Spring 5-1-2018

Get To The “Point”: Improving Location Services in Tippah County Mississippi

Haley Feather
University of Southern Mississippi

Get To The “Point”: Improving Location Services in Tippah County Mississippi

by

Haley Feather

A Thesis
Submitted to the Graduate School,
the College of Science and Technology
and the Department of Geography and Geology
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Master of Science

Approved by:

David H. Holt, Committee Chair
David M. Cochran
George T. Raber

Dr. David H. Holt	Dr. David M. Cochran	Dr. Karen S. Coats
Committee Chair	Department Chair	Dean of the Graduate School

May 2018

Published by the Graduate School

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

ABSTRACT

Point location using geographic information systems (GIS) technology has become integrated into everyday society and daily decision-making by utilizing addresses to provide goods and services. A need exists at a national, state, and local level for an address database. The objectives of this study were to [1] determine the most suitable address data model to be used in Mississippi, [2] determine how positional accuracy changes between urban and rural areas, and [3] determine spatial variations in aerial imagery. Address data model comparisons were conducted using match rates between street, parcel, and point address models. Positional accuracy was determined for urban and rural areas using GPS points and margin of error. A mean center and standard distance calculation were performed using one standard deviation. [1] The point address data model (93% matched) and parcel data model (93% matched) outperformed the street data model (06%). [2] The results show that 65% of the average mean points fell within 13 feet – 38 feet from the structure. The average distance from mean was 27.87 feet in urban areas and 82.98 feet in rural areas [3] 75% of the total points fell within the margin of error in urban areas and 80% of the total points in rural areas.

Match rates were influenced by both the quality of reference and input address datasets. Using an average point location is acceptable for addressing in urban and rural areas. There was no significant shift or change between the 2006 and 2015 imageries. Address collection using the point address data model and high-resolution aerial imagery is an accurate, cost-efficient way to build an address database.

ACKNOWLEDGMENTS

This work would not have been possible without the continued support of The University of Southern Mississippi Geography and Geology Department faculty and staff. I am especially indebted to Dr. David Holt, my thesis committee chair, for the guidance, knowledge, patience, and support on this journey to obtain my degree. I would also like to thank my other committee members, Dr. David Cochran and Dr. George Raber, for their encouragement and guidance throughout my graduate school career.

A special thanks to Mississippi 811 Inc. for the opportunity to continue my education. I would also like to thank the company President, Mr. Sam Johnson, and my manager, Amanda Russell, for providing the resources to complete this project. I am grateful to all of those with whom I have had the pleasure to work with during this and other related projects. The completion of this project could not have been accomplished without the support and resources of Tippah County E-911, Everything Is Somewhere, LLC, and Tippah County Electric Power Association.

Nobody has been more important to me in the pursuit of my degree than my family. I would like to thank my parents; whose love and guidance are with me in whatever I pursue. Thank you to everyone who has been with me on this journey. I am grateful for the experiences and opportunities this journey has afforded me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

ASPRS	American Society for Photogrammetry and Remote Sensing
CB	Citizens Band
DSF	Delivery Sequence File
DOP	Dilution of Precision
E-911	Emergency 911
ESRI	Environmental Systems Research Institute
FEMA	Federal Emergency Management Agency
FGDC	Federal Geographic Data Committee
FHWA	Federal Highway Administration
GPS	Global Positioning System
GIS	Geographic Information Systems
HARN	High Precision Accuracy Reference Network
MAF	Master Address File
MARIS	Mississippi Automated Resources and Information Systems
MSAG	Master Street Address Guide
NAD	National Address Database
NSSDA	National Standard for Spatial Data Accuracy
NGAC	National Geospatial Advisory Committee

PDOP	Positional Dilution of Precision
PHMSA	Pipeline and Hazardous Materials and Safety Administration
RMSE	Root Mean Square Error
UPS	United Parcel Service
US	United States
USPS	United States Postal Service

CHAPTER I - INTRODUCTION

Get to the "Point," Mississippi? Jack Dangermond (ESRI) once stated that "Knowing where things are, and why, is essential to rational decision making." GIS has adequately provided the answer to the "where" questions due in part to the advent of the Internet and the World Wide Web (Longley et al. 2005). Almost every location on Earth can be found by simply searching for an address or the name of a place. Everything is just one click or finger touch away, but what happens when the destination does not exist on a web map or is not found in a GPS unit? Who is held responsible for an individual's death because emergency responders could not accurately locate a person in need? When does the cycle of duplication end for creating address databases because address data is needed by all branches of government and varying levels of society? Addresses are arguably the most prominent widely used type of geographic information used in society. A study in March 2017 showed a list of Google services that were used by consumers. Google Maps ranked second behind Gmail with 66% of consumers using the service occasionally. Google Maps surpassed YouTube, Google Chrome (browser) and Google.com (search engine) (Statistica 2018) (Figure 1.1). Addresses designate the location of existing infrastructure such as homes and businesses, and their accurate depiction is critical to a variety of business processes and operations. Inaccurate addresses and database inconsistencies among a host of users potentially incur a cost to individuals, life and property loss, inefficient routing, and other challenges. A single, highly-accurate and comprehensive address database would improve the efficiency and effectiveness of stakeholders that currently use address data (Table 1.1).

1.1 Brief History of Addressing on a National Level

In 2012, the National Geospatial Advisory Committee (NGAC) assessed the need for the development of a National Address Database (NAD) in a white paper. The white paper included contributions from and access by all sectors of the economy, aggregating and integrating local address data, and conducting a formal benefit-cost analysis to identify the best development options (NGAC 2014). In 2013, the NGAC asked the Federal Geographic Data Committee (FGDC) to develop a funding strategy to implement a NAD. The result of efforts created the vision for a NAD: “The National Address Database is an authoritative and publicly available resource that provides accurate address location information to save lives, reduce costs, and improve service provisions for public and private interests” (NGAC 2014). The following case studies are a set of compelling business cases that support and demonstrate the value and utility of a NAD.

1.2 Case Studies in Support of a National Address Database

1.2.1 Federal Government

The United States Census Bureau requires continuous access to tribal, state, and local address data to update the Master Address File (MAF) used in the Census. The U.S. Census Bureau spent $444 million of taxpayers money and developed an independent MAF complete with geographic coordinates, but could not share with others because of federal law, Title 13 of the U.S. Code that is based around privacy issues (NSGIC 2010). Also, the 2010 Census spent $1.7 billion during the nonresponse follow up operations and vacant house checks (NGAC 2014). A NAD could save the Census

Bureau $196 million for the 2020 Census (Table 1.2). The Federal Emergency Management Agency (FEMA) uses site-specific address information in the preparation/creation of accurate exposure and impact assessments (NGAC 2014). No site-specific address information was a realized problem after Hurricane Katrina (2010) where rescue and recovery operations were slowed because there was no consolidated information source about where people lived. In 2012, rescue and recovery could have been expedited after Superstorm Sandy had FEMA not have to acquire and assemble granular address and remotely sensed data (NSGIC 2010; NGAC 2014).

1.2.2 United States Postal Service

Currently, the Census Bureau maintains a partnership with the U.S. Postal Service (USPS). Within this partnership, the Delivery Sequence File (DSF) of mailing addresses are shared with the Census Bureau, and the MAF is updated with new addresses. Again, under Title 13 US code, the Census Bureau cannot share any updated information back to the USPS, so the partnership is a one-sided partnership. This process is not as efficient as it could be because USPS cannot keep up with the 2 million addresses added each year by new construction and conversions of existing building into multiple occupancy units (NSIC 2010). They rely on updates from the local cities and local mail carrier offices, but data is inconsistent because of a lack of standards for addressing. A NAD would allow the Census Bureau to geocode addresses and share them at the address level and allow the USPS to validate the accuracy of their database and enhance mail delivery. Any updates made by the USPS could feed directly into the NAD and MAF.

1.2.3 State of Arizona

The state of Arizona needed a statewide address database when trying to identify the level of broadband services available throughout the state that was required by the Arizona Broadband Mapping Project (NGAC 2014). Through this project, Arizona was able to build a multi-jurisdictional address database that provided consistent and current address data to be used by all levels of government such as public safety, emergency response, and highway safety. This case study is an example of how local and state efforts can be developed to maintain a statewide address database.

1.3 Standards in Addressing

The current addressing system on the national level is a fractured system because of a lack of standards that are implemented from the federal level down to the local level of government. There is no recognized standard for address data, no central authoritative database, no feedback loop to address stakeholders of new addresses, and spotty capture of geographic coordinates (NGISC 2010). Over time agency databases diverge, and agencies become data hoarders who refuse to share which enables a cycle of duplicated efforts to achieve the same goal. The more participation of a regional area adhering to a standard, the higher possibility for full cooperation (MCCGRIS 2015). A purpose of standards is to create a universal framework that permits sharing of data and resources. Currently, an address system has several necessary components that when combined yield a unique description of the position (Figure 1.2). While there are variations to the components of an address, an address is the foundation of social, commercial, environmental, and political systems (NGAC 2014). The consequences of nonstandardization lead to dozens of systems and schemas unable to communicate with each other or be utilized by stakeholders (Figure 1.3). There are three types of standards that are important because each ensures a best practice guide for creating an address database. The first standard pertains to how an address is named and assigned. Most addresses are created by local government officials such as E-911 offices but are not consistent on a statewide level. The second standard involves how the address information gets recorded based on its address components. Addresses are either parsed out into individual fields or stored in a table as a single field of information. The last standard requires an exchange or use for how addresses are shared.

A statewide address system does not currently exist in Mississippi. The overreaching goal of this research is to determine the most suitable standards for address collection in Mississippi and which standards work best with existing local address systems. Among the many states that currently do not have a statewide database, Mississippi falls victim to the financial pitfalls of not having an accurate database to connect businesses, individuals, and government agencies to a standard, efficient workflow in providing services. The research explored ways to provide accurate geographic coordinates through three different data models and created an address dataset using the NAD standard for addresses. The study also sought to raise awareness of having duplicate, independent address databases and encourage local stakeholders to coordinate efforts and improve workflows.

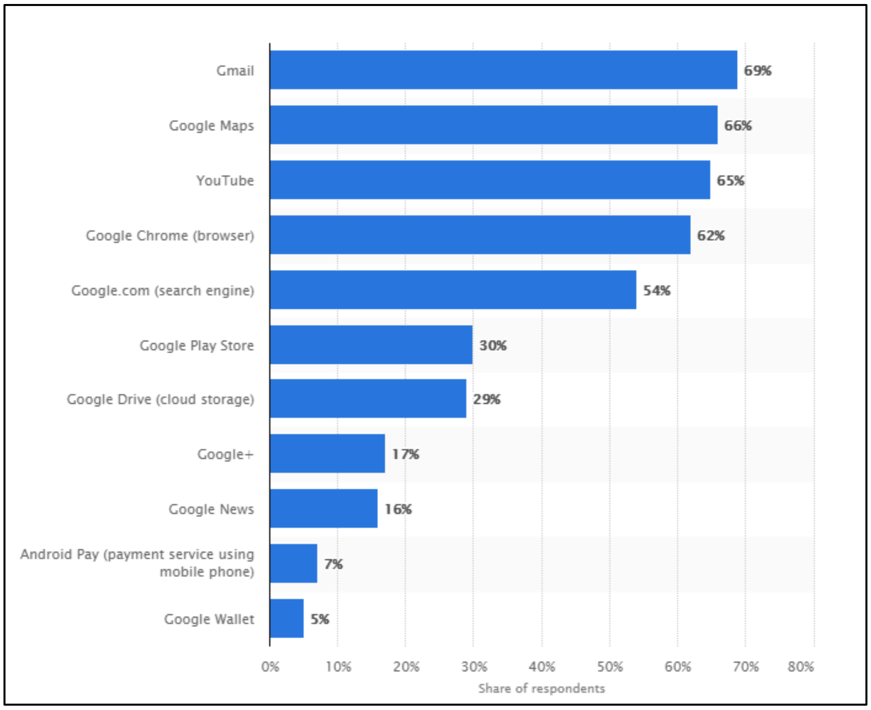


Figure 1.1 Google Services and Products that Consumers Used Occasionally

Figure 1.1 shows the percentage of respondents that used Google services and products occasionally. The Google Maps service ranked second (69%) above Google Chrome (62%) and Google.com (62%) (Statistica 2018).

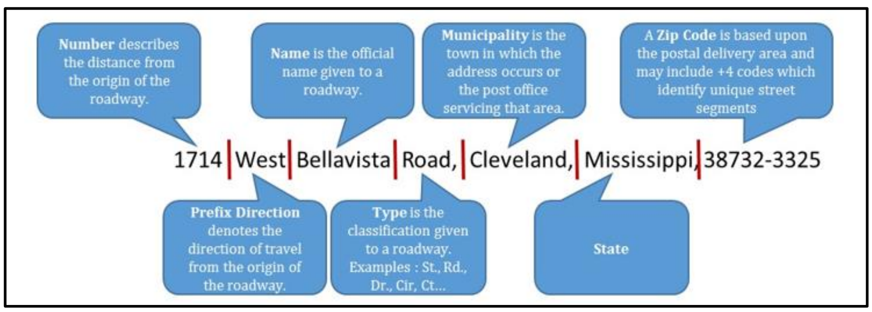


Figure 1.2 Address Components

Figure 1.2 is an example of an address breakdown by its components. Most addresses contain a number, prefix direction, name, and type. Additionally, a city/municipality, state, and zip code are useful when using a map service to find an address (MCCRSGIS 2015).

Table 1.1 Industries that Utilize Addresses

USERS	PURPOSE
Emergency Response, E-911	Police, Fire, Ambulance, Rescue
School Districts	School assignment, bus routing
Assessors and Taxation Offices	Building location
Recorders and Auditors	Property records
Voter Registration	Precinct assignment
Planning and Zoning Office	Building permit, planning studies
State Department of Revenue	Sales of tax collection and distribution
State Department of Transportation	Locate traffic accidents allowing access to Federal Highway Administration (FHWA) funding to improve dangerous non-state roads
State Department of Health and Human Services	Track medical benefits, disease, births/deaths, and vulnerable populations
U.S. Post Office, UPS, FedEx, etc.	Mail and package delivery
U.S. Census Bureau	Mail out census and survey forms, geocode responses
Federal Emergency Management Agency (FEMA)	Pinpoint disaster areas, provide relief
Department of Homeland Security	Locate and protect critical infrastructure
Utilities (public and private)	Locate, protect service areas, hookup, service calls, billing
Map and address companies (ex. HERE, TeleAtlas, Pitney Bowes)	Sell to insurance companies, location-based service companies, utilities, state and local government, etc.
Retail/Service	Delivery of goods and services
Internet maps (ex. Google Maps, Bing, Waze)	Navigation maps for public use

Table 1.1 lists the industries and purpose that would benefit from having a National Address Database, and the purpose shows the application and the need for a National Address Database. The Need for a National Address Database is essential to these industries (NGISC 2010).

Table 1.2 Census Bureau Costs by Operation

Census Operation	Total Cost in 2010 Census	Estimated Cost Avoidance with the NAD
Address Canvessing	$443,591,299	$35,733,480
Nonresponse follow-up	$1,589,397,886	$159,744,030
Non-ID	$3,725,555	$983,082
Total	$2,036,714,740	$196,460,592

Estimated cost avoidance for the Census Bureau with investment in the National Address Database (NGAC 2014).

CHAPTER II – LITERATURE REVIEW

2.1 GPS

A global positioning system (GPS) is a global constellation of thirty-one satellites that emit and receive positional information through trilateration[1] from a satellite to a receiver to determine a location on Earth (Kaplan and Hegarty 2006; Milner 2016). The system is broken into three different components: space, control, and user. The space segment incorporates the constellation of twenty-four core satellites and several backup satellites that fly in a medium orbit and circle around the Earth twice a day (NCO 2017). The control segment has global control and monitor stations that track GPS movement and satellite health (NCO 2017). The last segment, users, consists of the GPS receiver that accepts passive signals from GPS satellites and calculates the three-dimensional position and time (NCO 2014). Much like the Internet, GPS has grown extensively over the last decade as applications have become integrated into the global economy. An approximate 5 billion receivers are in current use across all technological platforms such as telecommunication, aerospace, agriculture, autonomous vehicles, mobile mapping, survey, defense, marine, and timing applications (Milner 2016). In 2011, the estimated value of global GPS was $9.1 billion (Milner 2016), and it is projected to triple as GPS technology continues to increase. Today, the purpose and capabilities of GPS are recognized and widely accepted. However, the history of GPS portrays a unique background of project abandonment, insufficient funding, and military branch independence.

[1] Trilateration – The process of determining position of earth by using distances instead of angles like triangulation (Penn State College of Earth and Mineral Sciences 2017)

Initially developed by the Department of Defense for precision weapon delivery and military navigation in the early 1970s, the idea of GPS came to fruition through several pre-existing military programs: Transit and Timation through the US Navy and 621B through the Air Force (Pace et al. 1995; Milner 2016). Transit was the first operational satellite-based navigation system that allowed users to measure location based on the Doppler shift[2], thus proving that space-based technology was reliable (Milner 2016). Transit provided satellite prediction algorithms, but it was slow, required long observation times and velocity corrections (Pace et al. 1995). It was not practical for aircraft or rapidly moving platforms like missiles. Timation, also a space-based program, focused on the development of high-stability clocks, time-transfer, two-dimensional navigation, and by using two experimental satellites, it demonstrated technology for three-dimensional navigation (Pace et al. 1995; Milner 2016). At the same time, the Navy was working on the Timation program; the Air Force was working on its version of three-dimensional navigation that provided continuous services with a vision of having a system with global coverage of satellites in geosynchronous orbits (Pace et al. 1995; Milner 2016). Unfortunately, the 621B program never progressed any further than the demonstration stages. In the late 1960s, the Navy, Air Force, and Army were all working independently on radio-navigation systems. The Department of Defense formed a joint committee involving all three services. The committee spent several years deciding on the specifics of a satellite navigation system and costs involved. Colonel Brad Parkinson, whom would later be recognized as the father of GPS, was put

[2] Doppler shift- The change in frequency for an observer that is moving relative to the frequency source (Example: The sound of an ambulance changes pitch as it approaches and passes by)

in charge over the joint program office to reach a compromise for the type of system that would benefit all services. Still, Parkinson faced adversity from budget cuts and funding problems to the Challenger disaster that was slated to carry future GPS satellites into orbit (Milner 2016). GPS revolutionized military combat operations beginning in the Persian Gulf War as well as Operation Desert Storm. President Ronald Reagan made GPS available to civilians after the downing of a Korean airplane over Russia (Pace et al. 1995; Milner 2016). Private companies, Hewlett Packard and Trimble, sought the rights to the GPS program and began creating GPS receivers for civilian use (Milner 2016).

2.2 Types of GPS Receivers

There are three grades of GPS receivers that have varying levels of precision and accuracy. Precision is the level of repeatability of measurement and accuracy is the proximity to the true value or accepted value of measurement (Figure 2.1). Consumer grade GPS receivers have an accuracy of 15-30 meters and can simultaneously track up to 12 satellites using the GPS antenna (UNC-Chapel Hill 2007). Consumer grade receivers are more prone to experience errors such as multi-path or signals reflected off buildings or walls which is why sometimes a GPS receiver will show the wrong current location for a user. A mapping grade receiver has built-in software to resist multi-path error and user-defined options for positional dilution of precision (PDOP)[3]. Mapping grade receivers allow for field collection, but post-processing of the data must be done on a computer in the office. Mapping receivers can correct the positional error with the use

[3] PDOP – Error caused by the relative position of the GPS satellites. The more signals a GPS receiver can "see", the more precise the GPS reading.

of differential GPS coordinates as ground reference stations. Mapping grade receivers have sub-meter accuracy. Survey grade receivers are like mapping grade receivers but have an accuracy of 5-30 millimeters in the horizontal/vertical direction. Survey grade receivers also have built-in software to help eliminate multi-path errors. Every GPS receiver is not perfect when capturing GPS locations. GPS receivers can have errors. Satellite geometry error is associated with the level of dilution of precision (DOP). A Multi-path error is when the GPS signal reflects off buildings, cars, trees, etc. Atmospheric effects can cause unwanted errors of +/- 5.5 meters in data collection if there are clouds and water vapor in the troposphere or electromagnetic interference in the ionosphere. Errors in satellite orbits, clock inaccuracies rounding errors can also cause GPS error of +/- 2.5 meters. Methods to eliminate errors include:

(1) Using software and antennas designed to resist multi-path interference
(2) Avoiding use of high powered CB radios because of frequency
(3) Setting GPS receivers so that data cannot be collected when PDOP is greater than 6
(4) Setting elevation mask to track only satellites 15 degrees above the horizon
(5) Using differential GPS for real-time broadcast of solutions

2.3 Geocoding Process

Geocoding is the process of transforming a location description, an XY coordinate pair, an address, or place to a location on the earth using a reference dataset (Zandbergen 2008; Goldberg 2009). The geocoding process involves three processing components, address processing, feature matching, and feature interpolation, as well as a reference dataset or multiple reference datasets for input addresses to be compared against (Figure 2). Address processing is parsing the input address into individual address components and fields (such as street name, street type, etc.). Because there are several ways to input an address, a standardized method is performed to format the input, so it can be matched and indexed against the reference dataset to return the best match. The geocoding process spans across many types of geocoding applications, there are frequent problems that cause the match rate to be poor and the result to be incomplete. Part of the feature matching algorithm uses both probabilistic and deterministic approaches. Probabilistic record linkage is the process of matching two sets of data under certain conditions of uncertainty (Zandbergen 2008). Probabilistic record linkage tries to link records which represent commonalities such as events, businesses, an institution, or address through fuzzy logic to score how records match or do not match. Deterministic record linkage assumes an error-free approach finding and linking records that match exactly with the reference data (Zandbergen 2008). To account for human misspellings a phonetic indexing system, Soundex, is used. Soundex indexes information based on how the word sounds rather than how it is spelled (Zandbergen 2008). The last address component of the geocoding process is feature interpolation. Feature interpolation is when the address location is interpolated from the reference data and street range on where an address falls along the road.

2.4 Street Network Data Model

Street names are often used in defining the location, so street names are used in address geocoding tools. As a result, commercial vendors have created custom geocoding tools and reference data as well as web-based address engines such as Google Maps, Bing Maps, and Yahoo Maps. Zandbergen (2008) investigated the foundations of the geocoding process, geocoding address data models and geocoding quality in an empirical comparison study. The three address data models that were evaluated were a street network, parcel boundaries, and address point address models. The street network data model is the most widely employed address model. All commercial vendors and most GIS geocoding software rely on street geocoding. The street network data model incorporates storing different names and address ranges to interpolate addresses where there is no address house number (Figure 2.3). The limitations to street geocoding is that positional error increases in rural areas and local statistics for detecting clusters (Burra et al. 2002; Cayo and Talbot 2003). Another limitation is having a good reference dataset (Zandbergen 2008). The second type of address data model is parcel boundaries.

2.5 Parcel Boundaries Data Model

Parcel geocoding are the most spatially accurate data with address information available (Rushton et al. 2006; Zandbergen 2008). However, the matching process against parcel plots or centroid of the polygon is much lower than by street or point address data models. A match in parcel geocoding is only a match to a single parcel address with one house number while street geocoding has an address range (Zandbergen 2008). Limitations to parcel boundary geocoding are address validation within an area. An address could have a non-standard reference listed in the house number field. Also, parcel geocoding does not account for multiple addresses within one parcel such as apartment complexes (Zandbergen 2008). The last address data model to overcome the limitations of both parcel and street geocoding is point address points.

2.6 Point Address Data Model

A point address data model is derived from a master address file (MAF). An example of a master address file is the master street address guide (MSAG) from E-911 for emergency purposes. Address point data can be derived from several existing data layers such as parcel data by creating a centroid point for all occupied parcels. Then, an address point can be moved to cover the primary structure, front door, or driveway as well as have points added for sub-addresses like apartment units, duplexes, etc. which do not have a separate parcel boundary (Zandbergen 2008). Field collection or verification using aerial imagery or driving to the address location can further supplement the point address data. However, the positional error can be compromised during the addition process using aerial imagery alone. Address points can be added or mislabeled to sheds or barns instead of the primary structure. Some subaddresses can still be overlooked when one structure is present using aerial imagery. Field verification of verifying addresses decreases the positional error and attribute quality of the point address data. Several commercial firms have started geocoding in the U.S. for selected urban areas but is not very widespread at this time (Zandbergen 2008).

2.7 Geocoding Studies

Previous studies by Cayo and Talbot (2003) determined the positional error in automated geocoding of residential addresses. Residential address data, parcel data, census tract data and high-resolution aerial orthoimagery were obtained for the research. The parcel data was used in conjunction with the census tract data to assign each address classification of urban, suburban, or rural based on population density. A valid street number, street name, and zip code were required for the geocoding process. MapMarker Plus Version 6.0 software was used to match the residential address to the software's street reference files. Mapmarker successfully geocoded 81% of the addresses in the research. The positional error was measured by obtaining the true location for a random sample of 1,000 addresses that geocoded correctly with MapMarker. A true location point was created as a third dataset. The true location was defined as the center of structure that was visually represented using 1-meter resolution on the high-resolution aerial orthoimagery. Straight-line distances were calculated between the true location and the automated geocoded points to compute the positional error. Cayo and Talbot (2003) concluded substantial differences in positional error between the automated geocoded points and the true location points amongst the different residential classifications with a high error in rural areas over urban and suburban areas. Parcel coordinates significantly reduced the mean positional error in rural areas. Zandbergen (2008) compared the three different address data models by obtaining data from six different databases for the same area. The databases had to meet the following criteria:

(1) Database had to be publicly available
(2) Database had to be recently updated
(3) Database had to be available for the entire state to allow for comparisons
(4) Sufficient sample size was needed

Address point data, parcel data, and street centerlines data were obtained and used as reference data for the six databases. Each of the reference datasets contained several different attributes for the address, but all had the minimum attributes: number, prefix direction, street name, street type, and suffix. Address locators were created for the three reference datasets that included the number, prefix direction, street name, street type, and suffix. The thresholds for spelling sensitivity and match scores were set to identical thresholds with the minimum match score set to a value of 60 (out of 100). This technique was used so that an address that was not a perfect one-to-one match or did not fall within a street range, the maximum score received based on the geocoding algorithms was 52. Ties[4] were permitted in the analysis but identified separately from the results. Zandbergen (2008) analyzed the number of perfect matches and ties, the number of additional matches and ties and the number of unmatched candidates. Zandbergen (2008) concluded that match rates for address point geocoding are only slightly lower than for street geocoding, but the higher rate for street geocoding could be due to false positives. This result confirmed that extensive field validation is required to eliminate false positives. Overall, parcel geocoding was much lower but varied by the database and geographic area. The geocoding quality of this research is very much a function of the quality and consistency of local reference data (Zandbergen 2008). The current research combines methods from both recent literature of Cayo and Talbot (2003) and Zandbergen (2008).

2.8 Applications of Addressing

Over the last decade, GIS has become interwoven into everyday tasks. As aforementioned, every aspect of technology relies on the Internet and GPS. GPS is the most accurate timekeeping application in the world which allowed for the integration into different types of industry. Addressing and GPS complement each other because,

[4] Tie – The address has more than one candidate with the same match score but at different locations (ArcGIS 2016).

without one another, point location would not work. The use of GPS in emergency services is one of the fastest growing technologies due to the sophistication of spatial mapping in law enforcement (Ratcliffe 2004). E-911 uses GPS in dispatching first responders to the location of an emergency using a spatial mapping component and address information given to the dispatcher. First responders use GPS to get to the incident location. During emergency response events such as after a tornado or hurricane, GPS allows for coordinated efforts to occur between different agencies. In the same way that emergency services use GPS, utility companies also use it in locating, maintaining, and updating infrastructure. One of the first applications of GPS in the utility industry was with the 2.6 million miles of natural gas and hazardous liquid transportation pipelines that run across the U.S. countryside. The Pipeline and Hazardous Materials and Safety Administration (PHMSA) enforced regulations to ensure the safety of the design, construction, operation, and maintenance of pipelines by capturing GPS coordinates for all pipeline infrastructure as pertinent to national security. Similarly, other utility companies obtain GPS coordinates for customer meters and location of service areas. The location of service areas, especially those underground, are important regarding public safety.

2.9 Addressing in Mississippi

Mississippi does not currently have a statewide address database. Although there have been committees formed in the past to create one, the financial burden of a statewide project is the reason why no such database currently exists. The creation of a point address database has started on a local level with E-911 offices, GIS departments, and private firms building databases for one county (Figure 2.4), but fifty counties remain unaddressed. Some of those fifty counties substitute not having the funding to create a database by utilizing a web map service such as Bing Maps and Google Maps. However, geocoding is only as good as the reference dataset utilized to interpret locations. In rural areas, addresses can be hundreds of feet from the true location. For example, Figure 2.5 shows the point location for the address 1996 Tanyard Road, Hernando, MS. In Google Maps, the location of the house falls along the road. The true location of the house is located approximately 0.5 miles off the road. Though that is a limitation of street geocoding, the question of positional accuracy is repeatedly challenged because Mississippi has a higher population that live in rural areas than urban. There are only three states (Maine, Vermont, and West Virginia) that have a higher rural population than Mississippi (Logue 2011). In 2000, 59 counties had 50% or more rural population, and 21 counties were classified as 100% rural (Logue 2011).

Though there are current Mississippi counties with GISs in both government and private departments across the state, there is no concurrent process or methodology for address collection nor is there a statewide address database readily available for public download on the Mississippi Automated Resources and Information Systems (MARIS) website. Current methodologies and data collected lack address, location, and metadata information in data workflows as well as a consistent method for collection amongst the GIS community in Mississippi. The research did not restrict the collection approach to one address model type but sought to present a holistic approach to address collection and maintenance. The research added to the existing data collection workflows done in earlier literature and created the most cost-efficient workflow and a best practice reference guide for Mississippians. The following research questions and analysis are presented in Chapter Three:

(1) Which addressing method is most suitable in Mississippi?
(2) How does the positional accuracy change over space between rural and city designated areas?
(3) What is the spatial and temporal variation of the aerial imagery?

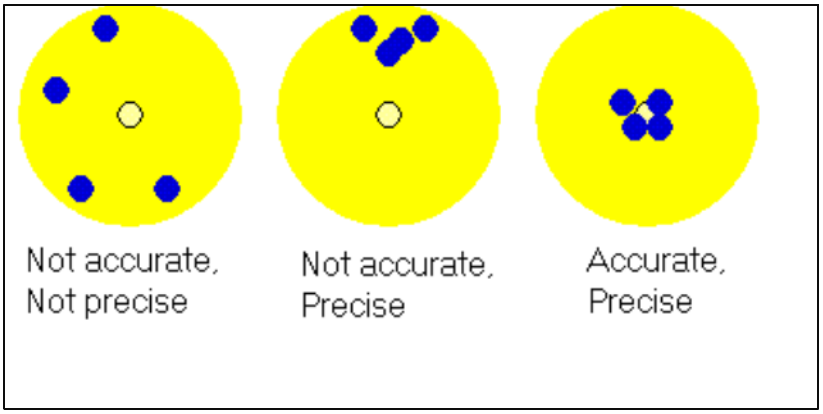


Figure 2.1 Precision versus Accuracy

Precision versus accuracy is important when using GPS receivers. If a GPS unit is not calibrated correctly, precision and accuracy can be affected. A GPS unit can be precise but not accurate during the collection process and skew the data results (UNC-Chapel Hill 2007).

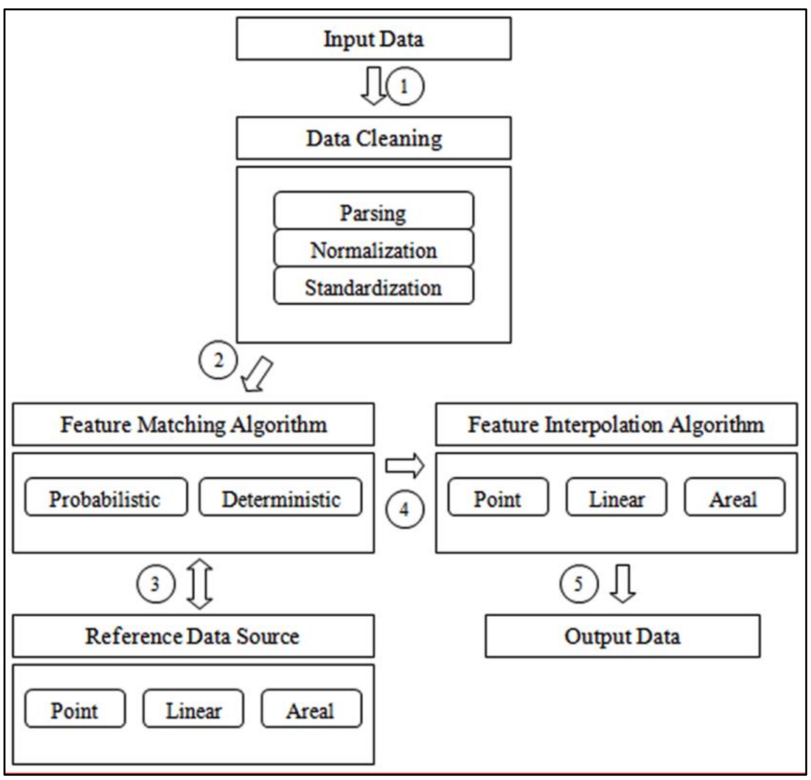
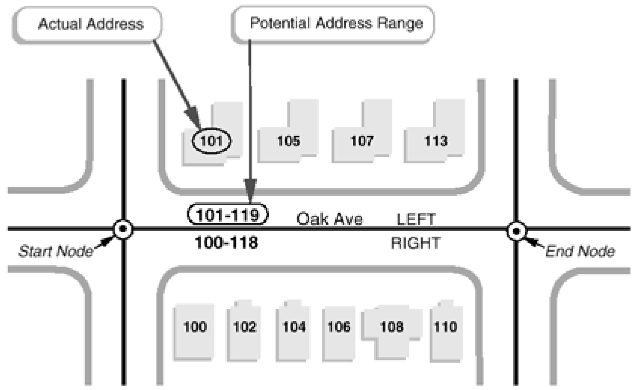


Figure 2.2 Geocoding Process

The geocoding process involves three processing components: address processing (step 1), feature matching (step 2) and feature interpolation (step 3). This process transforms a location description into an XY coordinate pair (Goldberg 2009).



Record Type 1 contains separate data fields for both the start and end of each address range.

Record Type 1				Address Range			
				Left side		Right Side	
				Start	End	Start	End
RT	TLID	FENAME	FETYPE	FRADDL	TOADDL	FRADDR	TOADDR
1	0007654320	Oak	Ave	**101**	**119**	**100**	**118**

Figure 2.3 Example of a Street Network Data Model

Of all three data models, the street network data model is the most widely employed. The street data model incorporates storing different names and address ranges to interpolate addresses where there is no address house number.

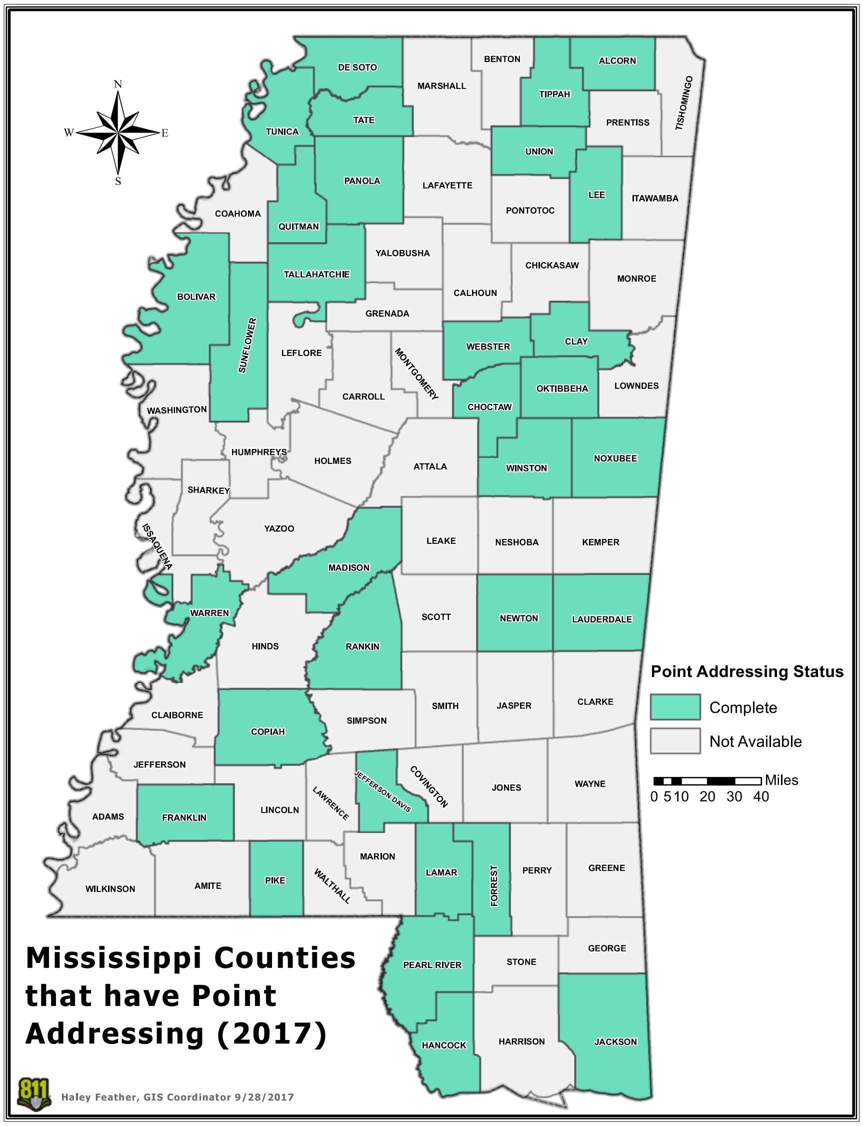


Figure 2.4 Total Number of Mississippi Counties that have Point Addressing

Mississippi does not currently have a statewide address database. The responsibility of creation has fallen on E-911 offices and local GIS departments. The number of counties that have already created an address database has a green color shade, and the counties that remain to be addressed have a grey shade.

Figure 2.5 Example of Street Network Data Model Versus Point Address Data Model

Web map services such as Google utilize the street network data model, which interprets an address based on a street range. In rural areas, addresses can be hundreds of feet from the true location. Figure 2.5 is an example of a common problem found when using street geocoding. For the address, 1993 Tanyard Road, Hernando, MS, Google Maps places the address close to the road (top left), but the true location is off of the road (top right). The large distance (0.5 miles) can present problems for industries that utilize addresses to provide goods and services (Google Maps 2017).

CHAPTER III – METHODOLOGY

3.1 Study Site

Reliable and available reference information for address point, parcel, and street geocoding was not available for all counties in Mississippi. As a result, this study employed extensive search criteria to gain access to address data of various types. The street centerline data was available for all 82 counties and had the required fields for geocoding. The parcel data was available for 10 counties, and 8 out of the 10 counties had complete address information for geocoding. There were only 32 counties that had a point address dataset, and their quality was considered acceptable for research purposes. For time restraints on the research, only one county was chosen (Figure 3.1). Tippah County has a land area of 457.82 $miles^2$, a population of 22,232, and is located in northeast Mississippi (US Census 2010) (Figure 3.2). Tippah County has 5 municipalities: Blue Mountain, Dumas, Falkner, Ripley, and Walnut (Table 3.1). The two cities chosen for the research were Ripley and Falkner because both cities had a large sample of structures in both the 2006 and 2015 aerial imageries.

3.2 Sources

Three different datasets were obtained for use in the comparison of the three address data models (Table 3.2). Addresses for electric power customers were obtained from the county electric power company in August of 2017 (n=14,833). The data were not standardized and only contained a latitude and longitude coordinate pair and the physical address. Though the dataset was not already standardized, it was chosen for this research as a dataset for its addresses in rural areas. The dataset was standardized and restructured to the National Address Database (NAD) schema (Appendix A). The addresses from the E-911 office were obtained in June 2017 (n=10,709). This dataset was created using the center of the structure and already standardized to the minimum street requirements (prefix, name, type, suffix). It was last updated in May 2017, so it was an adequate database to use in the research. The data was produced using both the parcel data model and the point address data model. A GPS was used in the field verification process, and it had a 12-foot margin of error. The addresses from the tax accessors office were obtained in January of 2017 from the county tax accessor (n=8,523). A polygon extraction was performed on the parcels that had that had an address (Figure 3.3), and a feature to point conversion was completed on the polygon layer to create a point at the center of every polygon (Figure 3.4). A physical address aided in the elimination of converting points for empty parcels with no structures. The original dataset met the minimum street requirements but was restructured to the NAD schema. The raster datasets used in the research were county flown aerial imagery from 2006 and 2015 with spatial resolutions of two feet and twelve inches, respectively. The 2006 aerial imagery was flown to be used for cadastral (tax) and infrastructure mapping purposes. The data met the National Standard for Spatial Data Accuracy (NSSDA) at a 1:400 map scale and had an RMSE value of < 2 feet. The data used a North American Datum of 1983 HARN projection. The 2015 aerial imagery was flown to be used to update county GISs. It met the American Society for Photogrammetry and Remote Sensing (ASPRS) class 1 accuracy for 1: 200 map scale and had an RMSE value of 2 feet. The data used a state plane coordinate system, NAD 1983 Mississippi East, projection.

A geodatabase was created to keep the data organized and allow for topology and attribute domains to limit errors. The feature datasets were created as administrative boundaries, cadastral, point addresses, temp, and transportation with a North American Datum 1983 2011 State Plane Mississippi East FIPS 2301 Feet US projection using ArcMap 10.3 mapping software from ESRI.

3.3 Which addressing method is most suitable in Mississippi?

The research sought to answer how many addresses within the parcel and electric power association datasets would match the reference dataset (E-911) as well as how the street ranges would match through automatic geocoding. The research assumed that one address was linked to one structure per property boundary, meaning a one-one relationship. No multiple addresses, mobile home parks, or duplexes were included in the analysis. Prior to building the street data model, the streets were verified in the field for spelling and correctness as well as compared to the Master Street Address Guide (MSAG) for address attribute completeness. In the ArcGIS mapping software, ArcCatalog, a dual range street locator was built using the street centerlines and the range fields. The E-911 address spreadsheet was geocoded using the street locator and ArcMap geocoding tool. Another address locator was built in ArcCatalog. A point locator for single houses was chosen, and the parcel data was used as the reference dataset. Using the ArcMap geocoding tool, the electric power association address spreadsheet was geocoded using the parcel locator. Then the geocoding steps were repeated, and the E-911 address spreadsheet was geocoded with the parcel locator.

3.4 How does the positional accuracy change over space between rural and city designated areas?

As previously stated, this research only looked at one-one relationships between an address and structure. Structures associated with multiple addresses were not included in the analysis. Initially, the parcel dataset obtained from the tax assessor's office was a polygon shapefile. The parcel data was extracted from the tax assessor's database and formatted to only include the following fields that applied to the study: Parcel_ID, ownername, address1, address2, address3, sub_num,neigh_code, street_num, street, and sub_name. In the ArcMap software, a feature to point tool was implemented on the parcel polygon layer. The output result was the creation of a point in the center of every parcel (Figure 3.4).

Two aerial imagery datasets were created by spatially digitizing or creating points in each dataset using a 1:500 map scale (Figure 3.9 and Figure 3.10). Points in each dataset were created where the front door was located based on visual cues on the imagery such as awnings, porches, and sidewalks. Using a convenience sampling technique 300 points (150 inside the city limits and 150 outside the city limits) were selected so that a control point/true location could be obtained in the field with a GPS unit. Using an Ike 3.0 rangefinder GPS unit points were collected in the field by shooting a laser at the front door of each structure. Using a laser allowed for the point collection to take place from the road rather than trespass onto private property. The threshold for sampling was a margin of error for the GPS unit of +/- 8 feet.

MapSight software was used to extract the field collected points from the GPS unit to a desktop computer. The field points were exported out to a shapefile. GPS point outliers were removed from the analysis due to field error. An 8-foot buffer was created around the GPS points to represent the threshold margin of error (MoE). A spatial location search was performed to see how many points from each of the datasets (E-911, Parcel, 2006 aerial imagery, 2015 aerial imagery, electric power association) fell within the GPS MoE and grouped by urban or rural. Next, a merge was implemented on the datasets (E-911, Parcel, 2006 aerial imagery, 2015 aerial imagery, electric power association) and grouped by full address. An average XY value was calculated using the mean center tool and grouped based on the full address (Figure 3.5 and Figure 3.6). Then, a standard distance buffer was created using the average mean center (Figure 3.7 and Figure 3.8). The buffer size of one standard deviation was created, and the output results were grouped by full address. A spatial location selection was completed to capture the number of points that fell within the average distance buffer from each dataset (E-911, Parcel, 2006 aerial imagery, 2015 aerial imagery, electric power association) and grouped by urban or rural. The standard distance was categorized into five groups using natural breaks (Jenks) classification. The output results were grouped by urban or rural.

3.5 What is the spatial and temporal variation of the aerial imagery?

The research investigated the spatial variation between the 2006 and 2015 imageries. Two copies of the E911 address dataset were exported out. Points within each dataset were moved to where the front door was likely located using awnings, porches, and sidewalks as an aid. Each dataset was digitized using a 1:500 map scale (Figure 3.9). A convenience sampling technique was used to obtain 300 points (150 inside the city limits and 150 outside the city limits). These points were selected for field collection with the GPS unit of the control point/true locations. An 8-foot buffer was created around the GPS field points to represent the margin of error (MoE). A spatial location selection was performed to see how many points from each of the datasets (2006 aerial imagery and 2015 aerial imagery) fell within the GPS MoE and grouped by urban or rural.

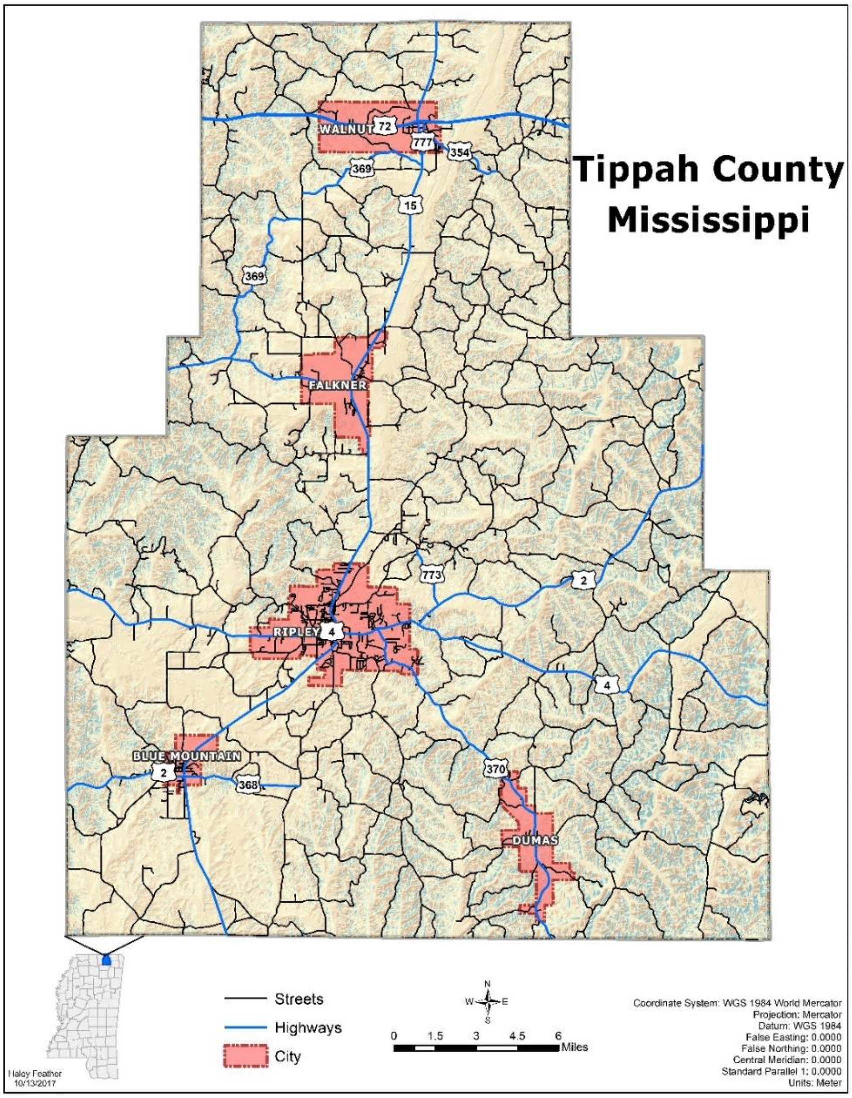


Figure 3.1 Location of Study County – Tippah County, MS

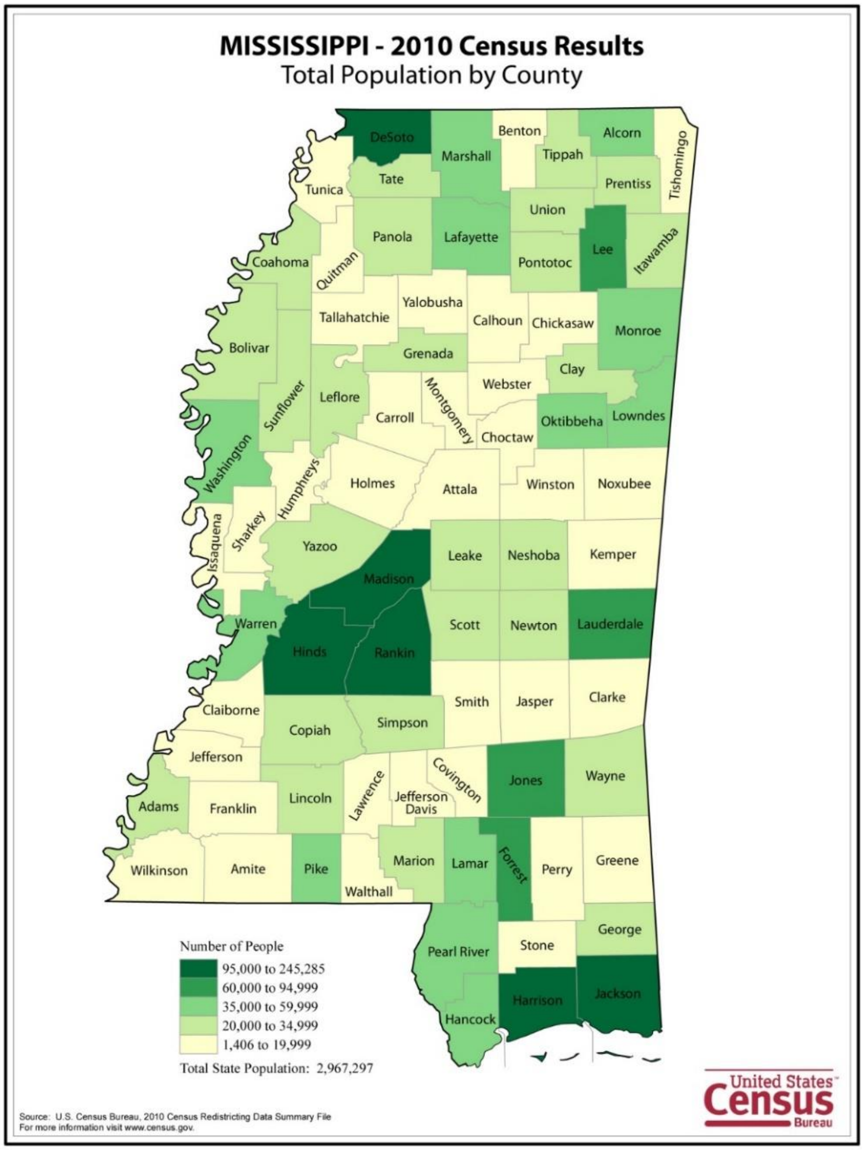


Figure 3.2 2010 Total Population Counts by County

Figure 3.2 shows the total population by county using the 2010 data summary file. Tippah County falls into the category of a total population between 20,000 – 34,999 represented by the light green color shade (US Census 2010). The population of Tippah County is 22,232.

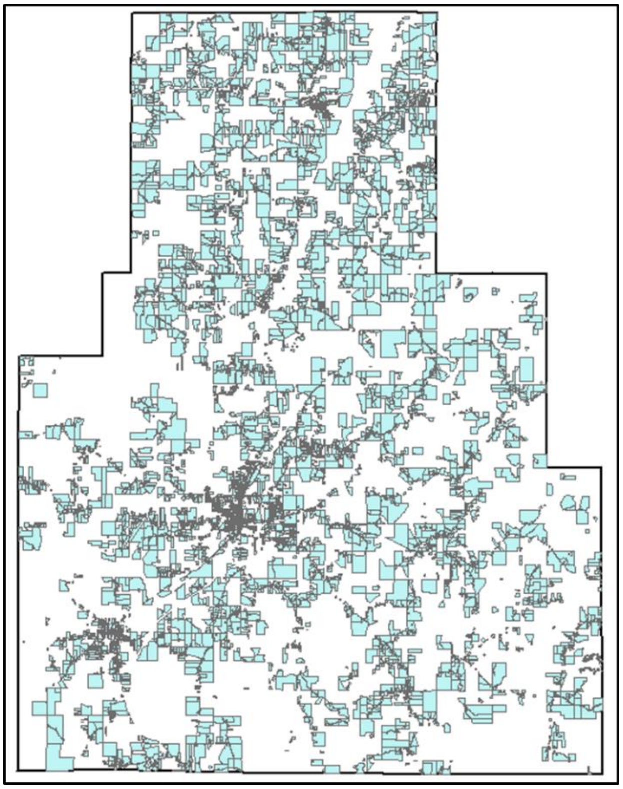


Figure 3.3 Parcel Extraction with Addresses

Parcel polygons that contained an address in the parcel tax roll were extracted to a new dataset. This step ensured that the parcel data model utilized the addresses provided in the tax roll attribute table.

Figure 3.4 Polygon to Point Conversion

A point centroid was created using the feature to point tool, and a point was created for every parcel that contained an address.

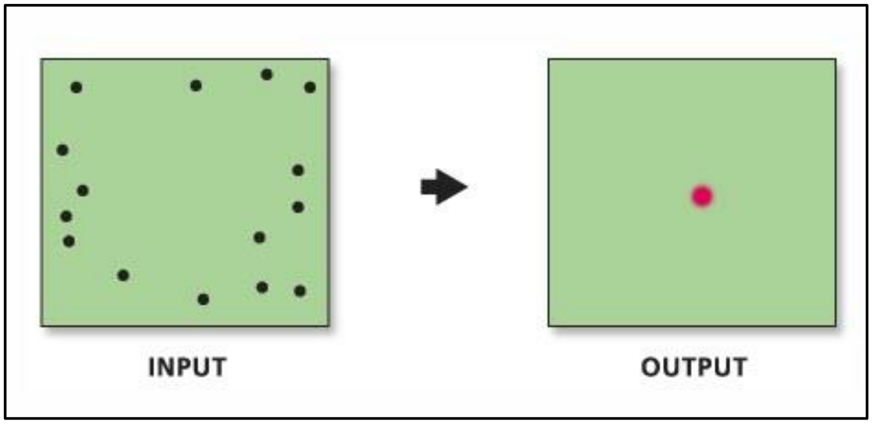


Figure 3.5 Illustration of Mean Center

Figure 3.5 is an example of how the mean center tool works (Mitchell 2005). For each cluster of points, the tool creates an average point in the output layer. The output created an average point for the E-911, parcel, 2006 aerial imagery, 2015 aerial imagery, and electric power association data that were associated with each structure and shared the same address.

The Mean Center is given as:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad (1)$$

where x_i and y_i are the coordinates for feature i, and n is equal to the total number of features.

The Weighted Mean Center extends to the following:

$$\bar{X}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad , \quad \bar{Y}_w = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \qquad (2)$$

where w_i is the weight at feature i.

The tool also calculates the center for a 3rd dimension if a z attribute exists for each feature:

$$\bar{Z} = \frac{\sum_{i=1}^{n} z_i}{n} \quad , \quad \bar{Z}_w = \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i} \qquad (3)$$

Figure 3.6 Mean Center Equation

Figure 3.6 shows the mean center equation used to calculate an average point (Mitchell 2005). An average XY value was created around each structure point and full address to determine if an average could be used as an addressing method for creating an address database.

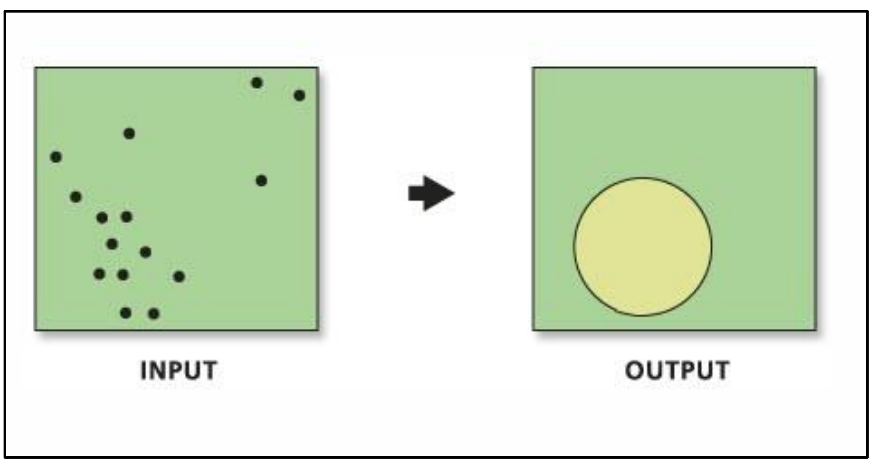


Figure 3.7 Illustration of Standard Distance

Figure 3.7 is an example of the inputs and outputs for the standard distance tools (Mitchell 2005). This tool created an average distance buffer around the points from each dataset to each structure based on full address.

The Standard Distance is given as:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} + \frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n}} \qquad (1)$$

where x_i and y_i are the coordinates for feature i, $\{\bar{X}, \bar{Y}\}$ represents the Mean Center for the features, and n is equal to the total number of features.

Figure 3.8 Standard Distance Equation

Figure 3.8 shows the calculation equation for the standard distance tool. The equation used the average mean for each structure and its related points from each dataset and created an average buffer circle. Each buffer was grouped based on a full address which allowed for individual buffers to be created around each structure (Mitchell 2005).

Figure 3.9 Front Door Point Placement for the City of Ripley Using 2006 Aerial Imagery

Figure 3.9 is an example of what GIS layers were used to create the 2006 aerial imagery dataset. The parcel polygons (red) were used as a spatial reference to identify structures and points (yellow) were created using the 2006 2-foot aerial imagery and positioned where the front door was likely located on each structure using awnings, porches, and sidewalks as an aid.

Figure 3.10 Front Door Point Placement for the City of Ripley Using 2015 Aerial Imagery

Figure 3.10 is an example of what GIS layers were used to create the 2015 aerial imagery dataset. The parcel polygons (red) were used as a spatial reference to identify structures, and points (green) were created using the 2015 12-inch aerial imagery and positioned where the front door was likely located on each structure using awnings, porches, and sidewalks as an aid.

Table 3.1 Population Counts and Population Density for Tippah Municipalities

City / Town	Area (sq km)	Population	Population Density (people per km^2)
Blue Mountain	4.3	920	216.2
Dumas	10.1	470	46.5
Falkner	13.1	514	39.1
Ripley	29.3	5395	184.1
Walnut	14.1	771	54.8

Table 3.2 Research Datasets and Sources

Data Set	Source	File Format	Spatial Reference	Year Collected
Parcels	Tri-State Consulting	Shapefile	NAD_1983_StatePlane_Mississippi_East_FIPS_2301_Feet	2016
2006 Aerial Imagery	Mississippi Automated Resource Information Systems	Raster	NAD_1983_HARN_StatePlane_Mississippi_East_FIPS_2301(US feet)	2006
2015 NAIP Imagery	Mississippi Automated Resource Information Systems	Raster	NAD_1983_StatePlane_Mississippi_East_FIPS_2301_Feet	2015
Streets	Navteq	Shapefile	WGS 84	2010
Point Addresses	Tippah County Electric Power Association	Excel	NAD_1983_StatePlane_Mississippi_East_FIPS_2301_Feet	2016
Point Addresses	Everything is Somewhere, LLC	Shapefile	NAD_1983_StatePlane_Mississippi_East_FIPS_2301_Feet	2017

CHAPTER IV – RESULTS

4.1 Which addressing method is most suitable in Mississippi?

There were three-address data models that were used in the geocoding process to determine model suitability for Mississippi: street data model, parcel data model, and point address data model. There were three different categories of results returned after the geocoding process was completed: matched, unmatched, or tied. Each result described how well the address components in the reference dataset and the address input dataset were matched or geocoded.

4.1.1 Street Address Data Model

The result of the street data model using the street address locator did not yield a high number of geocoded matches when it used the address spreadsheet provided by E-911 (Table 4.1). Inside the urban areas, the number of matches was 14, and in the rural areas, only 4 out of a possible 150 points were matched. The XY coordinate information for the matches that were produced by the street locator fell along the road in correspondence with the street range.

4.1.2 Parcel Address Data Model

For the parcel data model, the original tax roll table was used as the input address dataset, and two separate address locators were used to geocode the parcel address information. The street address locator results were similar to those of the street data model (Table 4.2). In the urban area, 13 matches were returned, and 5 matches geocoded correctly in the rural area. However, the results of the point address locator for the tax roll were almost opposite from the street locator results. Within the urban areas, there were 145 matched addresses and 135 matched addresses in the rural areas. There was only 13 unmatched addresses total, and all were in rural areas. The number of tied records in the results were 6 with one address component, road type, not matching against the reference dataset. Two different input address datasets were used in the geocoding process, but the same street reference dataset was used in both address locators. The quality of the street reference dataset could have played a role in the match score.

4.1.3 Point Address Data Model

For the point address data model, the electric power association data was the input dataset, and the same address locators, street and point were used. The street locator resulted in 137 unmatched in the urban areas and 145 in the rural areas with more matched in the urban with 14 than in the rural area with only 5. There was only one tie out of both urban and rural areas, and the address component that was different was the prefix direction field. Comparable to the trend of higher match rates in point address locators, the number of matches inside the urban areas was 136, and unmatched were 8 addresses. In the rural areas, the number of matches was higher than in the urban areas with 142 matches and 136 rural matches. The electric power association dataset was the only dataset that yielded higher match rates in the rural areas than urban areas contradicting previous literature stating match rates are higher in urban areas than rural with geocoding. There were only 15 unmatched addresses in total between urban and rural but rural had one less unmatched than urban with seven. These results of higher match rates in the rural areas than urban could have occurred because the electric power association address attributes were more accurate than the parcel tax roll addresses in the attribute table. The street locator did not yield high match rates regardless of the input address dataset. The match rates were substantially higher by using a point address locator that utilized an address style for single houses. Both the parcel and the street geocoding data models did not depict the geocoded address point to be on the on the structure. Though the geocoding process utilized the quality of the attributes within each dataset, the determination of the positional accuracy was demonstrated by creating a control and average dataset.

4.2 How does the positional accuracy change over space between rural and city designated areas?

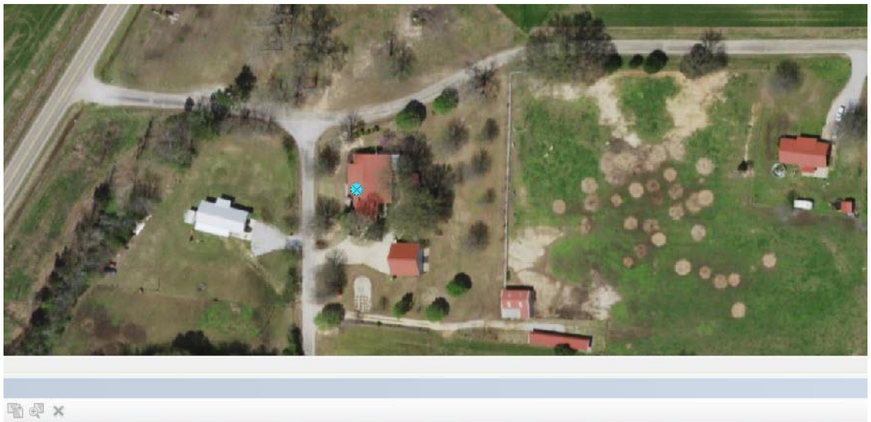
To determine the positional accuracy of each dataset: electric power association, E-911, aerial 2006 and 2015, parcel; a control point was obtained in the field using a rangefinder, laser GPS unit. Each control point was obtained by capturing the XY coordinate information of the front door to each structure in the sample dataset from the road (Figure 4.1). Figure 4.1 shows the type of attribute information collected by the GPS unit, which included a picture for each XY coordinate (Figure 4.2) visually displaying the front of each structure. Ten of the control points were removed from the analysis due to field error. Figure 4.4 is an example of dense tree coverage that lead to an inaccurate GPS point capture. Figure 4.5 and Figure 4.6 provide other tools such as Google Earth and picture images to help resolve errors. Commonalities of long distance captures of XY coordinate information and dense tree coverage were associated with all erroneous control points. Figure 4.7 and Figure 4.8 is an example of one control point that was captured at 609.15 feet. Each dataset was spatially mapped and displayed together on one map. Figure 4.10 shows the point datasets overlaying the 2015 12-inch aerial imagery. A visual overview of the spatial distribution of points within the urban cities of Ripley and Falkner show a clustered pattern around each structure (Figure 4.11). In the rural areas, clustering around structures present on the aerial imagery is found throughout the sample datasets when looking by group except for the parcel datasets (Figure 4.12). The size of parcel could have played a part in whether the parcel was clustered or an outlier from the rest of the datasets. These results were similar to what Zandbergen (2008) found in looking at parcel data when comparing address point, parcel, and street geocoding techniques. An average mean center of all datasets was created. Figure 4.13 shows the average point features and the control points overlain over the 2015 aerial imagery. The significance of the spatial position of the average point feature is further discussed in the next chapter. A margin of error (MoE) buffer was created around the control points to see how many features from the other datasets fell within the buffer (Table 4.6). The aerial 2006 and the aerial 2015 datasets had the highest number of features fall within the MoE buffer, and the parcel and electric power association datasets had the least. The aerial 2006 dataset had 122 features fall within the MoE buffer, and 67 of them were in urban areas (Table 4.7). The aerial 2015 dataset had the highest number of features, 139, that fell within the MoE buffer.

A standard distance was calculated from each XY coordinate in each dataset to the average mean center and grouped by urban or rural classification (Table 4.8). The average distance between point features and the average mean center was 27.87 feet in the urban areas and 82.98 feet in the rural areas. A discussion of the probable differences in the distance by classifications is explored in the Discussion Chapter. Table 4.9 describes the standard distance-by-distance category. The closest category of distance was a range from 13.18 feet - 37.54 feet, and 64% of the total average points fell within this category. There were only four average points that fell into the range with the most extensive distance range of 354.67 feet – 504.72 feet.

4.3 What is the spatial and temporal variation of the aerial imagery?

The research explored shifts in point features collected using the point address data model with aerial imagery. The research compiled two datasets using the front door approach method and the 2006 and 2015 aerial imageries that were flown by two separate companies nine years apart from each other. A distance calculation was performed on the two datasets, but the results were inconclusive. There was not enough change between the two datasets to produce a relevant distance calculation. However, the number of point features from the 2006 aerial imagery dataset, and the 2015 aerial imagery dataset had the highest count of features that fell within the control point margin of error. This result is useful given that the control point was taken at the front door. In the 2015 aerial dataset, 79 point features fell within the control point MoE in the urban classification and 60 point features for the rural classification. The results for point capture within the control point MoE for the 2006 aerial dataset were lower than the 2015 dataset but still significant with 65 point features within the MoE buffer in the urban classification and 57 in the urban classification.

The results of the research complement the findings of both Cayo and Talbot (2003) and Zandbergen (2008). Further exploration of the results and explanations are detailed in the next chapter.



Distance	TrueBeari	Date	Time	ImageFile	TargetLon	TargetLat	TargetAlt
69.29	98.1	2017:12:15	18:13:57	Images\Target Position_G100760_20171215_181357.jpg	-88.920543307	34.86271174	445.65

Figure 4.1 Example of Point Capture from GPS Unit

Figure 4.1 shows the attribute information collection in the field. Latitude and longitude coordinates were collected along with an image file, date, time, and distance.

Figure 4.2 Example of Picture Capture from GPS Unit

Figure 4.2 is the image file that corresponds to the other attribute information collected in Figure 4.1.

Figure 4.3 Example - KMZ Layer View in Google Earth

Figure 4.3 is the same example presented in Figure 4.1 and 4.2. It shows another capability of the MapSight software in that there are multiple export options to view the GPS collected information. A KMZ layer file can be used in other applications such as Google Earth.

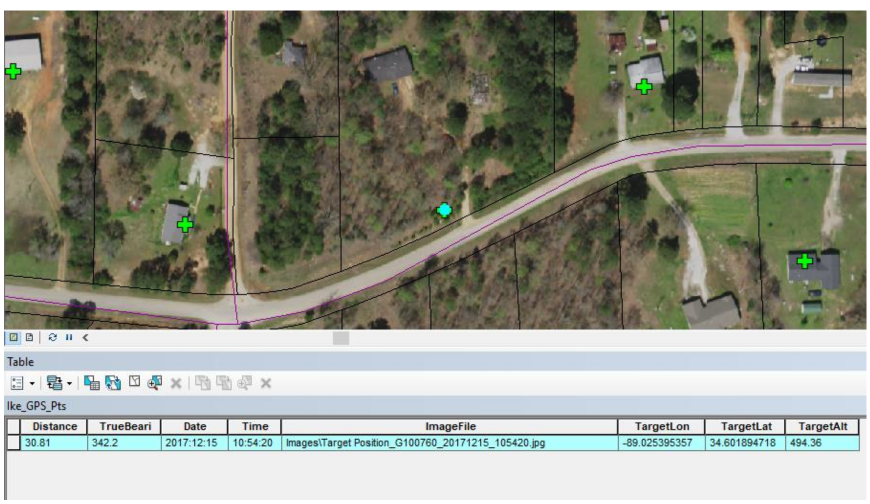


Figure 4.4 Example #1 of GPS Error

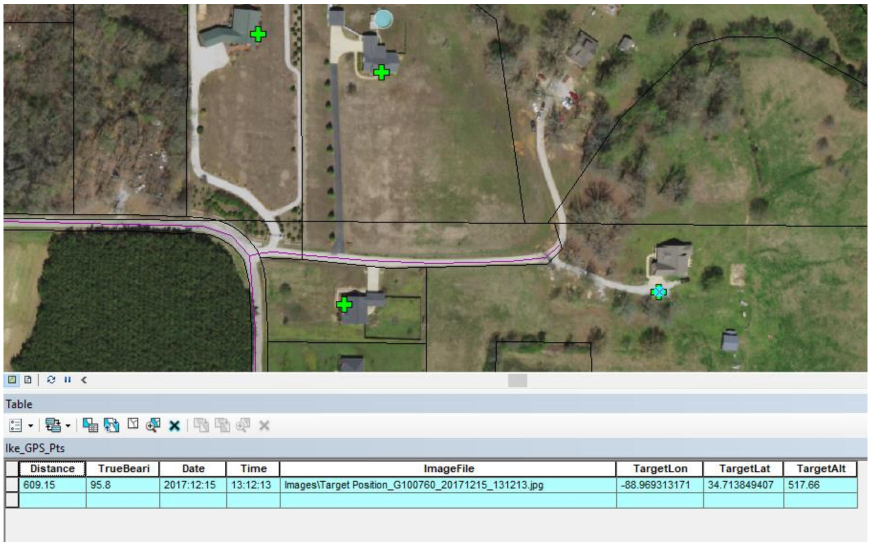
Figure 4.4 displays the GPS point spatially. The selected point (light blue) does not fall on or close to the structure. A comparison of the distance attribute and aerial imagery can be used to identify GPS error quickly. Knowledge of acceptable distances during field collection can quickly eliminate these types of errors.

Figure 4.5 Example #1 – Alternative Data View

Figure 4.5 displays an alternative view (Google Earth) of the GPS error in Example #1 from Figure 4.4. When trying to identify the reason for the error, aerial imagery can be used as a tool to help identify problems such as tree foliage blocking the front door.

Figure 4.6 Example #1- GPS Picture Image

Figure 4.6 is a visual representation of the field collected point from Example #1 in Figure 4.4. Pictures captured in the field can aid in the determination of errors. This picture suggests that the GPS error in Example #1 was a result of the densely wooded area in front of the structure.



Distance	TrueBeari	Date	Time	ImageFile	TargetLon	TargetLat	TargetAlt
609.15	95.8	2017:12:15	13:12:13	Images\Target Position_G100760_20171215_131213.jpg	-88.969313171	34.713849407	517.66

Figure 4.7 Example #2 of GPS Error

Figure 4.7 displays the GPS point spatially. The selected point (light blue) is displayed in the driveway next to the structure. The distance attribute (609.15 feet) suggests that the long-distance capture of the point could have played a part in the inaccuracy of the point.

Figure 4.8 Example #2 – Alternative Data View

Figure 4.8 displays an alternative view (KMZ layer in Google Earth) of the GPS error in Example #2 from Figure 4.7. The long-distance represented by the blue line and the tree cover from the aerial imagery suggests that both of these aided in the inaccuracy of the GPS point.

Figure 4.9 Example #2- GPS Picture Image

Figure 4.9 is a visual representation of the field collected point from Example #2 in Figure 4.7. Pictures captured in the field can aid in the determination of errors. This picture suggests that the GPS error in Example #2 could have been a result of the XY capture of a tree limb or tree trunk. This structure had tall windows on either side of the front door that could have been mistaken as the front door when standing 609.15 feet away.

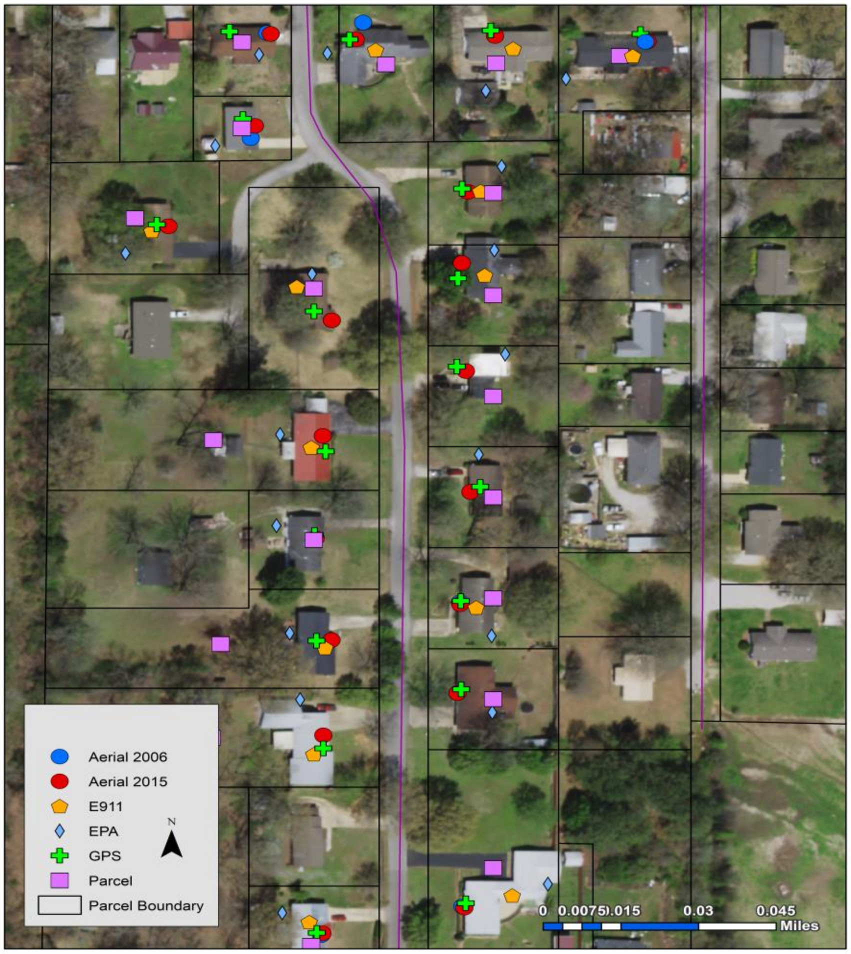


Figure 4.10 Point Locations for Research Datasets

Each dataset (2006 aerial imagery, 2015 aerial imagery, E-911, Electric Power, GPS, and parcel) was spatially displayed in the ArcMap software. Figure 4.10 demonstrates how the sample points from each dataset were clustered in pattern and fell close to each structure suggesting that any of the individual datasets could be spatially sufficient to use when creating an address database.

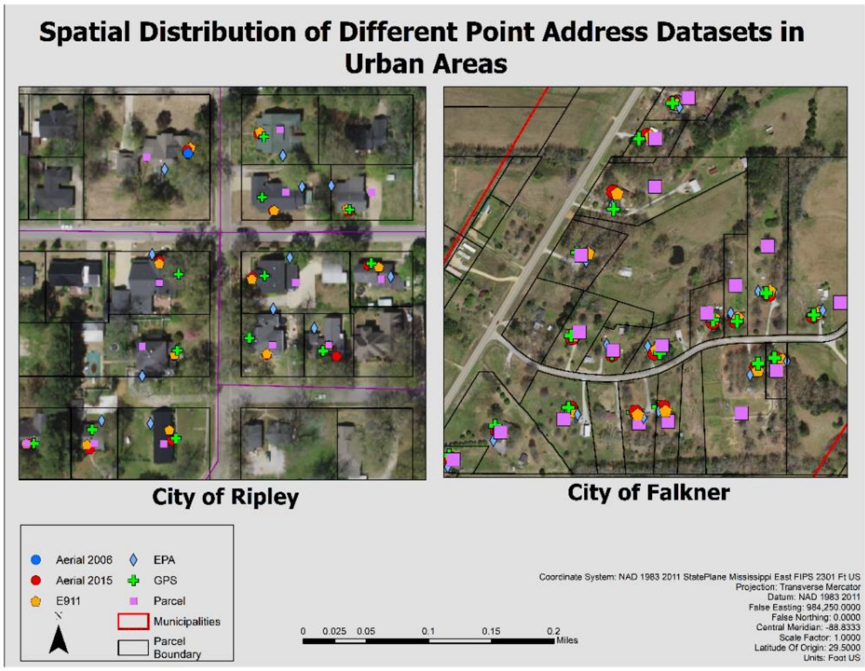


Figure 4.11 Spatial Distribution of Different Point Address Datasets in Urban Areas

Figure 4.11 shows a clustered pattern for each set of sample points within the cities of Ripley and Falkner.

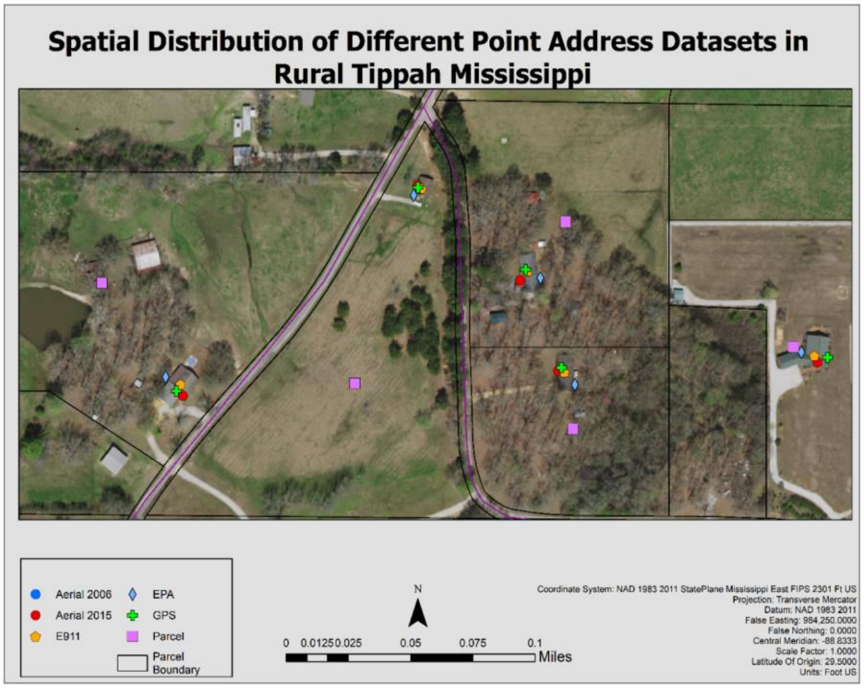


Figure 4.12 Spatial Distribution of Different Point Address Datasets in Rural Areas

Variance in the spatial accuracy of the parcel dataset (pink) in rural areas is displayed in Figure 4.12. The other datasets fall close to the structure.

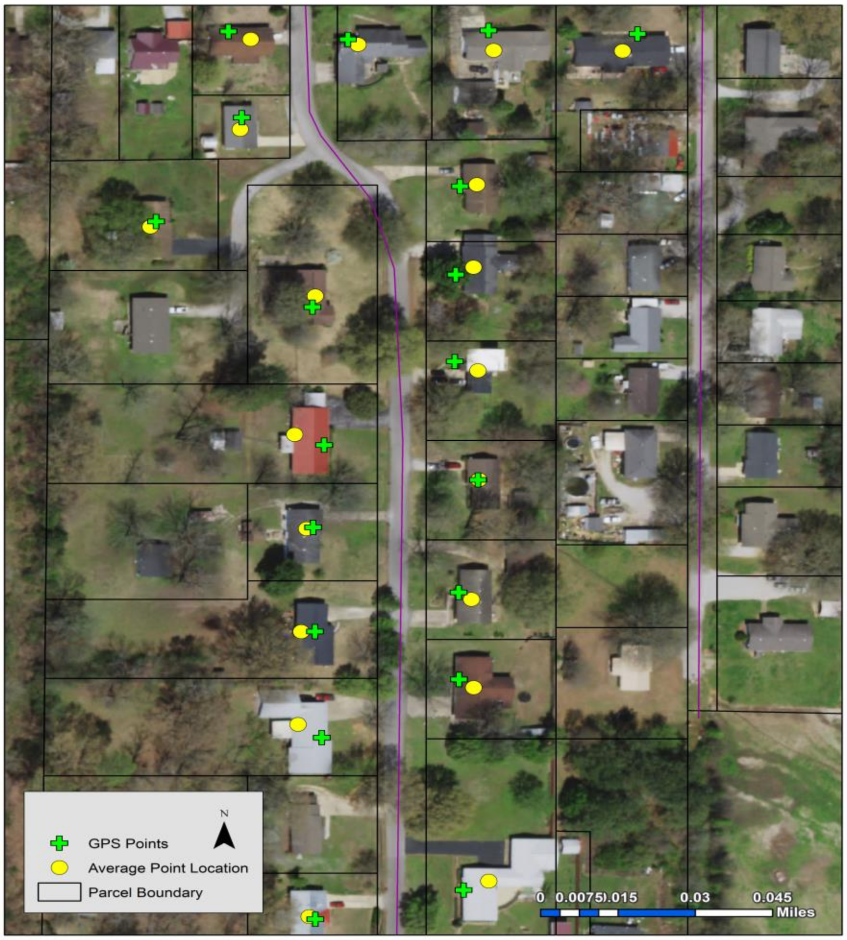


Figure 4.13 Average Point Location versus GPS Point Location

Figure 4.13 highlighted the average point location from each dataset in Figure 4.10 and compared it against the GPS field collected point. All the average points (yellow) fell on the structure of each house.

Table 4.1 Street Data Model Geocoding Result

Status	Urban	Rural	Total
Matched	14	4	**18**
Unmatched	135	146	**281**
Tied	1	0	**1**
Total	**150**	**150**	**300**

The result of the street data model using the street address locator did not yield a high number of geocoded matches when it used the address spreadsheet provided by E-911. There were more matches in the urban areas over the rural areas for the sample points.

Table 4.2 Parcel Data Model Geocoding Result with Street Ranges

Status	Urban	Rural	Total
Matched	13	5	**18**
Unmatched	136	145	**281**
Tied	1	0	**1**
Total	**150**	**150**	**300**

The parcel data model using street ranges produced similar results to those of the street data model. Most of the sample points were unmatched, but of the matched, there were more in the urban areas than rural.

Table 4.3 Parcel Data Model Geocoding Result with E-911 Data

Status	Urban	Rural	Total
Matched	145	135	**280**
Unmatched	0	14	**13**
Tied	5	1	**7**
Total	**150**	**150**	**300**

The parcel data model result for the E-911 data produced more matches in both the urban and rural areas.

Table 4.4 Point Address Model Geocoding Result with Street Ranges

Status	Urban	Rural	Total
Matched	12	5	**17**
Unmatched	137	145	**282**
Tied	1	0	**1**
Total	**150**	**150**	**300**

The point address data model using street ranges produced similar results to the parcel data model. Most addresses were unmatched, and there were more matches in urban areas than rural.

Table 4.5 Point Address Model with Geocoding Result with Electric Power Association Data

Status	Urban	Rural	Total
Matched	136	142	**278**
Unmatched	8	7	**15**
Tied	6	1	**7**
Total	**150**	**150**	**300**

The point address data model using the Electric Power data matched (93%) of the total number of sample points. There were slightly more matches (136) in rural areas than in urban areas (142).

Table 4.6 Total Number of Point Features within the Control Point Margin of Error (8.0 feet)

Dataset	Features within the Control MoE (8.0) feet	Percentage
Aerial 2006	122	36.09%
Aerial 2015	139	41.12%
EPA	0	0.00%
E911	33	9.76%
Parcel	4	1.18%
Average	40	11.83%
Total	**338**	**100.00%**

An 8-foot spatial buffer (MoE) was created around the control point. The majority of the aerial datasets (2006 and 2015) fell within the 8-foot buffer.

Table 4.7 Number of Point Features within the Margin of Error Buffer (8.0 feet)

Dataset	Urban	Rural	Total
Aerial 2006	65	57	**122**
Aerial 2015	79	60	**139**
EPA	0	0	**0**
E911	16	17	**33**
Parcel	4	0	**4**
Average	29	11	**40**
Total	**193**	**145**	**338**

There was a higher feature count in the urban areas than rural areas for features that fell within the 8-foot spatial buffer (MoE).

Table 4.8 Standard Distance of each Source Dataset to Average Mean Center

Area Classification	Distance (Feet)
Urban	27.87
Rural	82.98

The average distance in the rural areas (82.98 feet) was a greater distance than the average distance of the urban areas (27.87 feet).

Table 4.9 Average Mean Point by Standard Distance (Feet)

Standard Distance (Feet)	Number of Average Mean Points
13.18 - 37.54	192
37.55 - 84.69	66
84.70 - 171.69	25
171.70 - 354.66	13
354.67 - 504.72	4
Total	**300**

The standard distance was divided into 5 categories, and the number of features that fell within each distance was obtained. Most of the points (192) fell within a distance of 13.18 – 37.54 feet.

CHAPTER V – DISCUSSION

5.1 Research Question: Which addressing method is most suitable in Mississippi?

The research investigated three-address data models for suitability in Mississippi, and the research found the address data model (93% matched) and the parcel data model (93% matched) had the highest match rates, and the street data model had the lowest geocoding performance (06% matched). The most suitable method in Mississippi that will generate the highest match rates is to employ a composite locator encompassing multiple data models and use a mixed method approach to build an address database. The research utilized a sampling from datasets that originated from either a point, line, or polygon vector type and had a one-one feature match. For example, one address was associated with one parcel and one point in each of the other point datasets. No apartment complexes, multiple address structures per one parcel, or housing duplexes were included in the study. Address locators were built using the roads layer, parcel polygon layer, and E-911 layer. A reference dataset was created for each locator by utilizing the vital address component information from each of those datasets. The study measured the geocoding quality by assessing the match rate or the percentage of input addresses that produced a positive match against the reference dataset within each address locator. Results from table 4.3 using the parcel tax roll and table 4.5 using the electric power association information strongly suggests that geocoding match rates are influenced by the quality of both the input address data and reference data. Other studies also found that the quality of match rates was influenced by the quality of the reference data (Zandbergen 2008; Zandbergen 2009). The same point address locator was used to geocode two different input address datasets, and the match rate result percentage was the same at 93%. Though there is no consensus on a universal standard for acceptable match rates (Zandbergen 2009), a 93% match rate for individual geocoded locations is an acceptable percentage in the initial stages of building a point address database. Field collection and verification of addresses or attribute table correction of input address information could improve match rates. Again, when comparing Table 4.3 and Table 4.5, there were slightly more matches found in rural areas (94.6% matched) than in urban areas (90.6% matched) within the electric power association data. This result conflicts with research on automatic geocoding match rates in urban versus rural areas (Cayo and Talbot 2003; Dearwent et al. 2001), but the higher quality of the input address information could have played a part in the number of matched features in the electric power data.

The parcel data model was only as accurate as the reference data used in the address locators. The parcel data model was spatially accurate in the urban areas partially due to smaller parcel areas and the acknowledgment that structures are typically built in the center of parcels in urban areas. The larger parcels might have returned a geocoded match in the rural areas, but the point was not spatially accurate when mapped. Parcel databases were not initially created to geocode because typical database structure includes legal properties not necessarily the street address associated with the parcel. However, when a database does include complete street addresses as with this research, match rates typically fall within the 40% - 75% match range (Dearwent et al. 2001; Zandbergen 2008). In this research, the magnitude of the quality of the reference data significantly affected the match rates in the research to not fall within the percentage thresholds of previous studies. Also, the results from the parcel data model affirm that if the data is good quality and has complete street address information, it could be cost-effective on larger samples to initially use the parcel centroids for address points in the initial address database creation. Using GPS and aerial photography works better with smaller sample datasets because it is a manual, very time-consuming process to create the points. The parcel data model is an automated technique that will generate fast results. The downside to the parcel data model is that it does not handle multiple associated addresses within one parcel, larger parcels will have a higher spatial inaccuracy and most likely will not be located near or on the structure. The parcel data model would provide an excellent foundation database to build upon or a reliable supplemental database to use in conjunction with one of the other data models. The research found that the sole utilization of the parcel data model and parcel centroid does not provide a complete and spatially accurate database.

The street network address data model had the least number of matches and the poorest overall performance when used as a reference dataset to geocode the E-911, parcel, and electric power datasets. The street locator used the street address ranges from the 2010 Navteq dataset. The dataset that the research anticipated having high match rates for was the E-911 dataset (Table 4.1). Though the results in Table 4.1, Table 4.2, and Table 4.4 supported the previous studies on street geocoding in urban versus rural areas (Bonner et al. 2003, Cayo and Talbot 2003; Zandbergen 2008), 94% of the addresses in each dataset did not provide a matched result. Because the attribute information in the E-911 dataset was used as a reference dataset in the point locator and produced high matched results, inferences can be drawn that the quality of the input address dataset was not the cause for reduced match rates; rather poor match rates were a result of low-quality reference data in the street ranges. These findings contradicted previous studies that focused readily on street geocoding match rates with reported match rates of 30%, 77%, 78%, and 79% using four different commercial vendor reference datasets (Zhan et al. 2006). Considering the street network geocoding is the most widely employed address data model among private and commercial firms, the results could have been susceptible to complications that impeded the success of generating high match rates. The research built two street locators to answer the research question. The first street locator did not return any matches. An ESRI support document[5] presented a description of the cause and solution for the software bug dealing with dual range address locators. The new street locator did return match results but still did not return the percentage of matches that previous studies found. Several properties were updated within the street locator such as spelling sensitivity that allows for spelling variation and side offset that determines the distance from either side of a line feature where a matched location should be placed, but the matched results did not change. The original creation date of the roads layer, 2010, could have negatively impacted the geocoding results, or the quality of the street range attributes could have been incorrect because the point address locator produced higher match rates using the same input datasets.

5.2 Research Question: How does the positional accuracy change over space between rural and urban designated areas?

The research established that positional accuracy increased in urban areas and decreased in rural areas, especially with structures that fell on large properties. Positional

[5] https://support.esri.com/en/technical-article/000011688 (ESRI 2017)

Accuracy was determined by assessing the distance between the GPS, field-captured control points and the E-911, 2006 aerial imagery, 2015 aerial imagery, electric power association, and parcel datasets. Three hundred sample points were chosen through non-probability convenience sampling. The research chose a smaller sample size based on the previous studies on the positional error of geocoding (Bonner et al. 2003; Ward et al. 2005; Schootman et al. 2007). This method was chosen over random point sampling because of rugged, dirt roads in the rural areas in Tippah County. The research chose sample points that fell along Mississippi Highway 15 and county roads that intersected with the highway. A quantitative point count was done to confirm that there were corresponding points in each of the other five datasets before field collection occurred. The research chose the Ike 3.0 GPS unit because of its built-in laser rangefinder that would allow for front door capture and be the least intrusive on people's properties. Figure 4.1 shows what the GPS data looked like in the ArcGIS mapping software. The attribute table contained the distance from the road to the front door. The distance attribute was important during field collection because one could quickly assess if the captured point information were accurate. Using the distance field as an aid for correctness worked out the majority of the time, but without being able to look at the collected points in real-time, there were still errors in the GPS data during the processing phase. Ten points were removed from the analysis to avoid skewing the statistical analysis results because they were clearly errant points. The MapSight software allowed for additional layers such as KMZ layers to be generated when the points were extracted from the GPS unit to the desktop. The blue line in Figure 4.3 shows the straight line distance from where the point at the front door was collected to the road.

Many times, aerial imagery reveals insufficient data about a structure or property, so the built-in camera on the GPS unit helped supplement the aerial imagery (Figure 4.2). Pictures added visual representations of each structure to spatial XY coordinates and helped resolve most errors encountered in the field. The errors were likely linked to obstructive views of the front door due to trees, objects, or topography such as hills (Figure 4.6). These results suggest negative impacts on the overall positional accuracy assessment but also makes clear the subjectivity and sensitivity of GPS collection in the field.

This method of point collection was not cost-effective. Driving every road and shooting every front door with the highest precision and positional accuracy available proved to be time-consuming and full of errors that could have been eliminated through the use of just high-resolution aerial imagery (Cayo and Talbot 2003; Schootman et al. 2007; Zimmeran 2007). The research results agree with Zandbergen (2008) on a combined approach such as parcel, point, and GPS may increase geocoding match results and database completeness. Point collection using parcel or aerial imagery did not increase database completeness and repeatability for multiple address structures. Field verification is still needed to confirm addresses, housing unit counts, and address numbers.

Figure 4.12 shows the different point placements for each dataset along a county road in Tippah County. The E-911 dataset used the center of structure placement method. The electric power association dataset used a sub-meter GPS unit to capture the XY coordinate pair for the electric meter. Figure 4.10 shows that most of the point placement ended up near the back of every structure with the majority of the points close to the house. The parcel dataset was created using the parcel centroid and attribute information. Figure 4.12 gives an example of the variance of small and large parcels. The parcel centroids did not fall on the structures when the parcels were larger and contributed to the positional error statistics. Both of the aerial imageries' point data were created by placing a point near the presumed front door and showed that there was little spatial shift between the imagery from 2006 and 2015.

The research looked at positional accuracy change between rural and urban areas, and Figure 4.11 shows a clustered pattern among point placements in each dataset for the City of Ripley and City of Falkner. These initial findings were not a surprise when considering the geocoding results in research question one. Figure 4.12 shows the spatial distribution for a rural area in Tippah County. All of the datasets except the parcel centroid data were spatially located at or near the structure. There was concern that the parcel centroid might negatively influence the average point statistics. An average mean center was created for each group of points per structure. Figure 4.13 shows that the average point for the same area as Figure 4.10. Each average point location still fell on each structure including the areas where there were larger parcels present. The results show that an average mean center of available datasets could be a quick way to build a base for an address database quickly. Previous studies utilized median center instead of mean center (Whitsel et al. 2006; Schootman 2007; Strickland et al. 2007; Zandbergen 2008) because median center identifies the location and is not influenced by outliers such as parcel centroids on large parcels. In this research, the outliers did not have a significant impact on average mean, but they could impact studies with larger sample sizes.

A standard distance was calculated between point features and the control point. The average distance from the control point in the urban area was 27.87 feet. These results agree with what was recognized in the clustering patterns found among the datasets. The average distance in the rural areas was larger than in the urban areas at 82.98 feet. The parcel data could have affected the distance results, but the results align with previous research that positional error increases in rural areas (Cayo and Talbot 2003; Zandbergen 2008). Table 4.9 shows the standard distance categorized into five groups using natural breaks (Jenks) classification where 65% of the average mean points fell within 13.18 ft – 37.54 feet from the control point. The results show that regardless of what method is used (center of structure, front door), an average point location can be an acceptable dataset to use if there are errors in the source data. This approach could increase database completeness and spatial accuracy in a cost-efficient manner.

5.3 Research Question: What is the spatial and temporal variation of the aerial imagery?

The research results determined that the spatial shifts in imagery between the 2006 and 2015 aerial imageries had a minimum change in the distance making the shift negligible and useful for spatial studies on positional accuracies over time using high-resolution imagery that allowed for a margin of error to still include the structure. The positional accuracy was determined by using a front door GPS control point compared to points generated from two aerial imagery datasets (2006 and 2015) that inferred the location of front doors. The front door on each structure was typically inferred visually from the structure proximity to the road and identification of sidewalks, intersection paths, porches, and overhangs on the imagery that appeared to lead to the front door of each structure. Table 4.7 gave a descriptive count of the number of point features from each dataset that fell within the margin of error (MoE) buffer for the GPS control point. There were higher point counts in urban areas in both the 2006 and 2015 datasets. The two datasets combined accounted for 75% of the total points that fell within the 8-foot MoE in urban areas, which suggests that point creation using high aerial imagery is as accurate and precise as field collection with a GPS unit inside urban areas. The same conjecture can be made about the descriptive counts for the rural areas. Though there were not as many points from the other datasets to fall within the GPS MoE and produce high total counts, 80% of the total points found within the MoE in rural areas were from the aerial imagery datasets. The research affirms that using aerial imagery to collect the front door in point collection is relevant, precise, and potentially cost-efficient if the aerial imagery is readily available to use. Another question that these results raise is the need to collect points at the front door. Arguably, if address point collection is as precise and accurate as GPS point collection is from the field, can it be assumed that point creation using the center of structure could produce the same precision and accuracy? Using the front door method allowed the research to use a control point in the study to determine positional accuracy, but it weakened the repeatability of the study. The interpretation of the front door is potentially subject to change based on a researcher's inference, thus likely producing different outputs. Previous studies suggest that center of structure can also be used to assess positional error in geocoding (Cayo and Talbot 2003; Whitsel et al. 2006; Strickland et al. 2007; Zimmerman et al. 2007). The recommended best practice is to use that center of structure as the point location to hold the address attributes.

CHAPTER VI – CONCLUSION

The research methodology framework followed a combination of previous works by Cayo and Talbot (2003) and Zandbergen (2008) to try to determine the most efficient ways to create an address database by looking at various address data models and datasets in Tippah County. Table 4.5 showed the results from the point address data model to be the most suitable for Tippah County. There were a few deductions found in the comparison of each address data model that is supported by previous studies:

> (1) Match rates are influenced by both the quality of the reference datasets and the input address datasets (Zandbergen 2008)
> (2) Older, dated reference datasets are subject to be of lower quality and error-prone such as the 2010 dual range attributes in the Navteq data (Whitsel et al. 2006)
> (3) A combination of address data models is cost efficient to improve address completeness and match rates.

The point address geolocator performed the best because the E-911 reference dataset (2017) was the most accurate and complete database, and the electric power association address information did not have many errors in the address attributes which allowed for match rates to be 93%. Sole dependence on one address data model does not create a complete address database and highest match rates in the geocoding process; instead utilizing more than one address data model like parcel and point address data models will increase match rates (Table 4.3). The research also found the most cost-efficient mixed address data approach varies across time and space. The following is an example of how someone in Mississippi could use a mixed address data model approach to cut costs in building an address database by using high-resolution imagery. First, the parcel data model can be used to generate points that will give almost complete coverage inside urban areas when the assumption is that most parcels have structures inside city boundaries. The points can be moved to the center of the structure using the high-resolution imagery. Most of the work is done in office saving time and money. For rural areas, the framework might change slightly using parcel data, but the overall methodology is like that for urban collection. This framework does not consider multiple address structures such as duplexes or apartments or adding points to structures that have an associated address; these locations would require direct field validation. However, through the point address model using both aerial imagery and GPS, points can be created and validated as needed. There might be some instances when the quality of the source data is questionable, then GPS field verification can be utilized to fix address attribute information or verify existing information already in the address database. It is highly recommended that field checks be done to ensure address attribute integrity is valid and up to date, and there are several ways of doing so through field notes or with mapping software that will allow active edits to the attribute information in the field. GPS collection is better suited for small sample datasets such as mapping and addressing new structures. Using a handheld GPS unit like the one used in this research is not recommended to use on large sample datasets or use in the initial stages of building an address database because of the amount of time it would take to capture the center of structure accurately. A direct influence on positional error and precision is collection time. There is a greater chance for error when the collection process is rushed or hurried, and the collector is not allowing the proper time window for the GPS unit to find satellites to lower PDOP.

The last research question investigated the spatial shift of aerial imagery over time. A legitimate argument when creating an address database is if the aerial imagery is used in the initial creation, how is time saved and positional accuracy preserved if new aerial imagery becomes available? How well did the address points that were created using the 2006 imagery fair against the address points that were taken at where the front door was on the 2015 imagery? The research sought to calculate a standard distance between the two datasets to quantify any shift, but there were not enough points to spatially perform the tool, but a visual inspection showed the points were consistently close. The research gathered findings from table 4.7 and used descriptive counts from the number of point features that fell with the GPS control point MoE buffer. The research results for urban, 75%, and rural, 85%, of the point features fell within the 8-foot control point MoE and suggests that point creation using high-resolution, orthorectified aerial imagery is as accurate and precise as field collection with a GPS unit. Using the front door method created a control point in the research study and proved to be successful, but the center of structure can also be used as a control point in assessing positional error (Cayo and Talbot 2003; Whitsel et al. 2006; Strickland et al. 2007; Zimmerman et al. 2007).

6.1 Why an Address Database is Important in Mississippi

The importance of building an accurate address database is important on an individual level because the fundamental principles of knowing where something is located is so readily demanded through the technology of smartphones, computers, and mapping systems that few people realize how it influences their daily decision making.

Accurate addressing is also important on a local and state level. Without a standard address database, negative impacts such as loss of life and property, inefficient routing, and cost to individuals incur when address databases lack consistency and accuracy. The National Geospatial Advisory Committee (NGAC) recognized that the current addressing database on a national level is a fractured system (NSGIC 2010). The NGAC assessed the need for a National Address Database (NAD). A national schema standard that would create a systematic process for creating, maintaining, and updating address information. The NGAC and the Federal Geographic Data Committee (FGDC) included an in-depth look at case studies from various levels of federal, state, and local governments that would greatly benefit from a standard for address databases across multiple systems (NGAC 2014). One federal expenditure, the United States Census Bureau, cannot currently share any of the 2010 Census point address information it collected back to the state or local governments, leaving other agencies the costs of generating their own. With a NAD, the estimated cost of savings is $196 million (Table 1.2). On a state and local level, a statewide address database using the NAD framework would create a universal schema that would permit sharing of data and resources and help eliminate duplicated efforts.

Address information is used locally in school districts, taxation offices, planning and zoning, health and human services, retail/service, utilities, and mapping services. The most significant beneficiary of a statewide database and known "caretakers" of assigning and maintaining an addressing database is emergency response. Emergency responders use address information to locate where an emergency is taking place and dispatches the police, fire, ambulance, and rescue accordingly. E-911 coordinators do the best they can in communicating with neighboring counties, but each county maintains their data. Over time diverging of database conformity and loss of data occurs because there is no framework for sharing data and no statewide standard to organize all the different datasets. Nationally, emergency response has a national standard for GIS data, but in Mississippi, counties are assigning, addressing, and maintaining address systems based on the education of a predecessor or outdated methods. Some counties appropriate the necessary common fields in an address and leave the other fields off (Figure 1.2), but the field names all vary from county to county throughout the state. The NAD schema holds the most common address fields in high regard and would benefit the emergency response community or any stakeholder if it was chosen to be used as a statewide schema would not dissatisfy the emergency response community or any other stakeholder if it was chosen as a schema to be used statewide. The beginning of change starts at the local level with local Mississippi counties who want an address database and who want to be able to share local data with neighboring counties. A small change such as a statewide schema will impact other industries (Table 1.1) that rely on addresses to provide goods and services.

6.2 Limitations

The research provided new knowledge about GIS address collection in Mississippi and the types of problems that could arise when building an address database. There were limitations to this study. The study excluded multiple addresses, duplexes, mobile home parks, and apartments from the study. Only one-one relationships were considered, and it was not a holistic view of the types of structures that are typically found in towns and rural places in Mississippi. Although several other studies used a sample size less than what was used in this study, the larger sample size would have allowed for the standard distance to correctly work when determining the shift in distance between the aerial imagery datasets. Street geocoding is the most widely employed address data model, and the research was unable to produce viable match rates to be able to make a strong statement to prove or disprove other studies. The last limitation was the research looked at address data models individually instead of looking at a composite locator that combined datasets to increase match rates. For example, if street ranges are good quality, a composite locator can be built that the geocode process looks at multiple datasets at one time. A mixed combo or include all three such as E-911, street range, and parcel data in one composite locator.

6.3 Future Studies

Future research efforts in addressing efforts in Mississippi should focus on refinements in the address framework for urban and rural areas such as the use of different address data models to build a database. Also, studies should focus on best practices for increasing match rates and database completeness in multiple address structures such as apartments, duplexes, and mobile home parks. Perhaps smart technology like a smartphone or iPad with a map-grade GPS unit and aerial imagery could suffice. Any study would improve quality control on the field-collected points to improve the reference datasets further.

Any new database takes time in the type of database, structure, types of features, properties of features. The NAD schema (Appendix A.1) has an existing geodatabase template that makes it easy to either import existing data into or create new data within. There are fifty Mississippi counties remaining (Figure 2.4) that need a spatial address database. An estimation of the total time it takes to build an address database will vary by county. Each county has diverse populations, different road lengths and road counts, and original dataset availability and quality. The number of people working on the project could shorten or lengthen the project time depending on how much data must be created by hand. All these components affect completion time. Again, it begins with change and the acknowledgment on the local level that an address database is needed. Once the need is presented, documentation and studies are readily available on how to begin. The expectation from the research completed in Tippah County is that this study was informative on what types of data can be used to begin building an address database and cost-efficient ways to be successful.

APPENDIX A – National Address Database Schema

Table A.1 National Address Database Schema

Field Name	Field Alias	Type	Length	Domain	Expected Use
State	State	Text	2		Always Used
County	County	Text	40		Always Used
Inc_Muni	Incorporated Municipality	Text	100		Commonly Used
Uninc_Comm	Unincorporated Community	Text	100		Commonly Used
Nbrhd_Comm	Neighborhood Community	Text	100		Commonly Used
Post_Comm	Postal Community Name	Text	40		Commonly Used
Zip_Code	Zip Code	Text	7		Always Used
Plus_4	Zip Code 4 Addition	Text	7		Occasionally Used
Bulk_Zip	Bulk Delivery ZIP Code	Text	7		Rarely Used
Bulk_Plus4	Bulk Delivery ZIP Plus 4 Addition	Text	7		Rarely Used
StN_PreMod	Street Name Pre Modifier (PRM)	Text	15		Commonly Used
StN_PreDir	Street Name Pre Directional (PRD)	Text	50	X	Commonly Used
StN_PreTyp	Street Name Pre Type (STP)	Text	25	X	Commonly Used
StN_PreSep	Street Name Pre Type Separator (STPS)	Text	20	X	Commonly Used
StreetName	Street Name (RD)	Text	60		Always Used
StN_PosTyp	Street Name Post Type (STS)	Text	15	X	Commonly Used
StN_PosDir	Street Name Post Directional (POD)	Text	50	X	Commonly Used
StN_PosMod	Street Name Post Modifier (POM)	Text	25		Commonly Used
AddNum_Pre	Address Number Prefix (HNP)	Text	15		Commonly Used
Add_Number	Address Number (HNO)	Text	6		Always Used
AddNum_Suf	Address Number Suffix (HNS)	Text	15		Commonly Used
LandmkPart	Landmark Name Part (LMKP)	Text	150		Occasionally Used
LandmkName	Landmark Name Part (LMKP)	Text	150		Occasionally Used
Building	Building (BLD)	Text	75		Commonly Used
Floor	Floor (FLR)	Text	75		Commonly Used
Unit	Unit (UNIT)	Text	75		Commonly Used

Table A.1 (continued)

Field Name	Field Alias	Type	Length	Domain	Expected Use
Room	Room (ROOM)	Text	75		Rarely Used
Addtl_Loc	Additional Location Infro (LOC)	Text	225		Rarely Used
Milepost	Milepost	Text	50		Rarely Used
Longitude	Address Longitude	Float	12		Always Used
Latitude	Address Latitude	Float	11		Always Used
NatGrid_Coord	National Grid Coordinates	Text	50		Always Used
GUID	GUID	GUID			Always Used
Addr_Type	Address Type	Text	50	X	Commonly Used
Placement	Address Placement	Text	25	X	Commonly Used
Source	Address Source	Text	75		Always Used
AddAuth	Address Authority	Text	75		Commonly Used
UniqWithin	Unique Within	Text	75		Occasionally Used
Last/Update	Date Last Updated	Date	26		Always Used
Effective	Effective Date	Date	26		Occasionally Used
Expired	Expiration Date	Date	26		Occasionally Used

Table A.2 Address Placement Domain

CODE	DESCRIPTION
STRUCTURE - ROOFTOP	STRUCTURE - ROOFTOP
STRUCTURE - ENTRANCE	STRUCTURE - ENTRANCE
STRUCTURE - INTERIOR	STRUCTURE - INTERIOR
PARCEL - CENTROID	PARCEL - CENTROID
PARCEL - OTHER	PARCEL - OTHER
LINEAR GEOCODE	LINEAR GEOCODE
PROPERTY ACCESS	PROPERTY ACCESS
SITE	SITE PLACEMENT
OTHER	OTHER
UNKNOWN	UNKNOWN

Table A.3 Street Type Pre-Directional and Post-Directional Domain

CODE	DESCRIPTION
NORTH	NORTH
SOUTH	SOUTH
EAST	EAST
WEST	WEST
NORTHEAST	NORTHEAST
NORTHWEST	NORTHWEST
SOUTHEAST	SOUTHEAST
SOUTHWEST	SOUTHWEST

Table A.4 Address Type Domain

CODE	DESCRIPTION
RESIDENTIAL	RESIDENTIAL (HOUSING)
COMMERCIAL	COMMERCIAL (OFFICE, RETAIL, RESTAURANT, BANKING)
MULTI	MULTI-USE (MIXED COMMERCIAL/RESIDENTIAL)
OPEN	OPEN SPACE (FOREST, VACANT, CEMETERIES)
INDUSTRIAL	INDUSTRIAL
GOVERNMENT	GOVERNMENT/PUBLIC SERVICES (FIRE/POLICE, LIBRARY, GOVERNMENT OFFICES)
RELIGIOUS	RELIGIOUS
RECREATION	RECREATION (BALL FIELDS, PARKS, GOLF COURSES, SKI AREA)
EDUCATIONAL	EDUCATIONAL (SCHOOLS, UNIVERSITIES)
INSTITUTIONAL	INSTITUTIONAL (HOSPITALS, GROUP HOMES, PRISONS, ETC)
OTHER	OTHER
UNKNOWN	UNKNOWN

Table A.5 Street Pre-Type and Post-Type Domain

CODE	DESCRIPTION
ALY	ALLEY
ANX	ANEX
ARC	ARCADE
AVE	AVENUE
BCH	BEACH
BG	BURG
BGS	BURGS
BLF	BLUFF
BLFS	BLUFFS

Table A.5 (Continued)

CODE	DESCRIPTION
BLVD	BOULEVARD
BND	BEND
BR	BRANCH
BRG	BRIDGE
BRK	BROOK
BRKS	BROOKS
BTM	BOTTOM
BYP	BYPASS
BYU	BAYOU
CIR	CIRCLE
CIRS	CIRCLES
CLB	CLUB
CLF	CLIFF
CLFS	CLIFFS
CMN	COMMON
CMNS	COMMONS
COR	CORNER
CORS	CORNERS
CP	CAMP
CPE	CAPE
CRES	CRESCENT
CRK	CREEK
CRSE	COURSE
CRST	CREST
CSWY	CAUSEWAY
CT	COURT
CTR	CENTER
CTRS	CENTERS
CTS	COURTS
CURV	CURVE
CV	COVE
CVS	COVES
CYN	CANYON
DL	DALE
DM	DAM
DR	DRIVE
DRS	DRIVES
DV	DIVIDE
EST	ESTATE

Table A.5 (continued)

CODE	DESCRIPTION
ESTS	ESTATES
EXPY	EXPRESSWAY
EXT	EXTENSION
EXTS	EXTENSIONS
FALL	FALL
FLD	FIELD
FLDS	FIELDS
FLS	FALLS
FLT	FLAT
FLTS	FLATS
FRD	FORD
FRDS	FORDS
FRG	FORGE
FRGS	FORGES
FRY	FERRY
FT	FORT
FWY	FREEWAY
GDN	GARDEN
GDNS	GARDENS
GLN	GLEN
GLNS	GLENS
GRN	GREEN
GRNS	GREENS
GRV	GROVE
GRVS	GROVES
GTWY	GATEWAY
HBR	HARBOR
HBRS	HARBORS
HL	HILL
HLS	HILLS
HOLW	HOLLOW
HTS	HEIGHTS
HVN	HAVEN
HWY	HIGHWAY
INLT	INLET
IS	ISLAND
ISLE	ISLE
ISS	ISLANDS
JCT	JUNCTION

Table A.5 (continued)

CODE	DESCRIPTION
JCTS	JUNCTIONS
KNL	KNOLL
KNLS	KNOLLS
KY	KEY
KYS	KEYS
LAND	LAND
LCK	LOCK
LCKS	LOCKS
LDG	LODGE
LF	LOAF
LGT	LIGHT
LGTS	LIGHTS
LK	LAKE
LKS	LAKES
LN	LANE
LNDG	LANDING
LOOP	LOOP
MALL	MALL
MDW	MEADOW
MDWS	MEADOWS
MEWS	MEWS
ML	MILL
MLS	MILLS
MNR	MANOR
MNRS	MANORS
MSN	MISSION
MT	MOUNT
MTN	MOUNTAIN
MTNS	MOUNTAINS
MTWY	MOTORWAY
NCK	NECK
OPAS	OVERPASS
ORCH	ORCHARD
OVAL	OVAL
PARK	PARK
PASS	PASS
PATH	PATH
PIKE	PIKE
PKWY	PARKWAY

Table A.5 (continued)

CODE	DESCRIPTION
PL	PLACE
PLN	PLAIN
PLNS	PLAIN
PLZ	PLAZA
PNE	PINE
PNES	PINES
PR	PRAIRIE
PRT	PORT
PRTS	PORTS
PSGE	PASSAGE
PT	POINT
PTS	POINTS
RADL	RADIAL
RAMP	RAMP
RD	ROAD
RDG	RIDGE
RDGS	RIDGES
RDS	ROADS
RIV	RIVER
RNCH	RANCH
ROW	ROW
RPD	RAPID
RPDS	RAPIDS
RST	REST
RTE	ROUTE
RUE	RUE
RUN	RUN
SHL	SHOAL
SHLS	SHOALS
SHR	SHORE
SHRS	SHORES
SKWY	SKYWAY
SMT	SUMMIT
SPG	SPRING
SPGS	SPRINGS
SPUR	SPUR
SQ	SQUARE
SQS	SQUARES
ST	STREET

Table A.5 (continued)

CODE	DESCRIPTION
STA	STATION
STRA	STRAVENUE
STRM	STREAM
STS	STREETS
TER	TERRACE
TPKE	TURNPIKE
TRAK	TRACK
TRCE	TRACE
TRFY	TRAFFICWAY
TRL	TRAIL
TRLR	TRAILER
TRWY	THROUGHWAY
TUNL	TUNNEL
UN	UNION
UNS	UNIONS
UPAS	UNDERPASS
VIA	VIADUCT
VIS	VISTA
VL	VILLE
VLG	VILLAGE
VLGS	VILLAGES
VLY	VALLEY
VLYS	VALLEYS
VW	VIEW
VWS	VIEWS
WALK	WALK
WALL	WALL
WAY	WAY
WAYS	WAYS
WL	WELL
WLS	WELLS
XING	CROSSING
XRD	CROSSROAD
XRDS	CROSSROADS

REFERENCES

Bonner, Matthew R., Daikwon Han, Jing Nie, Peter Rogerson, John E. Vena, and Jo L. Freudenheim. "Positional Accuracy of Geocoded Addresses in Epidemiologic Research." Epidemiology 14, no. 4 (July 2003): 408-12. doi:10.1097/01.ede.0000073121.63254.c5.

Burra, Tara, Michael Jerrett, Richard T. Burnett, and Mark Anderson. "Conceptual and Practical Issues in the Detection of Local Disease Clusters: A Study of Mortality in Hamilton, Ontario." The Canadian Geographer/Le Géographe Canadien 46, no. 2 (June 28, 2002): 160-71. doi:10.1111/j.1541-0064.2002.tb00737.x.

Cayo, Michael R, and Thomas O Talbot. “Positional Error in Automated Geocoding of Residential Addresses.” *International Journal of Health Geographics*, vol. 2, no. 10, 19 Dec. 2003, doi:10.1186/1476-072X-2-10.

Dearwent, Steve M., Robert R. Jacobs, and John B. Halbert. "Locational Uncertainty in Georeferencing Public Health Datasets." Journal of Exposure Science & Environmental Epidemiology 11, no. 4 (2001): 329-34. doi:10.1038/sj.jea.7500173.

ESRI. "ArcGIS Help." About Geocoding a Table of Addresses. Accessed November 17, 2017. http://desktop.arcgis.com/en/arcmap/10.3/guide-books/geocoding/geocoding-a-table-of-addresses-about.htm.

General Electric Company. GE MapSight User Manual. 2014.

Google Maps. "Google Maps." Map. In Google Maps. 2017.

"GNSS Frequently Asked Questions - GPS." *Federal Aviation Administration Navigation Programs*, Federal Aviation Administration, 20 Dec. 2016, www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservicen/gnss/faq/gps/#1.

Kaplan, Elliot D, and Christopher J Hegarty. 2006. *Understanding GPS Principles and Applications.* 2nd. Norwood, Massachusetts: Artech House, Inc.

Logue, Barbara J. Urban and Rural Populations in Mississippi: A Comparison. Report. 2011. 1-6.

Longley, Paul A., Michael F. Goodchild, David J. Maguire, and David W. Rhind. Geographical Information Systems and Science. 2nd ed. John Wiley & Sons, 2005.

Milner, Greg. 2016. *Pinpoint: How GPS Is Changing Technology, Culture, and Our Minds.* 1st. New York, New York: W. W. Norton & Company.

Mississippi Coordinating Council for Remote Sensing and Geographic Information Systems. Executive Summary Map Mississippi. Working paper. 2015. 1-13.

Mississippi Coordinating Council for Remote Sensing and Geographic Information Systems. Executive Summary Map Mississippi. Working paper. 2016. 1-13.

Mitchell, Andy. The ESRI Guide to GIS Vol 2 Spatial Measurements & Statistics. Redlands, CA: ESRI Press, 2005.

National Coordination Office for Space-Based Positioning, Navigation, and Timing. 2017. Control Segment. November 3. Accessed January 31, 2018. https://www.gps.gov/systems/gps/control/.

—. 2014. *GPS Applications.* November 24. Accessed January 31, 2018. https://www.gps.gov/applications/.

—. 2017. *The Global Positioning System.* June 6. Accessed January 31, 2018. https://www.gps.gov/systems/gps/.

National Geospatial Advisory Committee. The Need for a National Address Database - Use Cases. Report. 2014. 1-21.

National Geospatial Advisory Committee. The Need for a National Address Database - Use Cases. Report. 2014. 1-24.

National States Geographic Information Council. Addressing Coordination Issues.

Report. 2010. 1-4.

Pace, Scott, Gerald Frost, Irving Lachow, David Frelinger, Donna Fossum, Donald Wassem, and Monica Pinto. 1995. *The Global Positioning System: Assessing National Policies.* Santa Monica, CA: Critical Technologies Institute.

Penn State College of Earth and Mineral Sciences. 2017. Chapter 11: Trilateration. Accessed February 1, 2018. https://www.e-education.psu.edu/natureofgeoinfo/c5_p12.html.

Ratcliffe, Jerry H. "Geocoding Crime and a First Estimate of a Minimum Acceptable Hit Rate." International Journal of Geographical Information Science 18, no. 1 (2004): 61-72. doi:10.1080/13658810310001596076.

Schootman, Mario, David A. Sterling, James Struthers, Yan Yan, Ted Laboube, Brett Emo, and Gary Higgs. "Positional Accuracy and Geographic Bias of Four Methods of Geocoding In Epidemiologic Research." Annals of Epidemiology 17, no. 6 (2007): 464-70. doi:10.1016/j.annepidem.2006.10.015.

Statistica. Google Services and Products Used at Least Occasionally. PDF. 2018.

Strickland, Matthew J., Csaba Siffel, Bennett R. Gardner, Alissa K. Berzen, and Adolfo Correa. "Quantifying Geocode Location Error Using GIS Methods." Environmental Health 6, no. 1 (2007). doi:10.1186/1476-069x-6-10.

The National States Geographic Information Council. Addressing Coordination Issues. Report. 2010. 1-4.

UNC-Chapel Hill. Types of GPS Receivers. PPT. 2007.

United States Census Bureau. 2010 Census Redistricting Data Summary File. Report. 2010.

Ward, Mary H., John R. Nuckols, James Giglierano, Matthew R. Bonner, Calvin Wolter, Matthew Airola, Wende Mix, Joanne S. Colt, and Patricia Hartge. "Positional Accuracy of Two Methods of Geocoding." Epidemiology 16, no. 4 (2005): 542-47. doi:10.1097/01.ede.0000165364.54925.f3.

Whitsel, Eric A., Miguel Quibrera, Richard L. Smith, Diane J. Catellier, Duanping Liao, Amanda C. Henley, and Gerado Heiss. "Accuracy and Repeatability of Commercial Geocoding." Epidemiologic Perspectives & Innovations 3, no. 8 (July 20, 2006): 1-12. doi:10.1186/1742-5573-3-8.

Zandbergen, Paul A. “A comparison of address point, parcel and street geocoding techniques.” *Computers, Environment and Urban Systems*, vol. 32, no. 3, 2008, pp. 214–232., doi:10.1016/j.compenvurbsys.2008.11.006

Zandbergen, Paul A. “Geocoding Quality and Implications for Spatial Analysis.” *Geography Compass*, vol. 3, no. 2, 2009, pp. 647–680., doi:10.1111/j.1749-8198.2008.00205.x..

Zhan, F. Benjamin, Jean D. Brender, Ionara De Lima, Lucina Suarez, and Peter H. Langlois. "Match Rate and Positional Accuracy of Two Geocoding Methods for Epidemiologic Research." Annals of Epidemiology 16, no. 11 (2006): 842-49. doi:10.1016/j.annepidem.2006.08.001.

Zimmerman, Dale L. "Estimating the Intensity of a Spatial Point Process from Locations Coarsened by Incomplete Geocoding." Biometrics 64, no. 1 (2007): 262-70. doi:10.1111/j.1541-0420.2007.00870.x