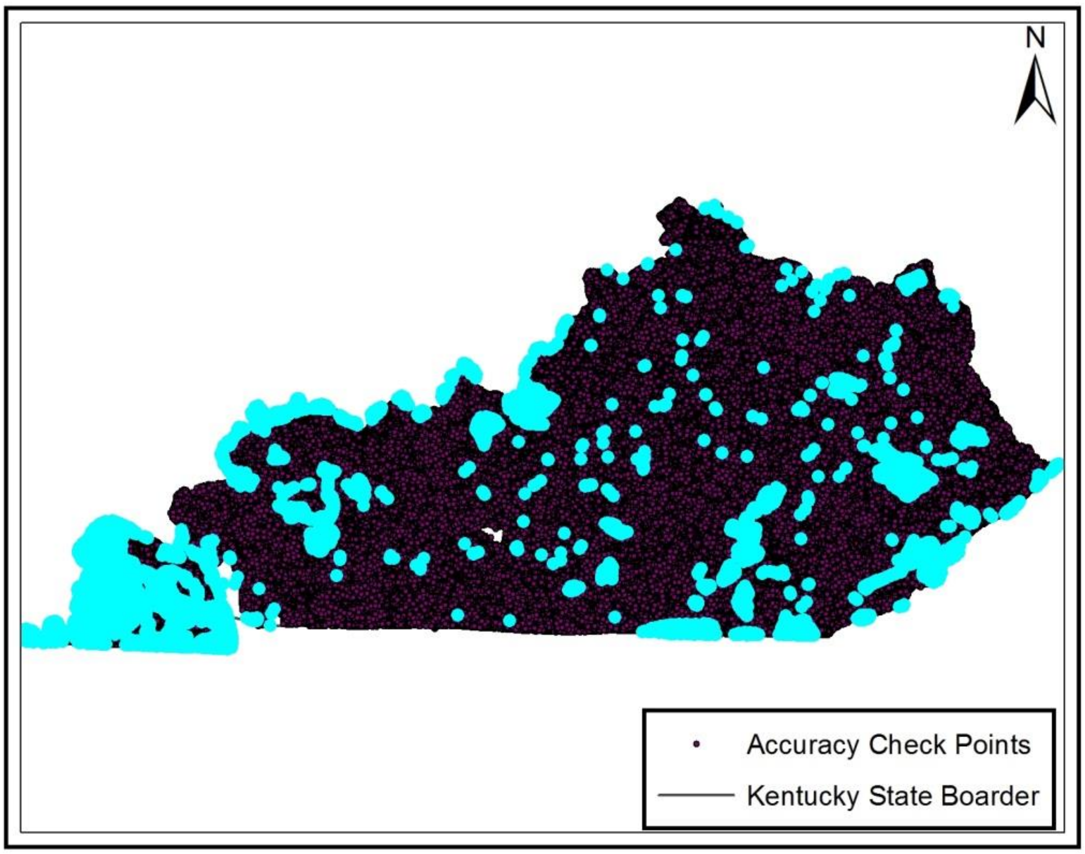


FIGURE 59: Accuracy checkpoints eliminated

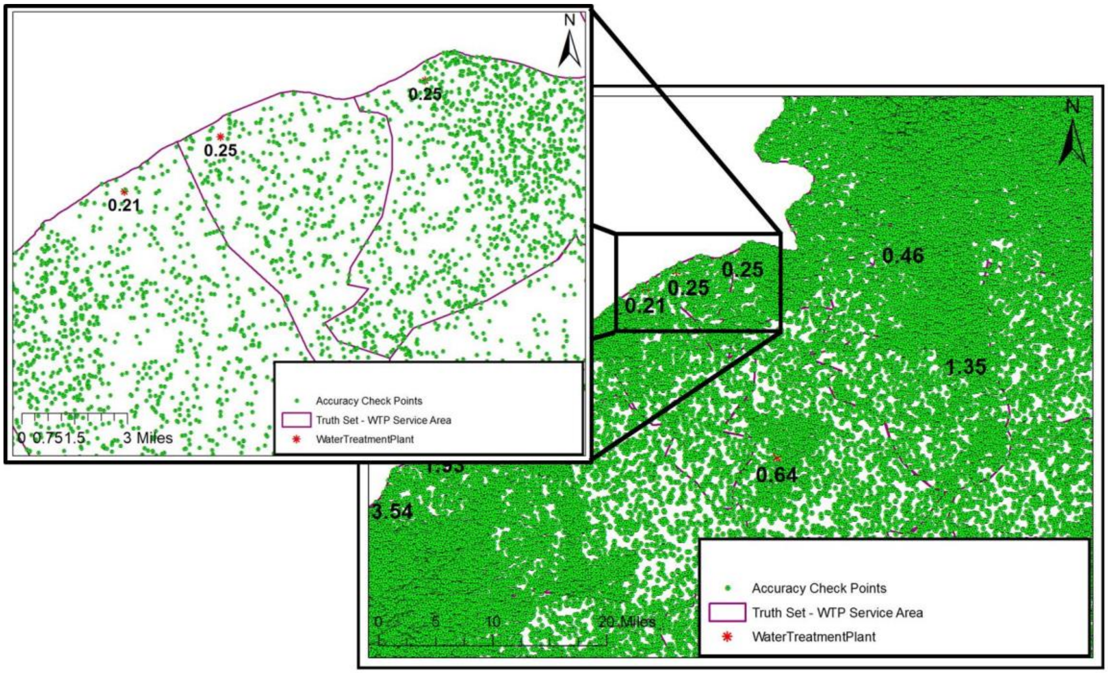


FIGURE 60: Accuracy checkpoints randomly distributed in each census block group based on the number of household units

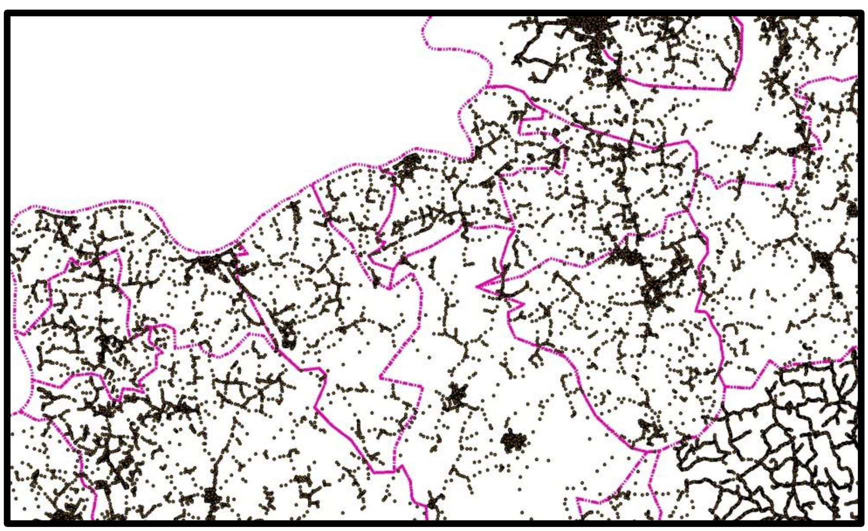


FIGURE 61: Accuracy checkpoints placed on water lines

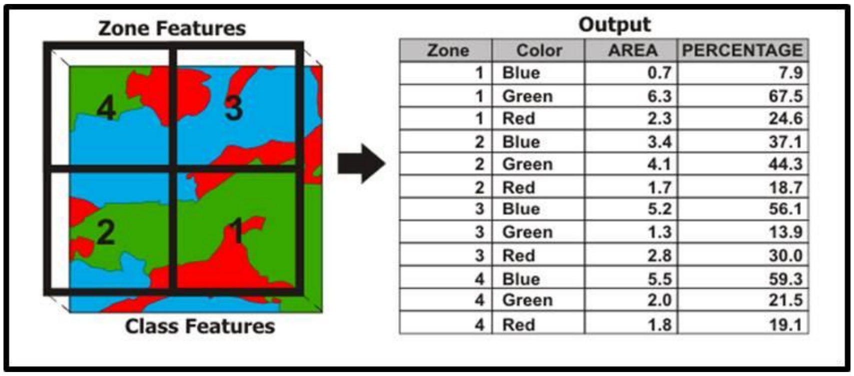
7.3 Accuracy Assessment

In this study I test the reliability of various methods in approximating service areas for water treatment plants. Similar to the previous accuracy analysis, two types of accuracy assessments were implemented.

Aggregate impact analysis compares the agreement or disagreement between the total areas defined by the reference dataset and the method in question. I created an error matrix for each layer using the amount of area – as per the previous study. In addition I also created an error matrix for each layer using the amount of area agreement and disagreements.

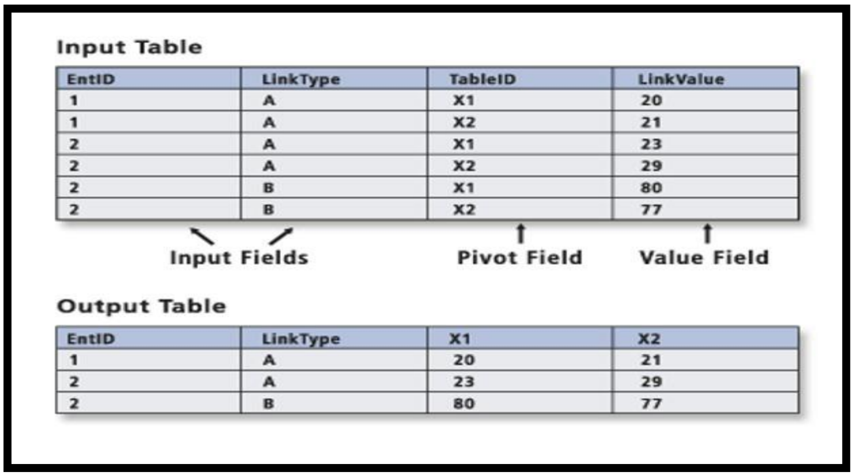
Creating the error matrices from GIS data required multiple steps to overlay layers to transfer attributes and some database table manipulation (Congalton and Green, 2008). I used ESRI's ArcGIS 10.2.1 "tabulate intersection" tool to compute the intersection between two feature classes and cross-tabulates the area, length or count of the intersecting features. See FIGURE 62 for an example of the functioning of the "Table Intersect" tool. I include this figure and reference to ensure replicability of this work. I defined the zone field as the reference dataset for WATP service areas. Then the accuracy checkpoint layer went through a spatial overlay process to get the attributes of the service area approximation layers. For aggregate impact analysis, the service area approximation layer was defined as the class feature layer. For point accuracy analysis, the point accuracy checkpoints with the service area approximation layers attributes were defined as the class feature layer. Output of this table was used as an input for the "Pivot Table" tool to create the final error matrix. An example of this process is shown in FIGURE 63.

Details of the error matrix creation process for aggregate impact analysis can be seen in FIGURE 64, and details of the error matrix creation process for point impact analysis can be seen in FIGURE 65.



Zone	Color	AREA	PERCENTAGE
1	Blue	0.7	7.9
1	Green	6.3	67.5
1	Red	2.3	24.6
2	Blue	3.4	37.1
2	Green	4.1	44.3
2	Red	1.7	18.7
3	Blue	5.2	56.1
3	Green	1.3	13.9
3	Red	2.8	30.0
4	Blue	5.5	59.3
4	Green	2.0	21.5
4	Red	1.8	19.1

FIGURE 62: Tabulate intersect tool overview



EntID	LinkType	TableID	LinkValue
1	A	X1	20
1	A	X2	21
2	A	X1	23
2	A	X2	29
2	B	X1	80
2	B	X2	77

EntID	LinkType	X1	X2
1	A	20	21
2	A	23	29
2	B	80	77

FIGURE 63: Pivot table example

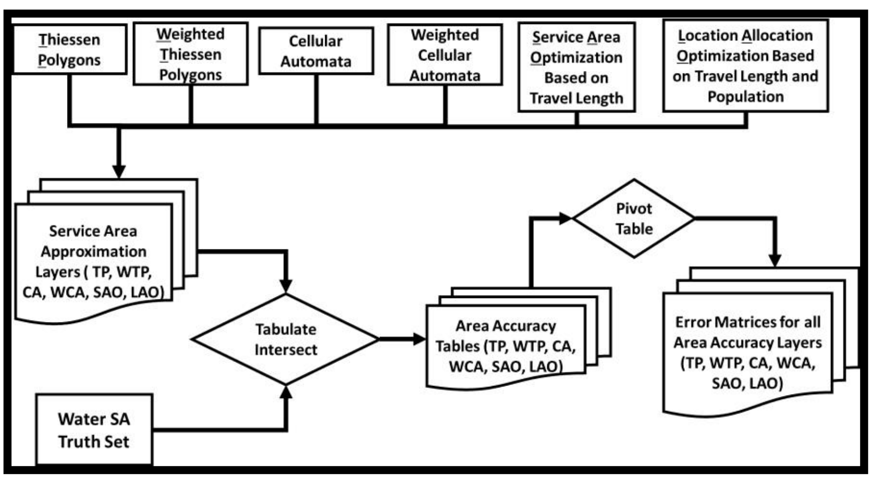


FIGURE 64: Creation of error matrices for aggregate impact analysis

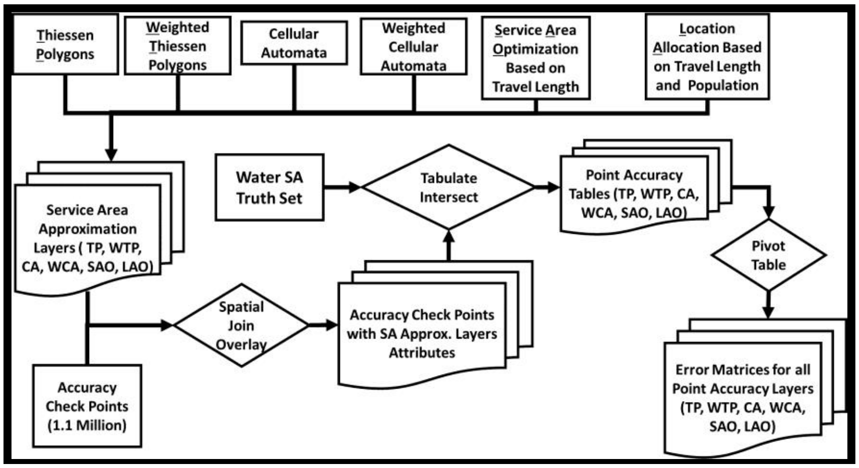


FIGURE 65: Creation of error matrices for point impact analysis

Once I finished producing the error matrices, I calculated overall accuracy values for aggregate (area) and point accuracies. I did this first using the accuracy point dataset created by points placed across the area and then using the points placed on the water lines. For both the aggregate and point impact accuracy analyses, the commission and omission errors are calculated and presented as producer's and user's accuracies for WATP service areas.

Overall accuracy is calculated using the approach described in (Congalton and Green, 1999, 2008), as follows:

$$OverallAccuracy = (\sum_{i=1}^{k} n_{ii})/n$$

where k is the total number of water treatment plants, n is the number of sample points and n_{ij} is the variable that represents each cell in the matrix where i and j represents columns and rows respectively. As shown on an example error matrix in FIGURE 66, the columns are the reference values and the rows are the values for a particular method (e.g., WCA). I create four of these matrices, one for each method, to assess the accuracy of each calculated dataset when compared to the reference dataset. For each sample point, I determine which reference polygon it falls on (*i*) and which calculated polygon (*j*) it falls on. If both of these polygons represent the service area for the same water treatment plant, the ID fields match (i=j), which adds one to the n_{ij} cell in the matrix.

		Reference (Ground Truth) Service Area Data							
		Substation 1	Substation 2	Substation 3	Substation 4	....	.....	Substation *i*	Row Total
Service area data for each method	Substation 1	n_{11}	n_{21}	n_{31}	n_{41}				
	Substation 2	n_{12}	n_{22}	n_{32}	n_{42}				
	Substation 3	n_{13}	n_{23}	n_{33}	n_{43}				
	Substation 4	n_{14}	n_{24}	n_{34}	n_{44}				
	....								
	....								
	Substation *k*							n_{kk}	
									Overall Accuracy: $(\sum_{i=1}^{k} n_{ii}) / n$

FIGURE 66: Example error matrix

The producer's and user's accuracy values as well as Kappa analysis for significance are calculated as follows, after (Congalton and Green, 1999, 2008).

Producer's accuracy can be calculated by

$$\Pr oducer' sAccuracy_j = \frac{n_{jj}}{n_{+j}}$$

And the user's accuracy can be calculated by

$$User' sAccuracy_i = \frac{n_{ii}}{n_{i+}}$$

7.3.1 Kappa Analysis

Kappa analysis is a discrete multivariate technique. It is one of the analyses used for statistically determining if one error matrix is significantly different than another matrix (Bishop et al., 2007). The KHAT statistic is the result of Kappa analysis and it is another measure of agreement of accuracy (J. Cohen, 1968). It is a measure of agreement between the actual agreement in the error matrix and the chance agreement. Actual agreement in the error matrix is indicated by the major diagonal in the matrix and represents the agreement between service areas that each approximation method created and the reference dataset.

Let p_{ij} denote the proportion of the samples in the (i,j)-th cell corresponding to n_{ij}. Let p_{i+} and p_{+j} be defined by

$$p_{i+} = \sum_{j=1}^{k} p_{ij} \quad \text{and} \quad p_{+j} = \sum_{i=1}^{k} p_{ij}$$

Also, let

$p_o = \sum_{i=1}^{k} p_{ii}$ be the actual agreement and $p_c = \sum_{i=1}^{k} p_{i+} p_{+j}$ be the chance agreement.

Then the maximum likelihood estimate of Kappa is calculated by

$$\hat{K} = \frac{p_o - p_c}{1 - p_c}$$

For each error matrix, the computed KHAT value is a measure of agreement or accuracy. The values can range from +1 to -1. There should be a positive correlation between the truth set and the methods because they both overlap the water treatment plants, so negative values are not possible in this analysis. Possible ranges for KHAT values are categorized in to three main groupings (Landis and Koch, 1977): a value greater than 0.8 (80%) represents strong agreement, a value between 0.4 and 0.8 (40-80%) represents moderate agreement and a value below 0.4 (40%) represents poor agreement.

Calculating Z value is a means for testing the significance of the KHAT statistic for a single error matrix. Through this I can determine if the agreement between the produced service area dataset and the reference dataset is significantly greater than zero (i.e., better than a random dataset). I can also apply the Z test to have pairwise comparison of two error matrices. With this I can determine if these two error matrices are significantly different or not.

Let $\hat{K}_1$ and $\hat{K}_2$ be the kappa statics for error matrix #1 and #2. Let also var ($\hat{K}_1$) and var ($\hat{K}_2$) be the estimates of variance. In this context, the Z value is used for testing the significance of a single error matrix as expressed by:

$$Z = \frac{\hat{K}_1}{\sqrt{\mathrm{var}(\hat{K}_1)}}$$

For pairwise comparison of error matrices, the Z value is computed as follows:

$$Z = \frac{|\hat{K}_1 - \hat{K}_2|}{\sqrt{\mathrm{var}(\hat{K}_1) + \mathrm{var}(\hat{K}_2)}}$$

More detailed information about above listed equations and equations for calculating error matrix variance values can be found in (Congalton and Green, 1999, 2008).

7.4 Transportation-based service area estimation Discussion and Conclusion

In this section, I report accuracy analysis results for aggregate impact analysis (AIA) and point impact analysis (PIA). The sample size for aggregate impact analyses is 203, representing the area of the spatial extent of Kentucky. On point impact analysis, my sample size is 1.07 million points that are categorized in 203 classes. On the high fidelity version of the point impact accuracy dataset, where the accuracy points are placed on the water lines, my sample size was over 1.1 million points. It would be logical to consider point impact analysis as the more accurate of two throughout this study because it has a very high sample size, and it is based on exact spatial locations rather than the amount of area comparison that aggregate impact analyses use. I created polygon datasets that approximate the service areas for water treatment plants in Kentucky, using tested six different methods: Thiessen Polygons (TP), Weighted Thiessen Polygons (WTP), Cellular Automata (CA), Weighted Cellular Automata (WCA), Service Area Optimization (SAO) and Location Allocation Optimization (LAO).

In TABLE 2, TABLE 4 and TABLE 5, I present overall accuracy values for all the methods. Table 4 and Table 5 show point impact analysis for two different point accuracy fidelities. Table 4 is created using the points placed across the study area whereas Table 5 is created using points specifically placed on the water lines. The accuracy results for those fidelities are very similar, so I use the numbers from the lower fidelity one for comparison as the conclusions from both fidelities are the same.

In both AIA and PIA, the LAO method has the highest accuracy values (72.32 and 83.08). At almost 10 points behind LAO, SAO and WCA had the second best accuracy values. Accuracy values for standard versions of TP and CA are around 60 (AIA) and 70 (PIA), with TP having slightly better accuracy than CA. WTP accuracy is the lowest in all the tests. Examination of the resulting polygons illustrates the difference of WTP compared to others as shown in FIGURE 47. This is likely due to the particular implementation of this method and an outlier in the dataset with a very large daily water output value. The WTP implementation uses raster graphics to do the weighting and this may allow the outlier to overtake most of the space in the study area. The problem can arise when the data is converted from raster format to vector format.

Producer's and User's accuracy tables provide accuracy values for each water treatment plant. Due to the large size of these tables I include user's and producer's accuracy tables for location allocation Point Impact Analysis and cellular automata aggregate impact analysis in "APPENDIX E: EXAMPLE PRODUCER'S AND USER'S ACCURACY TABLE FOR AGGREGATE IMPACT ANALYSIS."

KHAT values are a measure of agreement or accuracy and they can range from +1 to -1. Since a value greater than 0.8 (80%) represents strong agreement (Landis and Koch, 1977), I can conclude that with KHAT value of 82.21, the LAO point impact analysis shows strong agreement with the reference dataset. All the remaining accuracy numbers show moderate agreement with the reference dataset except TP, which shows poor agreement.

Pairwise Z values show the agreement between a pair of methods. A higher Z value indicates more significant differences in the error matrices of the pair in question. See TABLE 7 for calculated Z values comparing all the weighted methods to their standard counterparts. Since WTP has low accuracy, but TP has decent numbers, it was expected for their difference to be high. TABLE 8 shows pairwise comparison Z values for all the possible pairs.

TABLE 2: Kentucky water SA accuracy based on aggregate impact analysis

	Thiessen Polygons	Weighted Thiessen Polygons	Cellular Automata	Weighted Cellular Automata	Service Area Optimization	Location Allocation Optimization
Overall Accuracy	61.66	12.29	57.63	62.14	64.68	**72.32**

TABLE 3: Kentucky water SA KHAT and Z values based on aggregate impact analysis

	Thiessen Polygons	Weighted Thiessen Polygons	Cellular Automata	Weighted Cellular Automata	Service Area Optimization	Location Allocation Optimization
KHAT	61.40	11.10	57.33	61.81	64.42	72.32
Z value	241.90	69.44	223.44	241.11	258.78	1674.21

TABLE 4: Kentucky water SA overall accuracy based on point impact analysis

	Thiessen Polygons	Weighted Thiessen Polygons	Cellular Automata	Weighted Cellular Automata	Service Area Optimization	Location Allocation Optimization
Overall Accuracy	72.32	42.33	70.11	74.32	73.40	83.08

TABLE 5: Kentucky water SA accuracy based on high fidelity point impact analysis

	Thiessen Polygons	Weighted Thiessen Polygons	Cellular Automata	Weighted Cellular Automata	Service Area Optimization	Location Allocation Optimization
Overall Accuracy	71.8	47.33	70.3	72.4	76.8	**84.7**

TABLE 6: Kentucky water SA KHAT and Z values based on point impact analysis

	Thiessen Polygons	Weighted Thiessen Polygons	Cellular Automata	Weighted Cellular Automata	Service Area Optimization	Location Allocation Optimization
KHAT	71.09	35.43	68.68	72.18	72.19	**82.21**
Z Value	1603.86	823.51	1549.54	1668.14	1645.13	2186.19

TABLE 7: Kentucky water SA Z values comparing weighted to non-weighted methods

Pairwise comparison	TP vs WTP	CA vs WCA	SAO vs LAO
Z Value	577.27	69.07	173.48

TABLE 8: Kentucky water SA error matrix Z values for all pairs comparisons

Error Matrix Z value matrix	**TP**	**WTP**	**CA**	**WCA**	**SAO**	**LAO**
TP		577.3	38.4	30.4	17.7	191.4
WTP	577.3		538.3	612.0	598.2	818.7
CA	38.4	538.3		69.1	56.3	232.8
WCA	30.4	612.0	69.1		12.8	160.0
SAO	17.7	598.2	56.3	12.8		173.5
LAO	191.4	818.7	232.8	160.0	173.5	

Hypotheses H3-1 and H3-2, that SAO outperforms common methods in both point and area accuracy, can be accepted because the KHAT values are higher for both SAO and LAO accuracies.

Hypothesis H3-3 that LAO weighted using capacities and demands, will produce more accurate point and area results than all other methods can be accepted because LAO has better accuracy and KHAT values than SAO. This difference is significant because a pairwise comparison of the error matrices produced a high enough number, greater than 1.96 (Congalton and Green, 2008), to state that these two error matrices are statistically different at a 95% confidence level.

Hypothesis H3-4 that weighted methods outperform non-weighted methods on the water network for the state of Kentucky has to be rejected because WTP performed significantly worse than TP in terms of overall accuracy and KHAT values. The difference is significant with a Z value of 577, within the 95% confidence value. It is likely that the implementation of this method impacts the result for the approach. To determine Thiessen polygons, the software implementation does an approximation of the Voronoi algorithm based on an axis-aligned grid with interpolated values for source points on the axis-aligned grid. However, the minimum grid size usable in this software implementation (set for performance) was not small enough for an accurate conversion between the grid calculations and the finalized Thiessen polygons; this conversion was considerably impacted by differing weights in the weighted approach. Therefore, this particular software setup may not be sufficient to judge the performance of weighted TP.

The limitations of this study include: (1) the ground truth was estimated based on public dataset that is not comprehensive for the service area, (2) the software had a flawed implementation of the WTP algorithm, (3) demand is estimated based on population (zoning is not taken into account), and (4) the method requires considerable time for a GIS and service area allocation expert to merge diverse data sources, match projections, and filter out anomalous data.

In this chapter, I introduced a new type of service area approximation based on transportation networks. Such approaches have not been applied previously for critical infrastructure service area approximation. On the water network in the state of Kentucky, the transportation-network-based methods outperformed all other methods, with the weighted LAO transportation-network-based method having the best accuracy.

Visual inspection showed a high degree of alignment between the road networks and water pipelines (FIGURE 42). One exception to that is on rural areas where residents have their own private water wells. As shown in FIGURE 67, the layer for water lines (in blue) is placed on top of the layer for roads (in red). This layer arrangement shows the roads in red show more often in rural areas where there are no water lines accompanying them. Another outcome of the visual comparison was that interstates and major highways do not typically have water lines that run in parallel with them. Excluding the interstates and big highways from network optimization could be a good design choice as a refinement to the approach. With other types of CI networks, there is likely to be overlap as there is with water pipelines. However this would be specific to the CI network type and depend on their inherent structure that requires them to be aligned with roads. I expect substantial overlap between road network and other utilities such as electric power, natural gas, cable services, landline communication, etc. However, networks such as wireless phone, wide area Wi-Fi and specific other transportation networks would not be as conforming.

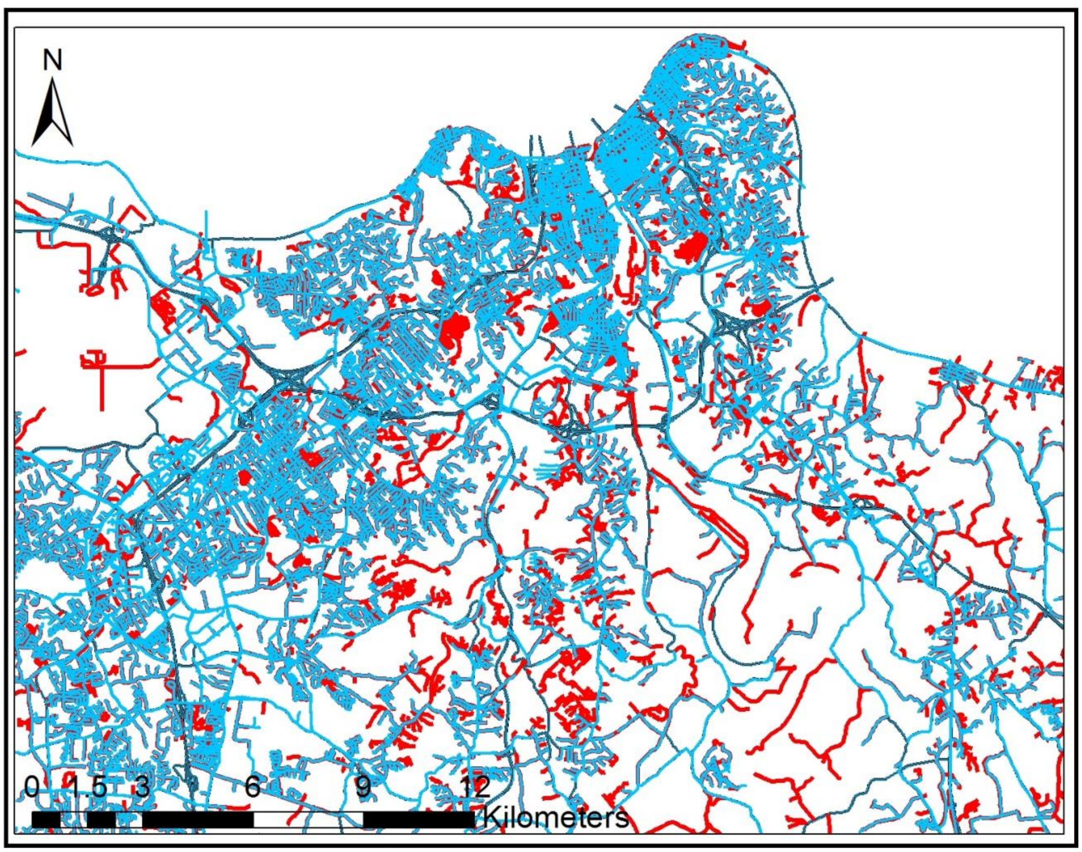


FIGURE 67: Water and road network overlap compared (blue: water, red: road)

CHAPTER 8: TRANSPORT-BASED APPROACHES FOR POWER NETWORK

Chapter 6 introduced and evaluated commonly employed distance- and cell-based techniques for service area approximation, which were comparatively evaluated in the context of a power dataset. Chapter 7 introduced two service area approximation methods based on road network optimization called Service Area Optimization (SAO) and Location Allocation Optimization (LAO). These were evaluated in the context of a larger dataset – the Kentucky state water distribution network, in comparison with the previously studied distance and cell-based techniques and SAO and LAO performed very well. In order to make a fuller comparison of results for discussion, here I evaluate the two new transport-based methods in the context of substation service area estimation for the same mid-size city power network that was studied in Chapter 6. This evaluation enables a more complete comparison of the possible differences in service area estimation methods across CI networks of different size and type, and will help to inform CI analysts about the considerations that must be taken in these different situations.

8.1 Data and methods

I obtained population data from the US Census Bureau website in TIGER GIS format for census blocks. Similarly, I also obtained the road data GIS layer from the U.S. Census Bureau website (Bureau, 2012). I used ArcGIS Network Analyst to create the network structure for driving directions to be used for the transport-based service area estimation methods SAO and LAO. LAO requires additional information about the capacity of the substations, while SAO does not take this into account.

To create the service areas, I first defined the facility locations and the road network. Then, to insure proper coverage, I set the maximum distance for the search reach to 25 km. ArcGIS Network Analyst then created service area polygons. As a last step for the SAO polygon layer, I clipped the resulting polygon set with city boundary layer. See FIGURE 68 for the SAO polygon layer along with the reference dataset for a small portion of the study area.

To create service areas based on LAO, I defined the substations as the source points with the "Load" output in megawatts as the capacity/weight value. I also defined census block points as the demand points. I set up the demand amount for each block point by multiplying the number of people in that block with the average per capita power usage value for the service area. Running the ArcGIS Network Analyst Location Allocation tool with the "Maximum Coverage Allocated" option and the settings noted above for source and demand points produces the optimized source-sink relationships. Carrying the relationship over to census block point areas and resolving based on the source facility ID produces the final service areas based on LAO. See FIGURE 69 for the service area polygons created with LAO, along with the reference dataset. FIGURE 70 shows a sample of the SAO service areas with assigned road lines, and FIGURE 71 shows an example of the LAO source-sink relationships used for service area creation. The load values shown in these figures are all in Megawatts.

I applied the same process used in Chapter 6 to compute the point data impact accuracy. First, I created a point dataset with 10K points randomly placed across the city's study area. The flowchart in FIGURE 72 shows the overall process – the attributes from the reference dataset and the SAO and LOA methods were transferred onto the point data layer through a series of spatial joins. To determine the overall accuracy, I divided the number of points where the method and reference sets agreed on the source by the total number of points (10K). Sample overlays of the accuracy check points for substation service areas created using SAO and LAO are shown in FIGURE 73 and FIGURE 74.

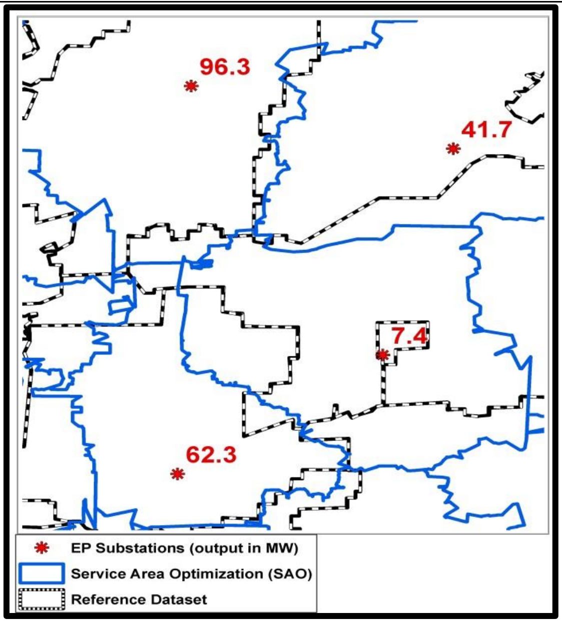


FIGURE 68: SAO EP service area polygons

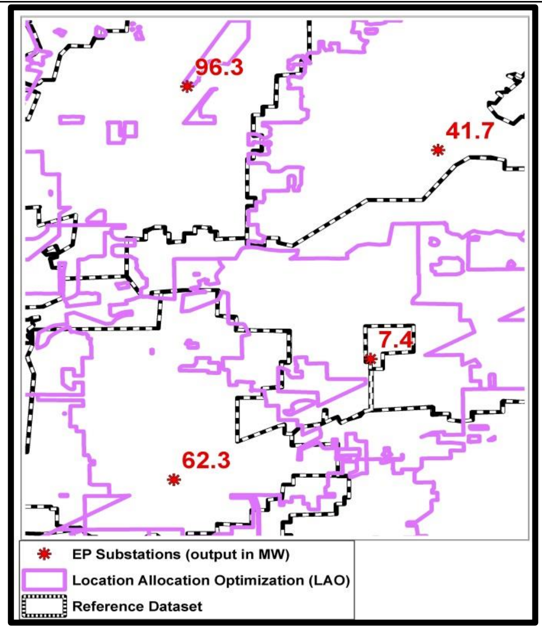


FIGURE 69: LAO EP service area polygons

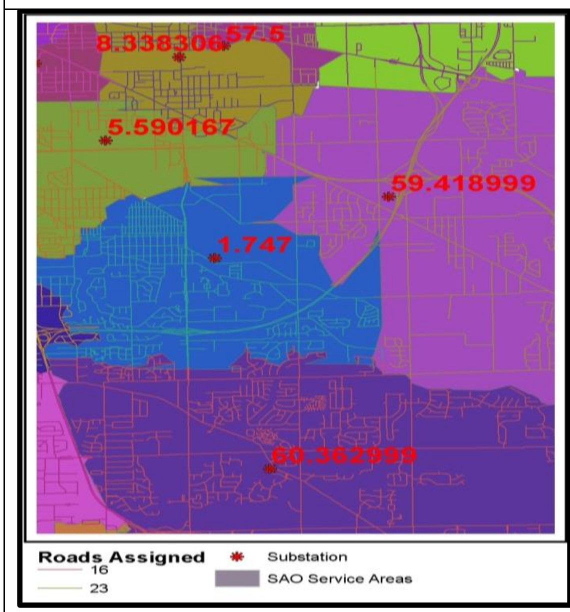


FIGURE 70: Road network & SAO EP service areas

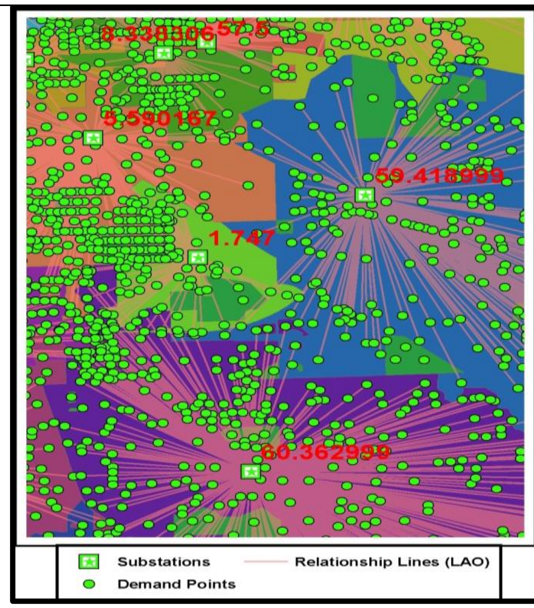


FIGURE 71: Source-sinks in LAO EP service areas

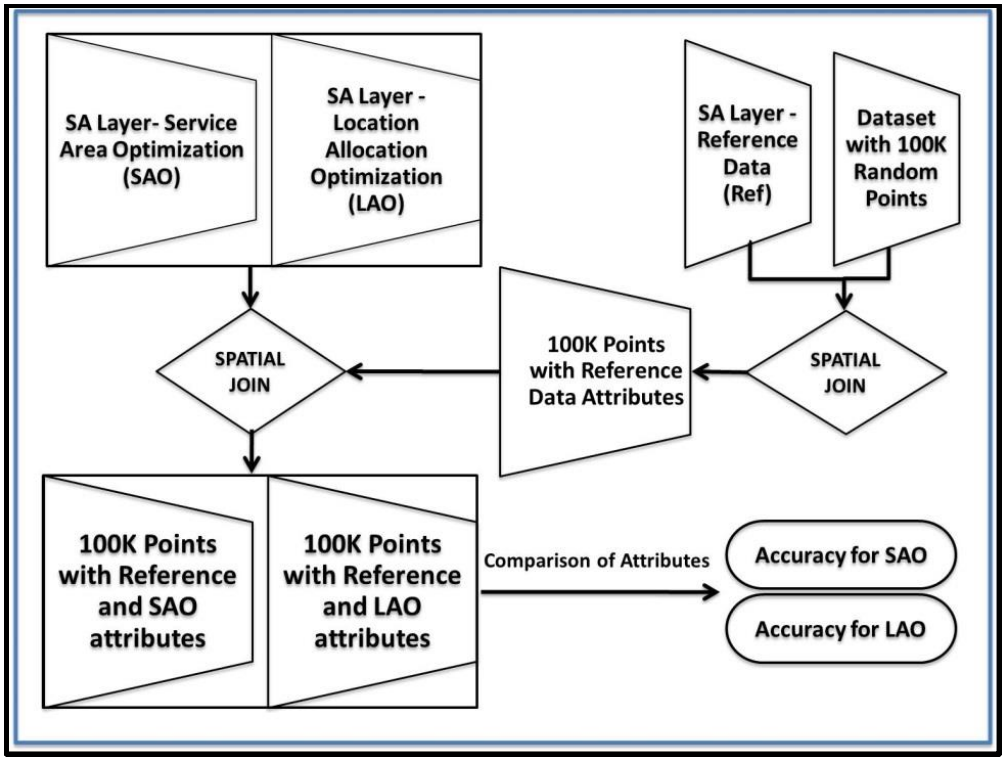


FIGURE 72: Point accuracy assessment flowchart for SAO and LAO

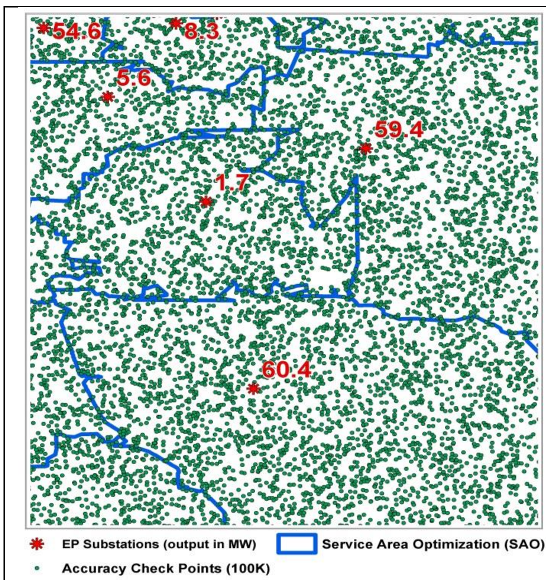


FIGURE 73: Sample SAO EP service areas with distribution of accuracy check points

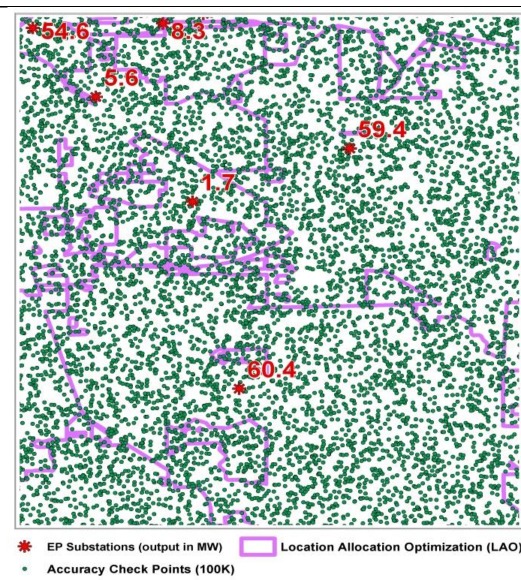


FIGURE 74: Sample LAO EP service areas with distribution of accuracy check points

8.2 Accuracy results for city-wide power network service area estimation

Overall accuracy results were lower than expected, with a point impact accuracy of 46.3% for SAO and 49.1% for LAO. This confirms the hypothesis H3-4 that the weighted method (LAO) outperforms the non-weighted method (SAO) for transport-based methods on electric power data at city scale. As shown in TABLE 9, when comparing the LAO and SAO accuracies for electric power service area estimation to those for distance- and cell-based methods as presented in Chapter 6, I found that the accuracies for LAO and SAO are lower than all the other methods. To understand the overall lower accuracy, I performed a visual inspection of the data from SAO and LAO overlaid with accuracy check points colored according to agreement with the reference dataset. FIGURE 75 and FIGURE 76 show the distribution of points with positive agreement on the source substation (green) and negative agreement (red).

TABLE 9: Accuracy of distance, cell, and transport methods on EP service area estimation for a mid-size midwestern US city

Method	WTP	WCA	TP	CA	LAO	SAO
Accuracy	68.9%	59.5%	54.1%	52.3%	49.9%	47.0%

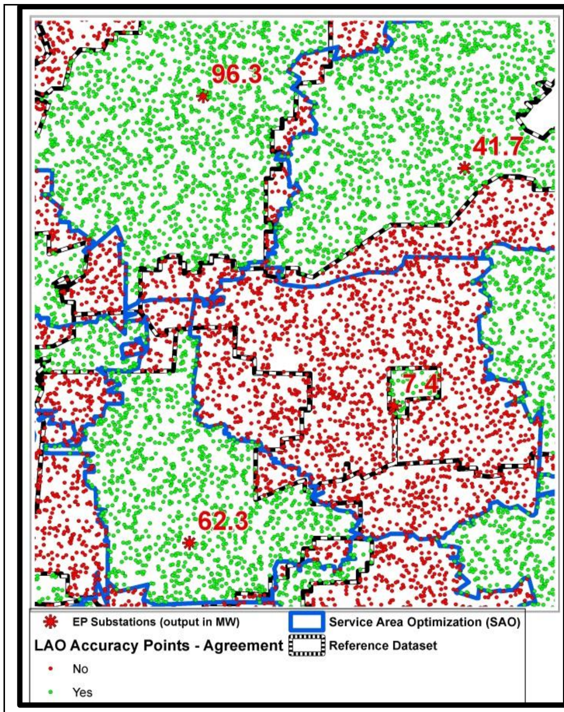


FIGURE 75: Sample of the visual point impact analysis for SAO

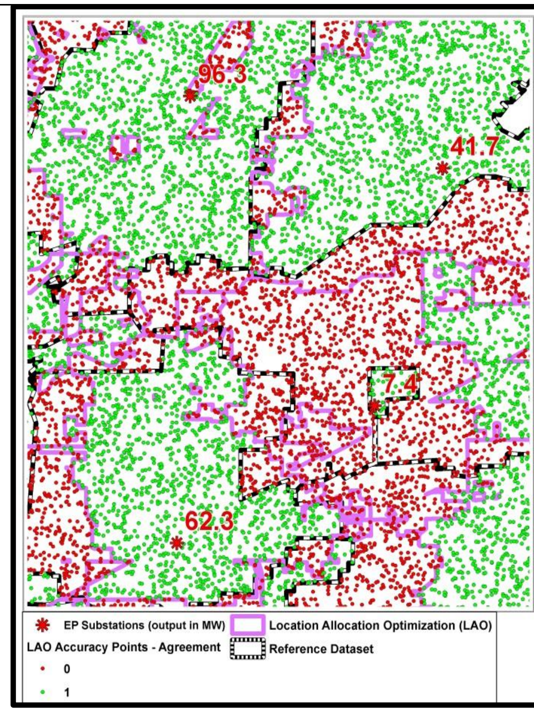


FIGURE 76: Sample of the visual point impact analysis for LAO

The primary reason for the lower accuracy of methods using road network for estimation appears to be the presence of industrial sites where there are the substations with a high power output used over a small amount of land. In addition, there are many large roads in proximity to industrial facilities, leading to higher numbers of source-sink connections between the substations presumably serving industry, and the residential data analyzed for accuracy. Because of this, the error introduced by industrial complexes using large amounts of power seems to greatly impact the accuracy of transport network based service area approximation methods. These results highlight the need to include industrial power usage to create source-sink relationships, instead of only population data combined with average power consumption per person. Other ways to mitigate this impact could be to eliminate the substations dedicated to industrial sites, or put barriers around them to reduce the amount of error they may be introducing.

Another contributing factor to differences in accuracy may be in the makeup of the study area. A statewide water network makeup is different than a city's power network makeup in terms of density and structure. In most states, the statewide water network includes cities and small towns with long distances between them. On the other hand, a city's electric power network is a tight grid of power lines and roads. Therefore, a transport-based method will have more difficulty with service area estimation in a city, where there are significantly more paths between sources and sinks. In a city, there are also fewer fluctuations in power output across the substations. The distribution of the substations and their power outputs are more uniform than the distribution and outputs of statewide water treatment plants. In a state water network, water treatment plants in one or two big cities provide two more orders of magnitude more water than the majority of the other water treatment plants. Methods based on the transportation network using source and demand amounts and locations compensate for this well, as the high density population in the big cities absorbs the source output. On the other hand, distance- and cell-based methods do not take the demand into account cannot contain outlier source points with very high outputs.

CHAPTER 9: COMPARISON OF SERVICE AREA ESTIMATION METHODS ACROSS NETWORKS

This chapter compares the accuracies for distance-, cell-, and transport-based service area estimation across the power and water network studies to understand how the methods differ across network types and study area sizes. In Chapter 7, RQ3 asked “Will applying metrics for transport optimization to service area estimation improve accuracy in comparison to common techniques?” Chapter 7 results showed transport-based methods to be more accurate for a state-wide water network, but follow-up results in Chapter 8 showed that accuracy was not improved using transport-based methods in the context of a city power network.

Research question RQ2 was a more general question about the differences in effectiveness among various service area estimation techniques for CI enablement scenarios. In this chapter, I compare the accuracy results across both study areas and network types noting that weighted SA estimation techniques produce more accurate results compared to their non-weighted counterparts across both network types and study areas. I expected the accuracy of the methods to be highest for transport-based, then cell-based, then distance-based as we saw in the Chapter 8 for the state water network. However, this chapter shows that changing the characteristics of the study area, network type, and the geographic extent can change the results dramatically; this ordering is reversed for electric power network service area estimation in a city.

9.1 Results of comparison across networks and methods

TABLE 10 summarizes the overall accuracies of the compared methods in the studies of water and power networks and FIGURE 77 shows these accuracies in a bar chart.

TABLE 10: Overall point accuracies for service area estimation methods

Network and study area type	**TP**	**WTP**	**CA**	**WCA**	**SAO**	**LAO**
Electric Power - City	54.1	68.9	52.3	59.5	47.0	49.9
Water - State	71.8	47.3	70.3	72.4	76.8	84.7

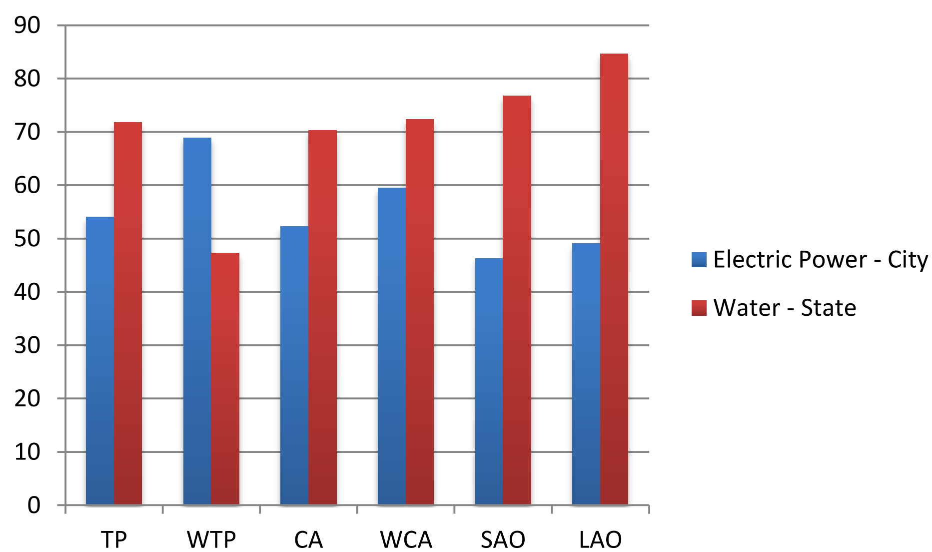


FIGURE 77: Chart of overall point accuracies for service area estimation across networks

9.2 Discussion of comparative accuracy across methods and networks

In Chapter 6 I hypothesized that cell-based service area (SA) estimation techniques produce more accurate results compared to distance-based ones (H2-1, rejected) and weighted SA estimation techniques produce more accurate results compared to their non-weighted counterparts (H2-2, accepted). In Chapter 7, the scope of H2-2 was extended to consider also the SAO and LAO techniques (H3-4) hypothesizing that the weighted SA estimation techniques produce more accurate results compared to their standard counterparts. However, H3-4 had to be rejected in the context of the water network for the state of Kentucky, because WTP performed significantly worse than TP in terms of overall accuracy and KHAT values.

In Chapter 7 I hypothesized that road-network-based service area optimization (SAO) will produce more accurate point impact results (H3-1, accepted) and more accurate aggregate impact (area estimation) results (H3-2, accepted) compared to cell-based (CA) or distance-based (TP) estimations. In addition I also hypothesized that road-network-based location allocation optimization (LAO), weighted using capacities and demands will produce more accurate point and area results than all other methods (H3-3, accepted). In considering the extension of these hypotheses from the water network context to include the power network context, the hypotheses must be rejected for the city-wide EP dataset, since accuracy values for LAO are lower than the other previously used methods.

Comparing the methods across both primary datasets, there are no clear-cut winners in service area estimation. For the water network across the state of Kentucky, the transport-based methods LAO and SAO outperform all other methods. However, for the electric power network in a mid-size Midwestern US city, these methods are the least accurate. This is likely because of the differential impact of industrial utility usage, and could also be affected by the large difference in the sizes of the study areas investigated. Much of the water used by industrial applications does not require treatment, so therefore the demand for these applications is not included in the average water usage per person and is also not included in the outputs of the water treatment plants. On the other hand, industrial usage of power does come from the power network, but is not accurately localized by using census data.

Another difference is apparent when comparing transport methods to standard methods: the improvement in accuracy is marked for weighted distance- and cell-based methods for the electric grid in a mid-size US city, but is less pronounced for the transport methods. Conversely, the difference in weighted methods for the water network is more pronounced than the differences between weighted and non-weighted distance and cell-based estimates for water across the state of Kentucky. These differences likely arise from the difference in the size of the study area, spatial and structural makeup of the datasets, and limitations on the implementation of some of the methods.

For the statewide water network, there are more source water treatment plans, with only a few very large plants and many substantially smaller ones. This disparity minimizes the impact of weighting over such a large area for the distance and cell-based methods. In other words, there is likely an averaging effect over the large areas, equalizing weighted and non-weighted distance- and cell-based methods. However, the transport networks likely reflect the population distribution quite well, so the weighted transport method was likely more accurate because of the alignment of the roads, water networks, and census block data, improved with source capacity data.

For the citywide electric power network, on the other hand, weighted distance and cell-based methods were significantly more accurate than their standard counterparts. This is likely due to the smaller area considered, making it important to take source capacities into account. Both transport network-based methods suffer from the lack of industrial usage data. In addition, for an electric power network, the source output distribution is more uniform, with no large outliers, helping distance and cell-based methods work effectively. On the other hand, transport-based methods are not as efficient in approximating the service areas because of the dense road and EP networks in a tight grid form.

Differences in performance also arise for distance (TP, WTP) and cell-based (CA, WCA) methods across the two networks. Within the same network, the standard CA and TP methods have similar accuracies. For the electric power service area estimation, the weighted methods outperformed the standard methods considerably. I note that the water network accuracy results for WTP are likely due to the implementation rather than the method. WCA did not improve CA results for the water network drastically.

9.3 Cross-network service area estimation accuracy conclusions

Based on the results presented in Chapters 6, 7 and 8, we can conclude that, across the study areas and networks, weighted methods are more accurate, with the exception of weighted distance-based methods on the water network. This discrepancy is likely due to implementation details in the tools used rather than the method itself. We also can reasonably conclude that the ordering of accuracies for the statewide water network is transport-based, then cell-based, then distance-based, but for the city power network, the accuracies are in reverse order. This difference is likely due to the impact of the lack of industrial zoning data for electric power demands while their usage is still included in source EP substation capacities. However, water usage from water treatment plants is not as impacted by this zoning difference because plants are likely not providing large volumes of water for industrial use, since this kind of water does not need the same treatment as drinking water. Another likely reason for the difference in the results is due to the differences in spatial distribution of the source points, their service output amounts and structural makeup of the networks across the study areas.

The comparison presented in this chapter is of particular importance in the design and development of future CI analysis tools, which will include multiple networks with potential cross-infrastructure dependencies. The results highlight the need to choose service area estimation methods that are appropriate for the network type and size, the availability of population and zoning data, not to mention geospatial feature data. In other words, these results show that applying the same method across networks may not be advisable. In particular, CI analysts will need to consider the purpose of the analysis, the size and features of the geographic study area, and the features of the utility network in selecting service area approximation methods. In addition, availability of source capacities, detailed population data and road network data must be a decision factor while selecting a method. All the weighted methods require the use the source capacities as a parameter. In addition to source capacities, both transport-based optimization methods require detailed road network data and weighted transport-based method (LAO) also requires detailed demand data. The results here also provide a starting point for reference in terms of design guidance for analysis considerations in electric power and water distribution networks at city and state scales.

CHAPTER 10: CONTRIBUTIONS, LIMITATIONS, AND CONCLUSION

This chapter summarizes the contributions, limitations, and future directions of my dissertation research in support of decision-making and analysis for critical infrastructures.

10.1 Summary of research contributions

In this dissertation, I have sought to investigate methods for effective decision support for critical infrastructure managers, including aspects of user support and accuracy of service area estimation methods, both of which are of high importance in making correct, timely decisions in the face of emergencies. I have developed a Decision Recommender Tool (DRT) framework and a prototype Critical Infrastructure Explorer (CIE) decision support platform described in Chapter 4 for use during infrastructure reconstitution. The DRT framework models users, targets (e.g. economic impact), and a simulation engine in an interactive geovisualization environment. A major benefit of this framework is that it accounts for cross-infrastructure interdependency.

Chapter 5 showed that decision makers preferred the prototype Critical Infrastructure Explorer (CIE) over common GIS tools for reconstitution tasks, and CIE allowed decision makers to make better decisions with (1) less time and (2) less cognitive load. The study confirmed our hypothesis that tools tailored to support critical infrastructure decision-making are needed for more effective and efficient decision-making.

In CI modeling and analysis, data of good quality are difficult to access for many reasons. The result is a reliance on estimation methods to analyze impacts of specific outages on populations, regional economies and other critical infrastructure elements. During the study of CIE, GIS experts expressed that accuracy estimation methods were of critical importance, greatly impacting the trust that experts would place in a decision support system. Therefore, the dissertation research studies that followed had a focus on evaluating the accuracy of service area estimation methods. Choosing the correct estimation methods to provide the highest possible accuracy for decision makers is critical in establishing trust and ensuring that decisions will not have unexpected negative impacts. Inaccuracy in service area estimation could be particularly important when making decisions that involve cross-infrastructure effects.

Chapter 6 studied the accuracy of two commonly used estimation methods on a power network dataset from a midsize Midwestern US city, demonstrating that weighted approaches are more accurate, as hypothesized. Weighted cellular automata (WCA) was the strongest performer estimating aggregated area impacts, however WTP had better accuracy for estimating point impacts. I found that cell-based methods have limited ability to balance capacities within a bounded service region. Adjustments to cell-based methods to handle service area boundaries may make them more accurate.

Chapter 7 studied the accuracy of common methods and two new transport network-based methods on the water network for the state of Kentucky, confirming the hypothesis that accuracy is improved by using transportation network optimization, and that weighted approaches were generally still more accurate. However, this study revealed a limitation of the ArcGIS implementation of the weighted distance-based method on the large area of a state. I also applied Kappa analysis for comparison of classification methods to CI analysis. This provides greater confidence in results through significance testing for pairwise comparison of service area estimation methods across networks and study areas.

Chapter 8 applied the new transportation-based methods to the electric power dataset used in Chapter 6, showing dramatically different results. While the water network was reflected accurately by the road network, transportation methods were much less accurate than distance- and cell-based methods in the context of a power network. I believe that the reason for this is that there was no zoning information included in the power reference datasets. The impact of this source of inaccuracy is not large for water treatment plants, since industrial uses of water such as cooling do not necessarily require water treatment and as such that unrepresented demand may also not be represented in the water treatment plant data. However, for electric power networks, most power is provided by the utility, so a dataset that does not include the location of high-demand points for industry may have very skewed results.

In Chapter 9, I compared the overall point accuracies across the methods and networks. Applying all six methods to a city electric power network and a state water network showed potential tradeoffs. In general, weighted methods outperformed non-weighted methods. Cell-based methods fell in the middle in both cases. This suggests that cell-based methods may be less sensitive to changes in area size than distance- or transport-based methods. Transport-based methods considerably outperformed other methods in the water network, while they performed the most poorly in the citywide electric power network.

The arc of this dissertation highlights the importance of providing accurate, tailored support for critical infrastructure decision-making. With a decision recommender tool, decision makers can make decisions more quickly, more correctly, and with less cognitive load, given that the system is relying upon accurate data. Because of very limited access to critical infrastructure service area data, it is of the utmost importance to understand the accuracy of service area estimation methods for decision support recommendation. To this end, I have developed two novel models for service area estimation based on transportation network optimization, showing them to be very accurate for state-wide water network service area estimation. I have demonstrated that, in general, weighted methods are more accurate, across network type and study area size. Transport-based service area estimation methods using population sink data did not work well for electric power service area in a city, where uniform populations and transport networks belie possible missing demand information from industrial and commercial zones.

Overall, the outcomes of this research provide insight for researchers and practitioners in geospatial information systems, and particularly for developers and users of applications for critical infrastructure modeling, analysis, and decision-making. Analysts and decision makers can benefit from tools and environments that enable new kinds of interaction, which can support improved analysis and thereby decision-making. And since a key enabling element is accurate knowledge about the infrastructures themselves, helping system developers and analysts understand the tradeoffs involved in modeling infrastructures and infrastructure interactions in different contexts can lead to more accurate modeling, enabling improved analysis and decision-making. Results of the accuracy analysis here indicate that decision makers and system designers should weight tradeoffs and select carefully when applying different service area approximation methods to CI source-sink analysis.

The practical application of this research is grounded in two primary ways. Much of the service area estimation work was conducted and published in collaboration with critical infrastructure analysts and researchers at Los Alamos National Lab, who employ these kinds of tools in response to real critical infrastructure scenarios. In addition the team at the Kentucky Infrastructure Authority expressed interest in making use of the byproducts of this research.

10.2 Limitations and Future Work

Results obtained through the experiments performed through this dissertation research also include some limitations that need to be listed here. For instance, CIE explained and Chapter 4 and implemented and tested in Chapter 5 is a prototype and needs many improvements and additions to make it useful for real life decision support during a disaster. Also accuracies of different methods varies according to the data and location it is applied to and therefore more research is required to have deeper understanding of their functioning. Here is a concise list of the primary limitations of this work:

1. The CIE tool is a prototype. Many more features are needed to make such a framework feasible for actual emergency decision support, particularly support for multiple networks, multiple criteria, prioritized appropriately for diverse emergency scenarios, and ideally real-time adaptation to changing situations and having live connections to dynamic data sources.

2. More studies on the accuracy of service area approximations are needed to understand the robustness of the findings presented, particularly to understand the impact of diverse CI networks, geographical features, or land-use type (such as residential/ industrial/ commercial).

3. More access to real datasets is needed for other researchers to replicate the kinds of analyses done here.

4. Integration for accuracy assessment and scenario simulations are needed for decision makers to practice before emergencies occur, and this practice could allow researchers and developers to discover new needs or considerations for CI decision-making software support.

Although our results are quite promising, the prototype that I created and described herein is only a first step in supporting decision-making and analysis for critical infrastructure. What is needed is a complex system that is able to analyze multiple critical infrastructure networks using multiple ranking factors to provide recommendations and support for the decision maker. Since CIE was only an experimental tool, it does not include functionality to run scenarios across multiple networks, or to take into account more than one priority criteria for decision support.

An ideal CI decision support system would help CI analysts and decision makers by accurately and quick describing the situation and recommending scenarios that best target areas for recovery. It is also important to provide transparency in suggested prioritizations and in the strengths and limitations of the service area approximation methods. This transparency will help decision makers be more efficient and effective while being aware of the possible pitfalls of using particular tools and methods.

More research is needed to describe the potential trade-offs between distance-, cell-, and transport-based methods for service area estimation as applied to diverse critical infrastructure network types and to study areas of varying sizes. This research could benefit from both empirical comparisons and formal probability-based methods. In reality some networks, like EP are dynamic in nature and realistically would need to be modelled in a dynamic fashion to have more real life applications. Adding estimation approaches such as agent-based modeling and simulation would allow for more dynamic modeling that can adapt to changing EP loads due to load balancing and fluctuations in demand, therefore service areas.

10.3 Conclusion

Integrated CI analysis tools can support improved decision-making, but rely on the accuracy of the data and estimation techniques they are built upon. The studies in this dissertation have shown that straightforward application of an individual technique is not possible, but that CI analysts and decision makers will need to be apprised of the limitations of each service area estimation technique, or the support software will need to select from among techniques depending on network type, study area size, study area population composition and availability of additional data such as road network and demand locations. Below is a brief list of selection criteria based on the results of this dissertation:

- Transport-based methods seem to perform better overall in state-wide sparse networks such as water, if detailed road network data is available. If the demand data and source capacities are also available, use of location allocation is more appropriate.

- When road network and demand data are not available, weighted versions of cellular automata are likely the best option in state-wide sparse networks such as water.
- Thiessen polygon methods weighted by source capacity are expected to perform best in city-scale power networks with many substations, considering point impact analysis criteria.
- Cellular automata methods weighted by source capacity are expected to perform best in city-wide power networks with many substations, considering aggregate impact analysis criteria. Weighted CA methods tend to perform better than their standard counterparts.

However, care must be taken to assure that the implementation of the estimation methods can handle the study area size. More study may show that transport-based methods, weighted using source capacities and information on sinks both from population and zoning information will be most accurate. Overall, this dissertation has shown that both integrated decision support and careful contextualization of service area models across different types of CI networks are essential to address the needs for complex, multi-dimensional analysis of critical infrastructure networks.

REFERENCES

Akabane, K., Nara, K., Mishima, Y., & Tsuji, K. (2002). Optimal geographical allocation of power quality control centres by Voronoi diagram. In *Proceedings of the 14th Power Systems Computation Conference*.

Al-Rasheed, K., & El-Gamily, H. I. (2013). GIS as an Efficient Tool to Manage Educational Services and Infrastructure in Kuwait. *Journal of Geographic Information Systems,* 5(1), 75-86.

Andersson, G., Donalek, P., Farmer, R., Hatziargyriou, N., Kamwa, I., & Kundur, P. (2005). Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. *IEEE Transactions on Power Systems*, 20(4), 1922-1928.

Apostolakis, G. E., & Lemon, D. M. (2005). A screening methodology for the identification and ranking of infrastructure vulnerabilities due to terrorism. *Risk Analysis*, 25(2), 361-376.

Aradau, C. (2010). Security that matters: critical infrastructure and objects of protection. *Security Dialogue*, 41(5), 491-514.

Argany, M., Mostafavi, M. A., Karimipour, F., & Gagné, C. (2011). A GIS based wireless sensor network coverage estimation and optimization: A Voronoi approach. *Transactions on Computational Science XIV* (151-172). Springer.

Arianos, S., Bompard, E., Carbone, A., & Xue, F. (2009). Power grid vulnerability: A complex network approach. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(1).

Aurenhammer, F. (1991). Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23(3), 345-405.

Authority, K. I. (2013). Water Resources Information System. From http://kia.ky.gov/wris/

Baloye, D., Adesina, F., & Kufoniyi, O. (2010). A spatial decision support system approach to sustainable physical development planning. *Journal of Geography and Regional Planning*, 3(9), 219-233.

Barto, A. G. (1975). Cellular automata as models of natural systems. Technical Report. The University of Michigan Engineering Library.

Belton, V., & Stewart, T. (2002). *Multiple Criteria Decision Analysis: An Integrated Approach*: Springer.

Berto, F., & Tagliabue, J. (2012). Cellular Automata. In Zalta, E. N. (ed.). *The Stanford*

Encyclopedia of Philosophy (Summer 2012 Edition).

Bhaduri, A., & Kloos, J. (2013). Getting the Water Prices Right Using an Incentive-based Approach: An Application of a Choice Experiment in Khorezm, Uzbekistan. *European Journal of Development Research*, 25(5), 680-694.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (2007). *Discrete multivariate analysis: theory and practice*. Springer.

Bochicchio, C., Fletcher, C., Dyer, M., & Smith, T. (2009). Reef-top sediment bodies: windward O'ahu, Hawai'i. *Pacific Science*, 63(1), 61-82.

Boots, B. N. (1986). *Voronoi (Thiessen) Polygons (Vol. 45)*. Geo books. Norwich, UK.

Bottero, M., Comino, E., Duriavig, M., Ferretti, V., & Pomarico, S. (2013). The application of a Multicriteria Spatial Decision Support System (MCSDSS) for the assessment of biodiversity conservation in the Province of Varese (Italy). *Land Use Policy*, 30(1), 730-738.

Bright, E. A., Coleman, P. R., Rose, A. N., & Urban, M. L. (2012). *LandScan 2011* [digital raster data]. From: http://www.ornl.gov/landscan/

Bureau, U. C. (2012). TIGER Products. 2014, From https://www.census.gov/geo/maps-data/data/tiger.html

Burrough, P. (1990). Methods of spatial analysis in GIS. *International Journal of Geographical Information Systems*, 4(3), 221-223.

Bush, B. (2005). Interdependent Energy Infrastructure Simulation System - IEISS Technical Reference Manual: *Los Alamos National Laboratory Technical Report*. LANL-D4-05.

Buttyán, L., Gessner, D., Hessler, A., & Langendörfer, P. (2010). Application of wireless sensor networks in critical infrastructure protection: Challenges and design options [Security and Privacy in Emerging Wireless Networks]. *Wireless Communications*, IEEE, 17(5), 44-49.

Castongia, S. M. (2006). *A Demand- Based Resource Allocation Method for Electrical Substation Service Area Delineation*. University of North Carolina at Charlotte.

Cavdaroglu, B., Hammel, E., Mitchell, J. E., Sharkey, T. C., & Wallace, W. A. (2013). Integrating restoration and scheduling decisions for disrupted interdependent infrastructure systems. *Annals of Operations Research*, 203(1), 279-294.

Cetinkaya, E. K., Broyles, D., Dandekar, A., Srinivasan, S., & Sterbenz, J. P. (2010). A comprehensive framework to simulate network attacks and challenges. In *Proceedings of the 2010 International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*.

Chopard, B. (2012). Cellular automata modeling of physical systems. *Computational Complexity: Theory, Techniques, and Applications*, 407-433.

Chu, C. Y. (2009). A computer model for selecting facility evacuation design using cellular automata. *Computer‐Aided Civil and Infrastructure Engineering,* 24(8), 608-622.

Clinton, W. (1998). Presidential Decision Directive 63. The White House, Washington, DC.

Coffrin, C., Van Hentenryck, P., & Bent, R. (2011). Strategic planning for power system restoration. In *Proceedings of the International Conference on Vulnerability and Risk Analysis and Management*.

Coffrin, C., Van Hentenryck, P., & Bent, R. (2012). Last-Mile Restoration for Multiple Interdependent Infrastructures. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213.

Cohen, J. P. (2010). The broader effects of transportation infrastructure: Spatial econometrics and productivity approaches. *Transportation Research Part E: Logistics and Transportation Review,* 46(3), 317-326.

Collins, M., Petit, F., Buehring, W., Fisher, R., & Whitfield, R. (2011). Protective measures and vulnerability indices for the Enhanced Critical Infrastructure Protection Programme. *International Journal of Critical Infrastructures,* 7(3), 200-219.

Congalton, R. G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment,* 37(1), 35-46.

Congalton, R. G., & Green, K. (1999). *Assessing the accuracy of remotely sensed data: principles and practices*: CRC Press.

Congalton, R. G., & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*: CRC press.

Cressie, N. (1991). *Statistics for Spatial Data*. John Wiley and Sons Inc., New York, NY.

De Berg, M., Van Kreveld, M., Overmars, M., & Schwarzkopf, O. C. (2000). *Computational Geometry*: Springer.

Densham, P. J. (1991). Spatial decision support systems. *Geographical Information Systems: Principles and Applications,* 1, 403-412.

Dong, P. (2008). Generating and updating multiplicatively weighted Voronoi diagrams for point, line and polygon features in GIS. *Computers & Geosciences,* 34(4), 411-421.

Dowell, L. J., & Maheshwari, S. (2000). Simulating earthquake damage to the electric-power infrastructure: a case study for urban planning and policy development. *Los Alamos National Laboratory Technical Report* LA-UR-00-3777, Los Alamos, NM.

Dueñas-Osorio, L., & Vemuru, S. M. (2009). Cascading failures in complex infrastructure systems. *Structural safety,* 31(2), 157-167.

Ebrahimy, R. (2014). *Investigating SCADA Failures in Interdependent Critical Infrastructure Systems*. arXiv preprint arXiv:1404.7565.

Ehlen, M. A., Loose, V. W., Vargas, V. N., Smith, B. J., Warren, D. E., Downes, P. S., Eidson, E. D., & Mackey, G. E. (2010). Economics definitions, methods, models, and analysis procedures for Homeland Security applications. *Technical Report No: SAND2010-4315*. Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States).

England, G. (2005). Military Support for Stability, Security, Transition, and Reconstruction (SSTR) Operations. *Department of Defense Directive*, 3000, 28.

ESRI (2014). ARCGIS Resource Center. From http://resources.arcgis.com/en/home/

Ewers, M. (2008). FastEcon Tool Summary Report Fiscal Year 2008: Los Alamos National Laboratory.

Fazel Zarandi, M. H., Davari, S., & Haddad Sisakht, S. A. (2011). The large scale maximal covering location problem. *Scientia Iranica,* 18(6), 1564-1570.

Fenwick, J. W., & Dowell, L. J. (1999). Electrical substation service-area estimation using cellular automata: an initial report. In *Proceedings of the 1999 ACM symposium on Applied Computing*, New York, NY.

Fenwick, L. J., & Lyne, M. C. (1999). The relative importance of liquidity and other constraints inhibiting the growth of small-scale farming in KwaZulu-Natal. *Development Southern Africa,* 16(1), 141-155.

Ferretti, V. (2011). A multicriteria spatial decision support system development for siting a landfill in the province of Torino (Italy). *Journal of Multi‐Criteria Decision Analysis,* 18(5-6), 231-252.

Finley, D. L., & Sanders, M. G. (1994). High transfer training (HITT): Instruction development procedures and implementation strategies. In *Proceedings of the Interservice/Industry Training, Simulation & Education Conference*.

Flowerdew, R., & Green, M. (1994). Areal interpolation and types of data. *Spatial*

Analysis and GIS, 121, 145.

Foody, G. M. (2010). Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sensing of Environment*, 114(10), 2271-2285.

Franco, D. O., Yang, L. I., & Hammer, A. E. (2012). *Critical Infrastructure and Economic Impact Considerations for Recovery from Chemical, Biological, and Radiological Incidents*. DTIC Document, United States Department of Homeland Security. Office of Science and Technology.

Gahegan, M. (2000). The case for inductive and visual techniques in the analysis of spatial data. *Journal of Geographical Systems*, 2(1), 77-83.

Gahegan, M. (2000). On the application of inductive machine learning tools to geographical analysis. *Geographical Analysis*, 32(2), 113-139.

Gahegan, M., & Ehlers, M. (2000). A framework for the modelling of uncertainty between remote sensing and geographic information systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, 55(3), 176-188.

Gahegan, M., & Lee, I. (2000). Data structures and algorithms to support interactive spatial analysis using dynamic Voronoi diagrams. *Computers, Environment and Urban Systems*, 24(6), 509-537.

Gahegan, M., Takatsuka, M., Wheeler, M., & Hardisty, F. (2000). GeoVISTA Studio: a geocomputational workbench. In *Proceedings of the GeoComputation Conference*.

Garbolino, E., Lachtar, D., Sacile, R., & Bersani, C. (2013). Vulnerability and Resilience of the Territory Concerning Risk of Dangerous Goods Transportation (DGT): Proposal of a Spatial Model. *Chemical Engineering Transactions*, 32.

Geoffrion, A. M. (1974). *Lagrangean relaxation for integer programming*: Springer.

George, S. M., Zhou, W., Chenji, H., Won, M., Lee, Y. O., Pazarloglou, A., Stoleru, R., & Barooah, P. (2010). DistressNet: a wireless ad hoc and sensor network architecture for situation management in disaster response. *Communications Magazine*, IEEE, 48(3), 128-136.

Getis, A. (1994). Spatial dependence and heterogeneity and proximal databases. In Fotheringham, S., and Rogerson, P.(eds.). *Spatial Analysis and GIS*, 105-120. Taylor and Francis Ltd, London.

Ginevan, M. E. (1979). Testing Land-Use Map Accuracy - Another Look. *Photogrammetric Engineering and Remote Sensing*, 45(10), 1371-1377.

Gomes, E., & Lins, M. E. (2008). Modelling undesirable outputs with zero sum gains data envelopment analysis models. *Journal of the Operational Research Society*, 59(5), 616-623.

Gomes, E. G., & Lins, M. P. E. (2002). Integrating geographical information systems and multi-criteria methods: a case study. *Annals of Operations Research,* 116(1-4), 243-269.

Goodchild, M. F. (1990). Keynote address: spatial information science. In *Proceedings of the Fourth International Symposium on Spatial Data Handling* (Zurich, July 1990).

Goodchild, M. F. (1993). The state of GIS for environmental problem-solving. *Environmental Modeling with GIS*, 8-15.

Goodchild, M. F., Anselin, L., & Deichmann, U. (1993). A framework for the areal interpolation of socioeconomic data. *Environment and Planning A*, 25(3), 383-397.

Greene, W. H. (2002). LIMDEP: Version 8.0: Econometric Modeling Guide. *Econometric Software*.

Gronlund, A. G., Xiang, W.-N., & Sox, J. (1994). GIS, expert system technologies improve forest fire management techniques. *GIS World*, 7(2), 32-36.

Guikema, S. D. (2009). Natural disaster risk analysis for critical infrastructure systems: An approach based on statistical learning theory. *Reliability Engineering & System Safety,* 94(4), 855-860.

Haimes, Y. Y. (2006). On the definition of vulnerabilities in measuring risks to infrastructures. *Risk Analysis*, 26(2), 293-296.

Haining, R. (1990). Models in human geography: problems in specifying, estimating and validating models for spatial data. *Spatial Statistics: Past, Present and Future*. Michigan Document Services Ann Arbor, 83-102.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload,* 1, 139-183.

Hauer, J. F., Bhatt, N. B., Shah, K., & Kolluri, S. (2004). Performance of "WAMS East" in providing dynamic information for the North East blackout of August 14, 2003. In *Proceedings of the Power Engineering Society General Meeting*, 2004. IEEE.

Havlin, S., Kenett, D., Ben-Jacob, E., Bunde, A., Cohen, R., Hermann, H., Kantelhardt, J., Kertész, J., Kirkpatrick, S., & Kurths, J. (2012). Challenges in network science: Applications to infrastructures, climate, social systems and economics.

European Physical Journal-Special Topics, 214(1), 273.

Held, D. (2004). Globalisation: the dangers and the answers. *Open Democracy*, 27.

Hines, P., Cotilla-Sanchez, E., & Blumsack, S. (2010). Do topological models provide good information about electricity infrastructure vulnerability? *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(3), 033122.

Holgate, J., Williams, S. P., & Hardy, C. A. (2012). Information Security Governance: Investigating Diversity in Critical Infrastructure Organizations In *Proceedings of BLED Conference*.

Hord, R. M., & Brooner, W. (1976). Land-Use Map Accuracy Criteria. *Photogrammetric Engineering and Remote Sensing*, 42(5), 671-677.

Jankowski, P. (1995). Integrating geographical information systems and multiple criteria decision-making methods. *International Journal of Geographical Information Systems*, 9(3), 251-273.

Jenelius, E. (2010). User inequity implications of road network vulnerability. *Journal of Transport and Land Use*, 2(3).

Jenelius, E., Westin, J., & Holmgren, Å. J. (2010). Critical infrastructure protection under imperfect attacker perception. *International Journal of Critical Infrastructure Protection*, 3(1), 16-26.

Johansson, J., & Hassel, H. (2014). Impact of Functional Models in a Decision Context of Critical Infrastructure Vulnerability Reduction. In *Proceedings of the Second International Conference on Vulnerability and Risk Analysis and Management*.

Jumadi, J. (2013). Web-Based Spatial Decision Support System (SDSS) For Flood Risk Management In Surakarta: A Preliminary Result. In *Proceedings of Seminar Nasional Pendayagunaan Informasi Geospatial*.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-y., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4), 348-354.

Kaunda-Bukenya, N., Tadesse, W., Fu, Y., Tsegaye, T., & Wagaw, M. (2012). Spatial Decision Support System (SDSS) for Stormwater Management and Water Quality Assessment. In Voudouris, C. (ed.). *Water Quality Monitoring and Assessment*.

Keen, P. G. (1980). Adaptive design for decision support systems. *ACM SIGOA Newsletter*, 1(4-5), 15-25.

Kenny, J. F., Barber, N. L., Hutson, S. S., Linsey, K. S., Lovelace, J. K., & Maupin, M. A. (2009). Estimated use of water in the United States in 2005. *US Geological Survey Circular* 1344, (52).

Koger, M. S., & Landry, B. J. (2010). Personal mobile computing devices-the new perimeter. *Association Of Business Information Systems*, 45.

Koonce, A. M., Apostolakis, G., & Cook, B. (2008). Bulk power risk analysis: ranking infrastructure elements according to their risk significance. *International Journal of Electrical Power & Energy Systems*, 30(3), 169-183.

Koua, E. L., Maceachren, A., & Kraak, M. J. (2006). Evaluating the usability of visualization methods in an exploratory geovisualization environment. *International Journal of Geographical Information Science,* 20(4), 425-448.

Kreimer, A. (1991). Reconstruction after earthquakes: Sustainability and development. *Earthquake Spectra*, 7(1), 97-106.

Kulawiak, M., & Lubniewski, Z. (2013). SafeCity—A GIS-based tool profiled for supporting decision making in urban development and infrastructure protection. *Technological Forecasting and Social Change*.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Linger, H., & Burstein, F. (2001). From computation to knowledge management: The changing paradigm of decision support for meteorological forecasting. *Journal of decision systems*, 10(2), 195-215.

Linger, S. P., & Wolinsky, M. (2001). Estimating Electrical Service Areas Using GIS and Cellular Automata. In *Proceedings of the Esri International User Conference*.

Liu, Y., Paciorek, C. J., & Koutrakis, P. (2009). Estimating regional spatial and temporal variability of PM 2. 5 concentrations using satellite data, meteorology, and land use information. *Environ Health Perspective*. 117(6): 886–892

Mac Aoidh, E. (2008). *Implicit Interpretation of User's Interactions with Spatial Data*. University College Dublin.

Mac Aoidh, E., Bertolotto, M., & Wilson, D. C. (2008). Understanding geospatial interests by visualizing map interaction behavior. *Information Visualization*, 7(3-4), 275-286.

Mackie, K., & Cooper, C. (2009). Landfill gas emission prediction using Voronoi diagrams and importance sampling. *Environmental Modelling & Software*, 24(10), 1223-1232.

Maguire, D. J., & Longley, P. A. (2005). The emergence of geoportals and their role in spatial data infrastructures. *Computers, Environment and Urban Systems*, 29(1), 3-14.

Mahmood, A. N., Hu, J., Tari, Z., & Leckie, C. (2010). Critical infrastructure protection:

Resource efficient sampling to improve detection of less frequent patterns in network traffic. *Journal of Network and Computer Applications*, 33(4), 491-502.

Masucci, V., Adinolfi, F., Servillo, P., Dipoppa, G., & Tofani, A. (2009). Ontology-Based Critical Infrastructure Modeling and Simulation. *Critical Infrastructure Protection III* (229-242): Springer.

Matisziw, T. C., & Murray, A. T. (2009). Modeling s–t path availability to support disaster vulnerability assessment of network infrastructure. *Computers & Operations Research*, 36(1), 16-26.

Mbowe, J. E., & Oreku, G. S. (2014). Quality of Service in Wireless Sensor Networks. *Wireless Sensor Network*.

McPherson, T., & Burian, S. (2005). The Water Infrastructure Simulation Environment (WISE) Project. In *Proceedings of the Impacts of Global Climate Change*, (1-8).

Memmott, J. L., & Hanks, J. W. (1992). *Evaluation of upgrade to standards requirements for FM highways in Texas*. Texas Transportation Institute.

Mendonça, D., Cutler, B., Wallace, W. A., & Brooks, J. D. (2014). Collaborative Training Tools for Emergency Restoration of Critical Infrastructure Systems *New Perspectives in Information Systems and Technologies, Volume 1* (571-581): Springer.

Meyers, D. L. (2001). Asset Location: Client Goals and Objectives. In *Proceedings of the AIMR Conference*.

Michaud, D., & Apostolakis, G. E. (2006). Methodology for ranking the elements of water-supply networks. *Journal of Infrastructure Systems*, 12(4), 230-242.

Moreira, A. A., Andrade Jr, J. S., Herrmann, H. J., & Indekeu, J. O. (2009). How to make a fragile network robust and vice versa. *Physical Review Letters*, 102(1), 018701.

Network, K. G. (2014). KyGeoportal Home. From http://kygisserver.ky.gov/geoportal/catalog/main/home.page

Newton, K., & Schirmer, D. (1997). On the methodology of defining substation spheres of influence within an electric vehicle project framework. In *Proceedings of the Environmental Systems Research Institute User Conference*.

Newton, P. (1997). Stand density management diagrams: Review of their development and utility in stand-level management planning. *Forest Ecology and Management*, 98(3), 251-265.

Nigim, K., Hipel, K., & Smith, G. (2006). An effective multiple criteria approach to infrastructure reconstruction in devastated countries. *Journal of Systems Science and Systems Engineering*, 15(2), 232-246.

Okabe, A. (2000). *Spatial Tessellations : Concepts and Applications of Voronoi Diagrams*. Wiley, c2000. 2nd ed.s. New York.

Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). Front Matter. *Spatial Analysis along Networks: Statistical and Computational Methods*, i-xviii.

Okabe, H., Imaoka, H., Tomiha, T., & Niwaya, H. (1992). Three dimensional apparel CAD system. In *Proceedings of the ACM SIGGRAPH Computer Graphics*.

Omitaomu, O. A., & Badiru, A. (2007). Fuzzy Present Value Analysis Model For Evaluating Information System Projects. *The Engineering Economist,* 52(2), 157-178.

Oreku, G. S., & Mbowe, J. E. (2014). Critical Infrastructure Protection. *In Proceedings of the The International Conference on Digital Security and Forensics*.

Ouyang, M. (2014). Review on modeling and simulation of interdependent critical infrastructure systems. *Reliability Engineering & System Safety,* 121(0), 43-60.

Ouyang, M., & Dueñas-Osorio, L. (2011). An approach to design interface topologies across interdependent urban infrastructure systems. *Reliability Engineering & System Safety,* 96(11), 1462-1473.

Pala, O., & Wilson, D. C. (2013, April). User Study Analysis of a Geovisualization Decision Support Environment for Critical Infrastructure Recovery. *In Proceedings of the AGILE 2013,* Leuven, Belgium.

Pala, O., Wilson, D. C., Bent, R., Linger, S., & Arnold, J. (2014). Accuracy of Service Area Estimation Methods Used for Critical Infrastructure Recovery. In *J. Butts & S. Shenoi (Eds.*), *Critical Infrastructure Protection VIII* (Vol. 441, 173-191): Springer Berlin Heidelberg.

Patrikalakis, N. M., & Maekawa, T. (2009). *Shape interrogation for computer aided design and manufacturing*. Springer.

Patterson, S. A., & Apostolakis, G. E. (2007). Identification of critical locations across multiple infrastructures for terrorist actions. *Reliability Engineering & System Safety,* 92(9), 1183-1203.

Persello, C., & Bruzzone, L. (2010). A novel protocol for accuracy assessment in classification of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3), 1232-1244.

Quirk, M. D., & Fernandez, S. J. (2005). Infrastructure robustness for multiscale critical missions. J*ournal of Homeland Security and Emergency Management,* 2(2).

Reinermann, D., & Joachim, W. (2003). Analysis of critical infrastructures: The ACIS methodology (Analysis of critical infrastructural sectors). *In Proceedings of the*

Critical Infrastructure Protection (CIP) Workshop.

Rosenfield, G. H., Fitzpatricklins, K., & Ling, H. S. (1982). Sampling For Thematic Map Accuracy Testing. *Photogrammetric Engineering and Remote Sensing,* 48(1), 131-137.

Rozenstein, O., & Karnieli, A. (2011). Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. *Applied Geography,* 31(2), 533-544.

Saaty, T. L. (1994). How to make a decision: the analytic hierarchy process. *Interfaces,* 24(6), 19-43.

Santella, N., Steinberg, L. J., & Parks, K. (2009). Decision making for extreme events: Modeling critical infrastructure interdependencies to aid mitigation and response planning. *Review of Policy Research,* 26(4), 409-422.

Scavo, C., Kearney, R. C., & Kilroy, R. J. (2008). Challenges to Federalism: Homeland Security and Disaster Response. Publius. *The Journal of Federalism,* 38(1), 81-110.

Schiff, J. L. (2011). *Cellular automata: a discrete view of the world* (Vol. 45): John Wiley & Sons.

Schintler, L. A., Gorman, S., Kulkarni, R., & Stough, R. (2007). Moving from protection to resiliency: A path to securing critical infrastructure. *Critical Infrastructure* (291-307): Springer.

Schintler, L. A., Kulkarni, R., Gorman, S., & Stough, R. (2007). Using raster-based GIS and graph theory to analyze complex networks. *Networks and Spatial Economics,* 7(4), 301-313.

Schulman, L. S., & Seiden, P. E. (1986). Percolation and galaxies. *Science,* 233(4762), 425-431.

Sinton, D. F. (1992). Reflections on 25 years of GIS. *GIS World,* 5(2), 1-8.

Sousa, P., Bessani, A. N., Dantas, W. S., Souto, F., Correia, M., & Neves, N. F. (2009). Intrusion-tolerant self-healing devices for critical infrastructure protection. In *Proceedings of the International Conference on Dependable Systems & Networks*. IEEE/IFIP.

Sterbenz, J. P., Cetinkaya, E. K., Hameed, M. A., Jabbar, A., Qian, S., & Rohrer, J. P. (2013). Evaluation of network resilience, survivability, and disruption tolerance: analysis, topology generation, simulation, and experimentation. *Telecommunication systems*, 52(2), 705-736.

Sulewski, L. (2013). *A Geographic Modeling Framework for Assessing Critical*

Infrastructure Vulnerability: Energy Infrastructure Case Study. University of South Carolina.

Tobler, W. (1979). Cellular Geography. *Philosophy in Geography* (379-386): Springer.

Todd, P., & Benbasat, I. (1992). The use of information in decision making: an experimental investigation of the impact of computer-based decision aids. *MIS Quarterly*, 373-393.

Tolone, W. J. (2009). *Interactive visualizations for critical infrastructure analysis*. International Journal of Critical Infrastructure Protection, 2(3), 124-134.

Tolone, W. J., Ahn, G.-J., Pai, T., & Hong, S.-P. (2005). Access control in collaborative systems. *ACM Computing Surveys*, 37(1), 29-41.

Tolone, W. J., Johnson, E. W., Lee, S.-W., Xiang, W.-N., Marsh, L., Yeager, C., & Blackwell, J. (2009). Enabling system of systems analysis of critical infrastructure behaviors *Critical Information Infrastructure Security* (24-35): Springer.

Tolone, W. J., Wilson, D. C., Raja, A., Xiang, W.-N., Hao, H., Phelps, S., & Johnson, E. W. (2004). Critical infrastructure integration modeling and simulation *Intelligence and Security Informatics* (214-225): Springer.

Tolone, W. J., Xiang, W.-N., Raja, A., Wilson, D. C., Tang, Q., & McWilliams, K. (2007). Mining Critical Infrastructure Information from Municipality Data Sets: A Knowledge-Driven Approach and Its Implications. *Emerging Spatial Information Systems and Applications* (310-325): IGI Global.

Toole, G. L., Linger, S. P., & Burks, M. W. (2001). Automated utility service area assessment under emergency conditions. In *Proceedings of the Society for Computer Simulation International Conference*, Seattle, WA.

Toole, G. L., McCown, A. W., & Voeller, J. G. (2008). Interdependent Energy Infrastructure Simulation System. *Wiley Handbook of Science and Technology for Homeland Security*: John Wiley & Sons, Inc.

Torrens, P. M. (2000). How cellular models of urban systems work (1. Theory). *Center for Advanced Spatial Analysis Working paper Series*. University College London.

Torrens, P. M. (2006). Simulating sprawl. Annals of the Association of American Geographers, 96(2), 248-275.

Torrens, P. M. (2009). Process models and next-generation geographic information technology. *GIS best practices: Essays on geography and GIS*, 63-75.

Torrens, P. M., & Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science*, 19(4), 385-412.

Torrens, P. M., & O'Sullivan, D. (2001). Cellular automata and urban simulation: where do we go from here? *Environment and Planning B: Planning and Design*, 28(2), 163-168.

Tortora, R. D. (1978). Note On Sample-Size Estimation For Multinormal Populations. *American Statistician*, 32(3), 100-102.

Toubin, M., Serre, D., Diab, Y., & Laganier, R. (2012). Brief communication: An auto-diagnosis tool to highlight interdependencies between urban technical networks. *Natural Hazards and Earth System Science*, 12(7), 2219-2224.

Usov, A., Beyel, C., Rome, E., Beyer, U., Castorini, E., Palazzari, P., & Tofani, A. (2010). The DIESIS approach to semantically interoperable federated critical infrastructure simulation. In *Proceedings of the Second International Conference on the Advances in System Simulation (SIMUL)*.

Vatcha, R., Lee, S.-W., Murty, A., Tolone, W., Wang, X., Dou, W., Chang, R., Ribarsky, W., Liu, W., & Chen, S.-E. (2009). Towards sustainable infrastructure management: knowledge-based service-oriented computing framework for visual analytics. In *Proceedings of the SPIE Defense, Security, and Sensing Conference*.

Visarraga, D., Bush, B., Linger, S., & McPherson, T. (2005). Development of a JAVA Based Water Distribution Simulation Capability for Infrastructure Interdependency Analyses *Impacts of Global Climate Change* (1-8).

Volkanovski, A., Čepin, M., & Mavko, B. (2009). Application of the fault tree analysis for assessment of power system reliability. *Reliability Engineering & System Safety*, 94(6), 1116-1127.

Von Neumann, J. (1953). A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contributions to the Theory of Games*, 2, 5-12.

Von Neumann, J. (1993). *The Computer and the Brain.* Yale University Press.

Wang, F. (2001). On power quality and protection. Chalmers University of Technology.

Werley, K. A. (2002). *Constrained Cellular Colonization (C3) for Estimating Service and outage areas in Electric Power Transmission Networks* (Rev 4), Report LA-UR-01-4845, Los Alamos National Laboratory.

Westervelt, J. D. (2002). Geographic information systems and agent-based modeling. *Integrating geographic information systems and agent-based modeling techniques for simulating social and ecological processes*, 83-103.

Wilson, D. C., Lipford, H. R., Carroll, E., Karr, P., & Najjar, N. (2008). Charting new ground: modeling user behavior in interactive geovisualization. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*.

Wilson, D. C., Pala, O., Tolone, W. J., & Xiang, W.-N. (2009). Recommendation-based geovisualization support for reconstitution in critical infrastructure protection. In *Proceedings of SPIE 7346, Visual Analytics for Homeland Defense and Security* Orlando, FL.

Wilson, D. R., & Martinez, T. R. (2000). Reduction techniques for instance-based learning algorithms. *Machine learning*, 38(3), 257-286.

Wolfram, S. (1994). *Cellular automata and complexity: collected papers* (Vol. 1): Addison-Wesley Reading.

Yan, G., Eidenbenz, S., Thulasidasan, S., Datta, P., & Ramaswamy, V. (2010). *Criticality analysis of Internet infrastructure. Computer Networks*, 54(7), 1169-1182.

Zio, E., & Golea, L. R. (2012). Analyzing the topological, electrical and reliability characteristics of a power transmission system for identifying its critical elements. *Reliability Engineering & System Safety*, 101, 67-74.

APPENDIX A: SUPPORTING DOCUMENTS FOR USER STUDY

Informed Consent for
Recommendation System for Critical Infrastructure Recovery
Project Purpose
You are invited to participate in a research study entitled “Recommendation System for Critical Infrastructure Recovery.” Critical Infrastructures are integral part of everyday life as they provide essential services to all segments of the population. During an outage of the service it is often hard for a decision maker to determine the priorities for resource allocation. This is especially true when more than one network(s) are involved in the outage and the effects of outage in one network can cascade down to other networks. Therefore I have proposed a recommendation framework and created a software tool based on this frame work. In this study, I aim to gain insight into how the software tool that I created based on the recommendation framework compares to state of the art GIS tools that are used for this kind of analysis. You will be given various scenarios and will be asked to prioritize the network elements’ enablement in each scenario.

Investigator(*s*)
Dr. David Wilson, Associate Professor, Software and Information Systems Department, UNC Charlotte
Okan Pala, PhD Student, Software and Information Systems Department, UNC Charlotte

Eligibility
Participants must be at least 18 years old.
Participants must be able to comfortably communicate in spoken and written English.
Participants must be at least intermediate level users of GIS suite tools (by ESRI) to be able to perform geographic analysis. Experience in ArcGIS Network Analyst Extension is a plus but not a must.

Overall Description of Participation
In carrying out the tests to which this consent applies, I will ask you to perform several analysis tasks on digital Critical Infrastructure spatial data using ArcGIS and Network Explorer. You will first be given a short questionnaire regarding your background. You will then be introduced to a map that visualizes several data sets and asked to perform several analysis tasks using the ArcGIS Tools and Network Explorer Tool that I have created. While using these tools, you will be asked to “think-aloud;” I ask that you express any thoughts you have during the activity so I can understand your mental process. Finally, I will interview you to gather your feedback on the various aspects of the task. Your software use and interview will be recorded. Audio recording of participants’ voice and video recordings of the computer screen will be transcribed and coded using transcription and coding software. Video recordings of the participants’ upper body will be recorded as well for qualitative measurements.

Length of Participation
Your participation in this study will take approximately 60 minutes.

Risks and Benefits of Participation
There are no known risks to participation in this study. However, there may be risks which are currently unforeseeable. I will keep identifiable information about you, nevertheless, identifiers will be randomly numbered, encrypted and will be kept separately from the study data There are no personal benefits to you; however your participation may benefit decision makers and general population during disaster recovery by providing better tools to help speed up the resource allocation process. After completion of the study, you will be compensated with a $15 Target gift card. If you withdraw before the end of the study, or 45 minutes, whichever comes first, you will not be eligible for the compensation.

Volunteer Statement
You are a volunteer. The decision to participate in this study is completely up to you. If you decide to be in the study, you may stop at any time. You will not be treated any differently if you decide not to participate in the study or if you stop once you have started.

Confidentiality Statement
Any information about your participation, including your identity, is completely confidential. The following steps will be taken to ensure this confidentiality:
Your records will be kept in locked files and only study staff will be allowed to look at them.
The recordings of your tasks and interview will be stored on password protected computers, with access only to study staff.
No personally identifiable information will be released or shared.
All the audio, video and paper records will be destroyed 5 years after the end of the study.
Recordings could be used for other related studies for comparison purposes within next five years. Your contact information will be kept in a secure computer. I will keep names and contact information for use in recruitment for future studies. Therefore, you might be contacted to as for your participation in future studies.

Statement of Fair Treatment and Respect
UNC Charlotte wants to make sure that you are treated in a fair and respectful manner. Contact the university's Research Compliance Office (704-687-3309) if you have questions about how you are treated as a study participant. If you have any questions about the actual project or study, please contact Okan Pala (704-687-8387, opala@uncc.edu) Dr. David Wilson (704-687-8585, davils@uncc.edu).

Participant Consent
I have read the information in this consent form. I have had the chance to ask questions about this study, and those questions have been answered to my satisfaction. I am at least 18 years of age, and I agree to participate in this research project. I understand that I will receive a copy of this form after it has been signed by me and the investigator of this research study.

____ _______________

Participant Name (PRINT) DATE

Participant Signature

Investigator Signature DATE

Demographic and Background Questionnaire

Age: ___________

Gender:
Male
Female

What is your major and/or profession?

How have you used mapping applications (such as ArcGIS, etc.)? Please list what applications you have used and for what purpose.

Personal usage

Professional/work usage

What courses or training classes, if any, have you taken that use mapping applications? Please list what applications you have used.

How many years experience do you have with mapping applications?

a. < 1 year
b. 1-2 years
c. 3-4 years
d. 4-6 years
e. 6 + years

How many years experience do you have with ArcGIS Software Suit?

a. < 1 year
b. 1-2 years
c. 3-4 years
d. 4-6 years
e. 6 + years

How would you classify yourself?

Computer

a. Novice
b. Intermediate
c. Expert

GIS

a. Novice
b. Intermediate
c. Expert

Network or Utility Extension

a. Novice
b. Intermediate
c. Expert

Critical Infrastructure Expertise:

a. Novice
b. Intermediate
c. Expert

Disaster Recovery Expertise:

a. Novice
b. Intermediate
c. Expert

Recommendation System for Critical Infrastructure Recovery – Post Survey

1) Did you have any problems answering any of the tasks? If so, please explain which task(s) and why.

2) Please comment on ease of use for each software tool. Provide comparison of two software tools as it applies.

3) Please comment on effectiveness of each software tool. Provide comparison of two software tools as it applies.

4) Please comment on mental demand that the tasks required and compare the level for each task (1,2,3, and 4)

July 8,

To Whom It May Concern,

We are sending you this message to request your participation in a research study that I are conducting to test a software tool that I have developed. The software tool is designed to help decision makers with enablement prioritization task for Critical Infrastructure elements, such as power substations or telecommunication relays, during disaster response process.

The experiment is conducted by College of Computing PhD student Okan Pala and his advisor Dr. David Wilson. I will award a Target gift card worth $15 to all the participants who finish the tasks and complete surveys OR spend at least 45 minutes participating. To participate in this study one would have to be:

- Intermediate or higher-level users of GIS suite tools (by ESRI) to be able to perform geographic analysis,
- At least 18 years old,
- Comfortably communicate in spoken and written English.
- Experience in ArcGIS Network Analyst Extension is a plus but not a necessity.

Subjects will not have any kind of benefits besides feeling good about testing a new software tool that might have decision makers in the future. Society would benefit in the long run as decision makers could use new software tools based on this research to make better decisions effectively and timely.

We do not foresee any discomfort to be experienced by the participants. They might have a slight stress while competing the tasks or slight frustration but this should not be at a level to create significant discomfort. I do not foresee any economic and legal harm or threat.

Please contact us by email (opala@uncc.edu) if you are interested in participating in this study.

Critical Infrastructure Research Study

A Research Study at the College of Computing and Informatics at UNC Charlotte is looking for participants.

Qualifications:

• Intermediate or higher level users of GIS suite tools (by ESRI) in order to be able to perform geographic analysis.

• At least 18 years old.

• Must be able to comfortably communicate in spoken and written English.

• Experience in ArcGIS Network Analyst Extension is a plus but not a necessity.

Participants will be awarded a $15 Target Gift Card

(Participants will have to spend at least 45 minutes or finish all the tasks and surveys in order to receive the gift card)

Contact Okan Pala by Email at opala@uncc.edu

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Task Demand Comparisons

For each of the following, please place a mark to indicate which of the two was more demanding for you.

1.	__________ Physical Demand	OR	__________Mental Demand
2.	__________ Temporal Demand	OR	__________Mental Demand
3.	__________ Performance	OR	__________Mental Demand
4.	__________ Frustration level	OR	__________Mental Demand
5.	__________ Effort	OR	__________Mental Demand
6.	__________ Temporal Demand	OR	__________Physical Demand
7.	__________ Performance	OR	__________Physical Demand
8.	__________ Frustration level	OR	__________Physical Demand
9.	__________ Effort	OR	__________Physical Demand
10.	__________ Temporal Demand	OR	__________Performance
11.	__________ Temporal Demand	OR	__________Frustration level
12.	__________ Temporal Demand	OR	__________Effort
13.	__________ Performance	OR	__________Frustration level
14.	__________ Performance	OR	__________Effort
15.	__________ Effort	OR	__________Frustration level

APPENDIX B: KENTUCKY GEOSPATIAL WATER DATA

Water.mdb							
GIS Feature Class Name	**Item Names**	**Item Width**	**Item Type**	**# of Decimals**	**Description - All Entries must be in uppercase**	**Domain/Valid Range**	**Name of Domain/Range**
Purchased Water Source-PURCHSRC - for ALL community water systems							
Purchased water must be measured through a meter. The point for the purchased water source should align with a meter. Generally either the purchaser or the seller owns the meter. Both may include data regarding the same meter, since both often read it. Consequently, there are two places to insert the meter id in this coverage. Be sure to check both the PURNAME and PMETR_ID against the OWNER and METR_ID in the Meter coverage and/or the SELNAME and METR_ID against the OWNER and METR_ID in the Meter coverage.							
	SELPWSID	7	Text		Seller's Public Water Supply ID from NREPC DOW database. If seller is out of state, choose the appropriate state abbreviation from the bottom of the list (ie Tennessee would be TN, Ohio would be OH, etc...)	PULL FROM MASTER LIST	PWSID
	SELNAME	75	Text		Seller's Name	PULL FROM MASTER LIST	WATOWNER
	METR_ID	75	Text		Unique ID or common name for each meter - Meter ID assigned by the SELLER. This could be a number, street name, or common location name where the meter is located. If you do use a name and there is more than one meter to at that given location use a meter number as well, so that you have a unique (to that water system) identifier ... for example, VINE ST #2. The METR_ID corresponds to METR_ID in the METER and in the WELLSRC coverages.	NO DOMAIN/RANGE	N/A
	PURPWSID	7	Text		Purchaser's Public Water Supply ID from NREPC DOW database. If purchaser is out of state,	PULL FROM MASTER LIST	PWSID
	PURNAME	75	Text		Name of Purchaser - entity that purchases water at this point	PULL FROM MASTER LIST	WATOWNER
	PMETR_ID	75	Text		Unique ID or common name for each meter - Meter ID or common name assigned by the PURCHASER. This could be a number, street name, or common location name where the meter is located. If you do use a name and there is more than one meter to a that given location use a meter number as well, so that you have a unique (to that water system) identifier ... for example, VINE ST #2. The PMETR_ID corresponds to the PMETR_ID in the METER coverage.	NO DOMAIN/RANGE	N/A
	AVAIL	10	Text		Availability	PERMANENT, SEASONAL, EMERGENCY, OTHER	PSAVAIL
	OTHAVAIL	50	Text		Description of OTHER category of AVAIL	NO DOMAIN/RANGE	N/A
	ESTVOL	12	Double	2	Total estimated available volume of seller (gallons)	0.0 - 1,000,000,000	ESTVOL
	CURPURCH	12	Double	2	Current purchase (gallons per day)	0.0 - 1,000,000,000	CURPURCH
	AVGWDRAW	12	Double	2	Average daily usage for last 12 months in gallons	0.0 - 1,000,000,000	PSAVGDRAW
	HWDRAW	12	Double	2	Highest daily usage for last 12 months in gallons	0.0 - 1,000,000,000	PSHWDRAW
	MAXAMNT	12	Double	2	Maximum contract amount, if applicable (gallons per day)	0.0 - 1,000,000,000	MAXAMNT
	RAWPRICE	12	Double	2	Price of raw water per 1,000 gallons	0.00 -999.00	RAWPRICE
	EFFPRICE	12	Double	2	Effective price of finished water per 1,000 gallons purchased	0.00 - 10,000.00	EFFPRICE
	PURCHCON	3	Text		Is there a written purchase contract? (YES/NO)	YES, NO	YESNO
	REQWAT	12	Double	2	Total annual volume of water required to be available under purchase contract agreement (gal)	0.0 - 1,000,000,000	REQWAT
	EXPIR_DATE	8	Date		Expiration Date of Contract	NO DOMAIN/RANGE	N/A
	SPC_COND	100	Text		Cite special conditions/restrictions	NO DOMAIN/RANGE	N/A

APPENDIX C: IEISS XML INPUT EXAMPLE: KENTUCKY WATER SYSTEM


```
<Project name= “Kentucky Water Treatment Plants”>
<Model name= “KY WTP”>
  <Entities>
        <water>
          <Junction>
                <name>BARREN LAKE WTP</name>
                <pressure>14.7</pressure>
                <elevation>0</elevation>
                <maximumPressure>1000</maximumPressure>
                <slack>false</slack>
                <id>JKY0050929</id>
                <inService>true</inService>
                -<location>-
                      <point>
                            <x>-86.063998096</x>
                            <y>36.899576012</y>
                            <z>1.5</z>
                      </point>
                </location>
          </Junction
          <DeliveryPoint>
                <name>BARREN LAKE WTP</name>
                <id>KY0050929</id>
                <consumptionRate>4.22</consumptionRate>
                <maximumPressure>1000</maximumPressure>
                <inService>true</inService>-
                <location>-
                      <point>
                            <x>-86.063998096</x>
                            <y>36.899576012</y>
                            <z>1.45</z>
                      </point>
                </location>
                <connections>
                      <id>JKY0050929</id>
                </connections>
          </DeliveryPoint>
</water>
</Entities>
<ServiceAreaParameters>
        ‘Detailed SA parameters are defined here
</ServiceAreaParameters>
</Model>
<Map> ‘Map presentation and layer details are stored here</Map>
</Project>
```

APPENDIX D: WATER USE BREAKDOWN FOR US STATES

[Values may not sum to totals because of independent rounding; Mgal/d, million gallons per day; gal/d, gallons per day; n/a, not applicable]

State	Self supplied						Public supply			Total use		
	Self-supplied population (in thousands)	Percent of total population	Withdrawals (in Mgal/d): By source: Ground-water	Withdrawals (in Mgal/d): By source: Surface water	Withdrawals (in Mgal/d): Total	Self-supplied per capita use (in gal/d)	Population served (in thousands)	Water deliveries (in Mgal/d)	Public-supply per capita use (in gal/d)	Total population (in thousands)	Water use (withdrawals and deliveries, in Mgal/d)	Total domestic per capita use (in gal/d)
Alabama	521	11	39.1	0	39.1	75	4,040	326	81	4,560	365	80
Alaska	235	35	13.4	.68	14.1	60	429	46.8	109	664	60.9	92
Arizona	218	4	27.2	0	27.2	125	5,720	802	140	5,940	830	140
Arkansas	200	7	17.8	0	17.8	89	2,580	254	99	2,780	272	98
California	2,710	7	429	57.2	486	179	33,400	3,980	119	36,100	4,470	124
Colorado	299	6	34.4	0	34.4	115	4,370	530	121	4,670	564	121
Connecticut	841	24	63.1	0	63.1	75	2,670	200	75	3,510	263	75
Delaware	80.4	10	6.43	0	6.43	80	763	44.6	58	844	51.1	61
District of Columbia	0	0	0	0	0	n/a	582	82.7	142	582	82.7	142
Florida	1,790	10	190	0	190	106	16,100	1,530	95	17,900	1,720	96
Georgia	1,600	18	120	0	120	75	7,470	727	97	9,070	847	93
Hawaii	74.0	6	12.2	0	12.2	165	1,200	198	165	1,280	210	165
Idaho	424	30	86.6	0	86.6	204	1,010	181	180	1,430	267	187
Illinois	1,130	9	101	0	101	90	11,600	1,050	90	12,800	1,150	90
Indiana	1,630	26	124	0	124	76	4,650	353	76	6,270	477	76
Iowa	531	18	34.6	0	34.6	65	2,440	158	65	2,970	193	65
Kansas	149	5	14.9	0	14.9	100	2,600	209	80	2,740	223	81
Kentucky	696	17	22.2	12.6	34.8	50	3,480	243	70	4,170	278	67
Louisiana	551	12	44.0	0	44.0	80	3,970	485	122	4,520	529	117
Maine	575	44	34.1	0	34.1	59	746	37.8	51	1,320	71.9	54

Source:
http://water.usgs.gov/edu/wateruse/pdf/wudomestic-2005.pdf

APPENDIX E: EXAMPLE PRODUCER'S AND USER'S ACCURACY TABLE FOR AGGREGATE IMPACT ANALYSIS

Location Allocation - Point Impact Analysis

Name of the Plant	Users Accuracy	Producers Accuracy
ALBANY WATER TREATMENT PLANT A	78.27	59.88
ALBANY WATER TREATMENT PLANT B	98.53	82.07
ALLEN WTP	71.91	60.90
ARLINGTON WTP	100.00	45.61
ASHLAND WTP	83.75	99.87
AUGUSTA WTP	87.55	82.70
BARBOURVILLE WATER PLANT	91.24	74.17
BARDWELL WTP	100.00	86.67
BARKLEY LAKE WATER TREATMENT PLANT	61.42	89.48
BARLOW WTP	85.71	33.33
BARNEY JOHNSON	42.77	97.79
BARREN LAKE WTP	88.19	64.29
BARREN RIVER WTP	98.54	96.28
BEATTYVILLE WTP PLANT B	85.19	70.92
BEAVER CREEK WTP	74.21	90.23
BEAVER DAM WTP	14.38	79.57
BEECH FORK WTP	90.46	67.80
BELL COUNTY FC	46.43	46.43
BENHAM WTP	71.17	85.29
BENTON WTP	97.28	86.03
BEREA WTP	93.50	61.72
BROWNSVILLE WTP	98.28	70.10
BULLOCK PEN WTP	28.00	71.51
BURKESVILLE WATER PLANT	94.12	79.52
BURNSIDE WATER TREATMENT PLANT	91.15	35.43
BUTLER COUNTY WTP	66.77	65.55
CADIZ WATER TREATMENT PLANT	59.65	51.76
CALHOUN WTP	81.70	40.31
CALVERT CITY WTP	97.78	60.86

CAMPBELLSVILLE	88.21	81.28
CAMPTON WTP	59.51	87.82
CARLISLE WATER PLANT	84.11	80.35
CARR CREEK WTP	36.34	84.83
CARROLLTON WTP	81.42	73.61
CAVE RUN WATER TREATMENT PLANT	38.55	30.73
CAWOOD WTP	47.14	92.45
CENTER RIDGE WTP	100.00	75.00
CITY SPRINGS	72.61	98.70
COLUMBIA/ADAIR WATER TREATMENT PLANT	89.06	83.42
COLUMBUS WTP	100.00	100.00
CRESCENT HILL WTP	97.64	99.49
CUMBERLAND MUNICPAL WATER WORKS WTP	98.49	54.48
CUMBERLAND RIVER PLANT	85.28	79.65
CUMBERLAND RIVER WATER TREATMENT PLANT	50.30	97.61
CUNNINGHAM WTP	70.37	100.00
DEEP WELLS WTP	94.12	100.00
EARLINGTON WATER TREATMENT PLANT	21.48	10.10
EDDYVILLE WATER TREATMENT PLANT	80.49	49.07
EVARTS MUNICIPAL WATER PLANT	89.16	88.50
FALMOUTH WTP	73.49	73.38
FANCY FARM WTP	48.89	100.00
FERN LAKE PLANT	99.74	93.92
FLEMINGSBURG WTP	50.55	26.44
FORT KNOX/CENTRAL WTP	83.97	31.88
FOURTH STREET WTP	91.36	96.40
FRANCIS WTP	41.51	8.15
FRANKFORT WTP	89.56	83.82
FRANKLIN WTP	98.19	71.28
FT THOMAS WTP	76.49	82.64
FULTON MUNICIPAL WTP	96.54	76.76
GALLATIN WTP	74.79	24.62
GEORGE ARNOLD WATER TREATMENT PLANT	49.99	98.89
GHENT WTP	51.19	79.96
GRAYSON COUNTY	87.69	79.46
GRAYSON WTP	68.14	77.24
GREEN RIVER VALLEY WTP	78.45	94.77
GREEN RIVER WATER TREATMENT PLANT	86.88	83.28
GREENSBURG WATER TREATMENT PLANT	67.97	76.10
GREENUP WTP	72.15	94.04
GREENVILLE WATER TREATMENT PLANT A	83.13	80.56

GUIST CREEK LAKE WTP	95.93	72.98
HARDEMAN WTP	39.81	100.00
HARDINSBURG REVERSE OSMOSIS WATER TREATMENT FACILITY	64.49	79.00
HARLAN MUN WTP	93.21	47.47
HARRODSBURG WTP	73.76	84.79
HARTFORD WTP	69.15	39.64
HAWESVILLE WTP	77.71	40.88
HAZARD WTP	92.84	70.59
HENRY CO WTP	50.91	75.11
HERRINGTON LAKE WTP	91.56	86.75
HICKMAN WTP	100.00	93.32
HICKORY WTP	66.23	72.18
HINKSTON CREEK WTP	52.25	56.18
HODGENVILLE	92.84	71.44
HOPKINSVILLE WATER TREATMENT PLANT	97.26	82.62
HYDEN LESLIE WTP	79.18	80.47
IMPOUNDMENT PLANT	91.76	92.13
IRVINE WTP	92.01	85.67
JACKSON COUNTY WATER TREATMENT PLANT	60.96	73.44
JACKSON WTP	63.86	89.99
JAMESTOWN WATER TREATMENT PLANT	98.20	89.08
JENKINS WTP	91.56	54.51
JONATHAN CREEK WTP	93.78	84.10
KENTUCKY RIVER STATION II/ HARDIN'S LANDING PLANT	77.95	29.50
KENTUCKY RIVER STATION WTP	65.09	86.21
KENTUCKY RIVER WTP	74.74	84.29
KEVIL WTP	77.54	59.44
KUTTAWA WATER TREATMENT PLANT	96.55	82.97
KY RIVER WTP	91.90	83.33
LACENTER WTP	100.00	88.39
LAKE LINVILLE PLANT	97.84	82.74
LAKE PEE WEE WATER TREATMENT PLANT	96.54	88.81
LAUREL LAKE PLANT	99.82	39.62
LAUREL RIVER PLANT	86.48	63.84
LAWRENCEBURG WTP	97.64	78.29
LEBANON	68.59	87.37
LEDBETTER WATER TREATMENT PLANT	88.99	57.96
LEITCHFIELD	68.99	80.35
LEWISPORT WTP	100.00	59.51
LIBERTY WATER TREATMENT PLANT	76.95	84.62
LICKING RIVER WTP	82.85	94.01

LIVERMORE WTP	100.00	19.94
LOUISA WTP	83.82	95.06
LOVELACEVILLE WTP	80.00	40.00
MANCHESTER WATER PLANT	81.75	89.72
MARION WATER TREATMENT PLANT	85.08	86.39
MARTIN CO WTP	99.22	74.44
MAYFIELD WTP	97.12	79.96
MAYSVILLE WTP	87.91	95.74
MCCREARY COUNTY WATER TREATMENT PLANT	93.88	97.70
MCCREARY COUNTY WATER TREATMENT PLANT 2	93.32	79.89
MCKEE RESERVOIR PLANT	62.83	12.46
MEADOW HILL WATER TREATMENT PLANT	72.95	65.45
MEMORIAL PARKWAY WTP	68.81	98.70
MONTICELLO WATER TREATMENT PLANT	78.13	99.03
MOREHEAD STATE UNIVERSITY WATER TREATMENT PLANT	70.79	57.01
MOREHEAD UTILITY PLANT BOARD WATER TREATMENT PLANT	69.88	87.64
MORGANFIELD WTP	94.42	78.26
MORGANTOWN WTP	31.03	2.35
MOUNT STERLING WATER TREATMENT PLANT	81.66	81.13
MULDRAUGH PLANT	70.43	65.21
MURPHY LANE WTP	61.52	66.60
MURRAY WTP	99.15	99.39
NICHOLASVILLE WTP	85.84	55.02
NORTH POINT TRAINING CENTER	70.14	26.31
NORTH WTP	85.50	100.00
NORTONVILLE WATER TREATMENT PLANT	96.56	57.84
OHIO COUNTY WTP #2	46.28	56.57
OLDHAM CO WTP	90.23	92.68
OLIVE HILL WTP	94.22	50.27
OWENTON WTP	73.91	60.84
PADUCAH WTP	95.43	99.31
PAINTSVILLE WTP	97.52	77.46
PARIS WTP	99.67	51.47
PIKEVILLE WTP	85.92	88.22
PINEVILLE WATER TREATMENT PLANT	74.41	98.93
PIRTLE SPRINGS	96.90	47.84
PRESTONSBURG WTP	56.32	75.36
PROVIDENCE WATER TREATMENT PLANT #2	98.99	51.28
RATTLESNAKE RIDGE WTP	56.53	44.84
REIDLAND WTP	90.18	77.70
RICHMOND RD STATION WTP	81.63	88.21

ROYAL SPRING WTP	75.75	79.91
RUSSELL FORK WTP	80.04	87.63
RUSSELL WTP	93.93	70.15
S FORK KY RIVER WTP	79.85	67.75
SALYERSVILLE WTP	57.96	97.20
SANDY HOOK WTP	53.84	58.71
SCOTTSVILLE WTP	57.06	90.72
SEDALIA WTP	71.43	8.62
SHEA FORK MINE WTP	60.97	67.49
SOMERSET WATER TREATMENT PLANT	93.30	96.71
SOUTH GRAVES WTP	27.93	92.25
SOUTH SHORE WTP	99.84	82.53
SOUTH WTP	60.14	10.73
SPRINGFIELD	97.73	59.81
STURGIS WTP	88.26	46.60
SYMPSON LAKE	96.60	79.80
SYMSONIA WTP	46.03	96.67
TATUMSVILLE WTP	63.75	78.77
TAYLOR MILL WTP	70.24	50.14
TOMPKINSVILLE WTP	94.57	71.62
TREATMENT PLANT-sm	57.87	59.94
TREATMENT PLANT-big	73.08	87.26
TRIMBLE # 1 WTP - BRAYS	43.65	83.95
TWO CITY RESERVOIR WTP	73.14	96.23
VANCEBURG ELECRIC PLANT BOARD WTP	93.61	86.22
WALLINS CREEK WTP	22.37	80.47
WARSAW WTP	50.62	94.77
WATER DISTRICT TREATMENT PLANT	69.23	58.98
WAX WTP	85.95	76.73
WEBSTER CO WTP	18.25	69.66
WEST LIBERTY WATER TREATMENT PLANT 519	46.22	69.91
WEST POINT	34.94	14.65
WESTERN FLEMING WTP	82.33	81.63
WESTERN LEWIS RECTORVILLE WTP	92.20	45.60
WHEELWRIGHT WTP	93.04	22.14
WHITE MILLS	91.82	54.64
WHITE PLAINS WATER TREATMENT PLANT	83.40	81.85
WHITESBURG WTP	72.19	89.00
WICKLIFFE WTP	100.00	86.05
WILLIAMSTOWN LAKE WTP	90.90	54.52
WILMORE WTP	96.65	35.05

WINGO WTP	88.00	40.00
WOOD CREEK LAKE PLANT	21.02	96.41
WOODSON BEND PLANT	75.00	4.55
WORTHINGTON WTP	69.64	28.67
WTP #1 (WELL HOUSE RD)	100.00	38.46
WTP #2	100.00	40.00

Cellular Automata - Aggregate Impact

Name of the Plant	**Users Accuracy**	**Producers Accuracy**
ALBANY_WATER_TREATMENT_PLANT_A	1.17	39.59
ALBANY_WATER_TREATMENT_PLANT_B	5.96	81.51
ALLEN_WTP	0.69	31.61
ARLINGTON_WTP	64.24	65.85
ASHLAND_WTP	51.31	100.00
AUGUSTA_WTP	52.74	99.37
BARBOURVILLE_WATER_PLANT	86.15	42.33
BARDWELL_WTP	100.00	27.00
BARKLEY_LAKE_WATER_TREATMENT_PLANT	30.66	94.70
BARKLEY_LAKE_WATER_TREATMENT_PLANT_PRINCETON	0.38	100.00
BARLOW_WTP	100.00	100.00
BARNEY_JOHNSON	81.92	64.66
BARREN_LAKE_WTP	40.13	83.04
BARREN_RIVER_WTP	82.17	75.32
BEATTYVILLE_WTP_PLANT_B	67.32	92.88
BEAVER_CREEK_WTP	66.56	54.07
BEAVER_DAM_WTP	74.40	5.81
BEECH_FORK_WTP	80.19	62.22
BELL_COUNTY_FC	100.00	4.41
BENHAM_WTP	74.88	33.99
BENTON_WTP	86.40	65.00
BEREA_WTP	87.16	36.24
BROWNSVILLE_WTP	95.39	65.45
BULLOCK_PEN_WTP	79.24	25.85
BURKESVILLE_WATER_PLANT	94.29	54.61
BURNSIDE_WATER_TREATMENT_PLANT	99.97	2.05
BUTLER_COUNTY_WTP	79.77	66.69
CADIZ_WATER_TREATMENT_PLANT	100.00	5.39

CALHOUN_WTP	41.78	28.21
CALVERT_CITY_WTP	97.60	35.63
CAMPBELLSVILLE	54.72	82.11
CAMPTON_WTP	86.99	41.20
CARLISLE_WATER_PLANT	42.54	85.14
CARR_CREEK_WTP	51.41	73.56
CARROLLTON_WTP	72.97	43.92
CAVE_RUN_WATER_TREATMENT_PLANT	26.43	50.67
CAWOOD_WTP	53.25	98.90
CENTER_RIDGE_WTP	100.00	48.37
CITY_SPRINGS	80.69	80.29
COLUMBIA_ADAIR_WATER_TREATMENT_PLANT	57.56	68.18
COLUMBUS_WTP	100.00	100.00
CRESCENT_HILL_WTP	57.96	93.90
CUMBERLAND_MUNICPAL_WATER_WORKS_WTP	53.06	92.41
CUMBERLAND_RIVER_PLANT	98.94	63.92
CUMBERLAND_RIVER_PLANT_KNOX	0.51	100.00
CUMBERLAND_RIVER_WATER_TREATMENT_PLANT	72.67	96.04
CUNNINGHAM_WTP	100.00	93.63
DEEP_WELLS_WTP	100.00	100.00
EARLINGTON_WATER_TREATMENT_PLANT	100.00	12.40
EDDYVILLE_WATER_TREATMENT_PLANT	86.37	32.70
EVARTS_MUNICIPAL_WATER_PLANT	62.02	65.06
FALMOUTH_WTP	84.75	47.14
FANCY_FARM_WTP	83.13	64.41
FERN_LAKE_PLANT	65.44	87.72
FLEMINGSBURG_WTP	100.00	1.60
FORT_KNOX_CENTRAL_WTP	100.00	1.30
FOURTH_STREET_WTP	57.72	85.19
FRANCIS_WTP	100.00	1.24
FRANKFORT_WTP	53.60	82.18
FRANKLIN_WTP	100.00	57.22
FT_THOMAS_WTP	7.84	88.72
FULTON_MUNICIPAL_WTP	90.32	81.74
GALLATIN_WTP	69.11	27.11
GEORGE_ARNOLD_WATER_TREATMENT_PLANT	40.20	97.55
GHENT_WTP	6.11	100.00
GRAYSON_COUNTY	91.93	44.92
GRAYSON_WTP	23.78	56.75
GREEN_RIVER_VALLEY_WTP	65.38	83.42
GREEN_RIVER_WATER_TREATMENT_PLANT	37.75	82.09
GREENSBURG_WATER_TREATMENT_PLANT	63.65	75.81
GREENUP_WTP	43.47	96.30
GREENVILLE_WATER_TREATMENT_PLANT_A	95.23	25.68
GUIST_CREEK_LAKE_WTP	99.85	52.80
HARDEMAN_WTP	88.41	34.90
HARDINSBURG_REVERSE_OSMOSIS_WATER_TREATMENT_FACILITY	41.09	83.11
HARLAN_MUN_WTP	72.80	15.49

HARRODSBURG_WTP	25.07	71.09
HARTFORD_WTP	16.48	15.55
HAWESVILLE_WTP	80.79	26.60
HAZARD_WTP	63.47	86.81
HENRY_CO_WTP	27.09	46.35
HERRINGTON_LAKE_WTP	66.41	81.22
HICKMAN_WTP	100.00	88.81
HICKORY_WTP	40.21	91.79
HINKSTON_CREEK_WTP	66.63	41.36
HODGENVILLE	100.00	32.01
HOPKINSVILLE_WATER_TREATMENT_PLANT	77.54	86.45
HYDEN_LESLIE_WTP	62.65	83.03
IMPOUNDMENT_PLANT	97.37	31.18
IRVINE_WTP	86.00	76.31
JACKSON_COUNTY_WATER_TREATMENT_PLANT	32.62	82.58
JACKSON_WTP	99.28	72.95
JAMESTOWN_WATER_TREATMENT_PLANT	89.04	79.05
JENKINS_WTP	65.80	78.92
JONATHAN_CREEK_WTP	69.62	84.73
KENTUCKY_RIVER_STATION_II__HARDIN_S_LANDING_PLANT	94.10	28.53
KENTUCKY_RIVER_STATION_WTP	0.07	100.00
KENTUCKY_RIVER_STATION_WTP_LANCASTER	56.88	75.77
KENTUCKY_RIVER_WTP	25.08	96.21
KENTUCKY_RIVER_WTP_RICHMOND	0.32	100.00
KEVIL_WTP	83.23	18.34
KUTTAWA_WATER_TREATMENT_PLANT	38.29	98.41
KY_RIVER_WTP	89.06	53.32
LACENTER_WTP	100.00	53.57
LAKE_LINVILLE_PLANT	91.68	72.42
LAKE_PEE_WEE_WATER_TREATMENT_PLANT	60.04	94.59
LAUREL_LAKE_PLANT	100.00	13.43
LAUREL_RIVER_PLANT	44.76	30.17
LAWRENCEBURG_WTP	45.00	63.41
LEBANON	58.04	80.76
LEDBETTER_WATER_TREATMENT_PLANT	55.21	77.65
LEITCHFIELD	20.48	65.66
LEWISPORT_WTP	99.21	43.89
LIBERTY_WATER_TREATMENT_PLANT	97.61	61.78
LICKING_RIVER_WTP	56.80	96.80
LIVERMORE_WTP	100.00	1.43
LOUISA_WTP	78.02	80.88
LOVELACEVILLE_WTP	100.00	4.60
MANCHESTER_WATER_PLANT	97.35	63.37
MARION_WATER_TREATMENT_PLANT	58.46	50.83
MARTIN_CO_WTP	98.96	58.10
MAYFIELD_WTP	13.79	88.13
MAYSVILLE_WTP	10.34	95.81
MCCREARY_COUNTY_WATER_TREATMENT_PLANT	56.26	98.34

MCCREARY_COUNTY_WATER_TREATMENT_PLANT_2	76.63	60.46
MCKEE_RESERVOIR_PLANT	99.94	6.76
MEADOW_HILL_WATER_TREATMENT_PLANT	64.51	65.04
MEMORIAL_PARKWAY_WTP	30.73	94.48
MONTICELLO_WATER_TREATMENT_PLANT	63.57	66.98
MOREHEAD_STATE_UNIVERSITY_WATER_TREATMENT_PLANT	100.00	2.03
MOREHEAD_UTILITY_PLANT_BOARD_WATER_TREATMENT_PLANT	20.03	86.45
MORGANFIELD_WTP	68.59	72.10
MORGANTOWN_WTP	27.39	14.01
MOUNT_STERLING_WATER_TREATMENT_PLANT	65.15	54.48
MULDRAUGH_PLANT	26.99	77.67
MURPHY_LANE_WTP	89.45	63.48
MURRAY_WTP	97.96	97.30
NICHOLASVILLE_WTP	72.53	47.39
NORTH_POINT_TRAINING_CENTER	44.51	7.15
NORTH_WTP	60.40	99.93
NORTONVILLE_WATER_TREATMENT_PLANT	99.88	9.13
OHIO_COUNTY_WTP__2	17.36	79.96
OLDHAM_CO_WTP	49.69	55.05
OLIVE_HILL_WTP	86.48	20.76
OWENTON_WTP	73.14	59.33
PADUCAH_WTP	31.66	93.28
PAINTSVILLE_WTP	55.75	84.39
PARIS_WTP	88.68	45.69
PIKEVILLE_WTP	39.56	83.45
PINEVILLE_WATER_TREATMENT_PLANT	30.14	98.77
PIRTLE_SPRINGS	73.98	66.76
PRESTONSBURG_WTP	63.71	82.91
PROVIDENCE_WATER_TREATMENT_PLANT__2	100.00	11.30
RATTLESNAKE_RIDGE_WTP	44.76	20.64
REIDLAND_WTP	63.97	21.26
RICHMOND_RD_STATION_WTP	37.16	48.47
ROYAL_SPRING_WTP	85.14	31.29
RUSSELL_FORK_WTP	99.60	49.34
RUSSELL_WTP	48.73	50.94
S_FORK_KY_RIVER_WTP	48.18	90.24
SALYERSVILLE_WTP	91.99	55.43
SANDY_HOOK_WTP	96.56	42.41
SCOTTSVILLE_WTP	60.33	78.51
SEDALIA_WTP	100.00	2.48
SHEA_FORK_MINE_WTP	60.95	87.76
SOMERSET_WATER_TREATMENT_PLANT	1.87	99.98
SOUTH_GRAVES_WTP	70.11	76.63
SOUTH_SHORE_WTP	99.29	43.70
SOUTH_WTP	31.78	15.83
SPRINGFIELD	91.79	49.09
STURGIS_WTP	100.00	7.76
SYMPSON_LAKE	84.66	76.40

SYMSONIA_WTP	100.00	19.40
TATUMSVILLE_WTP	48.88	86.27
TAYLOR_MILL_WTP	63.75	28.54
TOMPKINSVILLE_WTP	90.33	82.50
TREATMENT_PLANT_BIG	0.02	100.00
TREATMENT_PLANT_SM	99.10	44.52
TRIMBLE___1_WTP___BRAYS	16.47	97.61
TWO_CITY_RESERVOIR_WTP	57.65	84.86
VANCEBURG_ELECRIC_PLANT_BOARD_WTP	74.63	95.56
WALLINS_CREEK_WTP	94.79	20.40
WARSAW_WTP	55.78	51.68
WATER_DISTRICT_TREATMENT_PLANT	36.39	84.80
WAX_WTP	80.40	65.90
WEBSTER_CO_WTP	16.94	34.28
WEST_LIBERTY_WATER_TREATMENT_PLANT_519	16.04	65.09
WEST_POINT	97.06	5.42
WESTERN_FLEMING_WTP	55.78	80.29
WESTERN_LEWIS_RECTORVILLE_WTP	93.43	19.02
WHEELWRIGHT_WTP	100.00	2.72
WHITE_MILLS	69.76	67.42
WHITE_PLAINS_WATER_TREATMENT_PLANT	98.40	34.68
WHITESBURG_WTP	66.43	75.42
WICKLIFFE_WTP	100.00	100.00
WILLIAMSTOWN_LAKE_WTP	91.22	60.36
WILMORE_WTP	85.18	13.70
WINGO_WTP	100.00	13.88
WOOD_CREEK_LAKE_PLANT	25.58	72.27
WOODSON_BEND_PLANT	100.00	1.19
WORTHINGTON_WTP	91.72	8.45
WTP__1__WELL_HOUSE_RD_	95.36	94.94
WTP__2	90.90	31.90