SEPTIC SYSTEM AUTOMATED LOCATION TOOL (SSALT): A LOW-RESOURCE METHOD FOR POPULATING A SEPTIC SYSTEM DATABASE FOR COUNTIES USING REMOTE SENSING AND ARCGIS

by

BRANDI CARR

(Under the Direction of Nandita Gaur)

Abstract

Failing septic systems contribute to compromised surface and ground water quality. Counties and regulatory agencies seeking to develop strategies to mitigate this potential source of pollution require, but typically lack, a digitized septic system database which can be logistically expensive to develop. To support that effort, a toolkit to locate and date septic systems was developed and tested in Jackson County, Georgia. Multi-temporal, high-resolution aerial imagery was classified using a supervised support vector machine algorithm to identify buildings. The building classification resulted in a 79% accuracy, and the results were geoprocessed with necessary vector layers in a GIS framework to develop Septic System Automated Location Tool (SSALT). The validation of SSALT with geocoded septic permits for Jackson county showed that 83% parcels were correctly identified to be on septic systems, 7% were incorrectly assigned on septic systems while 10% of parcels identified on septic systems did not contain a building.

INDEX WORDS: Automated algorithm, ArcGIS Pro, septic system database, onsite wastewater, remote sensing, aerial imagery, supervised classification

SEPTIC SYSTEM AUTOMATED LOCATION TOOL (SSALT): A LOW-RESOURCE METHOD FOR POPULATING A SEPTIC SYSTEM DATABASE FOR COUNTIES USING REMOTE SENSING AND ARCGIS

by

BRANDI CARR

BS, The University of North Georgia, 2018

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2023

SEPTIC SYSTEM AUTOMATED LOCATION TOOL (SSALT): A LOW-RESOURCE METHOD FOR POPULATING A SEPTIC SYSTEM DATABASE FOR COUNTIES USING REMOTE SENSING AND ARCGIS

by

BRANDI CARR

Major Professor:	Nandita Gaur
Committee:	Krista Capps
	Matthew Levi

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor, Dr. Gaur, for her invaluable knowledge, advice, and patience during my master's research. I am also extremely grateful to my defense committee, Dr. Capps and Dr. Levi, for generously providing their expertise and knowledge. Additionally, many thanks to the lab manager extraordinaire, Matthew Thibodeaux, for helping me with all my EM data collection. Also, two big thanks to Dr. Abney for volunteering her backyard for my field site and to Lacey Tucker and Abby Moody for tirelessly scanning hundreds of septic permits. Thanks to the Jackson County folks for all they have done to help me with this project: Joel Logan, Kelli Hinson, and Gina Roy. I can't forget to add a special thanks to my son, Abel Moody, who has supported me throughout my academic career and has been my continuous motivation. Lastly, this work would not have been possible without generous financial support from Georgia Environmental Protection Division.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1
INTRODUCTION

Septic systems are a popular onsite wastewater treatment method that are employed in the United States (US). The US Environmental Protection Agency (USEPA) estimates that one-third of homes in the southeastern United States are on a septic system (USEPA, 2018) and more than 30% of new homes are being built on septic (USEPA, 2002). These systems treat over four billion gallons of sewage daily (USEPA, 2013). Specific to the upper Oconee water planning region in Georgia that represents 13 counties with a good mix of rural and urban land-use, approximately 55% of the municipal wastewater is treated through septic systems (Upper Oconee Regional Water Plan, 2017).

This method of onsite wastewater treatment is safe and efficient if appropriately maintained. However, a failed system or malfunctioning septic system that fails to appropriately treat the wastewater before effluent reaches the groundwater (Cogger, 1987) can be economically and environmentally disruptive at the household and the watershed levels since it could lead to untreated wastewater backing up into the house or contaminating surrounding surface and groundwater resources (Conn et al., 2012). System failures are often attributed to inadequate soil treatment capacity or system age (Dewalle 1981, Mancl & Slater 2000, & Vedachalam et al., 2013). In regions serviced by septic systems, a high density of septic systems and aging systems have also often been attributed to declining water quality under both baseflow and wet conditions. Lipp et al., 2001 showed that fecal indicator organisms were generally concentrated in an area with high septic system densities. Mallin and McIver, 2012 showed that septic leachate loaded nutrients and fecal microbes into park bodies in the sandy barrier islands of Cape Hatteras National Seashore, U.S.A. and recommended identifying alternate means of wastewater treatment. Using dye tracer tests, Murphy et al., 2020 found evidence of microbial and chemical contamination of private drinking water wells from septic systems in Southeastern Pennsylvania. Squillace and Moran (2007) studied volatile organic compounds in principal aquifers across the United States between 1996 and 2002 and found a positive correlation between septic system density and five organic compounds. Robertson et al., 1991 found evidence of septic effluent plumes over 130 m downgradient in groundwater flows. Several Total Maximum Daily Load documents generated by the Environmental Protection Division in Georgia list failing septic systems as a source of enterococci or nutrient contributions in 303(d) listed or polluted streams. For example, EPD Ogeechee River Basin TMDL report in September 2020 (Ogeechee River Basin TMDL, 2020) and the Lake Lanier TMDL for Chlorophyll-a both list septic system as a potential source of contamination. Besides its impact on water quality, the effluent from septic systems eventually makes it way to groundwater and becomes a contributor to streamwater flow under baseflow conditions as well. A study from 24 watersheds within Georgia (Sowah et al., 2017) showed that septic systems contribute significantly to groundwater recharge during baseflow conditions especially when their density exceeds 87 septic units/km^2. This could have significant impact on water planning especially under drought conditions.

While the need for efficient septic system management to manage water resources has been established in scientific literature, a key factor limiting the abilities of local, regional, and state resource managers to estimate system failures and the environmental and human health impacts of aging and obsolete septic systems is limited availability of spatial data. Empirical data pertaining to system function and system specific data, including age, condition, location, and density are exceptionally limited, and this has led to highly variable estimates of system failures (Capps, et al., 2020). For example, the Georgia Department of Health estimates failure rates between 1 – 5% (GA DPH, 2012), whereas the USEPA suggests failure rates are greater than 20% (USEPA, 2002). Conversations with county managers and environmental health officers also revealed that spatially explicit data for septic infrastructure is either not available for all counties or they lack the resources to generate it.

Additional tools are need to support local jurisdictions to map their septic infrastructure. When coupled with information about system-specific attributes such as ages of tanks and soil type, digitized location information could help identify systems or regions at risk of pollution due to failing systems before substantial damage to the water quality occurs. This project aims to develop a low-cost method using commonly-available spatial data to map septic systems at the county level, while also creating an efficient digital septic system data layer to assist with environmental and water quality needs in Jackson County, Georgia.

This chapter provides a background on the tools and geospatial processes required for developing this tool. In the first section, I will present a background on septic systems and parameters associated with septic systems that could be related to system failure and are of interest to environmental managers. In the next section, I will describe the current status of septic system databases across the country and the available methods to create such databases followed by remote sensing and GIS based tools and methods available to develop such databases.

1.1 SEPTIC SYSTEMS: AN OVERVIEW

Septic systems, also known as onsite wastewater treatment systems (OWTS), are designed to use soil for the treatment and dispersion of wastewater (sinks, toilets, showers, dishwashers, washing machines) from homes and businesses when a central public sewer service is not available (GA DPH, 2002). The size and design of an OWTS varies widely depending on environmental conditions such as soil type, topography, building size, and proximity to water bodies, and management jurisdictions through region specific local regulations (USEPA, 2021). Determination of these variables will dictate the type of septic system installed. The most commonly used system design is a Conventional System, which has existed for decades (USEPA, 2021). The subsurface workings of a conventional system consist of two main parts – the septic tank and the drain field (USEPA, 2021). The septic tank is a buried, watertight container built to receive and partially treat raw domestic sewage (USEPA, 2021). The heavy solids (sludge) settle to the bottom of the tank, while the rest of the fluid wastewater (effluent) is discharged into the drain field for additional treatment. The drain field consists of shallow, gravel-filled trenches with corrugated pipes to filter the effluent until gravity continues to transport it down through the soil, where the microbes further filter the effluent before completing its journey to the groundwater. Therefore, onsite systems can provide long-term, high-quality domestic sewage treatment with appropriate site designations, design, use, and installation (Cogger, 1987).

The scope of ground and surface water contamination occurring from septic systems depends on multiple factors, with the main being the soil properties (physical, chemical, and biological) within the drain field (Lusk et al., 2017; De and Toor, 2015; Karathanasis et al., 2006; /Wilhelm et al., 1994). Wastewater generated from households commonly contains several contaminants, such as phosphorus (P), Nitrogen (N), trace organic chemicals (TOrCs), and pathogens (Lusk et al., 2017). Once wastewater (effluent) leaves the septic tank, it goes through the secondary treatment within the drain field, leach field, absorption system, or subsurface wastewater infiltration, all of which consist of natural soil (USEPA, 2011; Lusk al., 2017). A drain field's adsorption and filtering production are affected by the biomat (biomaterial layer). Effluent flowing through the biomat into unsaturated soil has important involvement in reducing contaminants within the effluent. Biogeochemical processes in the soil can also reduce the effect of contaminants in both the septic tank and effluent (Lusk et al., 2017). For example, the nutrient N (nitrogen) is recognized as a source of potential pollution in the environment from septic systems (Lusk et al., 2017; Elliott et al., 2007; Hossain et al., 2010; Kaushal et al., 2011; Law et al., 2004; Oakley et al., 2010). The main source of N in raw wastewater (sewage) comes from urine and fecal matter (Lusk et al., 2017; Henze et al., 2002). In a conventional septic tank, organic N present in pH-neutral (~7) raw wastewater is mineralized to NH_4^+ (ammonium) through ammonification by heterotrophic bacteria. During this process, the soluble organic N is converted to C NH_3 (Carbon ammonia) and takes the Fithe NH_4^+ cannot be converted to NO_3^- through the aerobic nitrification process. Hence, the effluent contains primarily NH_4^+ with smaller percentages of organic N (Lusk et al., 2017). A quick and almost complete conversion of the NH_4^+ to NO_3^- is only expected if a sufficient unsaturated zone exists in the vadose zone (Lusk et al., 2017). When aged septic systems are working too close to the water table or when the drain field clogs and backs up due to a failed system, the system becomes anaerobic, and the transformation of NH_4^+ to NO_3^- becomes limited. Also, soil texture affects N loss from septic systems (Lusk et al., 2017, Long, 1995). A coarser textured soil (sandy) promotes rapid oxidation of C and N and, in turn, does not leave sufficient C for denitrification (NO_3^- converted to N gases) (Long, 1995). On the other hand, it was reported that there was no significant NO_3^-

adsorption in fine-textured soils (sandy clay, clay, & sandy clay loam) within the Georgia Piedmont (McVay et al., 2004, Bradshaw & Radcliffe, 2011, Lusk et al. 2017). In general, if a drain field becomes saturated, the possibility of NO_3^- leaching to ground and surface water is highly likely.

Septic systems can impair water quality from phosphorus (P) loss, although at a much lower concentration than N (Lusk et al., 2017). Phosphorus is more of a concern for surface waters than groundwater (Lusk et al., 2017), but can be problematic in both. Surface waters contaminated with excess P have been related to septic systems (Bowes et al., 2010; Brennan et al., 2015; Corbett et al., 2002; Edwards & Withers, 2008; Meinikmann et al., 2015; Lusk et al., 2017). It was found that septic systems contributed 4 – 55% of total P loads to lakes (Lombardo, 2006; Lusk et al., 2017). Wastewater from households contains P from toilets, baths, kitchens, and laundry, which mainly consists of human wastes and foods containing phospholipids, nucleotides, and certain sugars (Lusk et al., 2017). A septic tank's anaerobic environment converts those P forms to orthophosphate (Beal et al., 2005; Lusk et al., 2017). Therefore, the P that does not settle to the bottom of the tank with the sludge enters the drain field as effluent and contains mostly orthophosphate (85%) and only 15% organic P (Gold & Sims, 2001; Lusk et al., 2017). Once in the drain field, the majority of organic P gets mineralized to orthophosphate in the right conditions by both biological and biochemical processes, with only a small portion of organic P that might be sorbed to soil surfaces (Lusk et al., 2017; McGill & Cole, 1981). Depending on the soil components, orthophosphate can be sorbed to organic matter, clay minerals, hydroxides, and metal oxides. In soils that are acidic, P is precipitated along with Fe (Iron), Al (aluminum), and Mn (manganese), and in alkaline soils with Mg (magnesium) and Ca (calcium) (Lusk et al., 2017). Unlike N transformations, P processes can be reversed (McCray et al., 2009; Lusk et al., 2017), but usually, P removal is quite effective in the drain field due to the help of a biomat and soil retaining P through sorption and precipitation (Lusk et al., 2017). Multiple studies have observed that the different soil types vary in their ability to remove P from effluent because different soil components and mineralogy have different capacities for PO_4^- (phosphate) sorption (Lusk et al., 2017; Kimochi et al., 2004; Al-Shiekh Lhalil et al., 2004).

This understanding of how soils filter nutrients highlight the need to predict failures because installation errors, lack of maintenance, old tanks, and misuse can cause system failures, and thereby release unwanted nutrients to ground and surface water.

1.2 SEPTIC SYSTEMS REGULATION AND RECORD MAINTENANCE IN GEORGIA

Within the last two decades, regulations for OWTS have become much stricter (K. Hinson, personal communications, 2020). However, before these stringent regulations took effect, homeowners had little to no direction pertaining to system installation and often these systems were installed without soil tests, inspections, and permits. Currently in Georgia, the Georgia Department of Public Health (DPH) regulates OWTS. County boards of health are responsible for permitting, inspecting, and enforcement. If the system has a 10,000 gpd (gallons per day) or more flow rate, a change in the jurisdiction to a commercial system regulation is warranted. The Department of Natural Resources, Environmental Protection Division regulates commercial OWTS (NESC, 2020). The Georgia DPH has used multiple OWTS tracking devices throughout the years, such as spreadsheets and purchased databases (NESC, 2020). However, these methods have proved to be physically laborious and costly. Septic system data for most counties in Georgia is currently housed in environmental health departments' filing cabinets which makes scanning and digitizing such permits a high human resource task. However, knowledge of this data for

Georgia ignites the motivation to provide a more efficient means of locating and aging the OWTS for the state and possibly other jurisdictions to boot.

1.3 OVERVIEW OF STATE-LEVEL SEPTIC SYSTEM DATABASES IN THE U.S.

Currently, there are only a handful of septic system databases around the country. In Georgia, there is a database maintained by the Department of Public Health (DPH) for Well and Septic Tank Referencing and Online Mapping (WelSTROM). The data is compiled from septic permits and manually input into the Digital Health Database (DHD) by participating local health district in Georgia, which is then synced to the WelSTROM database (WelSTROM, 2022). Although, the entire extent of Georgia is not covered by all WelSTROM map layers, the goal of WelSTROM is to support a web-based application that obtains private well and sewage treatment installation enabling methods for query, reporting, and mapping of these installations (WelSTROM, 2022).

The state of Virginia has a septic system data management created by the Virginia Department of Health (VDH). The state-wide database is called Virginia Environmental Information System (VENIS). It is similar to the Georgia WelSTROM database in that local health districts (LHD) enter their septic system permits into the VENIS database (VDH, 2018). Although LHD enters the septic information, the database is only used by VDH personnel and not counties or localities (VDH, 2018). The system contains application tracking (new construction and repair), soil evaluation summaries, and inspection reports for dates received, approved/denied, etc. (VDH, 2018). The database handles all onsite sewage systems, sewage hauling permits, and complaints. The VDH is still working to gain a complete state-wide inventory within the database.

The Indiana State Department of Health (ISDH) created Indiana's Network for Tracking of Onsite Sewage Systems (iTOSS). This data system manages onsite sewage by providing reporting capabilities and web access to ISDH personnel (Mettler & Atwood, 2010). The mission of the state-wide system is to enhance program efficiency, effectiveness, and compatibility across counties, which will help minimize the possibility of transmission of diseases associated with sewage treatment practices (Mettler & Atwood, 2010).

1.4 ROLE OF REMOTE SENSING AND GIS IN DEVELOPING SEPTIC SYSTEM DATABASES

GIS and remote sensing could be of considerable help to the governing and regulating agencies to develop statewide databases. GIS and remote sensing software and data are widely available to public and private companies allowing access to the necessary information to use these platforms. Specific data such as raster layers (aerial and satellite images) and feature layers (tax parcels, sewer lines, and building information) are often available to counties or high resolution imagery can be purchased from commercial platforms like Planet Imagery (Planet, San Fransisco, CA 94107). These datasets can be used to generate building locations through image classification, such as a supervised classification algorithm. Moving forward, this process could allow for a low-resource method to automate septic system locations and ages to assist states and counties with their missions.

1.5 SUPERVISED CLASSIFICATION ALGORITHMS

Images and image classification have become valuable ways of finding and conveying information in people's work (Gaye et al., 2021), which we have made use in experimentation to detect and age septic systems. Generally, supervised image classification, is the practice of allocating land cover classes such as urban, water, agriculture, wetlands, and forest to pixels or objects by calculating the reflectance statistics for each one or, in the case of objects, for each group of pixels (segments). The classes are derived from the Anderson Classification System of Land Cover/Use that was developed in 1976 specifically for use with remotely sensed data (Anderson, et al., 1976). The system contains different levels of land cover categories available from the National Land Cover Dataset (NLCD). There are several methods of image classification that exist for the analysis of remote sensing imagery. The performance and accuracy of these methods depends upon the subject, area of interest, the satellite or drone's spatial resolution, and the data distribution. This overview will focus on the supervised method and discuss various machine learning classification algorithms such as maximum likelihood, random forest, convolution neural networks (CNN), support vector machines (SVM), and a previously created building extraction deep learning model for determining the best method for our study.

MAXIMUM LIKELIHOOD

A widely used supervised image classification method is maximum likelihood. It is based on Bayes' Theorem (Perumal & Bhaskaran, 2010) and assumes the data are normally distributed in order to calculate the probability density (likelihood) (D.R. Kiran, 2017). The advantages are that it is simple to use, easy to train, and ignores irrelevant features (Schott, 2019). The disadvantage is that it works best with large data sets (Schott, 2019) and can be misleading with substantial overlaps of classes. Although, it is the most powerful classification algorithm when accurate training data is provided (Perumal & Bhaskaran, 2010), it did not meet the needs of our data sets due to this algorithm's possible overlapping of classes.

RANDOM FOREST

The next conceivable supervised classification technique to detect septic systems was the random forest method. The random forest algorithm bases the classification of several decision trees, hence the name forest. Thus, it is considered an ensemble learning technique (Scholz & Wimmer, 2021). Decisions are made with a decision tree classifier at multiple levels (Kulkarni & Lowe, 2016). Each tree generates a prediction and is used as part of the voting scheme to make a final prediction (Classify—ArcGIS Pro, 2020). The final predictions are based not on one tree but the entire forest. This strategy helps avoid overfitting the model to the training data (Classify—ArcGIS Pro, 2020), making it a robust image classification method. Although limitations being that it can take even longer to train than other models and be computationally intensive, therefore, was not suitable to our project.

CONVOLUTION NEURAL NETWORKS

Convolution Neural Networks (CNN) is a form of deep learning under the umbrella of machine learning. The neural network is loosely modeled after the human brain, hence making it a more complex algorithm (Montantes, 2020). CNN uses layers of interconnected nodes that compute the final output (Montantes, 2020). Saito and Aoki (2015) compared multiple CNN models for building, and road detection from raw aerial imagery (raw pixel values) using small patches clipped from the imagery instead of creating segments and training classes as done with most other supervised classification algorithms. Neural networks have been proven to outperform other machine learning algorithms but at the cost of losing the ability to distinguish the input features/variables once the CNN has completed the classification (Montantes, 2020).

In most cases, this is not the desired outcome for sharing a land cover classification where the features need to stay intact, such as detecting buildings for septic system detection.

SUPPORT VECTOR MACHINE

Support vector machines (SVM) are especially appealing for several reasons. Mitra et al. (2004) describe how SVMs are distinctly created for active learning because an SVM classifier is distinguished by a small set of support vectors (SVs) that are effortlessly updated over consecutive learning steps. Another positive, unlike the maximum likelihood classification is that SVMs do not require the data to be normally distributed (Classify—ArcGIS Pro, 2020). They can handle small training datasets successfully while often producing higher classification accuracy (Kulkarni & Lowe, 2016). The SVM can also handle larger images than other supervised classification algorithms (Tessellations Inc., 2020), which is essential to this research, as the aerial images are quite large. Finally, this algorithm is less susceptible to discrepancies in the training sample (Tessellations Inc., 2020), which is helpful when dealing with the large images of a county-wide classification.

ESRI BUILDING DETECTION

Esri is known for the software applications of ArcGIS (desktop, Pro, online, and mobile), which are being used in this study. Consequently, the company created a deep learning model for extracting building footprints from imagery that is accessible when using ArcGIS Pro and ArcGIS Online (https://mediaspace.esri.com, accessed 11/10/22). The service is free for users with an ArcGIS Online account. The materials include a tutorial, a download for the Esri building extraction model from the Living Atlas website (https://www.arcgis.com/home/item.html?id=a6857359a1cd44839781a4f113cd5934, 01/02/23), and the necessary deep learning framework available through GitHub. The model was based on high-spatial-resolution multispectral satellite imagery (Esri, 2020). Esri (2020) has found that the model works well on a global scale but was designed to be at optimum performance within the subcontinent of the United States. The tool and its parameters are fairly easy to navigate while obtaining the necessary framework and the model could be somewhat intimidating to a user unfamiliar with acquiring geospatial data. Although, this method could be of useful for creating a building classification for a user for implementation into SSALT.

Only a handful of models available were discussed here, but there are more techniques to perform supervised image classification using remotely sensed multispectral or three-band imagery data. Each has its pros and cons, and it is up to the user to determine the best fit for the problem being addressed. There is no method that can be deemed as the best classification method universally and the choice of method is driven by the question, data, and the area of interest. Duda et al. (2001) describe it best, when mentioning that it's important to keep a healthy skepticism concerning studies that claim to demonstrate an overall precedence of a specific learning or recognition algorithm. Although to create confident results and guarantee a high-quality product, it is crucial to perform the correct steps for image classification and validation, whether for public or private applications. The multiple numbers of machine learning classification algorithms to utilize can be overwhelming, especially since the literature on creating an automated toolbox for building detection classification is few and far between as it is not a highly researched subject. Hence, it is essential to understand the algorithms differences to create the best data classification possible.

LITERATURE CITED

Aero-Stream Septic System Cost Analysis – Explore Your Options When Your Septic System Fails. (2019). https://www.aero-stream.com/septic-system-cost.html

Atwood, M. M. a. C. (2010, January 2010). *Indiana's Network for Tracking of Onsite Sewage Systems (iTOSS): Indiana State Department of Health's Data System for Managing Onsite Sewage System Data* Proceedings of the Water Environment Federation, https://www.accesswater.org/publications/proceedings/-297174/indiana-s-network-for-tracking-of-onsite-sewage-systems--itoss---indiana-state-department-of-health-s-data-system-for-managing-onsite-sewage-system-data

Cogger, C. G. (1987). *Basic Principles of Onsite Sewage*.

Conn, K. E., Habteselassie, M. Y., Denene Blackwood, A., & Noble, R. T. (2012).

Microbial water quality before and after the repair of a failing onsite wastewater treatment system adjacent to coastal waters. *J Appl Microbiol*, *112*(1), 214-224. https://doi.org/10.1111/j.1365-2672.2011.05183.x

Commission, G. R. (2020, 2020). *WelSTROM Well and Septic Tank Referencing and Online Mapping*. Retrieved December 22 from https://www.welstrom.com/help.html

Division, G. E. P. (2000). *MINIMUM STANDARDS FOR PUBLIC WATER SYSTEMS* https://epd.georgia.gov/sites/epd.georgia.gov/files/related_files/site_page/standards.pdf

Gaye, B., Zhang, D., & Wulamu, A. (2021). Improvement of Support Vector Machine Algorithm in Big Data Background. *Mathematical Problems in Engineering*, *2021*, 5594899. https://doi.org/10.1155/2021/5594899

Georgia Department, o., Public, Health. (2016). RULES OF THE DEPARTMENT OF PUBLIC HEALTH CHAPTER 511-3-1 On-Site Sewage Management Systems. https://casetext.com/regulation/georgia-administrative-code/department-511-rules-of-georgia-department-of-public-health/chapter-511-3-environmental-health-hazards/subject-511-3-1-on-site-sewage-management-systems/rule-511-3-1-03-general-requirements-for-on-site-sewage-management-systems

Health, V. D. o. (2018). *Virginia Example of Septic System Data Management*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.epa.gov/sites/default/files/2018-09/documents/virginia_septic_database.pdf

Lipp, E. K., Kurz, R., Vincent, R., Rodriguez-Palacios, C., Farrah, S. R., & Rose, J. B. (2001). The effects of seasonal variability and weather on microbial fecal pollution and enteric pathogens in a subtropical estuary. Estuaries, 24(2), 266-276.

Lusk, M. G., Toor, G. S., Yang, Y.-Y., Mechtensimer, S., De, M., & Obreza, T. A. (2017). A review of the fate and transport of nitrogen, phosphorus, pathogens, and trace organic chemicals in septic systems. *Critical Reviews in Environmental Science and Technology*, *47*(7), 455-541. https://doi.org/10.1080/10643389.2017.1327787

Mallin, M. A., & McIver, M. R. (2012). Pollutant impacts to Cape Hatteras National Seashore from urban runoff and septic leachate. Marine Pollution Bulletin, 64(7), 1356-1366.

Ogeechee River Basin Total Maximum Daily Report (2020) https://epd.georgia.gov/ogeechee-river-basin-tmdl-reports

Oliver, C. W., Radcliffe, D. E., Risse, L. M., Habteselassie, M., Mukundan, R., Jeong, J., & Hoghooghi, N. (2014). Quantifying the Contribution of On-Site Wastewater Treatment Systems to Stream Discharge Using the SWAT Model. *Journal of Environmental Quality*, *43*(2), 539-548. https://doi.org/https://doi.org/10.2134/jeq2013.05.0195

Planet Imagery, San Fransisco, CA 94107 https://www.planet.com/

Ravi, N., & Johnson, D. P. (2021). Artificial intelligence based monitoring system for onsite septic systems failure. *Process Safety and Environmental Protection*, *148*, 1090-1097. https://doi.org/https://doi.org/10.1016/j.psep.2021.01.049

Robertson, W. D., Cherry, J. A., & Sudicky, E. A. (1991). Ground ‐ water contamination from two small septic systems on sand aquifers. Groundwater, 29(1), 82-92.

Sowah, R. A., Habteselassie, M. Y., Radcliffe, D. E., Bauske, E., & Risse, M. (2017). Isolating the impact of septic systems on fecal pollution in streams of suburban watersheds in Georgia, United States. Water research, 108, 330-338.

Scholz, M., & Wimmer, T. (2021). A comparison of classification methods across different data complexity scenarios and datasets. *Expert Systems with Applications*, *168*, 114217. https://doi.org/https://doi.org/10.1016/j.eswa.2020.114217

Squillace, P. J., & Moran, M. J. (2007). Factors associated with sources, transport, and fate of volatile organic compounds and their mixtures in aquifers of the United States. Environmental science & technology, 41(7), 2123-2130.

Upper Oconee Regional Water Plan, (2017) https://waterplanning.georgia.gov/water-planning-regions/upper-oconee-water-planning-region/upper-oconee-regional-water-plan

USEPA. (1980). *Design Manual of Onsite Wastewater Treatment and Disposal Systems*.

US Environmental Protection Agency. (2002) Onsite wastewater treatment systems manual. In EPA-625/R-00-008, US Environmental Protection Agency: Washington, D.C., 2002. https://www.epa.gov/sites/production/files/2015-06/documents/2004_07_07_septics_septic_2002_osdm_all.pdf.

USEPA. (2003). Voluntary National Guidelines for Management of Onsite and Clustered (Decentralized) Wastewater Treatment Systems.

US Environmental Protection Agency. (2013) Decentralized wastewater program annual report. USEPA: Washington, D.C., 2013. https://19january2017snapshot.epa.gov/sites/production/files/2015-06/documents/scb_decent_ar_2013_final-508compliant.pdf.

USEPA. (2021). How Your Septic System Works. https://www.epa.gov/septic/how-your-septic-system-works

USEPA. (2022, August 23, 2022). *Types of Septic Systems*. Retrieved December 15 from https://www.epa.gov/septic/types-septic-systems#septictank

CHAPTER 2

SEPTIC SYSTEM AUTOMATED LOCATION TOOL (SSALT)

Carr, B., N. Gaur, K.A. Capps and M.R. Levi. To be submitted to Water Research

ABSTRACT

This study introduces a Septic Systems Automated Location Tool (SSALT) to identify parcels serviced by onsite wastewater treatment systems and estimate their age in the absence of digitized septic system information. Parcel identification was done by classifying buildings using high spatial-resolution three band aerial imagery and geoprocessing county-wide GIS layers within a decision-making, ArcGIS framework. The buildings were classified using a supervised object-based detection approach with the support vector machine (SVM) classification algorithm. The classification was run for two years, 2001 and 2020 to approximate age of septic systems resulting in a Kappa scores of 92% for 2001 and 78% for 2020. The very high resolution (0.075 m) of the 2020 as compared to the 2 m resolution of the 2001 image was found to be the reason of low accuracy owing to noise and the presence of shadows in the imagery. A comparison of the output from SSALT from 2001 with the digitized datasets for 300 sub-divisions in Jackson county showed that 83% parcels were correctly identified to be on septic systems, 7% were incorrectly assigned on septic systems while 10% of parcels identified on septic systems did not contain a building in 2001.

2.1 INTRODUCTION

Failing septic systems are one of the most significant threats to groundwater quality (USEPA, 2003). Failure of systems can be correlated with age of the septic system and the density of septic systems in a watershed can impair water quality in receiving water bodies (Lipp et al., 2001, Squillace and Moran, 2007, Mallin and McIver, 2012, Murphy et al., 2020). Most regulatory agencies do not have digitized information on system locations and ages, which makes it particularly difficult for counties and water managers to determine regions susceptible to septic system failures and assess vulnerability of watersheds to pollution from these systems. The current methods to locate and describe the condition of aging septic systems require intensive physical labor and significant time and financial resources. For example, county health inspectors must visit individual parcels to geolocate individual OWTS resulting in a considerable time commitment. Efforts are also impeded by the potential lack of access to systems owned by unwilling homeowners or the inability to utilize GPS because of dense tree cover. Hence, an automated method to digitize the location and condition of OWTS would support local, regional, and state infrastructure and environmental management efforts.

Despite the challenges of maintaining databases of septic systems, they are an effective method of treating household wastewater within the US by dispersing the effluent throughout the environment, thereby protecting communal health (Links, 1994). These systems are extensively used to treat and dispose of wastewater in locations where centralized sewer is unavailable (Gunady et al., 2015) and treat wastewater from household plumbing through a combination of biological and technological processes (USEPA, 2022). Hence, the proper functioning of septic systems benefits public health by reducing the risk of disease and human exposure to pathogens, environmental health by removing pollutant from surface waters, groundwaters, and recharging aquifers (USEPA, 2022). However, once a septic system is permitted and installed, the homeowner becomes responsible for the maintenance of the structure (USEPA, 2022). There is no involvement from any regulatory agency unless a failure and threat to public health is reported and hence solicits only a reactionary response to such an event instead of being able to proactively avoid it. System failures are often attributed to inadequate soil treatment capacity or system age (Dewalle 1981, Mancl & Slater 2000, & Vedachalam et al., 2013). A lack of basic knowledge of maintenance among homeowners, especially within middle income, low income, and renters and financial barriers for homeowners (Kecil, 2020) can also lead to improperly functioning systems.

Although septic systems are attributed to environmental degradation, the locations of defective systems are widely unknown (Withers et al., 2013). Typical methods to obtain these data are expensive, time-intensive, and they depend on access to private property. The unknown number of systems can be attributed to few requirements in reported data from unreliable households and inadequate government funding for investigations (Walton, 2015). Counties and regulatory agencies responding to failing systems could benefit from knowing whether these failures are systematic or clustered and identify causes of failure to prevent them. However, such information is near impossible to generate without digitized septic system databases owing to the widespread distribution of these systems.

To address this gap in geospatial data, we developed a low-cost a method for populating a septic system database at the county level using remote sensing and ArcGIS: SSALT (Septic Systems Automated Location Tool). The tool uses existing GIS data and current remote sensing images and techniques to create a simple, low-cost model for advancement in detecting failing septic systems by first detecting the locations and ages.

2.2 MATERIALS AND METHODS

The aim of this study was to evaluate and develop a remote sensing and GIS driven algorithm that would become an automated tool to locate and age septic systems for counties. We utilized data obtained from the area of interest of Jackson County, Georgia. The following section details the data and methods that were employed for implementation of the completed Septic System Automated Tool (SSALT).

STUDY AREA

The study was conducted in Jackson County in the Upper Oconee River Watershed in northeast Georgia, US (Figure 1). The county is 339.6 square miles and contains nine cities and towns, and several unincorporated small communities encompassing 339.6 square miles (US Census Bureau, 2021). As of 2019, the county's population was estimated at 72,977 (US Census Bureau, 2021), most of which is considered rural with some populous cities served by sewer and septic systems (Figure 2).

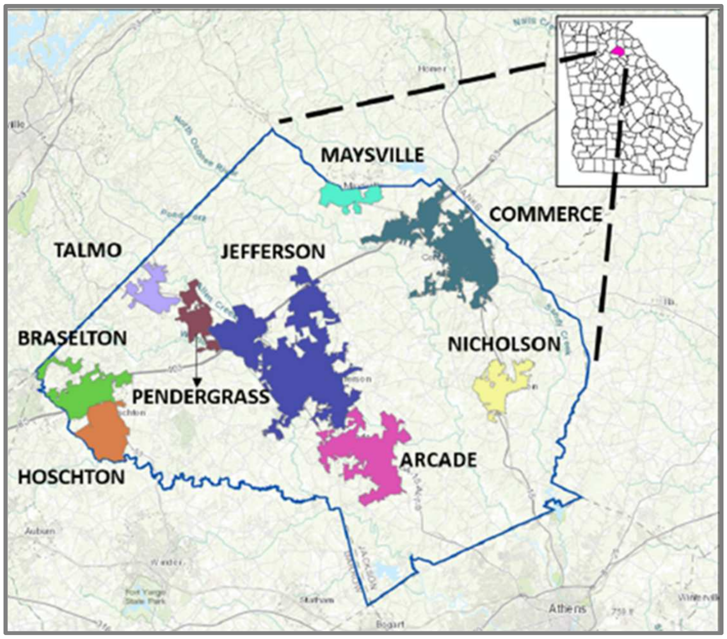


Figure 1: Jackson County, Georgia, showing the major cities and towns.

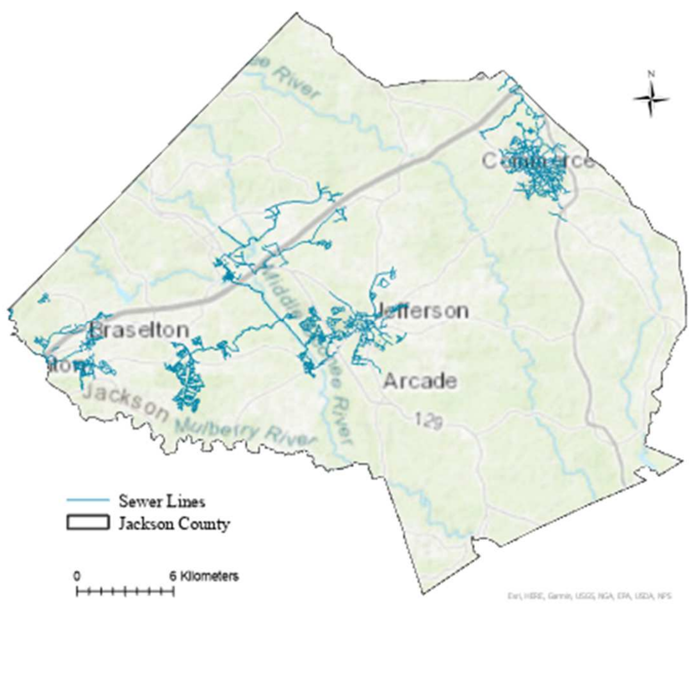


Figure 2: Jackson County, Georgia, shows the sewer lines clustered around the major cities.

REMOTE SENSING & GIS DATA

The automated digitization algorithm was developed using an object-based, time comparison classification within ArcGIS Pro 2.9 (Environmental Systems Research Institute (ESRI) in Redlands, California). Aerial imagery from the years 2001 and 2020 provided by Mr. Joel Logan (Jackson County, Georgia GIS manager), were classified using this algorithm. The resolutions of the images were 2 meters (2001) and 0.075 meters (2020). Existing GIS data from Jackson county's sewer lines and tax parcels was also used in the algorithm to create sewer buffers and identify numbers of parcels. The age of the septic systems was assumed to be the same as the building age. It was assumed that the septic system is simultaneously installed when a building is constructed. However, it is important to note that this assumption would fail in case of septic system repairs and replacements. Building age was estimated using a time comparison analysis of remote sensing imagery that was conducted to estimate the building construction year. The results of this algorithm were validated using geolocated septic system permits for 200 sub-divisions in Jackson County, Georgia to assess the accuracy of the tool outputs. In cases where a permit contained insufficient data such as a parcel address, we used an alternative method of validation based on using the qPublic website, a local government GIS for the web (qPublic/Beacon, 2022). Briefly, the tax parcels or addresses located from qPublic data were related to tables in the septic system map. This relation enabled unknown parcel information to be connected to points on the map.

SELECTION OF REMOTE SENSING IMAGERY

The initial step for designing the automated digitization of the septic systems algorithm was to obtain appropriate remote sensing imagery. Spatial information contributes a fundamental part in analysis for remotely sensed datasets (Aksoy et al., 2012). Tayara & Chong

(2018) asserted the importance for machine learning to distinguish between the background and foreground of an image which indicates the need for high spatial resolution. Hence, to detect buildings using an object-based supervised classification algorithm, high spatial-resolution imagery (<= 3 meters) was required, as Landsat and other open-source satellite mission data were too coarse for this objective. The primary benefit of having this high resolution is that tiny objects such as individual buildings can be undeniably identified (NASA EOS, 2019). However, high spatial-resolution images were not available through commercial satellites before the 21st century. Jackson County's GIS department collects aerial photographs (three-band orthoimage) for the entire county every three years. In addition to the orthoimages, the county routinely flies unmanned aircraft systems (UAS) between the three years to maintain the department's databases of all new developments. We used the three-band red, green, and blue (RGB) high spatial-resolution aerial images from 2001 (2 meters) (Figure 3) and 2020 (0.075 m) from the Jackson County GIS department to support detection of system locations and to estimate system age. The aerial images were both taken using a nadir (vertical) angle and have previously been orthorectified. Both years were utilized to produce a time comparison object-based supervised classification for analyzing locations and ages of buildings.

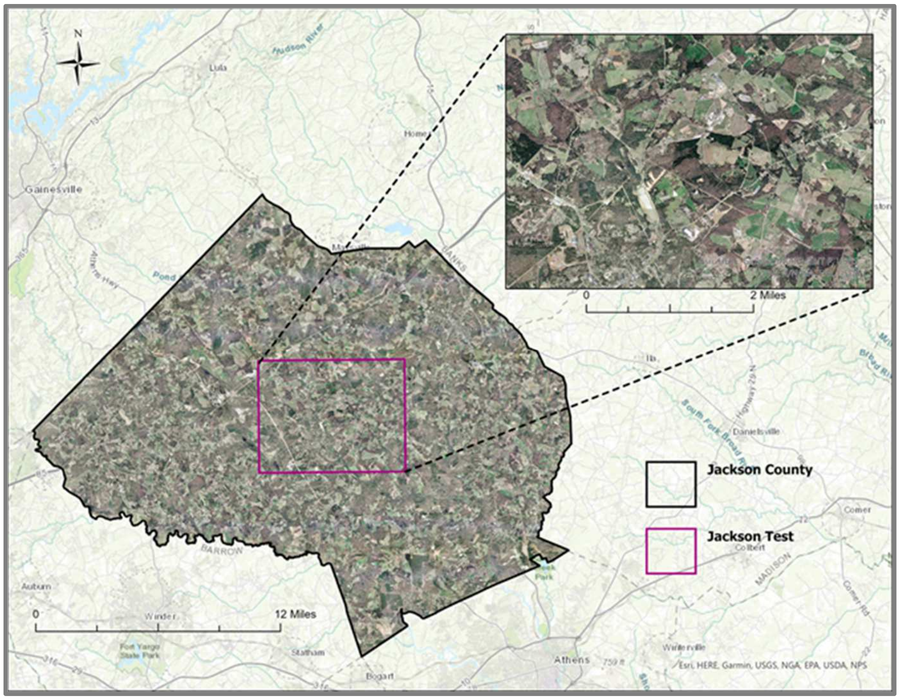


Figure 3: Aerial image of Jackson County in 2001 showing the zoomed-in test image worked within the classification.

IMAGE CLASSIFICATION

We used ArcGIS Pro 2.9 to visualize and analyze remote sensing imagery and create an automated model. The software is a GIS tool utilized by analysts around the globe. ArcGIS has ready-to-use geoprocessing tools for creating personalized models within the system which was used in this study to conduct an object-based supervised classification.

SEGMENTATION

The progress of image segmentation methods and object-based classification became widespread after high spatial-resolution remote sensing imagery was made available (Yongxue et al., 2006). It combines pixels based on shape and spectral characteristics to create objects that are used to train the object-based classification algorithm. Shape characteristics include rectangular and square houses, long thin rectangular roads, round areas denoting water, and large swaths of forests based on a spatial weight. It also uses the number of pixels (segments) an object contains to help extract objects further. Spectral characteristics were based on the true object color, such as green deciduous and evergreen forests, blue water, and bare orange soil since the imagery used in this study was a three band RGB true color imagery. These characteristics were assigned weights and these spectral, spatial, and segment weights determine the best fit for training the image by extracting texture and shape (Sun et al., 2009). These values were based on the resolution of the imagery to determine smoothing by a trial-and-error method (El-nagger, 2018). The chosen values (ranging from 1 -20) for the 2-meter 2001 aerial photo of Jackson County were as follows: spectral 14.5, spatial 17, and minimum segment size 20 (units in pixels). The values for the 2020 aerial photo were spectral 13, spatial 10, and minimum segment size 20. Again, due to the trial-and-error method, the values differ due to the spectral intensification of 2020. Even though the image was resampled to two meters, the spectral values were higher than in the 2001 image and required more smoothing. Another issue with the 2020 image was the large number of shadows in the image. The shadows may have been due to the high spectral or spatial resolution, or the time of acquisition (time of day) of the photo (Dare, 2005). As a general rule, the atmosphere is clearer in the morning than later in the afternoon (Dare, 2005). Lastly, the images were trained using the segmented raster (Figure 4) for detecting objects (object-based images).

Figure 4: Comparison of the 2001 aerial image and the output of the segmented image using the method stated above.

TRAINING

Training data from remotely sensed RGB images was used to develop predictions for specific land classes belonging to the image (Giannopoulos et al., 2020) and the NLCD. The training was performed on a small test polygon extracted from the county image (Figure 3). This was due to the computational intensity of running the entire county, which required a significant length of time (>= 12 hours). One advantage of using the chosen SVM algorithm is that it does not need a large training dataset (Gaye et al., 2021). There were small training samples for each category (8 classes: buildings, roads, water, evergreen, deciduous, mixed forest, planted/cultivated, and barren). We trained the image to identify and separate the segments into these different land cover/land use categories. More training sites available for input increases the quality of classification for the output (Perumal & Bhaskaran 2010). Therefore, we used training that consisted of creating representative areas to develop a numerical description of spectral signatures of individual land cover types.

Figure 5: Sequence of operations required for a supervised classification.

Supervised classifications typically have a series of three operations that need to be followed (Perumal & Bhaskaran, 2010) (Figure 5). The first is to create the training sites, which in our case was from the segmented image produced from the aerial image. Secondly, the software generates signature files, which are statistical characteristics of information from the digitized training sites, to use in the proceeding step. The last step is to apply the classification methods.

We implemented the object-based support vector machine (SVM) algorithm for classification of our study area. The findings from the Wu et al (2017) study support applying this method due to it producing the highest classification accuracy of buildings. Furthermore, the SVM was developed to work well with segmented rasters (object-based detection) and can handle small training datasets successfully while often producing higher classification accuracy (Kulkarni & Lowe, 2016). They are also well suited to high-dimensional data (Heumann, 2011), such as the case when classifying large images with high-spatial resolution (Zortea et al., 2007).

After the supervised classification sequence was completed, the SVM algorithm was validated for accuracy.

ACCURACY ASSESSMENT

A classification cannot be considered complete until it has been properly assessed (Congalton, 1991). Therefore, the aerial images of Jackson County were used to assess the accuracy by ground truthing the data to achieve an appropriate validation for the supervised SVM classification. For the 2001 and 2020 validation, an equalized random stratified sampling of 500 points (default parameter) was used. This sampling method was chosen since the focus was on the accuracy of the buildings class, therefore, allowing for the even distribution of fifty-six points in 2001 for nine classes and seventy-one points for eight classes in 2020 to check the model's accuracy. The other sampling methods (stratified and random) did not provide enough assessment points for the buildings class. A good rule of thumb is collecting at least fifty points per land class category for the confusion (error) matrix (Congalton, 1991). The number of samples per class can also be modified based on the relative importance of that class within the purposes of the classification (Congalton, 1991). It can also be helpful to concentrate the sample size on the classes of interest and increase their number of samples while reducing the sample size in the less important classes (Congalton, 1991). To enhance this theory, I assessed the minimum number of points to achieve acceptable accuracy. Since the main objective of this study was to classify buildings, this step was only performed for the buildings class. The number of points were increased in increments of 250 starting at 500 points, then 750, and once more for 1000 accuracy assessment points (Figure 6).

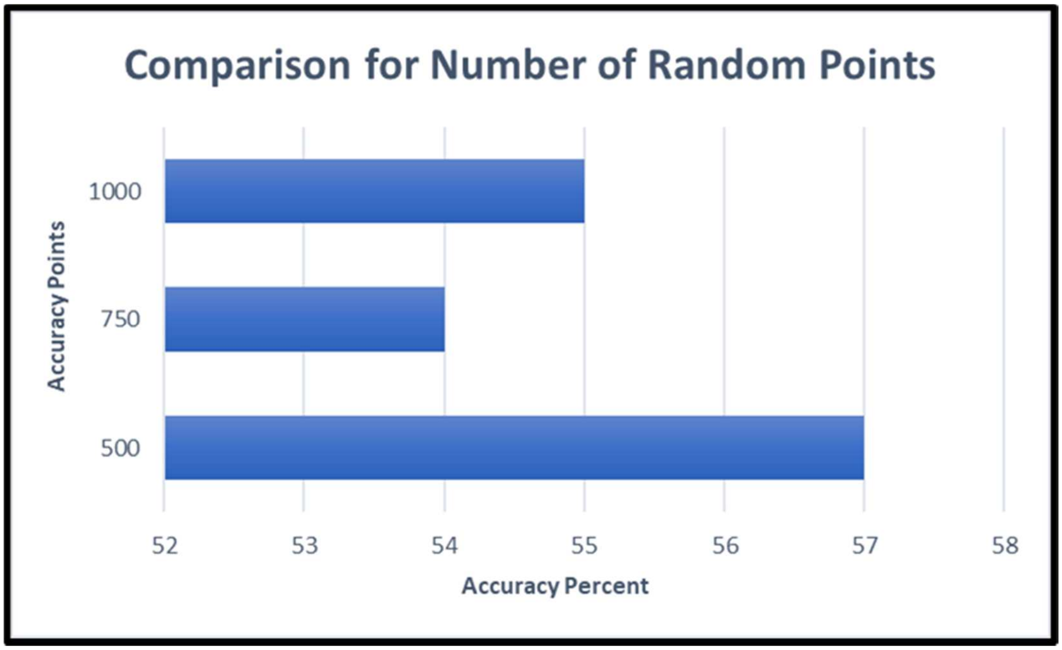


Figure 6: Graph showing Random number of points comparions for creating accuracy assessments.

FOREST MASK

Low accuracy in builidng accuracy indicated that the forest classes were creating errors of commission within the buildings class. With the observed data, a new method to refine the building accuracy was devised by applying a forest mask to the aerial image. Since the goal of this study was to identify buildings with high accuracy, features (forest classes) which could confound the classification were masked out. For instance, upon visual inspection of the classification, it was obvious that the roof colors were of browns and greens, which have close spectral signatures of the same browns and greens of trees. In addition, a spatial assessment of the segmented image determined segments of the forests had the same shape and size as the buildings. Hence, a novel idea to employ a forest mask was produced and applied on the image before further assessment of the building accuracy of the classified image.

Steps to produce a forest mask included using the 2001 NLCD for the 2001 aerial image, and the 2019 NLCD was used for the 2020 aerial image. The NLCD rasters are computed from Landsat data with a spatial resolution of 30 meters (Chander et al., 2009). Therefore, the first geoprocessing step (after clipping out the AOI) was to resample the NLCD to match the original county raster resolution of two meters. Then, the NLCD was reclassified, so the forested land and water bodies were NODATA. Once reclassified, the new NLCD that excluded the forests and water was converted to a polygon feature. Lastly, the NLCD polygon containing the urban and agricultural land was extracted by mask from the original two-meter aerial test images. This left a raster containing only the urban and agricultural land to be classified using the same segmented image and training feature class. The classification continued to utilize the forests classes because there were still trees scattered throughout the image. This new method (Figure 7) was performed for both test area classification years before additional building accuracy was assessed.

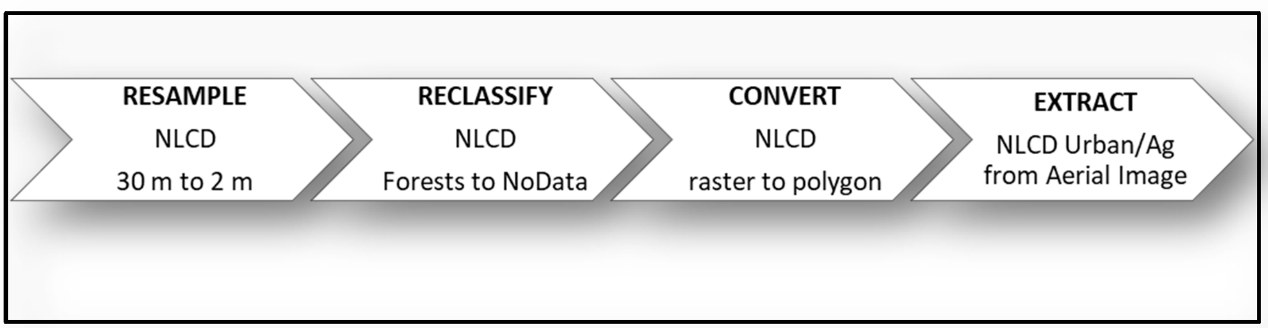


Figure 7: Flowchart for creating a new NLCD feature class containing only urban and agricultural land cover to use as the Forest Mask.

ASSESSING BUILDING CLASSIFICATION ACCURACY

Ten iterations of assessment points were performed for each of the four classification methods to obtain an average building accuracy score. Then, boxplots were created to compare the statistical values of these methods for both test area years (2001 & 2020) and for the entire classification of 2001 discussed later in the results section.

The four assessments for the test area classification are as follows: Original Classification, Forest Mask Classification, Original Classification using Parcels Layer, and Forest Mask Classification using Parcel Layer (Figure 8). Initially, the original classification was only considered correct when the point was precisely on a building using only the aerial image to ground-truth. Otherwise, it was considered inaccurate and assessed to the appropriate class. Subsequently, the Jackson County parcel layer was used to obtain ground-truth data since the objective of the algorithm was to identify parcels on septic systems. This method produced better results due to the larger area acceptable to be considered a building. If there was a building on the parcel, there was the possibility of a septic system also being on the parcel unless the building connected to a sewer line.

To create a complete accuracy assessment of the entire classification using all eight classes, water was excluded from the 2020 classification due to the water class having a minimal number of pixels, all 500 accuracy assessment points were ground-truthed using the aerial image. The highest building accuracy percentage was chosen from each of the four classification methods. The best iteration was sourced to create the total accuracy of all the classes from the computed confusion matrix.

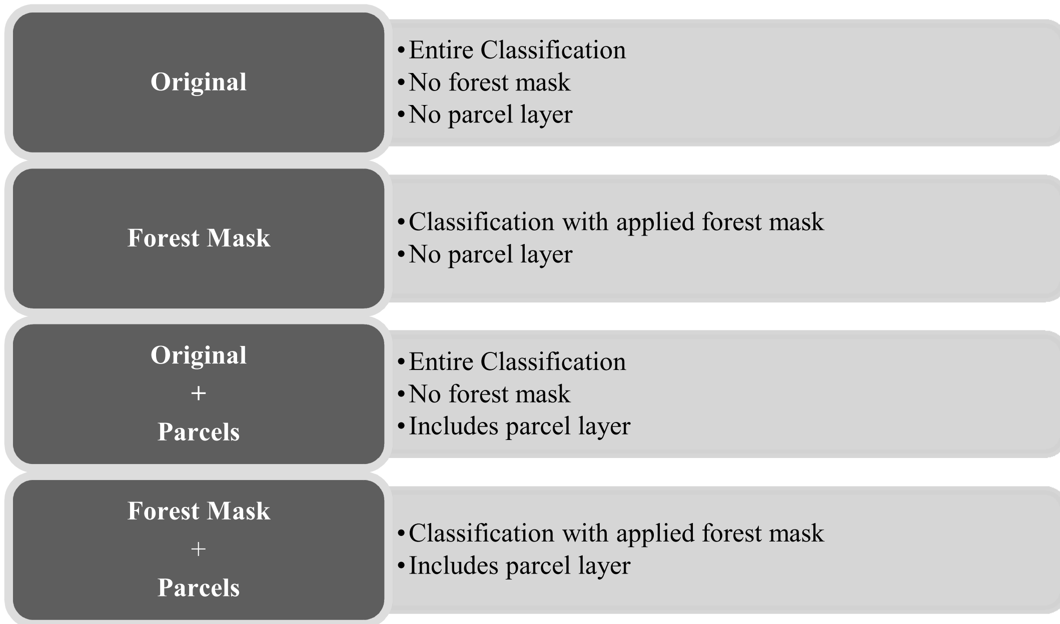


Figure 8: Flowchart showing the four classification assessments to determine the highest building detection accuracy.

CONFUSION MATRIX

The confusion matrix was the validation method used for the SVM algorithm. It is a proven method used to validate image classification outputs by comparing the reference (ground truth) data with the classification results. The diagonal values in the matrix are the correct assessment, while the non-diagonal values are the errors. The confusion matrix also generates a producer's accuracy, a user's accuracy, and a kappa coefficient. The producer's accuracy indicates the errors of commission. These errors occurred when the classification assigned pixels to an area where the pixels did not belong. The user's accuracy is an index for calculating the errors of omission. This error type happened when pixels belonging to one class were included in another (Lewis & Brown, 2001). Lastly, the kappa coefficient compares the classification results to the values assigned by chance (further described below; Lewis & Brown, 2001).

For the 2001 object detection classification, five-hundred points were made for the nine classes (56 points per class). Each point was ground-truthed by zooming into the image, assessing where the point was located, and compared with what/where the algorithm classified the point. The answers were input into the attribute table for each accuracy point. For example, the buildings are class 1, and the roads are class 2. If the point was on top of a road, 2 was input into the attribute table. If it was classified as a building by the model but was a road, this is an error of commission. After the entire five-hundred points were ground-truthed, the kappa score was calculated shown below in the results section. The kappa score is the index of the agreement for the overall accuracy of the classified data against the reference data (Congalton, 1991). Having a kappa score between 0.61 – 0.80 is adequate (Landis & Koch, 1977) since a kappa coefficients equal to 1 means a perfect agreement and a value near 0, would mean an agreement is no better than would be expected by chance (Rwanga & Ndambuki, 2017).

ARCGIS TOOLBOX

The creation of a toolbox with ArcGIS was accomplished using the model builder. The model strings multiple geoprocessing tools together, using the output of one tool to input a different mechanism (ArcGIS Pro Spatial Analysis, 2021). The final model was saved as a complete geoprocessing tool for other interested parties to use and apply the same analysis using relevant data layers.

The complete framework of the Jackson County data layers integrated with aerial imagery of Jackson was analyzed to characterize septic infrastructure at the county level. The GIS septic system location results were derived from the image classification process and input into the algorithm along with the tax parcels, sewer lines, and multiple geoprocessing tools. The outputs were exported into a building layer to detect septic systems' sites and ages further. The building/septic system layer was implemented to develop the automated digitization of the septic system algorithm (model) and combined with other significant layers (sewer, parcels, and roads) for Jackson County. The algorithm began with the 'Reclassify' tool with the classified image as the input to create an output raster layer that consists of only the building's class. The next tool in the algorithm was for converting the building's raster into a feature layer (vector data) with the 'Raster to Polygon' tool. The buildings' feature layer output was then merged with the Jackson County Parcel layer to create one feature layer that showed the tax parcels and buildings together. Next, buffered areas around the county sewer lines were produced with the geoprocessing 'Buffer' tool to implement the 200-foot (61 meters) threshold. The Georgia Department of Public Health determines this distance as an industry standard for the minimum length for OWTS (GA DPH, 2012), although, this threshold tends to vary dependent on the county (personal communication). Once feature layers for buffered sewer lines and buildings/tax parcels were created, the last part of the algorithm was to use the 'Erase' geoprocessing tool. Along with the input of the buffered sewer layer, this tool took the buildings/tax parcels and erased the areas that contain buildings on the sewer. One last step was to parameterize the input layers. The parameterization was added for the end user to insert their specific data. The parameterization inputs include the classified image, reclassification, county tax parcels, and county sewer lines. The final toolbox (model) and the final output was the 'Buildings on the Septic' feature layer, shown under the results. To apply the tool for the entire county, the data inputs were changed to include the whole county instead of the 400 square mile test area. This required the SVM classification and the buffered sewer line geoprocessing tools to be rerun for the entire Jackson County layers.

SEPTIC SYSTEM PERMITS

The validation approach for the location of the buildings was to incorporate septic system permits gathered from the Jackson County Department of Public Health (DPH). Permits for two-hundred subdivisions were collected and manually scanned into a computer to develop a digitized copy of each permit. A comma-separated value (CSV) file was generated using the physical addresses with their latitude and longitude coordinates of each building on a septic system. Using the CSV file and ArcGIS software, the addresses were geolocated from the coordinates of each address (input datum and projection). The buildings were then viewed on a map with the known addresses, which allowed the classified buildings to be validated with a geographic reference for the buildings on septic systems within Jackson County. Furthermore, the septic permits also validated the year the building and septic system was constructed. Sequentially, that yielded information about the ages of the septic system. Lastly, the digitized septic permits were added to the attribute tables of each geolocated building. This meant any person who viewed the septic system map could pull up the address of the building and read a copy of the permit containing all the system's information. The attachment of the permit allowed access if a user needed to know specifics on a particular building or address from the original septic permit.

qPublic.net DATASET

qPublic.net is a user-friendly website that allows the public online access to county and city data. Viewers can browse the website and use the interactive portal to obtain data for public records and GIS layers (Beacon / qpublic.net, 2022). It is also referred to as the local government GIS for the web.

Depending on the data needing to be collected for research will determine the method of acquisition. Since the website is user-friendly, most of the data acquisition is self-explanatory. Although, for this research, finding the year-built required more strategy. The year-built was not as obvious when searching by property, thus, it was determined through trial and error that a search by 'sales' resulted in a year-built for the building. The 'sales search' parameter was set for 'all' sales and 'land with building'. The desired number of years to search was entered into the 'sales date' and 'year built'. The data results were downloaded in an Excel format, edited, and added to ArcGIS as a .csv file. The addresses were shown on the map using the x and y coordinates. The addresses from qPublic were then related to the addresses from the algorithm results and the year built was joined to the attribute table for the property. This method of validation was suitable for certain scanned permits that were missing a year built. Therefore, it was a good back-up validation method for septic system permits/addresses with incomplete records.

2.3 RESULTS

BUILDING CLASSIFICATION ACCURACY OF JACKSON COUNTY 2001

The 2001 and 2020 test area images were prepared using four different methods to obtain the most accurate classification for the building class in the algorithm. It was solely based off the buildings class because the buildings were the highest priority for a high accuracy. The following results further explain the different methods and outcomes for a closer perspective.

ORIGINAL CLASSIFICATION

The 2001 Jackson County original building classification accuracy (forests included) resulted in the lowest accuracy values over all categories for both years. This group had a minimum accuracy of 46%, with the highest accuracy run being 66%, giving an average score of 57% (Figure 9). The main reason for the low accuracy values of building classifications was that the algorithm mistook forested areas for buildings because of similarities in spectral and spatial signatures. Therefore, the method of masking out the forests was utilized.

FOREST MASK CLASSIFICATION

Since the main issue regarding the inaccuracy of the original building classification for the 2001 Jackson County image was the forested areas being misclassified as buildings (due to the high density of trees in the then-rural county), forests were classified prior to running the entire classification and masked from the image leaving only the other eight classes. The 2001 Forest Mask classification resulted in higher accuracy than the original method for 2001, shown below in the boxplot (Figure 9). The forest mask classification received the lowest accuracy of 59% and the highest at 73%, averaging 64% accuracy. This method alone improved the building accuracy for the 2001 classification algorithm by 7%.

ORIGINAL CLASSIFICATION WITH PARCELS LAYER

To further enhance the accuracy of the building's classification, a new step was added to the ground truthing process. The county assesses people's property by parcels of land rather than buildings. Overlaying the GIS parcel layer on the classified image improved the accuracy of the original classification method by nearly 20% (Figure 9), which was a significant development. The algorithm's lowest accuracy out of ten runs was 68%, and the highest at 82%. One last process was necessary to complete the final building accuracy analysis. The forest mask classification previously discussed created a notable difference in accuracy from the original using only the aerial image to ground truth. Hence, it was presumed that the forest mask combined with the parcel layer would also boost the classification's accuracy.

FOREST MASK CLASSIFICATION WITH PARCELS LAYER

The final method for determining the best accuracy for the building's classification algorithm was to use the forest mask 2001 image with the county parcel to ground truth for building detection. This analysis resulted in the highest accuracy of all four methods, not only in 2001 but also in the 2020 classifications. The average score was 81%, with the lowest accuracy at 70% and the highest at 89% (Figure 9). Compared to the original classification sans parcel layer, it was a 15% increase in the overall accuracy mean. This method was only 5% higher in accuracy than the original classification that included the parcels layer, though still a slight overall improvement. It shows that the algorithm would depend on the intended image and whether there is heavy forestation or more urbanization. A comprehensive evaluation of the land cover of the study area would be essential for determining the appropriate method.

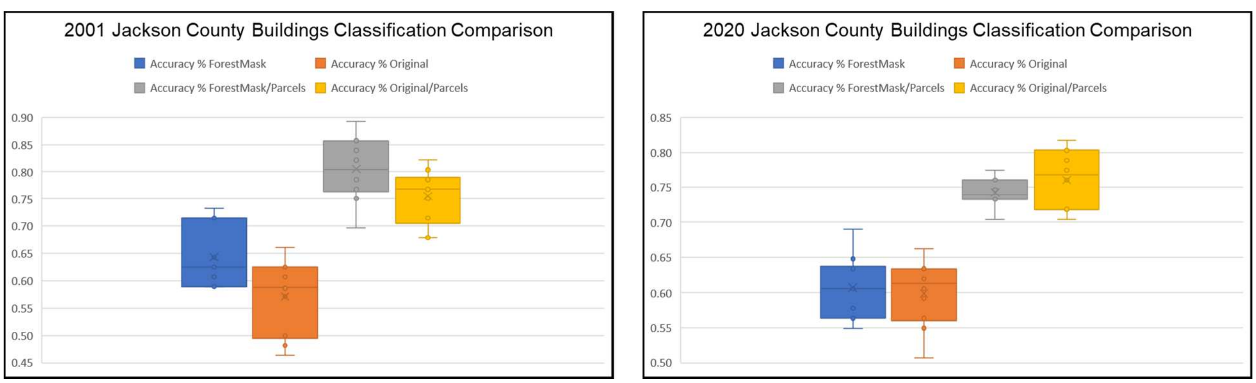


Figure 9: Comparison of boxplots for the 2001 and 2020 test areas of the Classification Methods Building Accuracy

BUILDING CLASSIFICATION ACCURACY OF JACKSON COUNTY 2020

ORIGINAL CLASSIFICATION

The results of the original classification accuracy for the 2020 image of Jackson County were similar yet marginally higher than that of 2001. As the boxplot shows for the 2020 comparison (Figure 9), the average score was 60%, with a low of 51% and a high of 66%. There were outliers as with the 2001 classification, but the range of values was smaller than the 2001 classification. Though the accuracy values were not as high as hoped, it is consistent with the year 2001, which means the algorithm stays consistent even with various remotely sensed imagery. This consistency is a good sign that the toolbox will be compatible with multiple users.

FOREST MASK CLASSIFICATION

The forest mask (FM) assessment for 2020 was narrowly more accurate than the original 2020 assessment. The overall average of 61% was just 1% higher than the initial 2020 assessment (Figure 9). This outcome contrasted with the 2001 accuracies, where the FM had a much more significant percentage of correct classifications. The lowest score of FM 2020 was 55% which is only 4% greater than the Original 2020. This result was unexpected as the overall accuracy values for the 2020 FM are lower than those of the 2001 FM runs. This could, in part, be due to less forested land cover nineteen years after the first assessment.

ORIGINAL CLASSIFICATION WITH PARCELS LAYER

There was a sizeable increase in the accuracy assessment of 2020 Original with Parcels. Compared to the 2020 Original Classification, the average score jumped 16% to a mean accuracy of 76% (Figure 9). The highest run was 82%, and the lowest accuracy did not drop below 70%. Out of the four methods for 2020, the Original Parcels Classification resulted in the highest accuracy mean. Interestingly, it is the same average accuracy as the 2001 Original Parcels Classification emphasizing the compatibility of the algorithm with various temporal and spatial resolutions.

FOREST MASK CLASSIFICATION WITH PARCELS LAYER

The building classification accuracy results from the FM with Parcels stayed in line with the results of the Original with Parcels, albeit marginally lower, with an average accuracy of 74% (Figure 9). The small range in values stands out with this particular assessment, which starts with the smallest at 70% and the largest at 77%. The 2020 FM with Parcels did not perform as well as the 2001 FM with Parcels with a decrease of 7% accuracy. This information could relate to differences in original resolution of the imageries since the 2020 imagery was more confounded by shadows as opposed to forests. The county had more urban land cover and fewer forests than in 2001. Based on further studies of the errors of commission, alternate landcover masks like shadows should be considered. Additionally, adding more training data would help create a higher classification accuracy for the year 2020.

BUILDING CLASSIFICATION ACCURACY OF ENTIRE JACKSON COUNTY 2001

The 2001 entire county was validated using the same method as the test area image. Although, upon observations of the previous methods for building classification, it was apparent that combining the parcel layer resulted in much better accuracy.   Therefore, the 2001 image was validated with only two of the four previously mentioned methods, the original classification with parcels and forest mask classification with parcels.

ORIGINAL CLASSIFICATION WITH PARCELS LAYER

The first method to validate the 2001 entire county was to use the original classification combined with the parcel layer. This combination method was recognized to obtain a higher building detection accuracy than just the original classification alone. Although, the assessment for the entire county (Figure 10) had a lower percentage range compared to the test area, with the lowest accuracy of 50% and the highest accuracy of 73% with an average building detection accuracy of 58%. While ground-truthing the original/parcels classification, it was concluded that the forest (trees) was a major issue for the lower building detection accuracy.

FOREST MASK CLASSIFICATION WITH PARCELS LAYER

Since the forest mask classification combined with parcels layer was a proven method to obtain the highest accuracy for the 2001 test area image, it was necessary to create a forest mask for the 2001 entire county. After completion of the accuracy assessment, the percentage increased significantly with the lowest value at 59%, albeit an outlier, and the highest value at 79%, with a mean accuracy of 71%. The majority of assessments fell within the 70th percentile. The entire county forest mask/parcel validation is on par with the validations performed for the test area classifications for both 2001(81% mean) and 2020 (74% mean). By removing the outlier from the assessment, the mean accuracy would equate to 73%. Issues with the SVM algorithm were a potential contributor to the lower accuracy, which will be discussed in more detail in the discussion section.

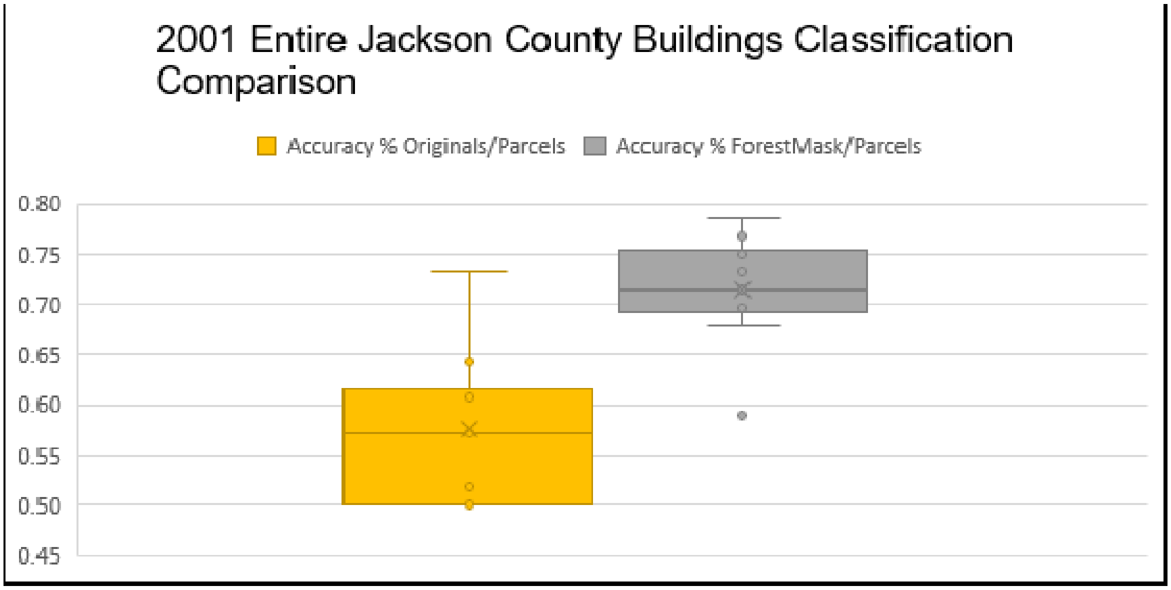


Figure 10: Boxplots for the 2001 Entire County Classification for Two Methods Building Accuracy

CONFUSION MATRICES 2001 & 2020

CONFUSION MATRIX FOR FOREST MASK/PARCELS - 2001

The 2001 Confusion Matrix, which was created using the best run of the building accuracy (89%), was implemented with the method of the Forest Mask combined with the parcels layer. The results were significant, with a kappa score of 92% (Figure 11). The confusion matrix was also computed with the water class excluded. It was selected from the attribute table and deleted to not skew the kappa score of the classification. It became apparent that the water class lowered the overall score. Since it does not affect the building accuracy or the classification otherwise, the best method was to exclude the entire class. On computing the confusion matrix for the 2001 classification, it was found that the forest does not get mistaken for buildings as much as the buildings get mistaken for the forests. The only time the forest is classified as a building is when it picks up the spectral signatures from the shadow (black) of the building as an evergreen forest (black) due to the similar dark color. In future research, it might be better to include a shadow class within the training feature. For this algorithm, the buildings' shadows are included in the building class. The reason for the decision was that the trees created shadows as well. Therefore, it seemed that creating a shadow class would not make a significant difference.

CONFUSION MATRIX FOR ORIGINAL/PARCELS – 2020

The 2020 confusion matrix was created based on the highest building accuracy from all four methods as was done for the 2001 confusion matrix. The one difference being that the best method for obtaining the highest building accuracy was using the complete original test image combined with the parcels layer instead of using the forest mask image. Unexpectedly, the 2020 classification accuracy was not as high as the 2001 having a kappa score of only 78% (Figure 12). As noted previously, there are multiple factors that contributed to the lower accuracies of the 2020 image- 1) higher noise to signal ratio, and 2), shadows creating mixed spectral signatures from beneath the trees and buildings (Katz et.al, 2020). The noise was attempted to be smoothed out by changing the parameters of the segmentation tool, but it did not even out to the effect of the 2001 image.

CONFUSION MATRIX FOR FOREST MASK PARCELS – ENTIRE COUNTY 2001

The Confusion Matrix for 2001 was created for the entire county from the highest accuracy iteration of the building class (79%). The best accuracy of 79% was executed from the Forest Mask/parcels method. In comparison to the test area confusion matrix for 2001, the entire county kappa score of 79% (Figure 16) was less significant. Whereas, compared to the 2020 test area confusion matrix, it was quite similar with a 1% increase in overall accuracy. As with the previous test area confusion matrices, the water class was removed due to the insignificance of the class (minimum pixels) and the fact that it obscured the accuracy of the classification. The roads class was determined to have the lowest user's accuracy score of 63%. The SVM algorithm tended to confuse the roads class with the cultivated and buildings' classes. The algorithm confusion of the buildings class with the road class was due to the similarities in shape, such as mobile homes and chicken houses being long and rectangular, as is a road. After ground-truthing, it was evident that the spectral similarities were the cause of misclassifications between the roads' class and the cultivated class. As mentioned before, the image was taken during the leaf off season (dormancy), therefore, fields and lawns in a dormant phase had comparable spectral signatures as the roads. Although the forest was masked out of the image, there were still areas that contained trees, hence, the need to keep the forests classes in the training sites for the classification. Per usual, the forests and barren classes fared well with high user's accuracy scores.

ClassValue	Buildings	Roads	Barren	Deciduous	Evergreen	Mixed Forest	Cultivated	Total	U_Accuracy	Kappa
Buildings	50	1	1	4	0	0	0	56	0.89	0
Roads	4	43	1	0	0	0	8	56	0.77	0
Barren	0	0	55	0	0	0	1	56	0.98	0
Deciduous	1	0	0	55	0	0	0	56	0.98	0
Evergreen	0	0	1	0	55	0	0	56	0.98	0
Mixed Forest	4	0	0	0	0	52	0	56	0.93	0
Cultivated	0	1	0	1	0	0	53	55	0.96	0
Total	59	45	58	60	55	52	62	391	0.00	0
P_Accuracy	0.85	0.96	0.95	0.92	1.00	1.00	0.85	0	0.93	0
Kappa	0	0	0	0	0	0	0	0	0	0.92

Figure 11: Confusion Matrix table for the 2001 test area classification accuracy assessment using the Forest Mask/Parcels combo with a kappa score of 92% and highlighting the user's accuracy of each class.

ClassValue	Buildings	Roads	Barren	Deciduous	Evergreen	Mixed Forest	Cultivated	Total	U_Accuracy	Kappa
Buildings	58	6	0	0	0	7	0	71	0.82	0
Roads	0	54	5	0	6	3	3	71	0.76	0
Barren	4	9	51	0	0	2	5	71	0.72	0
Deciduous	1	2	0	64	0	0	4	71	0.90	0
Evergreen	4	1	0	0	53	0	13	71	0.75	0
Mixed Forest	0	0	0	0	2	62	7	71	0.87	0
Cultivated	1	0	0	4	5	0	61	71	0.86	0
Total	68	72	56	68	66	74	93	497	0.00	0
P_Accuracy	0.85	0.75	0.91	0.94	0.80	0.84	0.66	0	0.81	0
Kappa	0	0	0	0	0	0	0	0	0	0.78

Figure 12: Confusion Matrix table for the 2020 test area classification accuracy assessment using the Originals/Parcels combo with a kappa score of 78% and highlighting the user's accuracy of each class.

ClassValue	Buildings	Roads	Barren	Deciduous	Evergreen	Mixed Forest	Cultivated	Total	U_Accuracy	Kappa
Buildings	44	3	0	0	0	4	5	56	0.79	0
Roads	8	35	3	0	0	0	10	56	0.63	0
Barren	1	1	53	0	0	0	1	56	0.95	0
Deciduous	0	0	0	49	0	0	7	56	0.88	0
Evergreen	2	0	1	0	42	0	7	52	0.81	0
Mixed Forest	1	0	0	0	0	40	15	56	0.71	0
Cultivated	0	0	0	0	0	1	55	56	0.98	0
Total	56	39	57	49	42	45	100	388	0.00	0
P_Accuracy	0.79	0.90	0.93	1.00	1.00	0.89	0.55	0	0.82	0
Kappa	0	0	0	0	0	0	0	0	0	0.79

Figure 13: Confusion Matrix table for the entire county 2001 classification accuracy assessment using the Forest Mask/Parcels combo with a kappa score of 79% and highlighting the user's accuracy of each class.

ISOLATED REGIONS ACCURACY

Observations of the classifications derived from the individual classes gave a better perspective of how well the algorithm performed. Certain types of classes fared better than others in the accuracy of the algorithm. Division of the classes into similar groups made the classification accuracy easier to visualize and understand. The values also demonstrated the ruralness of Jackson County. The chart below (Figure 14) indicates that forests held most of the land cover for 2001 (54%) and 2020 (50%). In the year 2020, the urban land cover increased by 9%. This result was characterized by a 5% loss of agricultural land and a 4% loss of forests.

The first group of classes being the urban class which consists of buildings and roads. Out of all three grouped classes, urban encompasses the smallest percentage of land cover at 15% in 2001 and 24% in 2020 (Figure 14). The charts below show user's accuracy (Figure 14) for buildings had a better accuracy over the roads class during both years. The roads class was nearly the same accuracy for both years at 77% and 76%, respectively. As Saito and Aoki (2015) observed, buildings and roads have a lot of shape variation and tend to be near trees and other obstructions making it a complex classification. This observation creates a better understanding of the difficulty for creating a high accuracy for the urban class and why the accuracy is slightly less than the forests and agriculture classes.

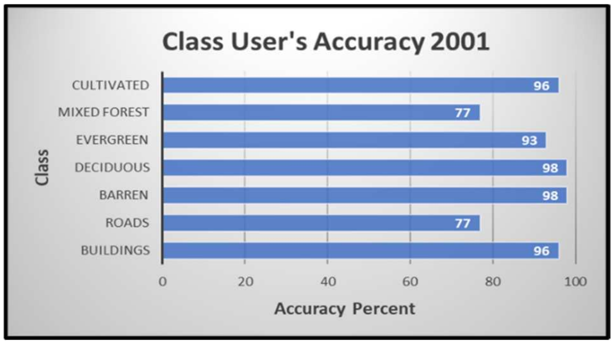
Included within the agriculture group is the cultivated and barren classes. Cultivated for obvious reasons, but this class also includes lawns and fields that may not necessarily be under cultivation. This class group is the second highest land cover at 31% in 2001 and 26% in 2020

(Figure 14). The cultivated class repeatedly had a high accuracy, although, it performed better in the 2001 classification with 96% compared to 86% in 2020 (Figure 14).

Barren land has a high percentage of area that are fallow fields which are used in cultivation. Therefore, this class is included within the agriculture. It is important to note however, that, there may also barren patches where new construction is taking place at the time of image acquisition. In the 2001 graph (Figure 14), it shows that the barren class had excellent accuracy at 98%. Then, the 2020 graph revealed a much lower accuracy of 72% for barren. This happens to be the lowest accuracy over all the land cover classes in both years. The low accuracy in 2020 for the barren class was an unexpected result. This theory was based off the fact that barren land has such a specific spectral signature within this region of the state due to the high red clay content within the soil that the algorithm would easily classify the barren correctly instead of mistaking it for another class. It is possible that the barren may be low in 2020 due to losing a higher spectral and radiometric resolution (8 bit) in turn for having a high spatial resolution (Frison et al., 2016, & Gov. Canada, 2016) giving the image a noisier effect.

The forests class was divided into three types of forests (mixed, deciduous, & evergreen). This method was used instead of one forest class because each type of forest has such a different spectral signature (Figure 15). It is also important to note that both aerial images of the county were taken during the leaf-off seasons, which makes distinguishing between the forest types easier. For both years, the forest held the majority of land cover within the AOI at 54% in 2001 and 50% in 2020 (Figure 16). Analysis of the accuracy assessments does not indicate any pattern of accuracy for the three forest types. The mixed forest received a higher accuracy in

2020, but the evergreen was much lower in 2020 and the deciduous was also less accurate in 2020 than in 2001 (Figure 14).



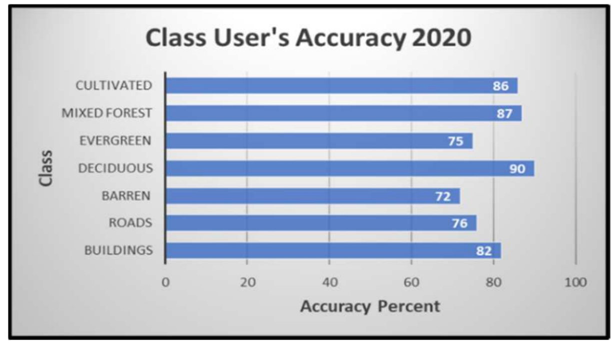


Figure 14: Bar graph showing results of the User's Accuracy (errors of commission) for each class within the training dataset in the building classification algorithm.

Figure 13: Image of 2001 Jackson county showing the difference between the evergreen (1), mixed (2), and deciduous (3) forests in a leaf-off season.

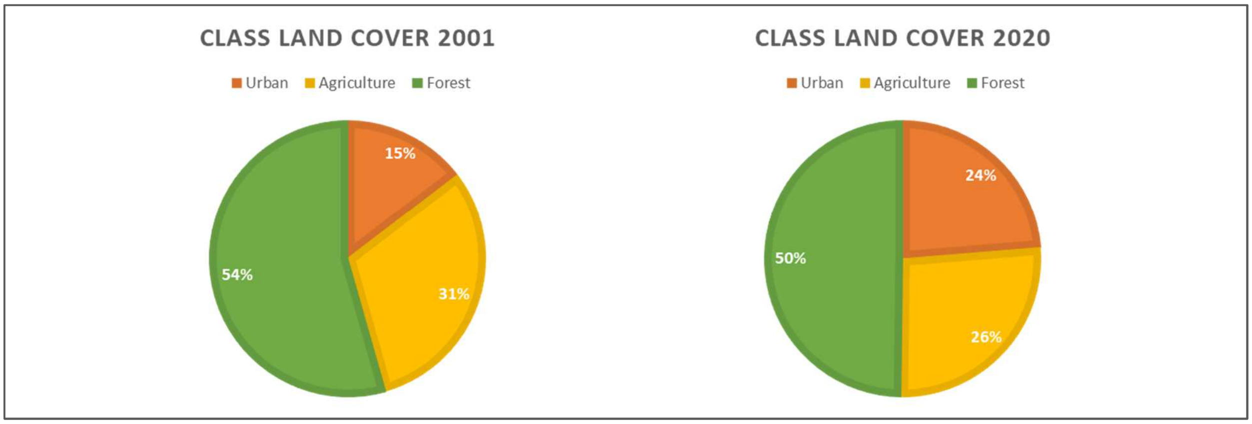


Figure 14: Graphs showing percentage of land cover groups for 2001 and 2020

ARCGIS AUTOMATED TOOLBOX

MODEL

We automated the results in an ArcGIS toolbox to support the need for a low-cost method to populate septic system databases. To run the toolbox as a geoprocessing tool, one only needs to have the appropriate data for the study area, such as a classified image, county tax parcels, county sewer lines, and the minimum distance to a sewer line. Since the algorithm is not 100% accurate, the final output of the toolbox will need to have a quality assurance (QA) assessment to determine the model's inaccuracies. Although, the geoprocessing tool was developed to be user friendly. The Septic System Automated Location Tool (SSALT) (Figure 17,18) eliminates the sewer (centralized) lines from the onsite (decentralized) septic systems and gives the final output of the located buildings (i.e., septic systems).

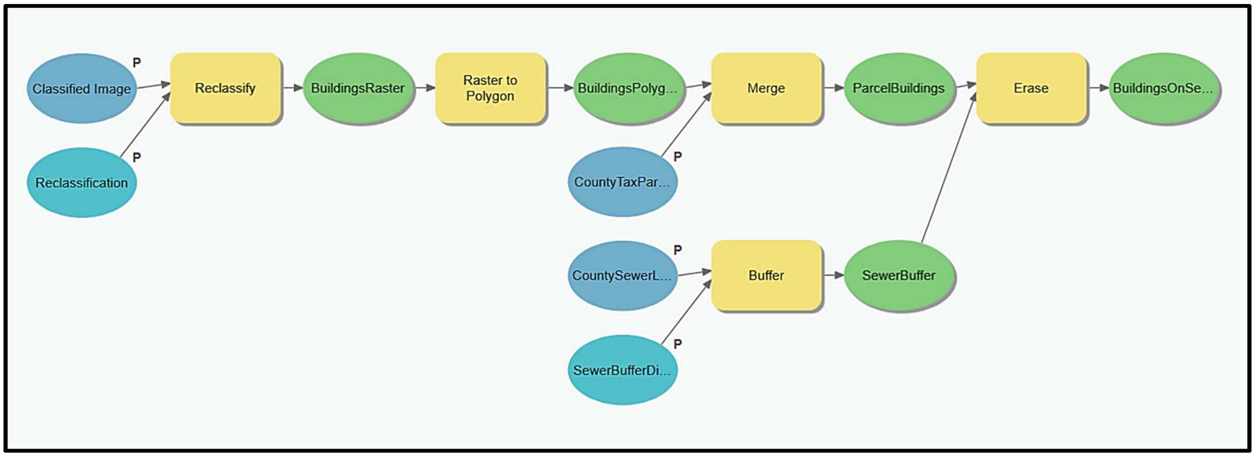


Figure 17: Visual representation of the automated ArcGIS toolbox created for SSALT. The blue ovals indicate the data inputs, the yellow rectangles are the geoprocessing tools, the green ovals are the outputs, and the turquoise ovals are the parameters needed for the specific tool.

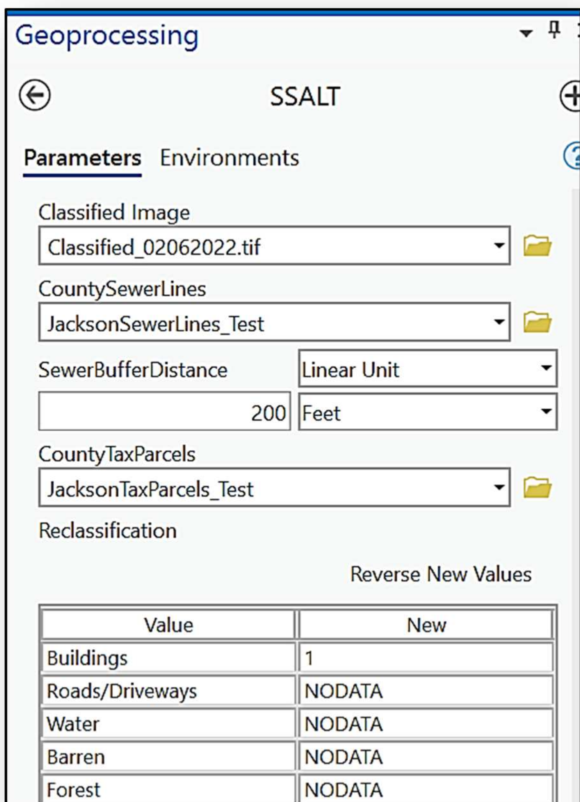


Figure 18: The geoprocessing tool for the automated toolbox (SSALT). This will be the interface seen and run by the users. The parameters will be adjusted to the needs of the user.

MAP OUTPUT OF TOOLBOX

After achieving a classification with higher accuracy of buildings and a satisfactory Kappa score of the 2001 image, the remainder of the analysis was a basic process of elimination. The GIS layers were combined to extract the buildings that belonged to sewers. This involved adding the county tax parcel layer and sewer layer. A buffer of 200 feet (61 meters) was created around the sewer line. This distance was defined by the Georgia Department of Public Health (DPH) as the maximum distance allowed to the sewer line from a building (Ga. DPH, 2016). Below, the final output of the map after the toolbox (SSALT) was run shows the extracted buildings/parcels and buffered sewer lines (Figure 19).

To understand the scope of the results: Within the 40 square miles clipped 2001 raster image of Jackson County (test image), there were 8,914 buildings. Out of the total number of buildings, 4,765 were detected on a septic system, with the remaining 4,149 buildings located on the city sewer lines.

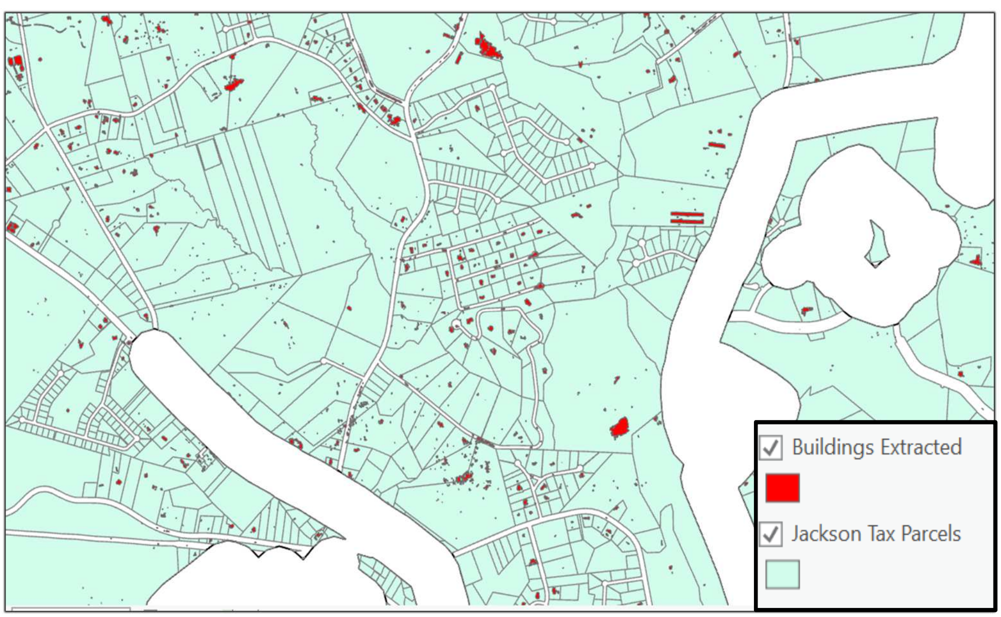


Figure 19: Final output after GIS layers is combined with the classification within the toolbox. In this map, all the white areas are on sewer, and the rest are considered septic.

SSALT was run on the entire 2001 Jackson County classification resulting in the output below (Figure 20). The map shows in detail the output of SSALT with tax parcels designated as septic (orange), tax parcels connected to sewer (green), and parcels that do not contain buildings (gray). The digitized septic system permits were geolocated to validate the results of SSALT to determine whether the parcels designated as septic do, in fact, contain a septic system. The attributes of the digitized permits also include the addresses and ages (permit year) of the systems (Table 1). The results of the final map output verify the total number of parcels on septic is 30,502 and the total number of parcels on sewer is 5,992 out of 36,494 total county tax parcels. As noted previously, the county has a higher land cover of rural, therefore, attributing to the high number of buildings on septic systems.

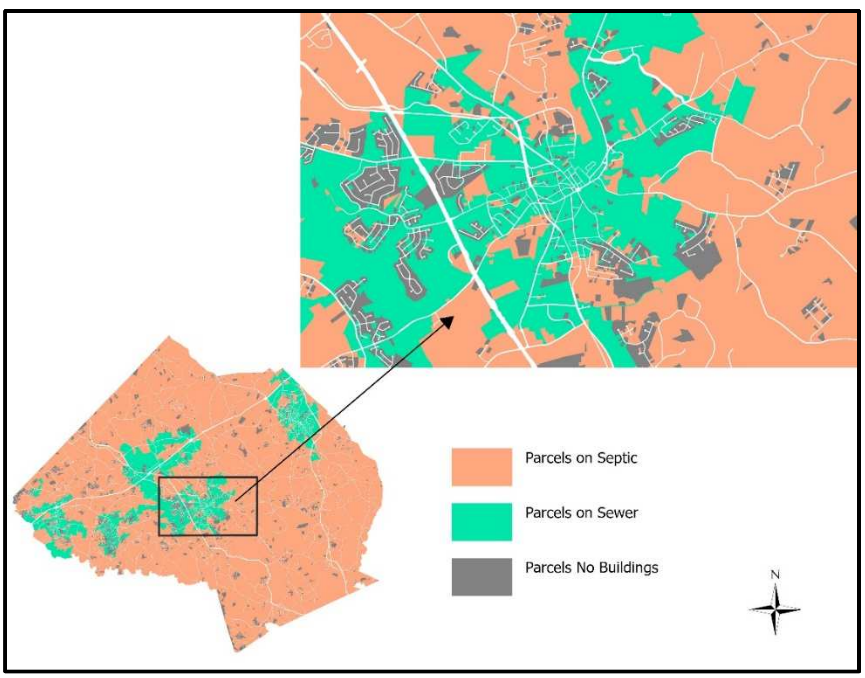


Figure 20: Output of SSALT showing the tax parcels in Jackson County by disposal method. Orange parcels are designated septic, green parcels are designated sewer, and gray parcels are without buildings.

Table 1: Data from points referenced in the map image above (Figure 23) of Academy Church Rd. in Jackson County. The table references the address, the permit year (age of system), and the disposal system (septic) of the address point.

Street Address	City	Zip	State	Permit Year	Disposal	Septic Size	Condition	Soil
1080 Academy Church Rd	Jefferson	30549	GA	9/30/2019	septic tank	1000	new	vance
1221Academy Church Rd	Jefferson	30549	GA	9/1/1999				
1111 Academy Church Rd	Jefferson	30549	GA	5/25/2016	septic tank	1000	new	
1373 Academy Church Rd	Jefferson	30549	GA	5/12/1987	septic tank	1000	new	gravel
1425 Academy Church Rd	Jefferson	30549	GA	1/4/2005	septic tank	1500	new	
1205 Academy Church Rd	Jefferson	30549	GA	4/24/1997	septic tank	1000	1	
1467 Academy Church Rd	Jefferson	30549	GA	4/28/2000	septic tank			wained

Lastly, we used the septic permits to validate the algorithm. There were approximately 1500 scanned septic permits (300 subdivisions) for Jackson County. Out of those permits, 648 were completed with the attributes and then geolocated. The map (Figure 21) shows the tax parcels containing accurate septic permits (green) at 83%, sewer parcels incorrect that contain septic permit records (red) at 7%, and parcels without buildings (gray) at 10%. The sewer parcels that were incorrectly classified could be a sign of a legacy septic system, which means the system was installed before the centralized sewer was in place. For the percentage of parcels that did not contain a building, this could be the result of a misclassified builidng or that the building was built after 2001.

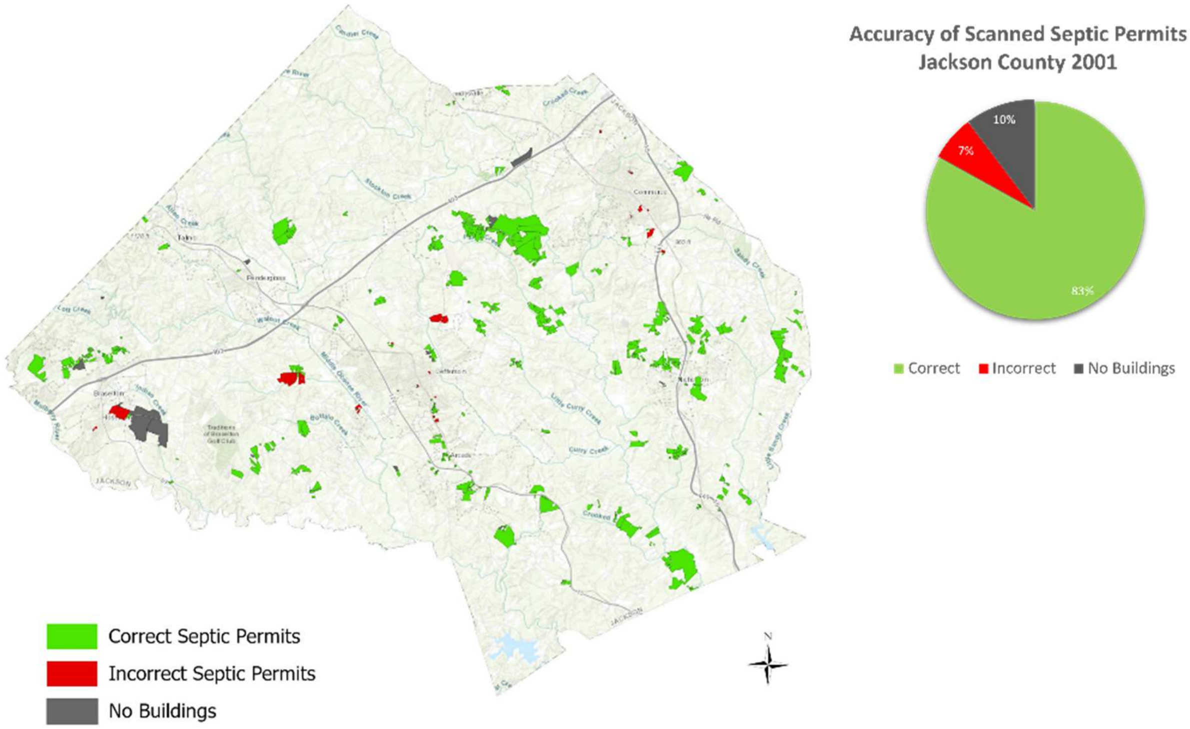


Figure 21: Map showing tax parcels containing accurate septic permits (green), sewer parcels incorrect that contain septic permit records (red) at 7%, and parcels without buildings (gray) at 10%

2.4 DISCUSSION

The objective of this study was to use remote sensing and GIS to develop a low-cost automated tool to identify parcels serviced by septic systems and estimate the age of these systems. Towards this effort, we developed SSALT which is a modular toolbox developed for ArcGIS. The toolbox works based on an image classification of aerial or other imagery to identify buildings locations and combines the results from this classification with GIS layers, namely, sewer line and tax parcel layers for its final output of parcels serviced by septic systems. The validation of SSALT for the 2001 imagery with geocoded septic permits from 200 sub-divisions in Jackson County showed that 83% of parcels were correctly identified to be on a septic system, 7% were incorrectly assigned to septic systems, while 10% of parcels were identified to not contain a building.

The accuracy of the tool was mainly influenced by the accuracy of the building classification but the results of spatial clustering in errors indicated that the decision framework applied for parcel on septic system detection was not sufficient. These results are discussed in detail in the following paragraphs.

The results for the image classification suggest that the classifications were significant, with all three classifications obtaining high kappa scores determined by the ranges of kappa from Rwanga & Ndambuki (2017): the 2001 test area image was 92%, the 2020 test area image was 78%, and the entire 2001 image was 79%. Mapping and land use detection is the most used application of remotely sensed imagery (Nehzak et al., 2022). Our initial classification findings were based on the NLCD land cover/use Anderson Classification system developed in 1976. The NLCD is extensively used by researchers and professionals from diverse fields of study, from fire to environmental management (Chamberlin & Tighe, 2009). Literature suggests that the early NLCD from 1992 – 2001 has excellent medium-scale accuracy but can lack higher accuracy at finer-scale projects (Chamberlin & Tighe, 2009). Therefore, this suggests that part of our study conducted for the 2001 image might not have had as high of accuracy for the land cover classification as needed. The performance of the developed methods throughout the conterminous U.S. for the 2016 NLCD had an overall accuracy assessment for 2016 at 91% accuracy (Dewitz, 2021). 2020 remotely sensed data used to develop the classification had better accuracy for the NLCD at 91% (Dewitz, 2021). Although this study indicates that the 2001 NLCD used for building detection classification was not as accurate as the 2020, our results disagree with this since the 2001 kappa score was better than the 2001 test area image and 2001 entire county image.

We also observed spatial clustering in the parcels incorrectly identified to be on sewer. These clusters were observed in major cities and could be attributed to legacy septic systems in older houses built before the establishment of a sewer line in that area. Such houses may or may not have transitioned to sewer. Such information could be gleaned from inspecting the utility bills of the houses to assess whether or not the house is connected to sewer or is still being serviced by the legacy septic systems. The toolbox could be expanded to flag spatial clusters of errors but would always require manual correction in the output to assign such parcels to sewer or septic.

A limitation of this study was the need for high-spatial resolution imagery. Classifying remotely sensed images of the high spatial resolution containing the three RGB bands using conventional methods can be challenging (Sameen et al., 2018). If the user does not have access to County imagery, a solution could be to source commercial satellites. These images come at a slight financial cost, but there are companies working on significantly reducing the cost of these satellites (Dial et al., 2003).

Future research could include combining LiDAR (Light Detection And Ranging) data to assist machine learning using a height dimension for further distinguishing land cover types (Chen & Gao, 2014, Qiong et al., 2017). This could potentially assist in distinguishing trees from buildings and achieve a higher accuracy (Kowalczuk, 2019). Also, the possibility of multi-spectral data acquisition to enable access to a false color image or NDVI (normalized difference vegetation index). A false-color image or NDVI could be an advantage for training a model to detect buildings by assisting in separating the vegetation from the buildings using the near-infrared band. Creating different parameters for the SVM and segmentation tools for higher building detection accuracy could also be assessed. The challenge is that there are so many possible methods, parameters, and data that it can be a continuous process.

2.5 CONCLUSION

The density of septic systems has often been identified by counties and environmental health agencies as the major factor leading to non-point source pollution in surface and groundwaters. Additionally, improperly sited systems or old systems increase the risk of failure for such systems and can be a direct threat to public health. In order to proactively address this issue, counties need to have a system to identify potentially failing systems (old or improperly sited). The main factor limiting the counties' ability in doing so is the absence of a digitized septic system database. This study presented a low-cost automated tool, Septic System Automated Location Tool (SSALT) to develop such a database. The tool was developed for ArcGIS Pro 2.9 and uses high resolution aerial imagery and GIS datasets. The tool was developed and tested in Jackson county, Georgia and provided an accuracy of over 80%. This tool has been built to accept user specified values for various parameters and can be easily adapted by other counties within and outside Georgia to develop a septic system database.

LITERATURE CITED

Alkhouri, S. (2014). Developing spatial information database for Jericho and Al Auja, West Bank, Palestine. https://doi.org/10.13140/RG.2.1.2426.1606

Beacon. (n.d.). qPublic.net Retrieved November 12 from https://qpublic.schneidercorp.com/

Bureau, U. S. C. (2021). Data Census.Gov. https://data.census.gov/cedsci/map?q=Jackson%20county%20georgia&tid=ACSDT1Y2019.B010&layer=VT_2020_050_00_PY_D1

Capps, K. A., Bateman McDonald, J. M., Gaur, N., & Parsons, R. (2020). Assessing the Socio-Environmental Risk of Onsite Wastewater Treatment Systems to Inform Management Decisions. Environmental Science & Technology, 54(23), 14843-14853. https://doi.org/10.1021/acs.est.0c03909

Chaitali Dhaware, M. K. H. W. (2016). Survey On Image Classification Methods In Image Processing. International Journal of Computer Science Trends and Technology (IJCST), 4(3), 246 - 248.

Cogger, C. G. (1987). Basic Principles of Onsite Sewage.

Conn, K. E., Habteselassie, M. Y., Denene Blackwood, A., & Noble, R. T. (2012). Microbial water quality before and after the repair of a failing onsite wastewater treatment system adjacent to coastal waters. J Appl Microbiol, 112(1), 214-224. https://doi.org/10.1111/j.1365-2672.2011.05183.x

Dewitz, J. (2021). National Land Cover Database (NLCD) 2019 Products. *USGS Digital Object Identifier Catalog*.

El-naggar, A. M. (2018). Determination of optimum segmentation parameter values for extracting building from remote sensing images. *Alexandria Engineering Journal*, *57*(4), 3089-3097. https://doi.org/https://doi.org/10.1016/j.aej.2018.10.001

Esri. (2020). *Classification accuracy assessment. Confusion matrix method. 50 North | GIS blog from Ukraine*. Retrieved December 6 from http://www.50northspatial.org/classification-accuracy-assessment-confusion-matrix method/#:~:text=A%20confusion%20matrix%20(or%20error,result%20and%20a%20reference%20image

Frison, P.-L., Jarlan, L., & Mougin, E. (2016). 3 - Using Satellite Scatterometers to Monitor Continental Surfaces. In N. Baghdadi & M. Zribi (Eds.), Land Surface Remote Sensing in Continental Hydrology (pp. 79-113). Elsevier. https://doi.org/https://doi.org/10.1016/B978-1-78548-104-8.50003-6

Gaye, B., Zhang, D., & Wulamu, A. (2021). Improvement of Support Vector Machine Algorithm in Big Data Background. *Mathematical Problems in Engineering*, *2021*, 5594899. https://doi.org/10.1155/2021/5594899

Georgia Department, o., Public, Health. (2016). RULES OF THE DEPARTMENT OF PUBLIC HEALTH CHAPTER 511-3-1 On-Site Sewage Management Systems. https://casetext.com/regulation/georgia-administrative-code/department-511-rules-of-georgia-department-of-public-health/chapter-511-3-environmental-health-hazards/subject-511-3-1-on-site-sewage-management-systems/rule-511-3-1-03-general-requirements-for-on-site-sewage-management-systems

GISGeography. (2022, August 17, 2022). Aerial Photography vs Orthophotography: What's the Difference? Retrieved 11/18/2022 from https://gisgeography.com/aerial-photography-vs-orthophotography/

Government, Of, & Canada. (2016, February 18, 2016). Radiometric Resolution. Retrieved November 20 from https://www.nrcan.gc.ca/maps-tools-and-publications/satellite-imagery-and-air-photos/tutorial-fundamentals-remote-sensing/satellites-and-sensors/radiometric-resolution/9379

Heumann, B. W. (2011). An Object-Based Classification of Mangroves Using a Hybrid Decision Tree—Support Vector Machine Approach. *Remote Sensing*, *3*(11), 2440-2460. https://www.mdpi.com/2072-4292/3/11/2440

Karathanasis, A. D., & Johnson, D. M. C. (2006). Stability and transportability of biosolid colloids through undisturbed soil monoliths. *Geoderma*, *130*(3), 334-345. https://doi.org/https://doi.org/10.1016/j.geoderma.2005.02.006

Karan, S. K., & Samadder, S. R. (2016). Accuracy of land use change detection using support vector machine and maximum likelihood techniques for open-cast coal mining areas. Environ Monit Assess, 188(8), 486. https://doi.org/10.1007/s10661-016-5494-x

Kecil, J. (2020). *Homeowner behavior change relating to septic system maintenance in coastal South Carolina* University of Michigan].

https://deepblue.lib.umich.edu/bitstream/handle/2027.42/163661/John_Kecil_Practicum.pdf?sequence=1

Khavarian Nehzak, H., Aghaei, M., Mostafazadeh, R., & Rabiei-Dastjerdi, H. (2022). Chapter 5 - Assessment of machine learning algorithms in land use classification. In H. R. Pourghasemi (Ed.), *Computers in Earth and Environmental Sciences* (pp. 97-104). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-323-89861-4.00022-1

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics, 33(1), 159-174. https://doi.org/10.2307/2529310

Lewis, H. G., & Brown, M. (2001). A generalized confusion matrix for assessing area estimates from remotely sensed data. *International Journal of Remote Sensing*, *22*(16), 3223-3235. https://doi.org/10.1080/01431160152558332

Links, T. (1994). Onsite wastewater treatment systems.

Liu, Y., Li, M., Mao, L., Xu, F., & Huang, S. (2006). Review of remotely sensed imagery classification patterns based on object-oriented image analysis. *Chinese Geographical Science*, *16*(3), 282-288. https://doi.org/10.1007/s11769-006-0282-0

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. (2019). Deep learning in remote sensing applications: A meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing, 152, 166-177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

NESC. (2015). Onsite Wastewater Installation Assessment: 2015 Report (Phase 1, Issue. https://www.nesc.wvu.edu/topics-of-interest/assessment-of-u-s-onsite-system-installations-2015-through-2018

NESC. (2020). Onsite Wastewater Installation Assessment:

Phase 2 (2015 – 2018) Report (Phase 2, Issue. https://www.nesc.wvu.edu/topics-of-interest/assessment-of-u-s-onsite-system-installations-2015-through-2018

Perumal, K., & Bhaskaran, R. (2010). Supervised Classification Performance of Multispectral Images. CoRR, abs/1002.4046. http://arxiv.org/abs/1002.4046

Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. *International Journal of Geosciences*, *08*(04), 611-622. https://doi.org/10.4236/ijg.2017.84033

Saito, S., & Aoki, Y. (2015). Building and road detection from large aerial imagery (Vol. 9405). https://doi.org/10.1117/12.2083273

Scholz, M., & Wimmer, T. (2021). A comparison of classification methods across different data complexity scenarios and datasets. *Expert Systems with Applications*, *168*, 114217. https://doi.org/https://doi.org/10.1016/j.eswa.2020.114217

Sun, X. g., Li, M. c., Liu, Y. x., Tan, L., & Liu, W. (2009, 20-22 May 2009). Accelerated segmentation approach with CUDA for high spatial resolution remotely sensed imagery based on improved Mean Shift. 2009 Joint Urban Remote Sensing Event, Tessellations, I. (2020, December 8). ArcGIS Pro: Supervised Classification. Retrieved November 28 from https://www.youtube.com/watch?v=YXfFq9ps7bc

USEPA. (1980). Design Manual of Onsite Wastewater Treatment and Disposal Systems.

USEPA. (2002). Onsite wastewater treatment systems manual. https://cfpub.epa.gov/si/si_public_record_Report.cfm?Lab=NRMRL&dirEntryID=55133

USEPA. (2003). Voluntary National Guidelines for Management of Onsite and Clustered (Decentralized) Wastewater Treatment Systems.

USEPA. (2021). How Your Septic System Works. https://www.epa.gov/septic/how-your-septic-system-works

Walton, B. (2015). Septic System Pollution Contributes to Disease Outbreaks. *WaterNews*. Retrieved 01/15/2023, from https://www.circleofblue.org/2015/world/septic-system-ease-outbreaks/

Withers, P. J., Jordan, P., May, L., Jarvie, H. P., & Deal, N. E. (2014). Do septic tank systems pose a hidden threat to water quality? *Frontiers in Ecology and the Environment*, *12*(2), 123-130. https://doi.org/10.1890/130131

Ye, Q., Xu, Q. H., & Xie, H. H. (2010). Removing shadows from urban aerial images based on

color constancy. Guangdianzi Jiguang/Journal of Optoelectronics Laser, 21, 1706-1712. https://doi.org/10.5194/isprsarchives-XXXIX-B3-525-2012

Zortea, M., Haertel, V., & Clarke, R. (2007). Feature Extraction in Remote Sensing High-Dimensional Image Data. *Geoscience and Remote Sensing Letters, IEEE*, *4*, 107-111. https://doi.org/10.1109/LGRS.2006.886429