IDENTIFICATION OF INLAND FRESHWATER LENSES IN CENTRAL FLORIDA USING REMOTE SENSING AND GEOSTATISTICS

by

YONESHA YASHELLE DONALDSON

(Under the Direction of Adam Milewski)

ABSTRACT

Today, over 2 billion people live in water-stressed or water-scarce countries. One possible way to mitigate water-related issues is through the exploration of untapped groundwater. Howbeit, spatially, not all groundwater is fresh; groundwater can sometimes be either brackish or saline. Here, groundwater in the form of brackish to saline can serve as an opportunity for hydrogeologists to study and further explore freshwater resources. This thesis develops an innovative way to identify the controlling mechanisms, occurrences, and origins of fresh groundwater surrounded by saline water within the Middle St. Johns watershed, Florida USA. In total eight conducive conditions for inland freshwater lens (IFL) formation were used as input parameters for inland freshwater lens potentiality mapping (IFLPM). High IFL potentiality is recognized in the southernmost part of the watershed deducing important formation mechanisms to recharge, confining layer thickness, elevation, precipitation, and lithology in the coastal aquifer system.

INDEX WORDS: hydrogeology; water resource; remote sensing; groundwater potential mapping

IDENTIFICATION OF INLAND FRESHWATER LENSES IN CENTRAL FLORIDA USING REMOTE SENSING AND GEOSTATISTICS

by

YONESHA YASHELLE DONALDSON

BS, Rutgers University-Newark, 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

2020

IDENTIFICATION OF INLAND FRESHWATER LENSES IN CENTRAL FLORIDA USING

REMOTE SENSING AND GEOSTATISTICS

by

YONESHA YASHELLE DONALDSON

Major Professor:	Adam Milewski
Committee:	Charlotte Garing
	Robert Hawman
	Lee Slater

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
August 2020

DEDICATION

I would like to dedicate this thesis to my parents, Jennifer Harris and Neville Donaldson. Thank you for always believing in me and trusting me to take my own paths. My love for you both is infinite.

ACKNOWLEDGEMENTS

Foremost, I would like to thank God, who is the head of my life. Everything that I am and have is because of you. Thank You. To my advisor, Dr. Adam Milewski, thank you for supporting my ideas and introducing me to new concepts to locate groundwater. I am a better scientist because of it. I know now that there is so much in store for me and water exploration. Thank you. Dr. Slater, you introduced me to the concept of hydrogeology and geophysics earlier on in my academic career, you invested your time and devoted your lab to me so that I can learn about the different geophysical techniques and their applications. You have opened up my mind to a world of ideas and what I want to do for the rest of my life. Thank you. Dr. Garing, thank you for always being a phone call away if I ever needed your advice in making my research better! Dr. Hawman, thank you for sharing your knowledge and expertise in geophysics and for adding a great component to this study.

Last, but not least, my lab mates! Thank you all for the laughs throughout the years. It was great learning and growing with you all. One Love.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES viii

LIST OF FIGURES ix

CHAPTER

1 INTRODUCTION 1

2 LITERATURE REVIEW 6

2.1 GIS and Remote Sensing 8

2.2. Random Forest (RF) machine learning algorithm 10

2.3. Geophysical signatures in variable-density groundwater 121

3 STUDY SITE 17

3.1. Geologic History 17

3.2. Land Relief 18

3.3. Hydrogeology and Aquifers Properties 19

3.4. Regional Climate 20

3.5. Middle St. Johns River watershed 21

4 RESEARCH METHODOLOGY 27

4.1. Objectives 27

4.2. Input Datasets and Pre-Processing 32

4.3 Description of models 44

4.4. Electrical Resistivity Tomography (ERT): Variable Density Forward Modeling 52

5 RESULTS AND DISCUSSION 57

5.1. WIOA GIS data-driven model 57

5.2. Randon Forest model 64

5.3. Electrical Resistivity 70

5.4. Future work 76

6 CONCLUSION 77

REFERENCES 80

LIST OF TABLES

Table 1. Weightage and score values for the thematic layers with inland freshwater lens development description in a coastal setting. 30

Table 2. Salinity zones and classes as it relates to total dissolved solids and electrical conductivities if the relationship is linear. Values adapted from (Williams and Kuniansky, 2019; Rhoades, 1992). 31

Table 3. Result summary of the OLS and GWR models. The results from the IFL potential zone with the lowest Akaike Information Criterion (AIC_c) were used to produce the predictive maps. 63

Table 4. OLS independent variable p-value results from the IFL potential map datasets............ 63

LIST OF FIGURES

Figure 1. The occurrence of global saline or brackish groundwater (TDS>1,000 mg/L), less than 500-meters below ground level. Image source: Werner and Laattoe, 2016. 15

Figure 2. Topographically induced depressions identified in Kuwait. Primary depressions are suggestive areas of untapped groundwater within the Raudhatain watershed............................. 16

Figure 3. Geographical map of the state of Florida, USA. 23

Figure 4. Thickness range of the Intermediate Aquifer System (IAS) across the state of Florida. 24

Figure 5. Study site map. The Middle St. Johns watershed located in central Florida................. 25

Figure 6. The areal extent of the Geneva lens located within the Middle St. Johns watershed, juxtapose Lake Jesup and Lake Harney... 26

Figure 7. Flow chart illustrating steps used to generate inland freshwater lenses potential zones. 29

Figure 8. High-resolution 30-year average (1981-2010) precipitation data. 36

Figure 9. Digital elevation model (DEM) showing elevations in feet... 37

Figure 10. Interpolated total dissolved solids (TDS) concentration, milligrams per liter for the top 50 feet of the Floridan aquifer system for the top 50 feet of the Floridan aquifer system. 38

Figure 11. Variable confinements of the hydrostratigraphic unit of the intermediate aquifer system. 39

Figure 12. Recharge zones classes to the upper Floridan aquifer.. 40

Figure 13. National land cover database 2011 (NLCD2011) 41

Figure 14. Unit age for each geological feature. 42

Figure 15. Interpolated transmissivity based on interpolation of ~1500 aquifer test of the Upper Floridan Aquifer, in units of feet squared per day. 43

Figure 16. Weight Overlay Index Analysis sensitivity test. 46

Figure 17. Random forest (RF) model training (70%-red) and validating (30%-green) data points. IFL formation factors are assigned to each data point. 50

Figure 18. Pre-modeling workflow done within ResIPy software. 55

Figure 19. Geological cross section of the study area along line B to B'. 56

Figure 20. GIS-data driven model output for IFL potential zones within the Middle St. Johns watershed. Areas in green marks areas for low, orange for moderate, and pink for high potential zones. 60

Figure 21. Left to right: GIS model observation and OLS geostatistical predictions. 61

Figure 22. Relationship between GIS model observations and OLS geostatistical predictions. 61

Figure 23. Left to right: GIS model observations and GWR geostatistical predictions. 62

Figure 24. Relationship between GIS model observation and GWR geostatistical prediction. 62

Figure 25. Optimization number of trees bases on OOB estimates of the error rate in the RF model. 66

Figure 26. Variable importance derived from the RF model. 67

Figure 27. ROC curve: (a) successive rate (b) predictive rate for RF model. 68

Figure 28. Inland freshwater lens potential map produced by Random Forest model using 30% testing dataset to make predictions. 69

Figure 29. Relationship between GIS model observation and machine learning algorithm, RF predictions. 69

Figure 30. Water class boundaries defined within the geoelectric model. Unsaturated zone not included (water table very close to the surface).. 72

Figure 31. Modeled resistivity values assigned to water classes as they relate to a one-dimensional and homogenous Earth. The top image shows the model used for the dipole-dipole array and the bottom image shows the model used for the Wenner array. 73

Figure 32. Inverted true resistivity values for water classes for dipole-dipole (top) and Wenner array (bottom). 74

Figure 33. Wenner configuration on part of the Geneva lens (model is not drawn to scale). 75

CHAPTER 1

INTRODUCTION

In coastal settings, freshwater scarcity is well known (Yao *et al.*, 2019) and is becoming a threat to the sustainable development of human society. With an increase in the local and global population, the dependency on groundwater for human and ecosystem survival will continue to increase. Populations living in coastal areas depend on groundwater for several reasons (Cellone, Tosi, and Carol, 2018). According to Mekonnen and Hoekstra (2016), the main driving forces of freshwater scarcity on a global level are an increasing world population, improving living standards, changing consumption patterns, and expansion of irrigated agriculture. Due to spatial and temporal variations of water demands, two-thirds of the global population (4 billion people) currently face severe water scarcity at least 1 month of the year (Mekonnen and Hoekstra 2016). Additionally, Vörösmarty *et al.*, 2010 studies projected that nearly 80% of the world's population is at high-level threat to freshwater scarcity because of human activities and the effects of climate change. Furthermore, there is an increasing need to balance all of the competing commercial demands on water resources so that communities have enough for their needs, especially since the economic growth in the commercial sector of the economy has far-exceed that in the residential and industrial sectors (Kim and McCuen, 1979).

Groundwater is a key source of drinking water that is essential to life here on Earth, supplying approximately 60% of the world's freshwater demands (Fetter, 2001). Groundwater is described as water in the saturated zone that fills the cracks or pore spaces of rock and mineral grains below ground level. Groundwater exists in many geological formations and is more reliable than surface water because it is naturally protected from direct contamination by rock and sediment layers acting as natural filters below the water table. Surface waters are more prone to seasonal variation, pollution, and anthropogenic activities, making groundwater mapping preferable to fulfill water supply needs. Although, groundwater serves as a significant freshwater supply for coastal regions, spatially, not all groundwater is fresh; groundwater can either be fresh (Total Dissolved Solids (TDS) < 1,000 ppm), brackish (TDS ranging between 1,000- 10,000 ppm), saline (TDS ranging between 10,000-35,000 ppm), or hypersaline (TDS > 35,000). Saline or hypersaline groundwater, here, serves as an opportunity for hydrogeologists to further study and explore freshwater resources to meet future water demands. It is the presence of saline and hypersaline groundwater on local and/or regional scales that typically salinize infiltrating freshwater forming a convex lens of fresh groundwater overlying the denser saline groundwater (Underwood, Peterson and Voss, 1992). The size and shape of the freshwater lenses are controlled by the geologic framework (e.g. porosity, permeability, hydraulic conductivity) and hydrodynamic processes (Schneider and Kruse, 2003). Identifying potential water resources in the form of freshwater lenses in coastal settings can be valuable to the growing population and future generations.

Freshwater lenses are common in coastal or island settings, where saline groundwater is derived from the sea (Werner and Laattoe, 2016; Wallis, Vacher and Stewart, 1991; Bugg and Lloyd, 1976; Stoeckl and Houben, 2012), but can also be found in inland or terrestrial settings due to the presence of remnant marine water and/or dissolution of evaporite deposits (Houben *et al.*, 2014; Rotz and Milewski, 2019; Milewski *et al.*, 2014; Laattoe *et al.*, 2017). There has been intensive amount of studies done to explore oceanic lens (Fetter 1972; Vacher 1988; Bear et al. 2010; Werner et al. 2012), but the current understanding of inland freshwater lenses (IFLs) is still limited. Recent discoveries of IFLs may be more widespread than previously thought (Houben *et al.*, 2014; Werner and Laattoe, 2017; Chongo *et al.*, 2015). One possible way to explore more of these unique valuable groundwater bodies is the identification of specific zones containing IFLs (e.g. Groundwater Potential Mapping (GPM)). GPM involves the exploitation of advanced remote sensing technologies based on data acquired from Geographic Information System (GIS). Most recently, machine learning models (e.g. boosted regression tree, classification and regression tree, random forest, etc.) have been integrating into GIS to produce groundwater potential maps (Naghibi and Pourghasemi, 2015). Application of these methods to the prediction of IFL potential zones is relatively new. Therefore, this research will create a simple GIS-based model and a random forest (RF) machine learning model integrated into the GIS-interface to produce an inland freshwater lens potential mapping (IFLPM) in a coastal aquifer system.

In the United States, Florida is the seventh largest rapidly growing population state and is expected to continue growing in the future (Smith, 2005). The Florida Peninsula is composed of thick marine limestones and dolomites deposited during the Cretaceous (~ 138 to 63 m.y. ago) and Tertiary (~ 65 to 2.58 m.y. ago) geologic periods. Topographic features are derived from shoreline features that formed during the regression and transgression of the sea 2 m.y. ago (Phelps and Rohrer, 1987). Within the Middle St. Johns watershed, in the northeast Seminole county, there is a 15 mi^2 isolated recharge area of the Floridan aquifer system that has formed a freshwater lens surrounded by saltwater. The Geneva “Bubble” lens, described as a true natural wonder, is roughly 22 square miles in diameter, 350 feet thick in the center, and enclosed by a 25 feet land surface altitude (Phelps & Rohrer, 1987). Water coming from this lens supplies drinking water for roughly 5,000 residents in Seminole county. There is considerable interest in additional development of the groundwater resources to serve the expanding needs of Geneva and its surrounding areas because excessive freshwater withdrawal could induce the invasion of brackish water into freshwater production wells due to the limited size and recharge rate of the freshwater lens (Panday *et al.*, 1993). Laattoe *et al.*, 2017 conducted a study recognizing the different formation mechanisms that facilitated known IFLs globally. The authors deduced these formation mechanisms associated with ephemeral surface water bodies, continuously losing perennial surface water bodies, continuously gaining perennial surface water bodies, oceanic island analogue, focused rainfall recharge, and anthropogenic effects. Unlike these formation mechanisms, the Geneva lens is a result of its host aquifer—The Floridan aquifer system (FAS). The downward hydraulic gradient from the surficial aquifer to the Floridan aquifer system and the absence of thick clay layers have allowed local freshwater to flush out the connate saltwater from sediments (Panday et al., 1993; Phelps and Rohrer, 1987), forming this unique water body. This lens is recharged by infiltration through surficial aquifer through a region of increased vertical permeability.

The major goal of this thesis is to identify favorable locations for inland freshwater lenses within the Middle St. Johns watershed using groundwater potential mapping. The objectives of this research are as followed:

I. Create two predictive-based models to identify inland variable-density groundwater flow conditions in a highly urbanized watershed in a coastal aquifer.
    a. Models:
        - Simple Weight Overlay Index Analysis (WIOA) GIS model
        - Random Forest (RF) machine learning algorithm

II. Apply spatial statistical models to quantify model performance and the impact each controlling factor has on IFLs formation mechanisms (e.g., ordinary least square (OLS), geographically weighted regression (GWR), and area under the receiving operating characteristic curve (AUROC)).

To achieve these objectives, a total of eight groundwater conditions that affect the IFL occurrence (e.g. elevation, confining layer thickness, lithology, land use and land cover, recharge mechanism, salinity, precipitation, and transmissivity) in Geneva were used as input parameters in all models. Furthermore, a comparative and statistical analysis of each IFL potential location was carried out in addition to identifying IFL formation parameter sensitivity. Finally, an intuitive software for complex geoelectrical inversion model was used to demonstrate electrical resistivity (ER) being a good tool to identify these untapped groundwaters in the field by demonstrating its ability to see a transition between fresh and saltwater. While IFLs cannot solve the problem of future water demands alone, they do provide drinking and agricultural water to indigenous people in remote areas and already serve as strategic and emergency water resources in many arid and hyper-arid lands (Kwarteng *et al.*, 2000; Young *et al.*, 2004).

The following chapters will contain a literature review of inland freshwater lenses, as well as the differences in formation factor between oceanic and inland freshwater lens in coastal zones. Geologic history, land relief, hydrogeology and aquifer properties, and regional climate information in Seminole county where known IFL exists are inferred. In the methodology chapter, specific details about the derivation of input parameters to the processing involved in deriving the final groundwater potential map. Finally, I report the results of model outputs, discuss those results and sensitivities, and conclude with remarks and future recommendations for this research.

CHAPTER 2

LITERATURE REVIEW

Inland, saline groundwater occurs at shallow to intermediate depths due to the presence of remnant marine water, dissolution of evaporate deposits, and/or evapotranspiration of rainwater (Laattoe et al., 2017; Van Weert, 2009). Therefore, freshwater lens development is possible inland given the proper geological, geomorphological, and hydrogeological settings that support this mechanism. Figure 1 identifies saline groundwater occurrences globally as wells as its derivative at these shallow and intermediate depths (<500 m). Florida saline water at such shallow depth is the result of dense connate water (Van Weert, 2009). Although IFLs are less common than oceanic lenses, several scientific publications describe the existence of IFLs throughout the world. For instance, a combination of geophysical surveys (Chongo *et al.*, 2015; Viezzoli, Auken and Munday, 2009; Barrett *et al.*, 2002), geochemical analyses (Langford, Rose and White, 2009), numerical modeling (Bauer *et al.*, 2006; Rotz, Milewski and Rasmussen, 2020) and physical modeling (Rotz and Milewski, 2019) have been employed to understand existing IFLs morphology (i.e. length, width, saltwater interface). As agriculture and population demands increase, identifying settings conducive for these unique groundwater phenomena is imperative. Thus far, hydrogeologists have discovered IFLs in Paraguay (Houben *et al.*, 2014), Australia (Werner and Laattoe, 2016), Zambia (Chongo *et al.*, 2015), Botswana (Bauer *et al.*, 2016), Kuwait (Parson's Corporation, 1961; Milewski *et al.*, 2014), New Mexico (Langford, Rose and White, 2009), Florida (Panday *et al.*, 1993), and Pakistan (Asghar et al., 2002) using techniques previously stated.

However, occurrences, origins, and controlling factor of IFLs have not yet been attempted in a coastal setting.

One-quarter of the global population lives in coastal regions, and continuously undergo an increase in population growth. Over pumping to meet these demands can lead to upconing, a resultant of small head gradients, high groundwater abstraction rates, and/or drain management of the landscape (Meyer, Engesgaard and Sonnenborg, 2019). Insight into the occurrence and sustainability of IFLs encourages water resources exploration and development in regions previously thought to contain no prospective new resources as well as information for water resource managers interested in water quality preservation and artificial aquifer recharge in the same regions. Previous studies done in the northeast part of Seminole Country, Florida, discovered fresh groundwater overlying brackish water in the Floridan aquifer. First studied in 1962, the Geneva lens formation was said to be related to recharge, piezometric head distribution, and land surface elevation (Barraclough, 1962). Post-Miocene to recent age sands formed an unconfined surficial aquifer, separated from the upper confining unit of the Floridan aquifer. Sedimentation and clay content vary with elevation in this area. On a topographic high, the area surrounding the Geneva lens has a thin clay layer which increases the vertical permeability, whereas, low lying areas have thicker clay units. A three-dimensional, density-dependent transport model shows that the lens is influenced by the saline water that discharges from the upper Floridan aquifer that the freshwater floats atop of (Panday *et al.*, 1993).

Drilling, hydrogeological tests, and geophysical models are commonly used for mapping groundwater potential zones. Although effective, these methods can be time consuming and expensive. Satellite remote sensing is a viable source of observations of land surface hydrologic fluxes and state variables, particularly in regions where *in-situ* networks are sparse. Over the last

10 years, the study of hydrology using remote sensing techniques has advanced greatly with the launch of satellite platforms and the development of more sophisticated retrieval algorithms. Precipitation, evapotranspiration, snow and ice, soil moisture, and terrestrial water storage variations are some variables in the land surface water that are now observable at varying spatial and temporal resolutions and accuracy via remote sensing (Tang *et al.*, 2009). Together, GIS and remote sensing technologies have been used as spatial analyst tools in investigations spanning environmental (Hinton, 1996), natural hazards (Temesgen, Mohammed and Korme, 2001), and hydrologic studies (Elewa and Qaddah, 2011; Elbeih, 2015).

2.1. GIS and Remote Sensing

For many decades, the occurrence of groundwater has been studied using aerial photo interpretation and geophysical techniques, but computer-based analysis of remote sensing (RS) data and geographic information system (GIS) has rarely been done (Elewa and Qaddah, 2011). Using an integration of different research tools and techniques, such as RS, GIS, geostatistical-based predictive models (e.g. ordinary least square, geographically weighted regression), and random forest machine learning technique, areas of potential untapped groundwater (IFLs) have been identified spatially for the first time in a coastal setting. Spatial representation of data used is critical to ground water-potential mapping, and GIS possesses the predictive and related analytical capabilities necessary to examine the complex problems of uncertainty. Still, there are uncertainties with each model regarding the minimum necessary validation levels that would ensure good potential mapping results (Elewa and Qaddah, 2011).

Groundwater conditioning factors that affect groundwater storage can be retrieved via remote sensing. For instance, A. Y. Kwarteng *et al.*, 2000 used aerial photographs, Landsat Thematic Mapper (TM) images, and Digital Elevation Models (DEM) for mapping paleo drainage patterns, large depressions, playas, and catchment areas. Madani and Niyazi (2015) defined lithology, rainfall, lineament density, drainage density, slope steepness, and land use/land cover hydrogeological parameters to relate to the groundwater. Elbeih (2015) considered lithology, stream network, lineament density, slope, drainage networks, and aquifer thickness for groundwater mapping in Egypt. For IFLs, Milewski *et al.*, 2014, identified settings conducive for topographically induced and focus recharged IFL formations in hyper-arid environment in northern Kuwait using remote sensing. By understanding the formation factors that facilitated the development of three existing IFLs within the Raudhatain watershed, the authors identified field (e.g. high infiltration capacities, moderate precipitation (~120 mm/day), and the presence of saline groundwater) and satellite base observation (e.g. DEM, AMSR-E, TRMM, Landsat TM) to map a total of 40 drainage depression for the potential development of freshwater lenses, where ~20 potential lenses where identified in the Raudhatain watershed basin alone (Figure 2). Often, there is little to no surface water present in such environments (Kwarteng et al. 2000; Laattoe et al. 2017). Therefore, many arid environments such as the Middle East depend solely on groundwater and desalinization for its natural freshwater resources (AlAli, 2008). Unfortunately, most aquifers within the Arabian Peninsula are saline due to the presence of underlying evaporitic deposits (e.g., Rus Formation). Therefore, freshwater in the subsurface, if present, is generally in the form of inland freshwater lenses (IFLs) (Kwarteng et al., 2000; Laattoe et al., 2017; Milewski et al. 2014; Rotz and Milewski, 2019). Similar to Kuwait, other states in the Arabian Gulf (e.g. Saudi Arabia, Bahrain, Qatar, Oman, and the

United Arab Emirates) are facing freshwater resource problems (Chafetz, McIntosh and Rush, 1988; Fallatah *et al.*, 2019) and rely on the presence of inland freshwater lenses groundwater to supply those needs. Even though the development of the IFLs in the Arabian Peninsula is a result of focused recharge in a topographic depression and defers from the development of the IFL located in Florida, the use of the technique is possible. For instance, a study was done in Mehran Region, Iran to explore the countries' groundwater potential zones. Water scarcity is the most limiting factor for the country's most economic section, agriculture. As a result, a groundwater potential map (GPM) was produced using GIS and other machine learning algorithms (e.g. random Forest and Maximum Entropy). To train and test the algorithms, the authors used groundwater data with high potential yield values. Additionally, groundwater conditioning factors generated from remote sensing data (DEM, Landsat Enhanced Thematic Mapper Plus) were used as input parameters for each model (Rahmati, Pourghasemi and Melesse, 2016). Moreover, satellite remote sensing datasets are not always readily available in these areas to create such models to access occurrences, origin, and controlling factors, therefore, this research will focus around an area where satellite data and imagery are more readily available.

2.2. Random Forest (RF) machine learning algorithm

GIS and machine learning algorithms are two new technologies used for hydrological mapping (Sameen, Pradhan and Lee, 2019). The machine learning methods that researchers have used in recent years include boosted regression tree (BRT), support vector machine (SVM), artificial neural networks (ANN), decision trees (DT), classification and regression (CART), general linear model (GLM), and random forest (RF) algorithms across multi-disciplinary fields

(Naghibi, Ahmadi and Daneshi, 2017). Compared to the mentioned machine learning methods, RF success rates and predictive rates outperforms or perform just as well as other predictive algorithms in multiple studies. RF can handle data from various measurement scales and makes no statistical assumptions, and is a useful tool for groundwater mapping, overcoming the limitation of artificial neural network (e.g. overfitting), assess important groundwater conditioning factor, and identifying the most important factors and reduce dimensionality (Rahmati, Pourghasemi and Melesse, 2015). The more the number of features, the more the chances of overfitting. Avoiding overfitting is a major motivation for performing dimensionality reduction. By reducing dimensionality, the features are not dependent on the data it was trained on and in turn results in good model performance on the data. The fewer features our training data has, the lesser assumptions our model makes.

RF has been widely used for environmental (Rodriguez-Galiano *et al.*, 2012), ecological modeling (Oliveira *et al.*, 2012), and other disciplines. For groundwater mapping, RF is fairly new. RF can be used for both regression and classification tasks. RF, however, is better at classification than it is for regression. Significant improvements in classification accuracy have resulted from growing an ensemble of trees and letting them vote for the most popular class (Breiman, 2001). In the case of regression, it does not predict outside the range of the trained data and may overfit data that may be too noisy, and lastly, we have very little control over what the model does. Compared to the overall performance of other machine learning algorithms (e.g. Support Vector Machine, Artificial Neural Network Maximum entropy), RF is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001).

2.3. Geophysical signatures in variable-density groundwater

Electrical resistivity (ER) quantifies how strongly a material opposes the flow of an electric current. The resistance of a material depends on what the material is made up of the shape of the material. Resistance is quantified in the following way: one ohm of resistance allows a current (I) of one ampere to flow when one volt (V) of electromotive force is applied. Ohm's law, equation (1), states that current is directly proportional to voltage and inversely proportional to resistance. Resistance not only depends on the material, but also the resistances of the resistor's length, cross-sectional area, and the material fundamental properties, resistivity denoted by ρ, equation (2). High resistivity indicates that the current does not readily go through the matrix, whereas, low resistivity indicates current goes through the matrix easier. The use of Electrical resistivity is suitable for groundwater exploration because there is a direct link between electrical methods and rock properties (Archie, 1942).

$$I = \frac{V}{R} \qquad \textit{Equation 1}$$

$$R = \rho \frac{L}{A} \qquad \textit{Equation 2}$$

In the subsurface, ER is controlled by the solid matrix of the soil and rock and the fluid contained in the pore space. Using Archie's empirical equation, the relationship between resistivity and porosity can be achieved (Archie, 1942):

$$F = \varphi^{-m} \qquad \textit{Equation 3}$$

where φ*, m, and F* represents the porosity, cementation factor, and formation factor, respectively. The formation factor is the ratio between the resistivity of the saturated rock, ρ_r, and the resistivity of the pore fluid, ρ_f. In a carbonate aquifer, the electrical resistivity for a given porosity can vary depending on the cementation factor. The cementation factor *m*, is strongly dependent on the shape of the grains and pores, type of grains and pores, specific surface area, tortuosity, and anisotropy (Garing *et al.*, 2014). The shape of grains and pores is a result of the degree of consolidation. The type of grains (lithological and mineralogical composition) can affect the surface conductance depending on whether specific clay mineral is present. Pore shapes can be intergranular, intercrystalline, or fractured. While the increase in fracturing decreases the value of *m*, an increase in surface area increases the value of *m* (Salem and Chilingarian, 1999). Tortuosity which is also used to link resistivity to porosity is the flow path of the electric current. Tortuosity is defined as the ratio of the actual or effective length of a flow path to the length of a porous medium. High values of tortuosity correspond to a high cementation factor. Lastly, the physical properties in the horizontal and vertical direction contribute to the orientation and variation of grain and pore properties (Salem and Chilingarian, 1999). Electrical current is conducted in porous media by means surface conductance and the ionic makeup of the pore fluid.

Water conducts electricity depending on the amount of total dissolved solids (TDS) present in water. TDS consists of inorganic and organic salts. Salts that dissolve in water break into positively and negatively charged ions; dissolved ions are conductors. Because dissolved ions increase salinity and conductivity, water with high values of TDS generally has a low electrical resistivity.

As a result of geoelectric investigation due to mineral matrix and pore fluid conductivity, the delineation of the freshwater lens (transition zone between fresh and saltwater) and lens thickness can be determined. Near-surface electrical resistivity has become increasingly popular over the last decade to map electrical properties to identify anomalies in the subsurface. Apparent resistivity is recorded through electrolytic conduction, where four metallic electrodes are planted into the ground and connected to a resistivity meter. The two outer electrodes A and B introduce current into the subsurface and the two inner electrodes M and N measure the potential difference (Raji, 2014). The apparent resistivity, equation 4, can then be calculated by the geometric factor times the electrode configuration and spacing. Because electrical resistivity is controlled by water content, conductivity of the fluid, presence of clay, porosity, permeability, and physical properties of the soil or rock particles, the geometry (e.g. width, length, thickness, transition zone) of these unique, untapped groundwater can be imaged.

$$\rho_a = \frac{2\pi\Delta V}{I}\left(\frac{1}{\frac{1}{AM}-\frac{1}{BM}-\frac{1}{AN}+\frac{1}{BN}}\right) \qquad \textit{Equation 4}$$

The literature review presented here demonstrated numerous techniques and applications used in groundwater potential mapping. In the current study, RF and GIS models were used to examine the relationship between IFL formation factors, IFL occurrences and to predict IFL potentiality. The results of the current study show promise in groundwater potential mapping to further explore these unique bodies of water given their formation mechanism and occurrences.

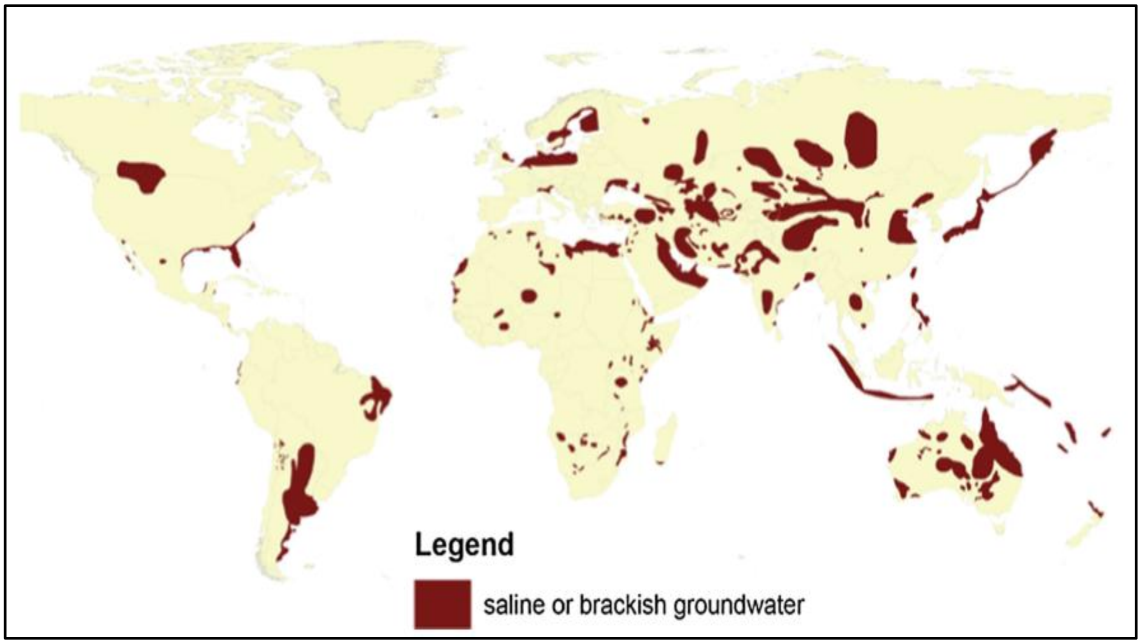

Figure 1. The occurrence of global saline or brackish groundwater (TDS>1,000 mg/L), less than 500-meters below ground level. Image source: Werner and Laattoe, 2016.

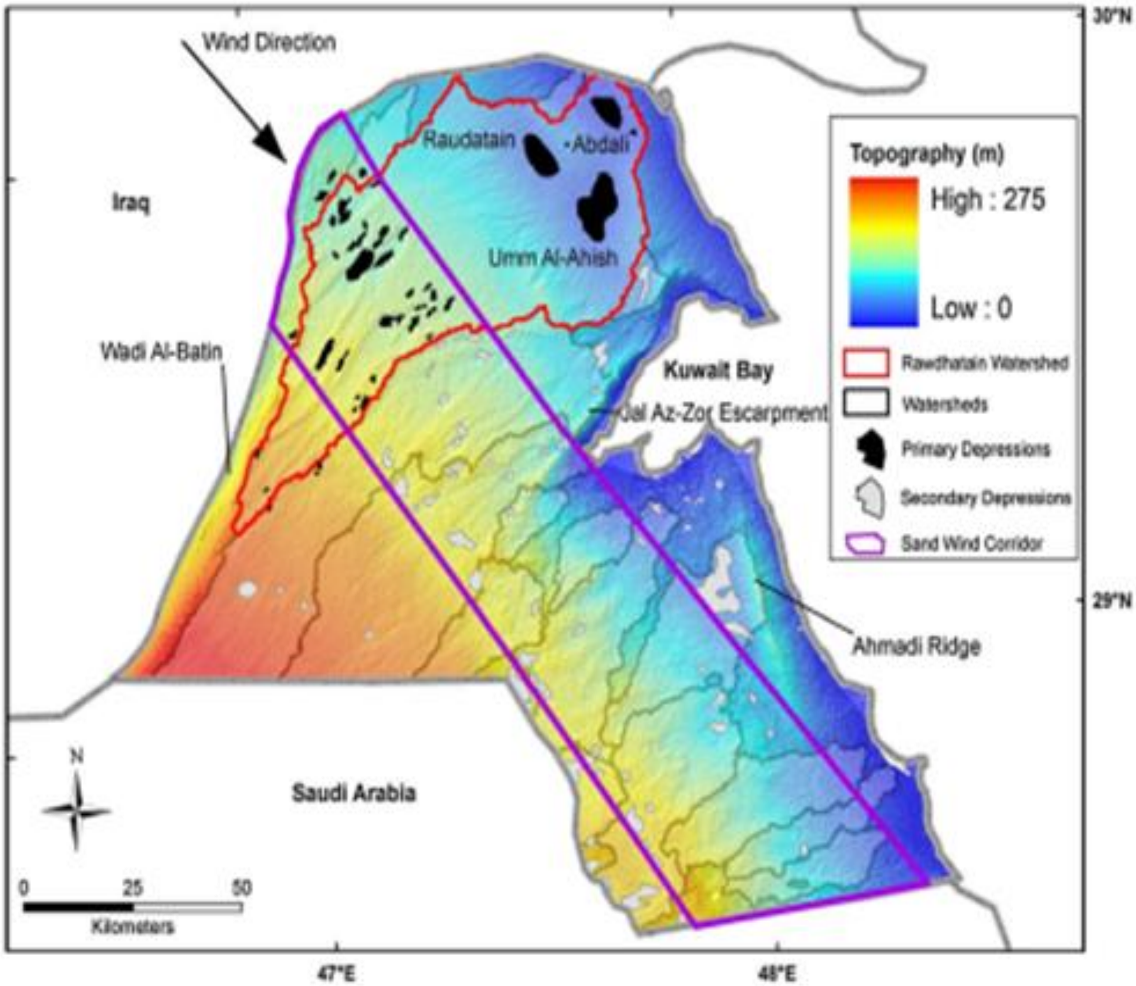

Figure 2. Topographically induced depressions identified in Kuwait. Primary depressions are suggestive areas of untapped groundwater within the Raudhatain watershed.

Image source: Milewski *et al.*, 2014.

CHAPTER 3

STUDY SITE

The state of Florida is the southeastern state of the United States separating the Atlantic Ocean and Gulf of Mexico waters, between 27.6648° N, 81.5158° W (Figure 3.). The state is a part of the major sub-region of the South Atlantic Gulf watershed and has 29 major watersheds. This research focuses on the only watershed within the state that has a known inland freshwater lens— the Middle St. John watershed. This watershed encompasses about 2,037 square miles of central Florida and includes the St. John River and confluence from the Ocklawaha River through Lake Harney.

3.1. Geologic History

Florida was formed about 530 million years ago by a combination of volcanic activity and marine sedimentation during the early Ordovician Period (Allen and Main, 2005). During this time, most of the world's land was aggregated in the supercontinent Gondwana. When Florida was part of the supercontinent Pangea, Florida was situated between North and South America and Africa. The deep bedrock that underlies Florida was originally a part of Africa. Eventually, Florida rifted from Africa's parent plate when the tectonic plate movement caused Pangea to split into Laurasia (North America, Europe, and parts of Asia) and Gondwana (South America, Africa, India, Australia, and Antarctica). Many of Florida's modern topographic features and surficial sediments were created or deposited during periods when the sea levels were high. As sea levels rose and fell, the calcium carbonate remains of the sea creatures and algae formed the sedimentary limestone bedrock. Today, the Florida Peninsula is composed of thick marine limestone and dolomite deposited during the Cretaceous (about 138 to 63 m.y. ago) and Tertiary (about 65 to 2.58 m.y. ago) geological period (Phelps and Rohrer, 1987). Rock layer thicknesses are a result of the constant deposition of calcium carbonate ($CaCO_3$) in a warm and shallow depositional environment. Marine organisms lived and died in shallow seas and their skeletal remains became concentrated, compacted, and later lithified into complex subsurface stratification of sedimentary rock formations.

3.2. Land Relief

Generally young with low-lying plains, Florida's topographic features were molded by running water, waves, ocean currents, winds, changes in sea level, and the wearing of limestone rocks by the process of dissolution. Dissolution potential for karst development in coastal carbonates derives partially from carbonic acid from atmospheric and soil carbon dioxide (CO_2), and from mixing of freshwater and saltwater (Fratesi, 2013) over the continuous carbonate rock, thus leading to the karstification of the upper surface and landforms (e.g. sinkholes, caves). The state's maximum elevation is at 105 meters (Britton Hill) and much of south Florida is mostly flat, with only a few meters of relief and lying only a few meters above sea level (Hine, 2008).

3.3. Hydrogeology and Aquifers Properties

Florida's hydrogeological regions consist of a surficial aquifer system (SAS), an intermediate aquifer system (IAS), the Floridan aquifer system (FAS), and a lower confining unit. Overlying the FAS, the surficial aquifer contains the water table and the uppermost hydrogeologic unit of post-Miocene (5 Ma to recent) and Miocene (23 to 5 Ma) chronostratigraphic units. Throughout the majority of the area, the surficial aquifer is thin, composing terrace and alluvial sands, that may aid in the temporary storage for groundwater that later recharges the underlying FAS. The surficial aquifer is thickest in the Biscayne aquifer in southern Florida and the sand and gravel aquifer in the westernmost part of Florida's panhandle. The thickness of the surficial aquifer ranges from 100-200 ft inland and exceeds 300 ft along the coastline (Reese and Wacker, 2009).

The upper confining unit of the FAS is the intermediate layers. The general thickness of the intermediate layer is thinnest in northern and central Florida as shown in (Figure 4). Interbedded in central and southern Florida, within the intermediate layer there exists a highly fractured zone called the Miocene Hawthorn Formation. This highly fractured and permeable zone is underlain by a less-permeable carbonate zone named the Ocala Avon Park. Underlain by the Ocala Avon Park is a lower permeable confining to semi-confining evaporitic (e.g. carbonates, sulfates, chlorides) and non-evaporitic rocks. Evaporitic rocks (e.g. carbonates, sulfates, chloride) are layered sedimentary rocks that form from brines in areas where water is lost through the process of evaporation. The Ocala Avon Park is overlain by the uppermost permeable carbonates of the FAS which includes the Suwannee and Ocala Limestone, and parts of the Hawthorn Group.

The Floridan aquifer system is the principal source of fresh water for agricultural, industrial, mining, commercial, and public supply in Florida (Williams and Kuniansky, 2016).

This aquifer system consists of low permeable limestone and dolomite beds of the Eocene (56 to 33 Ma) to Oligocene age (33 to 23 Ma). It also stretches across parts of Alabama, Georgia, and South Carolina covering approximately 100,000 square miles and is largely composed of carbonate rocks (Williams and Kuniansky, 2016; Maupin and Barber, 2015). The FAS is subdivided into two aquifer units: the Upper Floridan aquifer system (UFAS) and the Lower Floridan aquifer system (LFAS). If present, the middle confining unit composes low permeable rocks. However, in some instance, units in the middle confining layer is semiconfing, very leaky, or have the same hydraulic properties above and below the aquifer system. Limestone's porosity and its susceptibility to be enhanced by dissolution enable the Floridan aquifer to hold water.

The functionality and productivity of the FAS are dependent on its karstic features (e.g. sinkholes, springs). These karstic features, through the process of carbonate dissolution, create secondary porosity that is critical in controlling recharge and discharge (Tobin and Weary, 2004). The karst topography of northern and central Florida produces artesian springs, caves, and sinkholes. In areas where the aquifer system is unconfined or thinly confined, the dissolved rock increases transmissivity, whereas, in areas where the aquifer system is thickly confined the transmissivity is lower.

3.4. Regional Climate

The regional climate in central Florida is humid and subtropical, with cool, dry winters and warm, rainy summers (Willard *et al.*, 2007). Florida is divided into two climatic regimes; tropical (south) and subtropical (north) and is among the wettest states in the U.S. with most areas receiving at least 50 inches of rain annually. Summer months are between June and

September and winter months are between October and May. The rainy season runs from May 1 to late November. The average annual precipitation near the Orlando International Airport region for the period from 1981-2010 was 50.73 inches, with an annual mean temperature of 72.8°F (maximum temperature of 82.8 °F, minimum annual 62.89 °F) (Florida Climate Center: https://climatecenter.fsu.edu/products-services/data/1981-2010-normals/orlando). Hurricane season is from June to November, though September is the month during which they are most likely to occur. The average annual temperature ranges from 20 °C in the north to 25 °C in the south.

3.5. Middle St. Johns River watershed

There are numerous drainage basins with vast water networks within the state. These water networks are fed by the state's porous limestone substructure, which stores large quantities of water. The Middle St. Johns River watershed (Figure 5), includes the St. Johns River and extends from the Econlockhatchee River in Osceola, Orange, and Seminole counties northward into Lake and Volusia counties. The St. Johns River flows from south to north for about 300 miles from its headwaters in marshes in St. Lucie County to the Atlantic Ocean in Jacksonville. There are nine major tributary watersheds in the basin, including the watersheds for Lakes Harney, Monroe, and George. These large water bodies are part of and include the main stem of the St. Johns River. According to the St. Johns River Water Management District (SJRWMD), the other tributary watersheds are, from south to north, the Econlockhatchee River, Lake Jesup, Deep Creek, Wekiva River, Lake Kerr, and Alexander Springs.

Within the Middle St. Johns watershed, in the northeastern part of Seminole County, in the town of Geneva (Figure 6), there is an inland freshwater lens in the upper Floridan aquifer overlaying brackish water (Panday, *et al.*, 1993; Phelps and Rohrer, 1987; Barraclough, 1962; Tibbals, 1977; Laattoe et al., 2017) and used to supply a local population of ~5,000 inhabitant. The lens covers an area of 26 mi^2 (56.98 km^2) and is situated on a topographic high and exists in the phreatic aquifer. Sensitivity analysis from a density-dependent flow and transport analysis of the freshwater lens was done showing that the behavior of the IFL is significantly influenced by the discharges through the top of the upper Floridian, which also has a thin, leaky sandy clay confining unit with high vertical permeable which promotes the vertical recharge to the underlying aquifer (Panday *et al.*, 1993). The land surface area surrounding Geneva with an altitude greater than 7.6 meters (25 feet) is the dominant recharge area (Panday, *et al.*, 1993; Phelps and Rohrer, 1987); topographically low areas surrounds the recharge zones. A geological cross section of the study area can be seen in section 4.4 (Figure 19). The Geneva freshwater lens serves as a water supply for Geneva residents and farmers.

Figure 3. Geographical map of the state of Florida, USA.

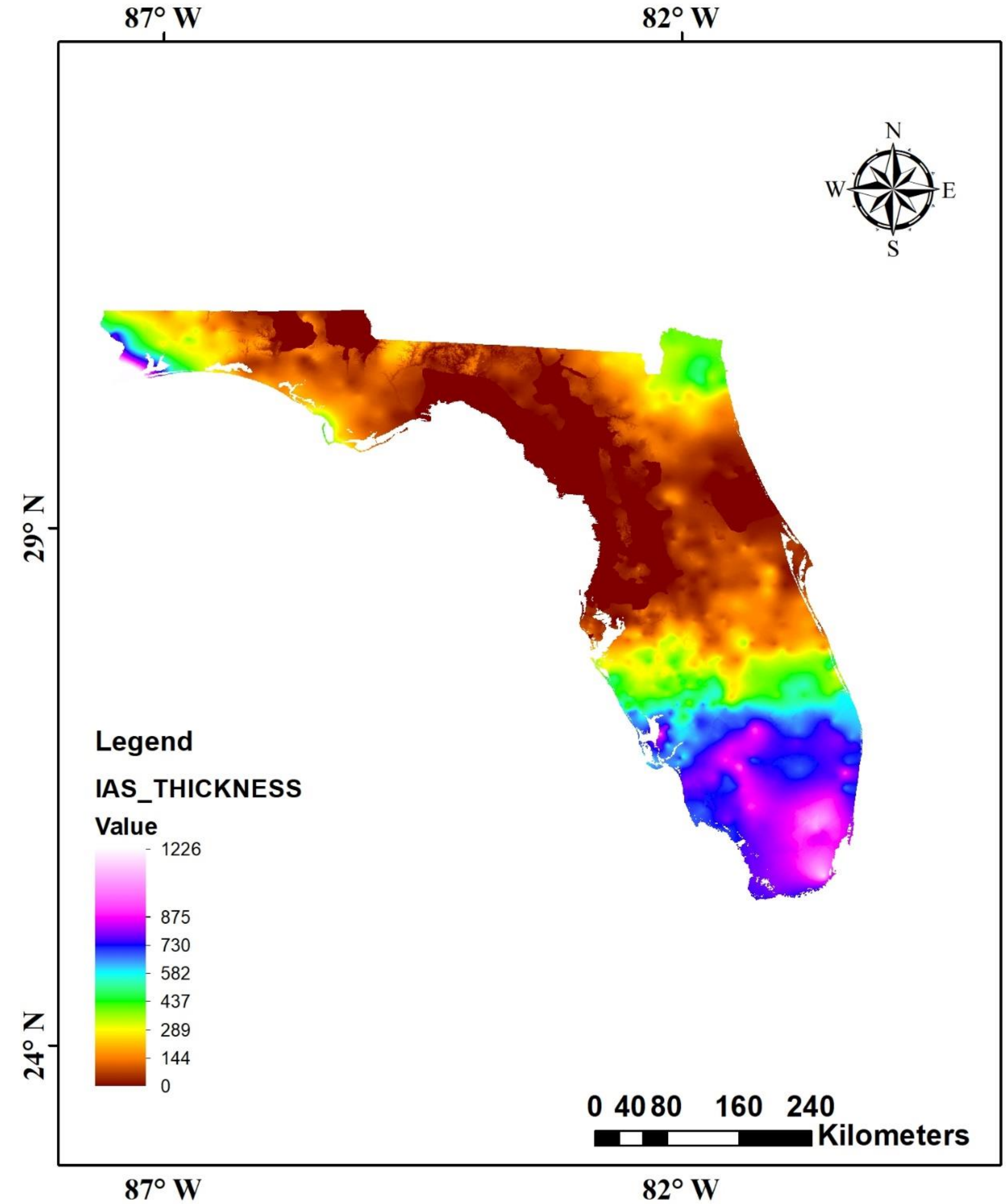

Figure 4. Thickness range of the Intermediate Aquifer System (IAS) across the state of Florida.

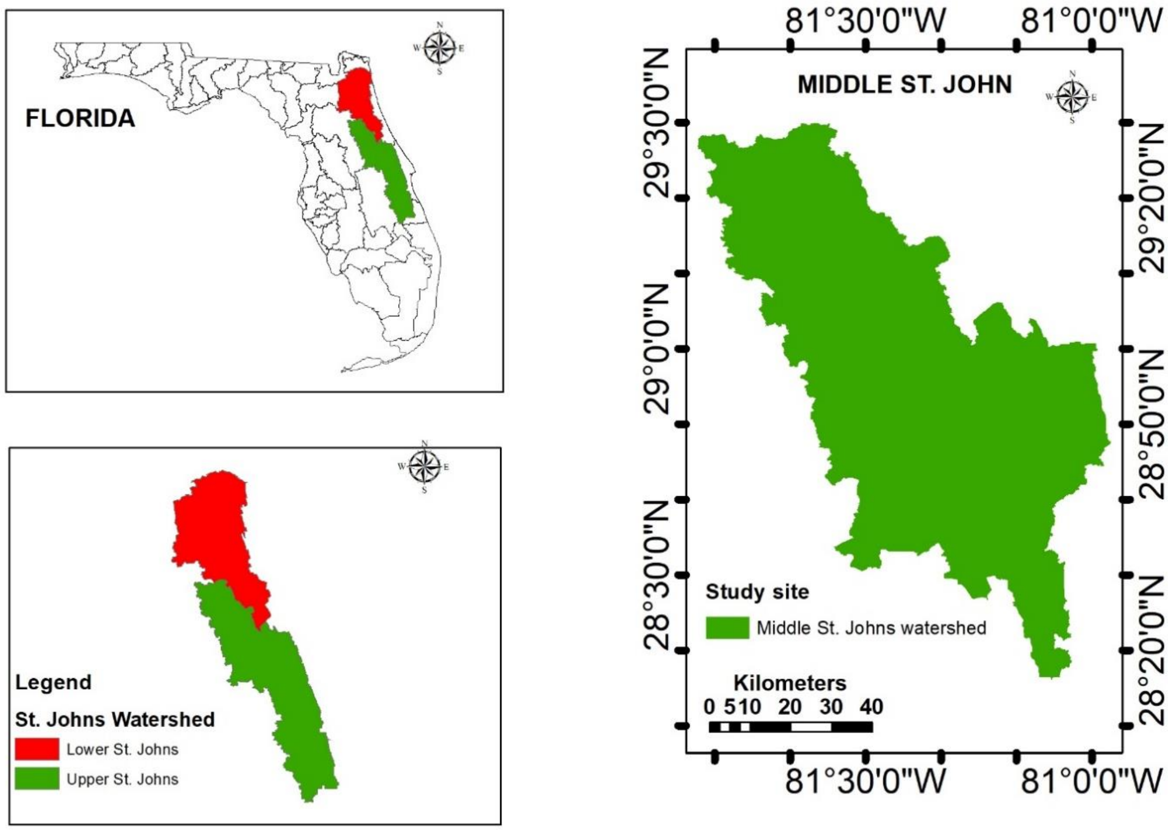

Figure 5. Study site map. The Middle St. Johns watershed located in central Florida.

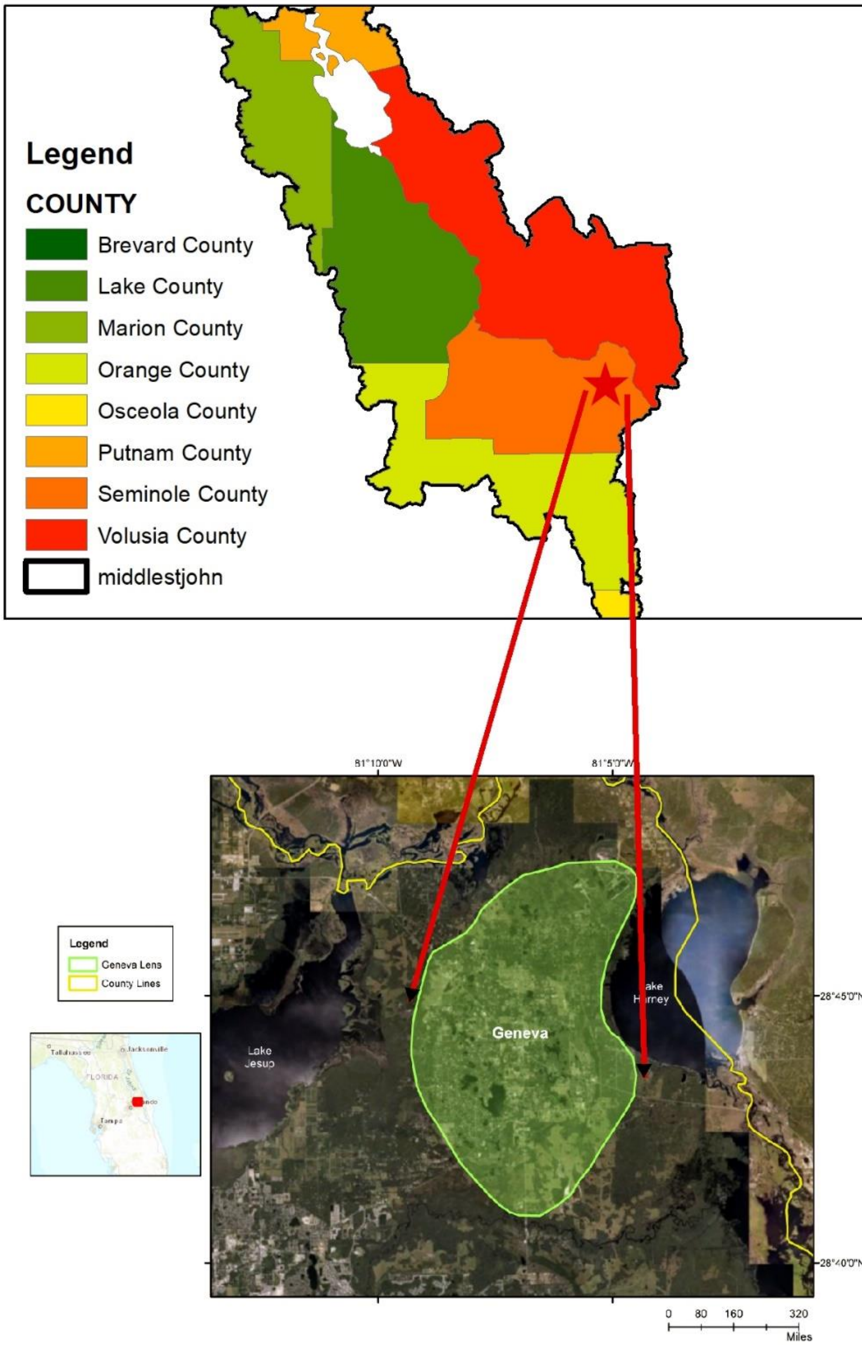

Figure 6. The areal extent of the Geneva lens located within the Middle St. Johns watershed, juxtapose Lake Jesup and Lake Harney.

CHAPTER 4

RESEARCH METHODOLOGY

4.1. Objectives

The main objective of this study is to produce an inland freshwater lens potential map (IFLPM) within the Middle St. John watershed using two models. Then later, apply ERT geoelectrical inversion and forward model to simulate apparent resistivity of the freshwater-saltwater interface. Eight thematic maps that contributed to the formation of the Geneva lens were integrated as input layers for IFLPM (Figure 7). Table 1 shows the weightage and score for the different thematic layers and their potentiality to IFL existence; weight value 10 indicates high potentiality. A simple GIS-based model was created to recognize zones of high potentiality with these layers. Secondly, the “randomForest” package in R was used for RF modeling, to produce another IFLPM. Moreover, sensitivity analyses were used to identify parameters uncertainties on the produced IFLPMs within each model. The IFLPM for each model was classified based on the classification techniques in GIS (e.g. Natural Breaks, Quantile, Equal Interval, and Geometrical Interval) for potential zones. Previous studies done with groundwater potential mapping found that classification technique, quantile is a good classifier in groundwater potential mapping (Nampak, Pradhan and Manap, 2014; Naghibi and Pourghasemi, 2015; Razandi *et al.*, 2015; Rahmati, Pourghasemi and Melesse, 2016). Therefore, this classification technique was used on all parameters, except for land use and land cover and recharge, where reclass field class_Name (e.g.

11-open water, 41-deciduous forest, etc.) and RECH_RANGE (e.g. discharge, low recharge, etc.) were used, respectively because the data values within the raster files were nominal and not ordinal like the other parameter files. Finally, geoelectric properties were modeled to predict the true resistivity of each water class using conceptual models, survey design, and noise levels (Table 2). The output of this research will provide a methodology to develop IFLPM that can be further explored and/or confirmed using this geophysical technique in the field. The confirmation of these untapped groundwaters, through ERT, can then be used for local uses, assessments, and protection.

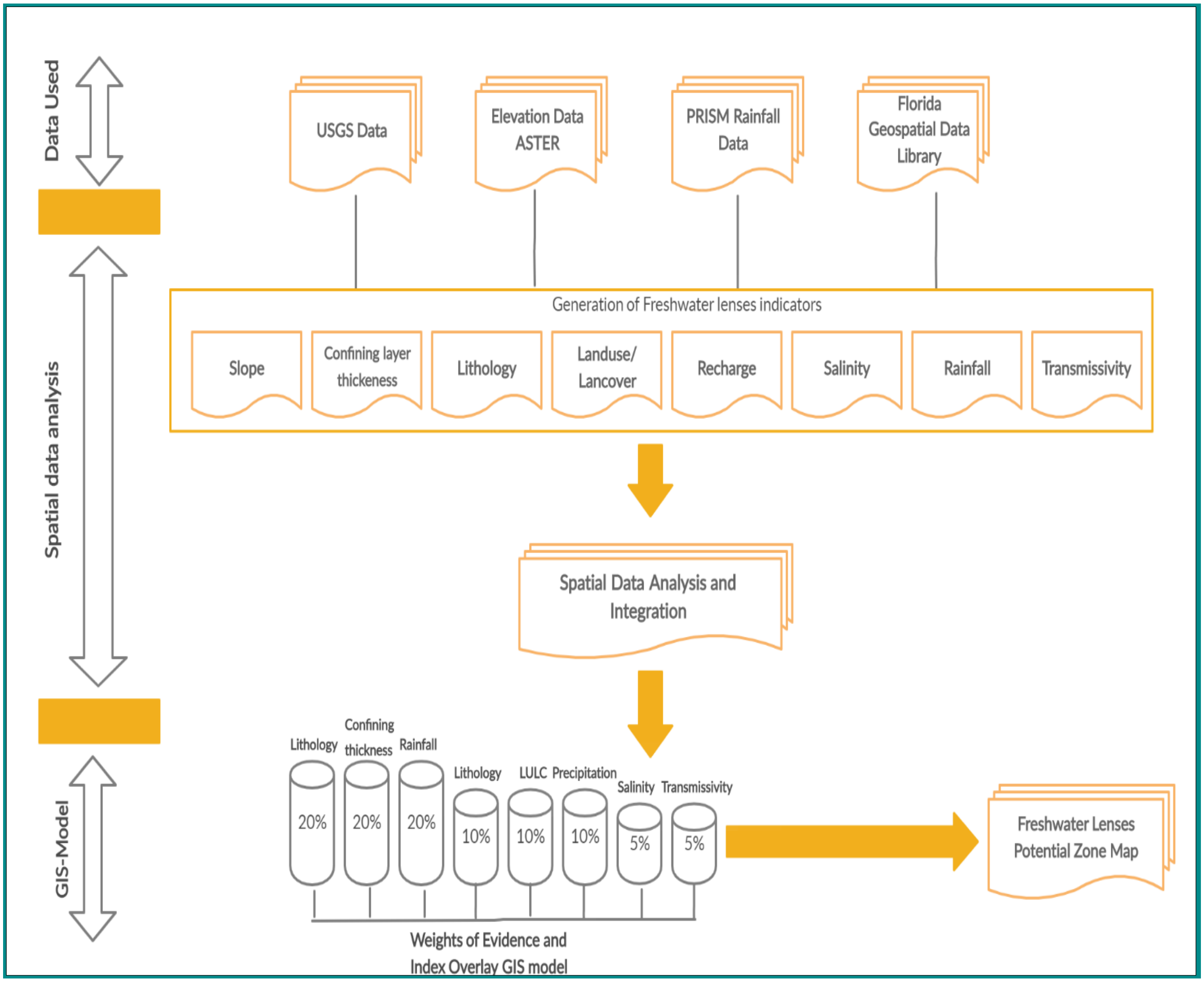

Figure 7. Flow chart illustrating steps used to generate inland freshwater lenses potential zones.

(LULC: land use and land cover)

Table 1. Weightage and score values for the thematic layers with inland freshwater lens development description in a coastal setting.

IFL formation in coastal setting (WIOA)	Class	Weight Value*	IFL development
Recharge Zone Score 20	Discharge Low Medium High	1 3 7 10	Sustains groundwater resources development and management. Recharge varies spatiotemporally.
Elevation (feet) Score 20	< 8 2-15 24-27 37	1 3 7 10	GW recharges are congruent with topographically high lands.
IAS thickness (feet) Score 20	>154 94-112 26-43 <15	1 3 7 10	Thin confining layer can allow saltwater to extrude from FAS
Precipitation (inches/year) Score 10	<48 50-51 54-55 >55	1 3 7 10	The infiltration of water into the subsurface allows for recharge.
Landuse/Landcover Score 10	Developed, High Intensity (24) Pasture/Hay (81) Developed, Open Space (21) Woody Wetlands (90)	1 3 7 10	Land must consist of minimal impervious surfaces to allow infiltration.
Lithology Score 10	Eocene Miocene Pliocene Holocene	1 3 7 10	The type of rock units and degree of weathering are good aspects for groundwater storage.
Total dissolved solid (mg/L) Score 5	<44 209-238 344-398 >727	1 3 7 10	Salinity create buoyancy effect for IFL to exist.
Transmissivity (ft^2/day) Score 5	<4,080.1-11,027.3 16,946.0-20,569.0 94,550.9-170,432.6 >748,512.1-243,767.6	1 3 7 10	The amount of water than can be transmitted through a rock unit.

*Note: Weight Value Key: 1- low potential, 10- high potential

Table 2. Salinity zones and classes as it relates to total dissolved solids and electrical conductivities if the relationship is linear. Values sourced from (Williams and Kuniansky, 2019).

Salinity Zone	**Salinity Class**	**Total dissolved solids concentration (mg/L)**	**Electrical conductivity (µS/cm) ***
Freshwater	Fresh	0-1,000	< 700
Brackish water	Slightly brackish	1,00-3,000	700 < EC < 2,000
	Brackish	3,000-10,000	
Transition zone	Moderately saline	10,000-35,000	2,000 <EC <10,000
Saline water	Saline	35,000-100,000	25,000 <EC<45,000

* Conversion factor: 1µS/cm = 10^{-4} S/m; 1 S/m = 1 ohm/m

4.2. Input Datasets and Pre-Processing

1. Precipitation

The main recharge mechanism to the Floridan aquifer is rainfall. On average, Florida receives 51 inches of rain per year, roughly 38 inches evaporates from water bodies and trees, thus on average, leaving about 13 inches to recharge the aquifer annually (SJRWMD). High-resolution 30-year average (1981-2010) precipitation data for the study area was retrieved from Parameter-elevation Regressions on Independent Slope Model (PRISM) Climate Group. PRISM uses modeling techniques to interpolate between stations through time and space. Datasets retrieved were “Norm81m” with spatial resolution 4km. This data was processed and clipped to the study area. Precipitation for this study area over a 30-year average is 48 inches/year to 55 inches/year (Figure 8).

2. Elevation

Based on previous studies by Barraclough (1962), the land surface elevation is a key component to the freshwater lens in Geneva. In topographically lower areas, thicker clay layers impede water exchange between the surficial and Floridan aquifer systems. The raster image of the digital elevation model (DEM) was retrieved from Florida’s Geographic Data Library Documentation (FGDL) to derive high elevation points within the watershed using ArcGIS 10.2 software. Elevation units are represented in feet. Elevation ranges from -13.7 feet to 36.8 feet within the study site (Figure 9).

3. TDS groundwater level

One way to mitigate water quality within the aquifer is to monitor the total dissolved solids (TDS). Total Dissolved Solids (TDS) consist of inorganic and organic salts. As a result, TDS can be used as an indicator of saline groundwater, which here is necessary for the bouncy effect to develop a freshwater lens. The USGS observed minimum depth to brackish (1,000-10,000 ppm) to highly (>10,000 ppm) saline groundwater. Estimation variation in TDS concentration came from geophysical logs (TDEM), and water sampling. Salinity mapping includes freshwater, brackish water, a transition zone, saline water, and brine water zones as previously mentioned. For this study, brackish zones, transition zone, and saltwater zones for the estimated total dissolved solids concentration, milligram per liter for the top 50 feet of the aquifer system were considered. The data was retrieved from (Qi, and Harris 2017). Dataset is in the Albers Equal-Area Conic projection. The TDS data collected from each well were placed in ArcGIS and interpolated across the study site using the Inverse Distance Weighting (IDW) interpolation in the spatial analyst toolbar. TDS ranged from 44.52 mg/L to 6,285.87 mg/L for the study area (Figure 10).

4. Intermediate confining layer thickness

Variable confinement of the hydrostratigraphic unit of the intermediate aquifer system was collected from the Florida Department of Environmental Protection Geospatial Open Data site. This grid was created from well core and cutting data, later interpolated using the kriging method across the study site. The dataset was brought straight into ArcGIS for further analysis. The intermediate confining layer ranged from 0 to 178 meters thick in the study area (Figure 11).

5. *Recharge zones*

Recharge to and discharge from the FAS occur through a variety of pathways and the most dominant characteristic of the system controlling these processes is the degree of confinement (Williams and Kuniansky, 2015). Unconfined and thinly confined areas are where direct recharge from precipitation enters the system (Miller, 1999). The data used to identify recharge zones was taken from the St. Johns River Management District (SJRWMD) geospatial open data source. This data consists of recharge to discharge zones of the upper Floridan aquifer (Figure 12).

6. *Land Use and Land Cover (LULC)*

The LULC map of the study area was obtained from the National Land Cover Database 2011 (NLCD2011). The NLCD uses Landsat imagery to categorize different land covers at 30m resolution. The data type is vector and formatted into a raster in Albers conical equal area projection (Figure 13).

7. *Lithology*

Lithology is considered as one of the most important indicators of hydrogeological features playing a fundamental role in both the porosity and permeability of aquifer materials (Rahmati, Pourghasemi and Melesse, 2016). The lithology map represents surficial and near-surface geology and was downloaded from the Florida Department of Environmental Protection (FDEP) geospatial open data portal. The study site was divided into 5 lithology units. In this study, in order of the most recent deposit, the lithology units were classified into Holocene (beach sand), Pleistocene/Holocene (clay or mud), Pliocene (sand), Miocene (sand), and Pliocene/Pleistocene (limestone) epoch (Figure 14).

8. Transmissivity

The position of the freshwater-saltwater interface is governed by large differences in transmissivity between the Upper and Lower Floridan aquifer and the confining units (Kuniansky, Bellino, and Dixon 2012). Where the aquifer is unconfined or thinly confined, infiltrating water dissolves the rock and transmissivity tends to be relatively high. Where the aquifer is thickly confined, less dissolution occurs and transmissivity tends to be lower. Data were retrieved from Kuniansky, Bellino, and Dixon (2012) study and interpolated across the study site using the kriging method. Kriging is a statistical model that includes the relationship among measured points, providing some measure of the certainty. Transmissivity values range from 4,080.12 to 748,512 ft^3/day within the watershed (Figure 15).

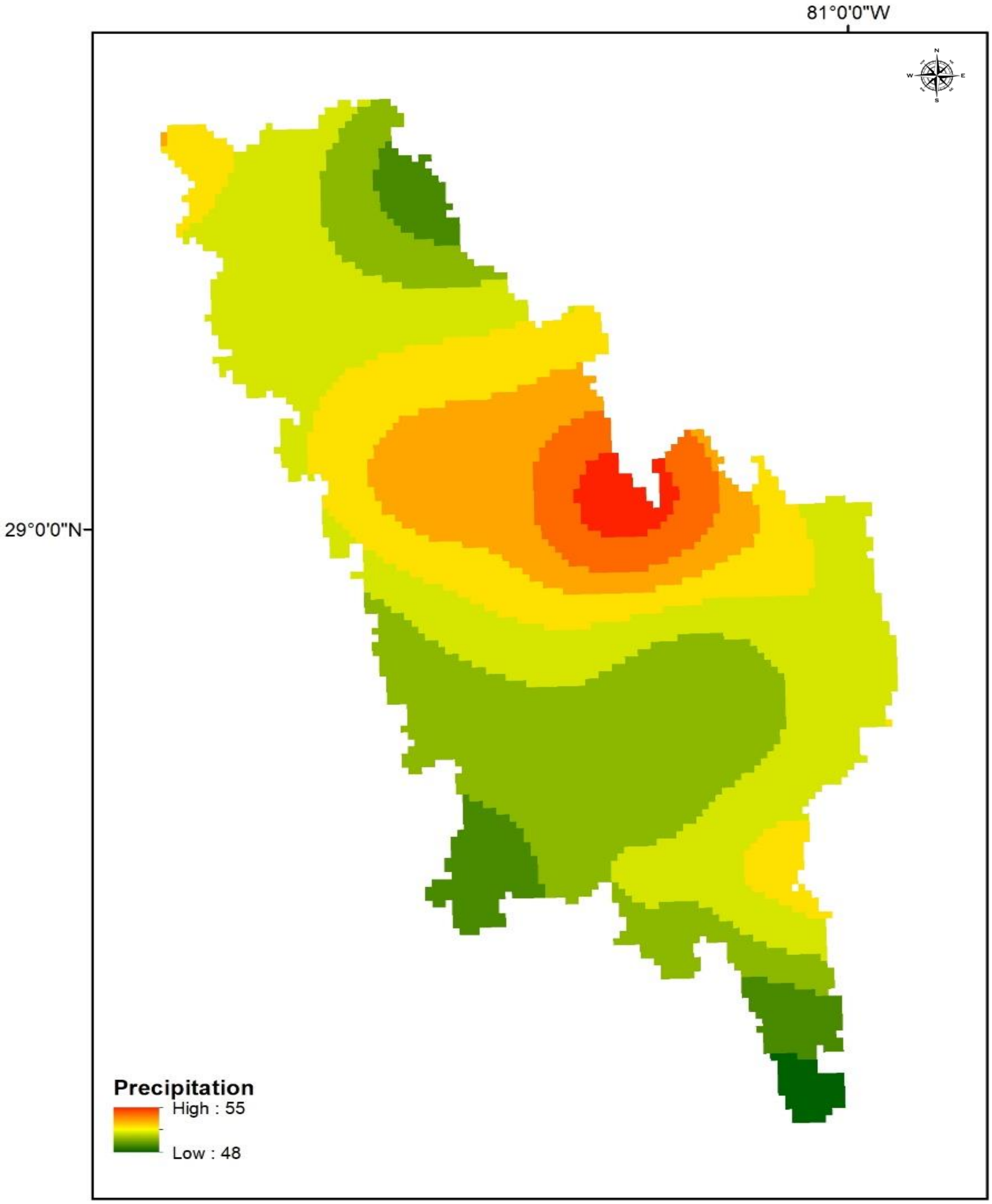

Figure 8. High-resolution 30-year average (1981-2010) precipitation data.

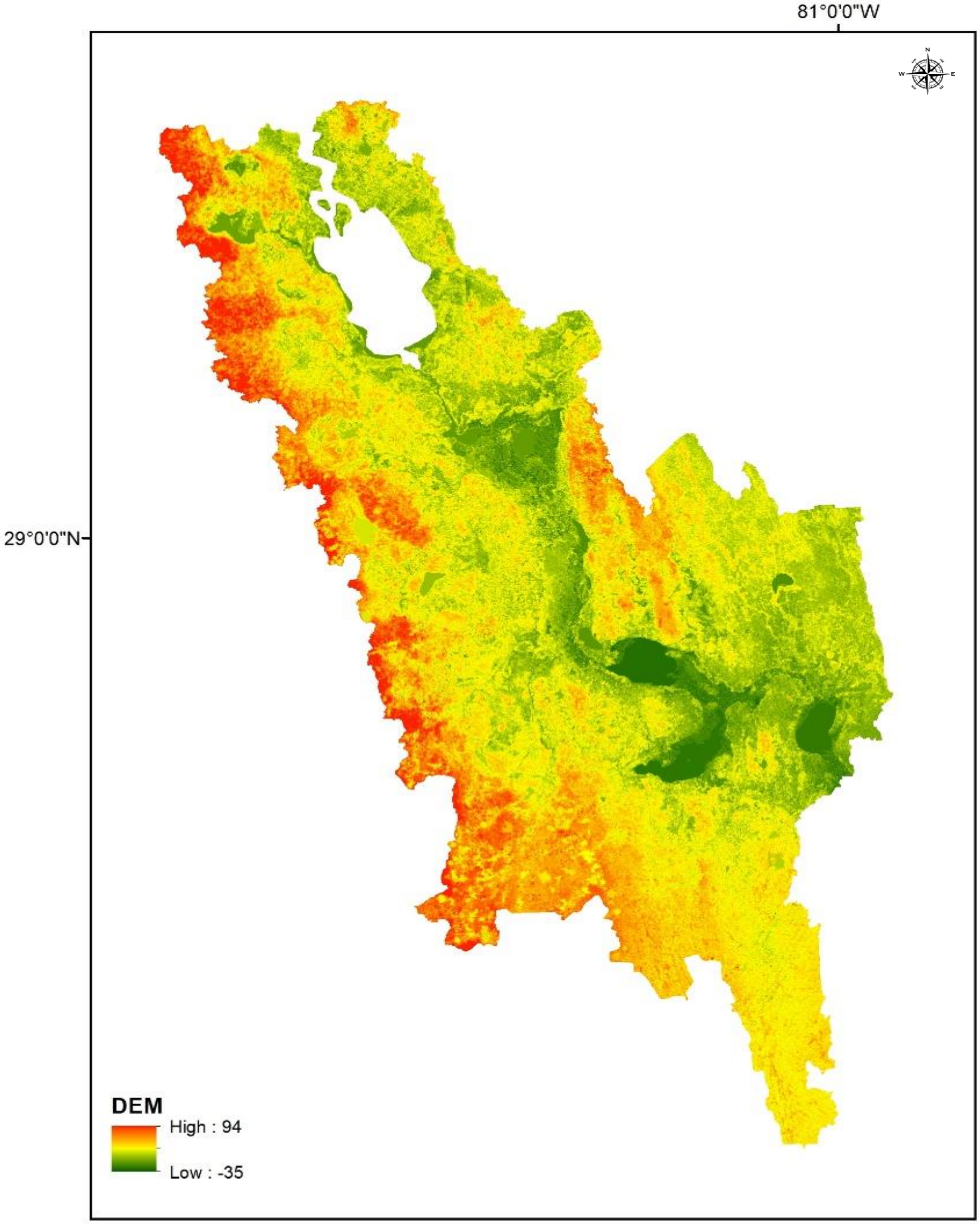

Figure 9. Digital elevation model (DEM) showing elevations in feet.

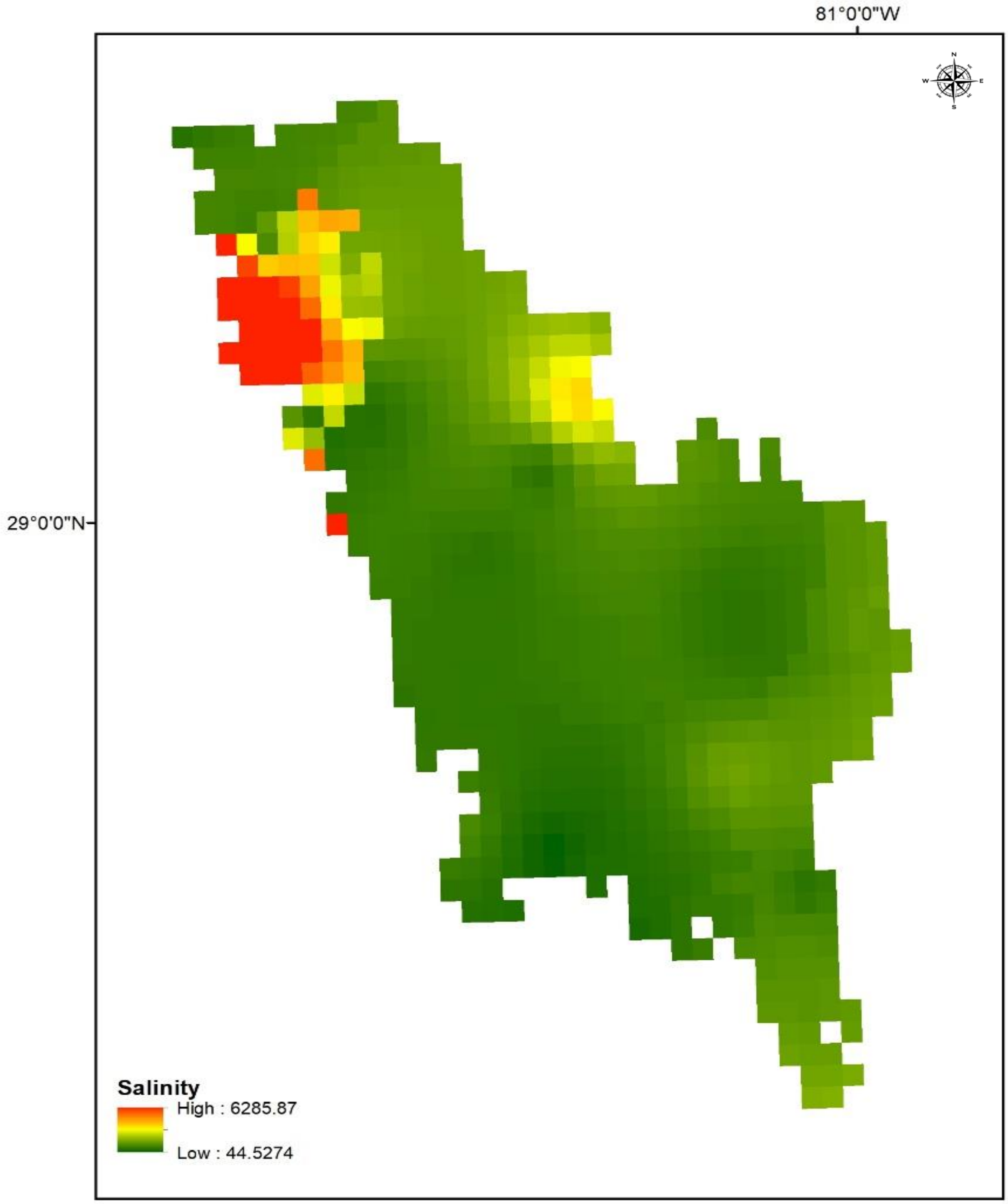

Figure 10. Interpolated total dissolved solids (TDS) concentration, milligrams per liter for the top 50 feet of the Floridan aquifer system for the top 50 feet of the Floridan aquifer system.

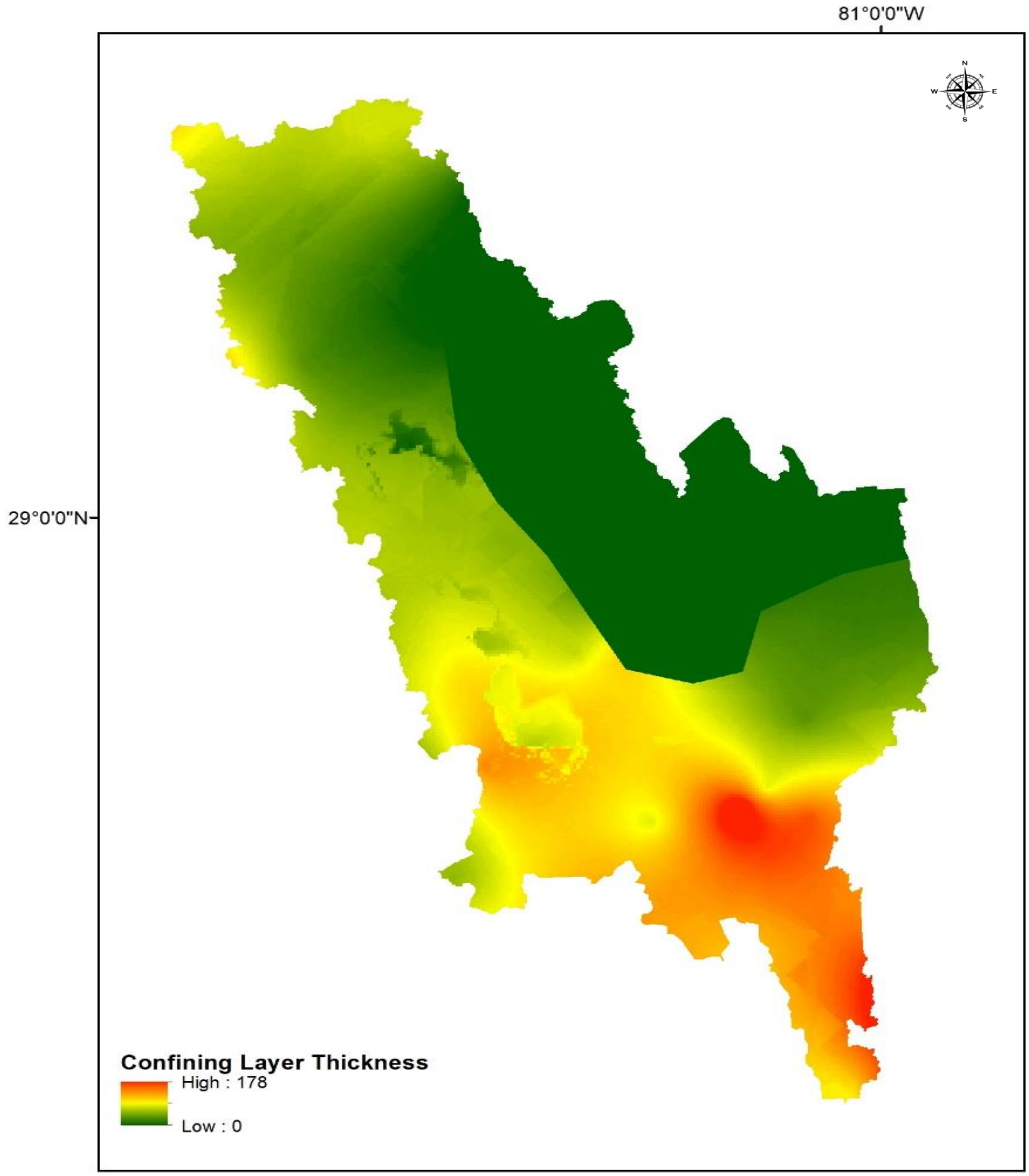

Figure 11. Variable confinements of the hydrostratigraphic unit of the intermediate aquifer system.

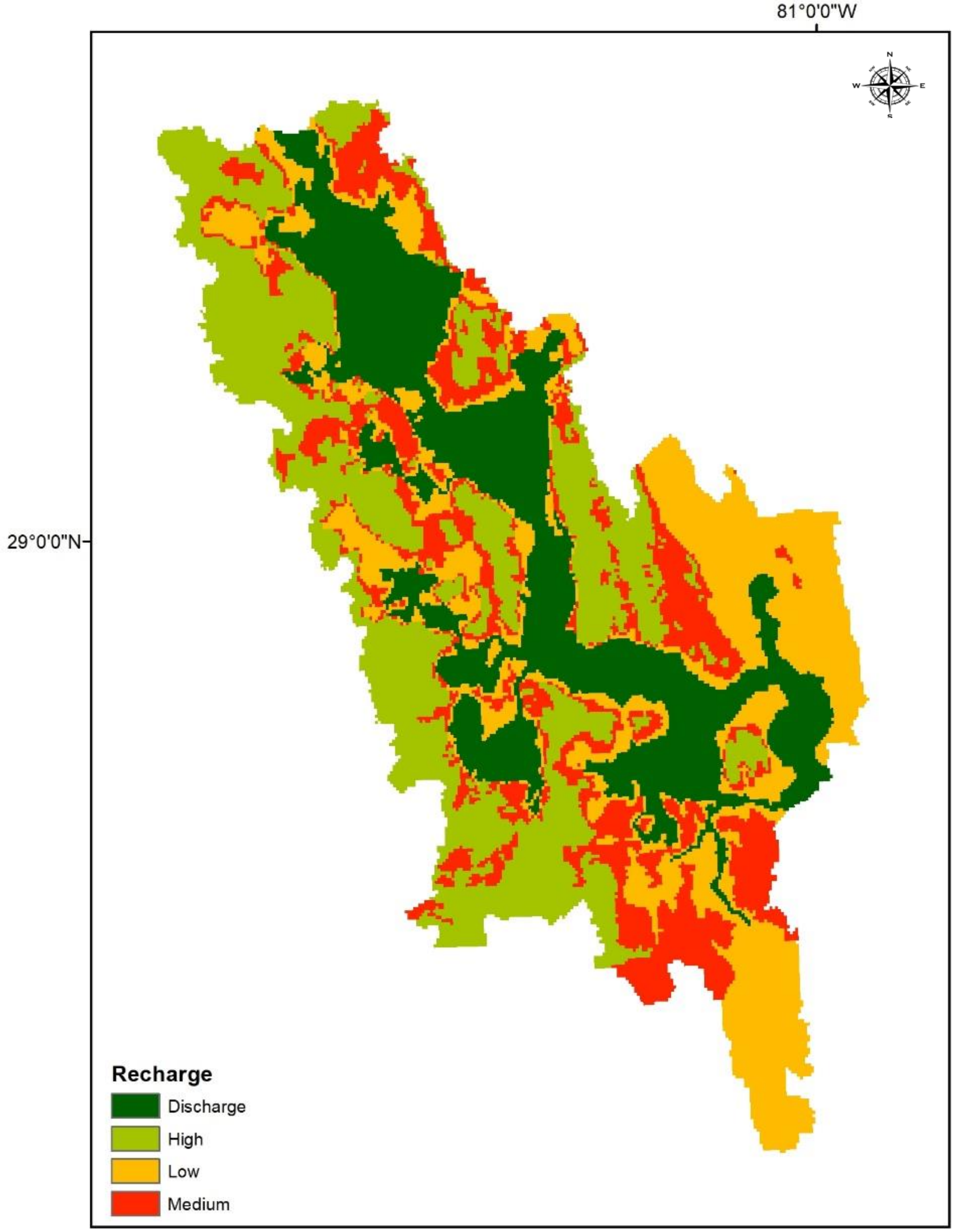

Figure 12. Recharge zones classes to the upper Floridan aquifer.

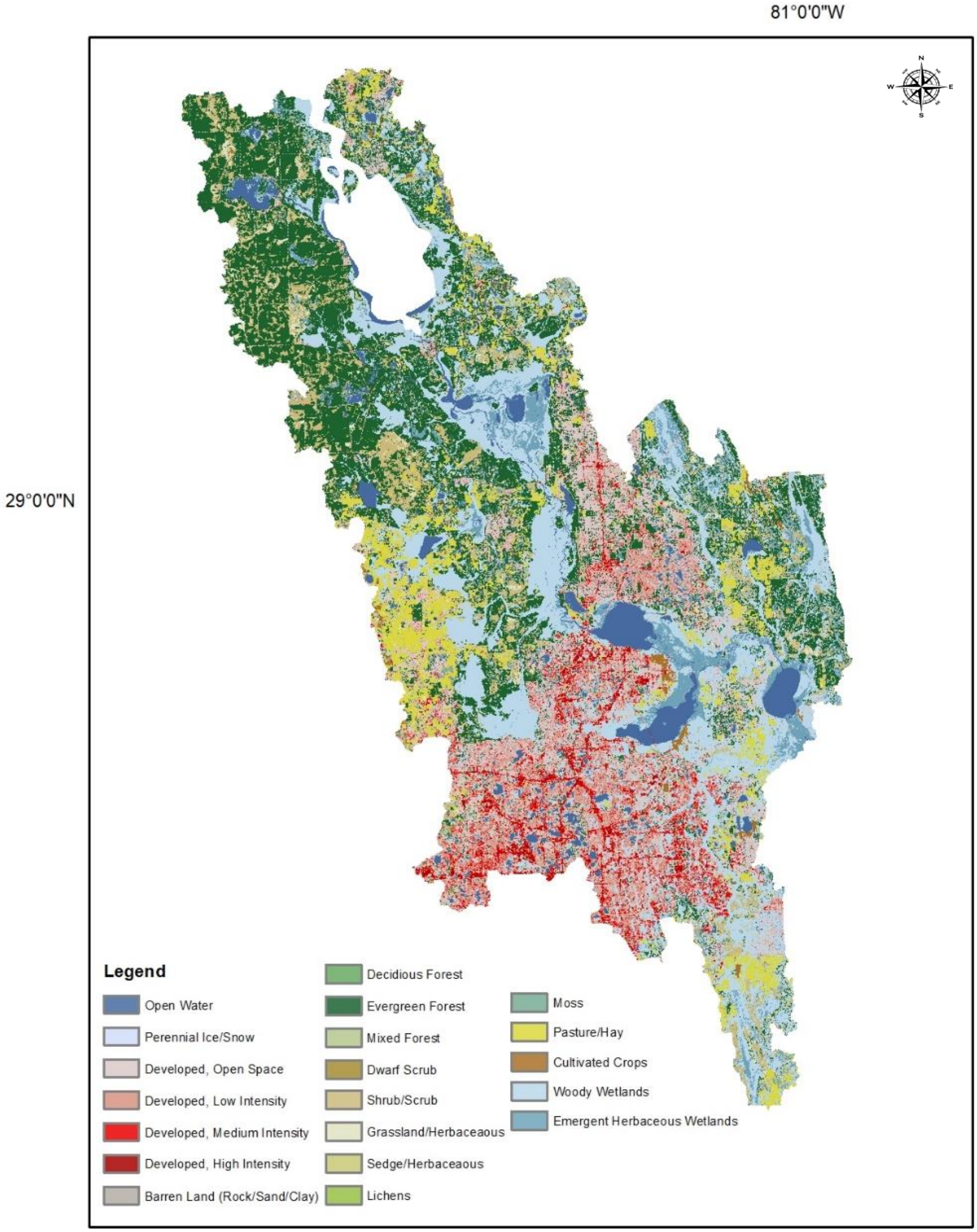

Figure 13. National land cover database 2011 (NLCD2011)

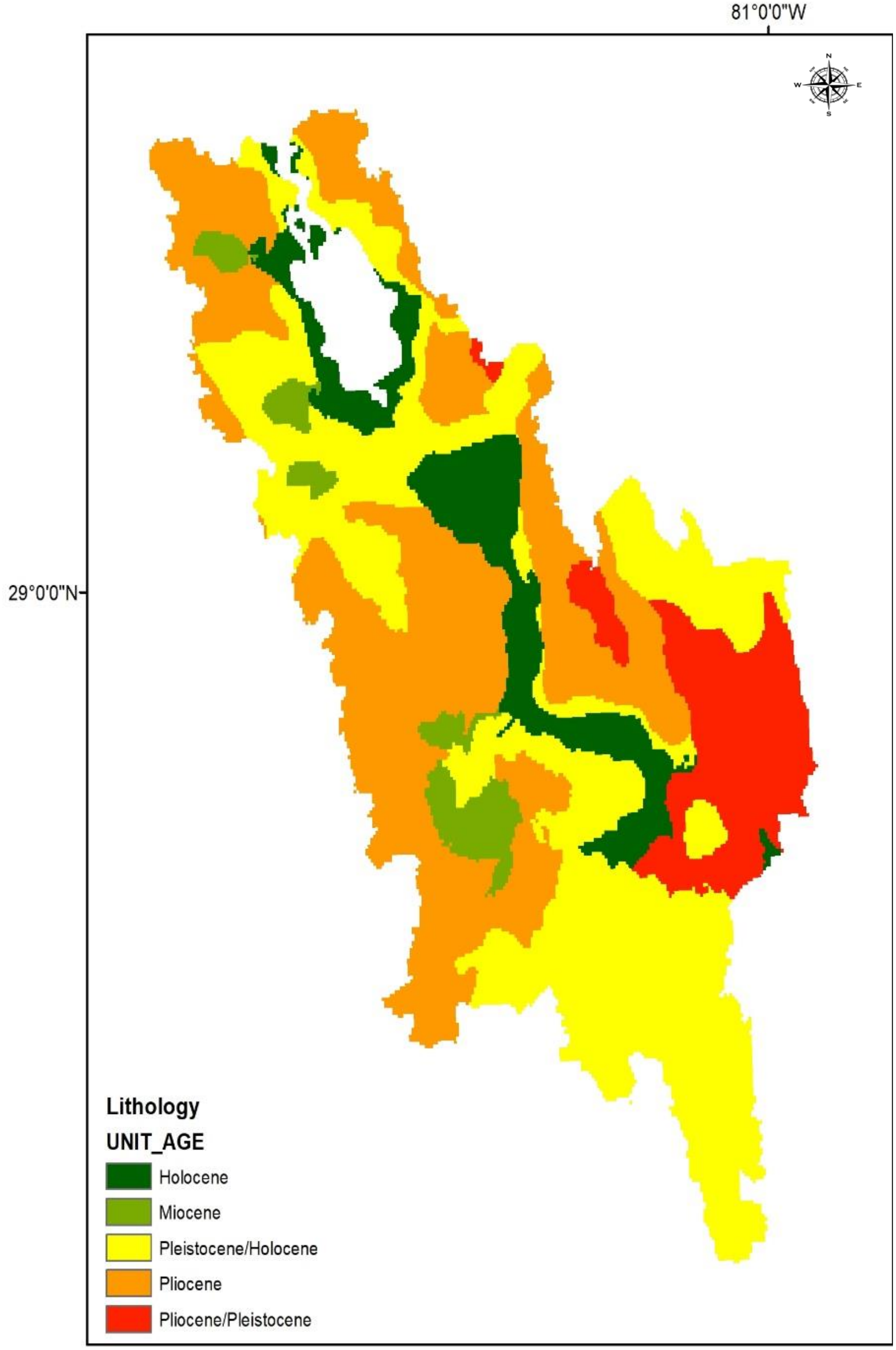

Figure 14. Unit age for each geological feature.

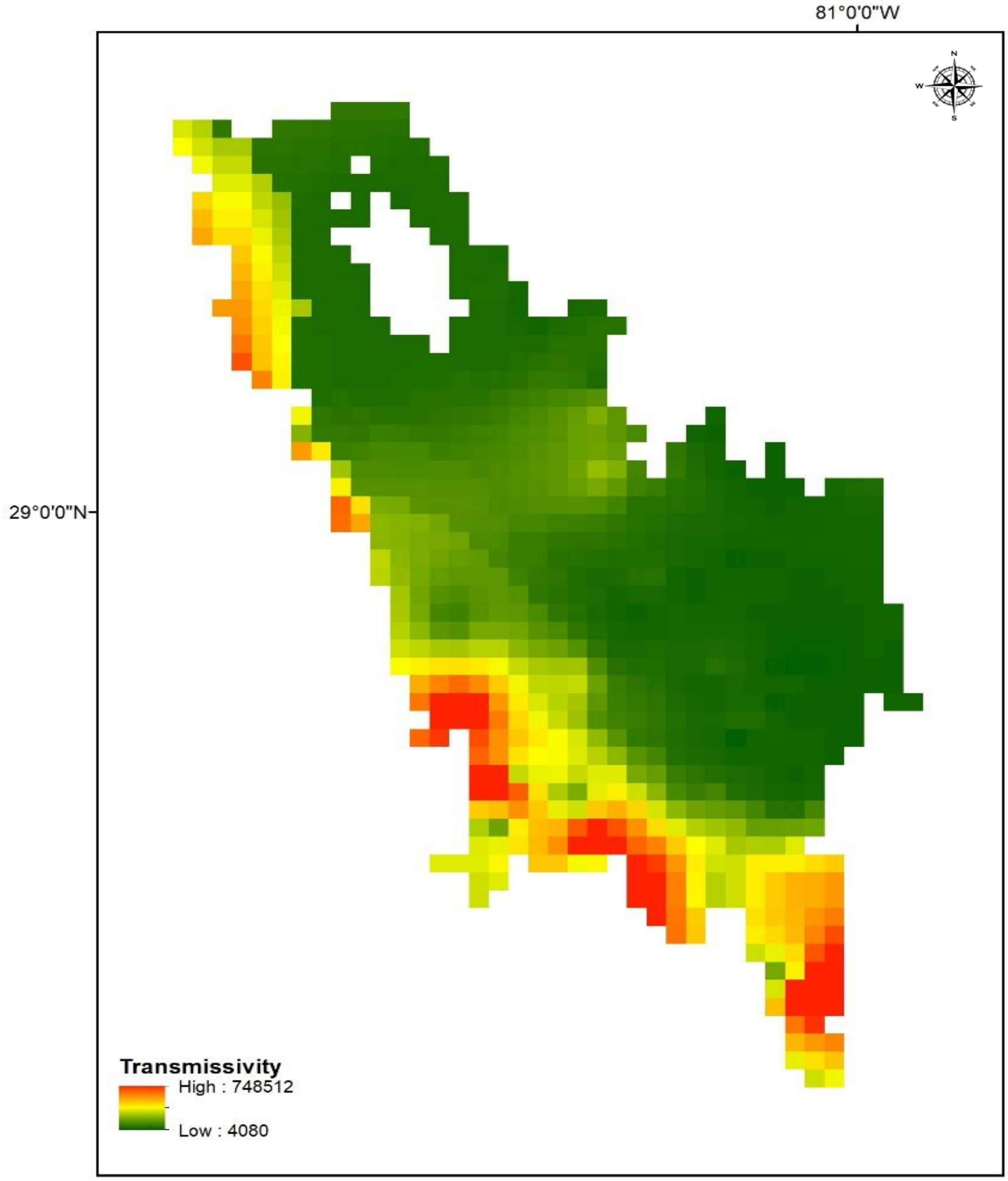

Figure 15. Interpolated transmissivity based on interpolation of ~1500 aquifer test of the Upper Floridan Aquifer, in units of feet squared per day.

4.3. Description of models

Model 1: WIOA GIS data-driven model

Firstly, datasets were integrated into an Arc-GIS based interface to analyze the spatial relationships. The reclassification tool was used to extract and reassign old values within each raster file to new values. The new values were ranked from one to ten; one being of lowest importance to ten being higher importance for IFL formation conditions.

After reclassification, the Weighted Index Overlay Analysis (WIOA) combined analysis of multi-class layers. Each parameter was overlaid and assigned a weight value ranging between 0 and 100% in the spatial analyst tool. The weight values range because not all parameters have the same level of contribution toward groundwater potentiality. In this study, scoring with the following percentages 20, 20, 20, 10, 10, 10, 5, 5 were assigned to elevation, confining unit thickness, recharge, lithology, land use and land cover, precipitation, depth to saline groundwater, and transmissivity layers, respectively was used to develop the IFLPM. Each factor considered is assigned a weight depending on the formation mechanisms leading to the development of the Geneva lens in central Florida. Consideration of the relative importance between the parameters leads to better representation of the actual ground. To see if the weights assigned to each parameter had any effect on the model output, a WIOA sensitivity test was done weighting each parameter differently (Figure 16). For the first sensitivity test, Figure 16a represent the main model. Figure 16b the following percentage 20, 20, 20, 8, 8, 8, 8, 8 were assigned to elevation, confining unit thickness, recharge, lithology, land use and land cover, precipitation, depth to saline groundwater, and transmissivity layers, respectively. Figure 16c equal weights were assigned to each parameter. In the present approach, the total weights of IFLPM were derived as the sum or product of the weights assigned to the different layers according to their level of contribution. Finally, the IFLPM was mathematically calculated using the raster calculator. Each ranked pixel values were extracted to points, assigned latitude and longitude for each field ID, and clipped to each parameter using the extraction tool in the spatial analyst toolbar.

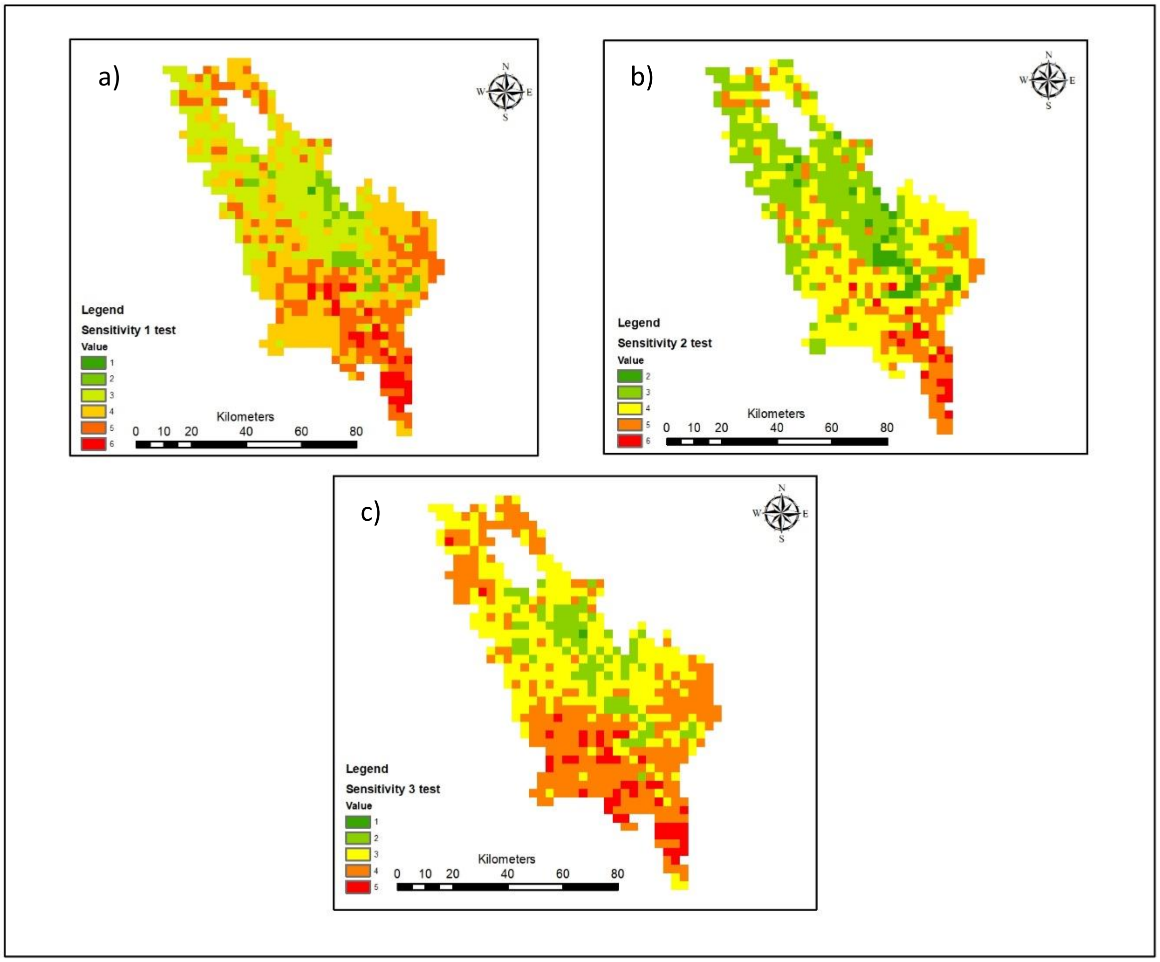

Figure 16. Weight Overlay Index Analysis sensitivity test.

GIS model validation using OLS and GWR

The validation step is the most important process of modeling and without it; the groundwater potential models would have no scientific significance (Chang-Jo and Fabbri, 2003). Several studies have explored uncertainty within a GIS-based data-driven model. Some studies have even found that the statistical methods used as well as the differences in data sources could lead to uncertainties within this model (Heikkinen et al., 2006). Ordinary least square (OLS) and geographically weighted regression (GWR) were performed on the IFL potentiality inventory map to assess the model's ability to predict where IFLs are located and the sensitivity of each IFL formation factor.

Ordinary least squares (OLS) regression was used to evaluate the performance of each inland freshwater lenses postulated location and verify that the model does not violate common regression model assumptions through diagnostic statistics assessments. OLS builds a global linear regression model for the entire study area to measure the overall fit of the data and produce predictions for a dependent variable (e.g. IFL ranks) based on relationships to assigned independent variables (e.g. recharge, elevation, etc.). OLS assumes variable relationships are constant in space, which prohibits regression coefficients to vary over space.

Also, due to the spatial heterogeneity of IFL ranks and individual IFL formation factor, a local spatial statistics technique that allows the regression coefficient to vary, such as GWR, would provide a more direct method of testing hypotheses that are the subject of spatially variation. GWR computes a unique regression equation for each point in a dataset using kernel type, bandwidth method, and optional weighting factors for individual features.

Model 2: Random Forest (RF) Machine Learning Algorithm Model

There is no known database containing IFL in Florida's coast, except the Geneva lens. Data values from the GIS-model of IFLs' presumptive locations were used to create a data frame of IFL locations and formation factors, to train and test random forest model to predict and classify IFL locations by ranks within the St. Middle Johns watershed. The outcome of the RF predictive model was compared to 30% of GIS initial model classification to further assess the model's ability to predict IFLs within the St. Middle Johns watershed and possibly elsewhere.

Random forest (RF) is a nonparametric technique (Breiman, 2001) that was developed as an extension of classification and regression trees (CART) and generates many classification trees (Breiman et al., 1984) to improve the prediction performance of the model. This algorithm constructs multiple trees based on random bootstrapped samples of the training dataset (e.g. 70%) and runs random binary trees that implement a subset of the observations over the bootstrapping approach, of the initial dataset a random choice of the training data was selected and implement to create the model. RF predicts the importance of a variable by looking at how much the error of prediction increases when out of bag data for the variables is permuted while all others are left fixed (Naghibi, Ahmadi and Daneshi, 2017). In general, the more trees in the forest, the more robust the decision tree, the higher the accuracy. This random forest model assesses IFL potential by IFL ranks (1-6); high ranking IFLs (e.g. 6) are strong indicators for IFL potential mapping. This study was conducted with an inventory of 662 IFL ranks generated by the GIS IFLPM randomly divided and split into 70% (e.g. 466) for training and 30% (e.g. 196) for testing and 8 explanatory groundwater conditioning factors (Figure 16).

After splitting the dataset into training and testing datasets, it was necessary to define two parameters: the number of variables/factors to be used in the tree building process (m_{try}) and the number of trees (n_{trees}). To minimize the generalization error, an internal RF function Tune RF was used to best optimized m_{try} and n_{trees} before training the model. By default the model is set to create 500 trees, to see if 500 was optimal for classification, as well as the frequency of IFL misclassification at each tree, an error rate for IFLs potential was plotted using the ggplot2 function within R. Each tree is generated by bootstrap samples and uses the out-of-bag (OOB) error to make validations. The OOB error is an unbiased estimate of the generalized error that the model made. With the OOB error, there are many advantages (e.g. no overfitting, low bias and variance, high predictive performance, low correlations among trees) (Prasad, Iverson and Liaw, 2006).

After running the model, parameters sensitive to IFL formation were identified. Parameter importance was measured on the mean decrease in Gini coefficient plot. Also, the OOB error rate from low to high potential zones classification was calculated from the training dataset. RF aims to identify the suitable model to analyze the relationship between independent variables and a dependent variable in the calibration phase (i.e. model building) to determine the weight value for each factor. In this study, IFL inventory of training dataset (i.e. 70% of the dataset), and 8 IFL conditioning factors were used as the dependent variables and independent variables, respectively.

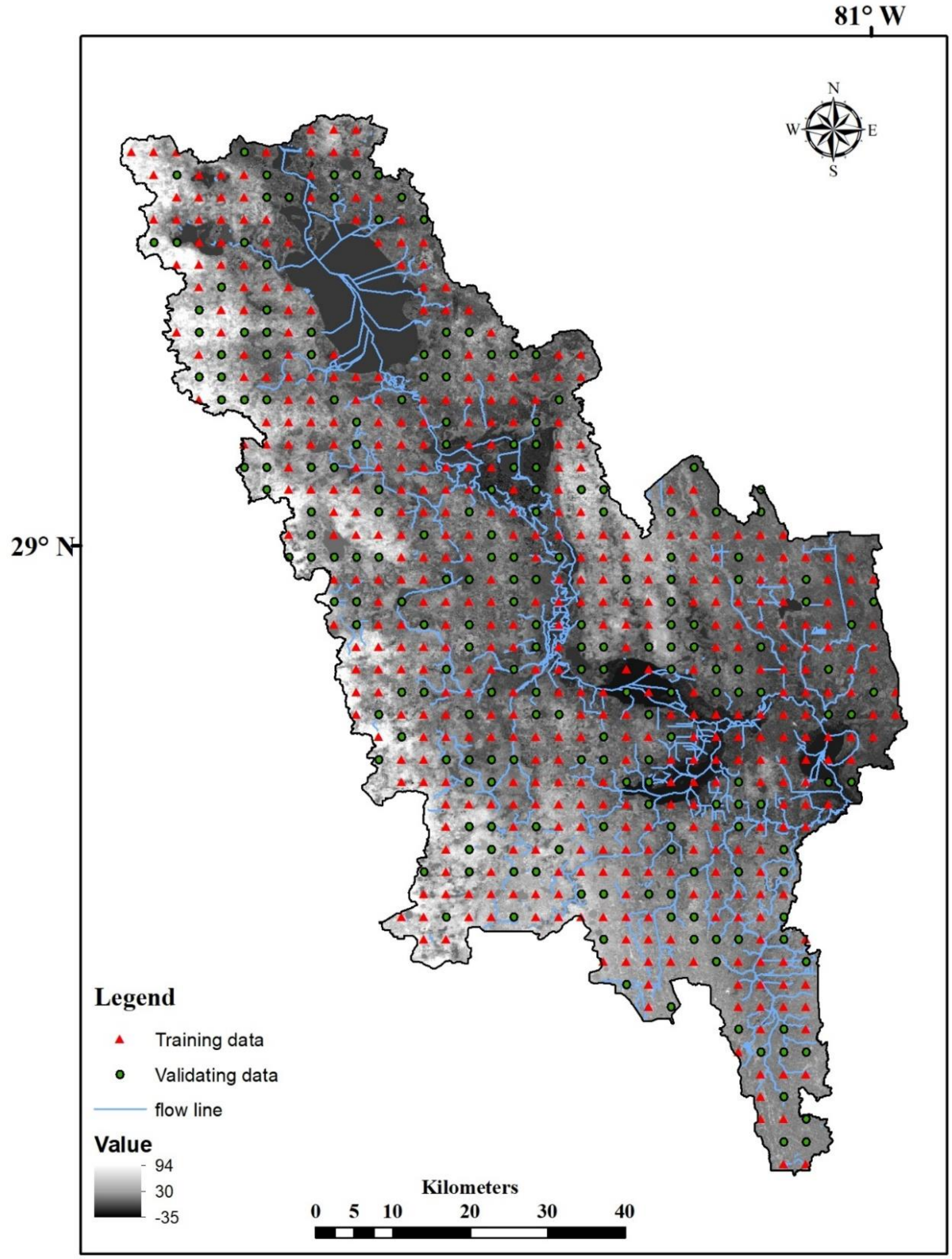

Figure 17. Random forest (RF) model training (70%-red) and validating (30%-green) data points.

IFL formation factors are assigned to each data point.

Random Forest Model Validation

Model validation was accomplished using the receiver operating characteristics (ROC) and area under the curve (AUC) to evaluate the model's efficiency. The receiver operating characteristic (ROC) curve has been used in many studies to evaluate the efficiency the model has on groundwater potential mapping. The ROC shows the performance of a classification model at all classification threshold. It is created by plotting the true positive, or sensitivity, rates against the false positive, or 1- specificity, rate. The true positive rate indicates what proportion of IFL rankings were correctly classified. The false-positive rate tells us when an IFL ranks correctly but get rejected by the model. The sensitivity calculated can be described by equation 5 and equation 6. ROC is a scientific technique that describes the efficiency of probabilistic and deterministic detection and forecast system (Swets, 1988). The area under the ROC curve (AUC) quantifies the uncertainty of the model, accounting for detected biases associated with those estimations. The uncertainty of the RF model has been investigated using AUC. To examine the efficiency and reliability of the IFLPM, the success and predictive rate curves were calculated. The quantitative-qualitative relationship between the AUC and prediction accuracy can be classified as follows: 50–60% (poor), 60–70% (average), 70–80% (good), 80–90% (very good), and 90–100% (excellent) Yesilnacar (2005).

$$\textit{True Positive Rate/Sensitivity} = \frac{True\ Positives}{True\ Positives + False\ Negatives} \qquad \textit{Equation 5}$$

$$\textit{False Positive Rate/(1-Specificity)} = \frac{False\ Positives}{False\ Positives + True\ Negatives} \qquad \textit{Equation 6}$$

4.4. Electrical Resistivity Tomography (ERT): Variable Density Forward Modeling

The final step to this research is to confirm the presence of these high ranking IFLs using subsurface imaging. This technique would be utilized in the field to verify high IFL potentiality based on the model predictions. For this thesis, what drives electrical resistivity signatures and a modeled cross section geoelectric forward and inverse modeling to derive the subsurface true apparent resistivity to delineate freshwater-saltwater interface is demonstrated. Data values retrieved from the literature are used to create a forward model to delineate freshwater-saltwater boundaries. Electrical resistivity quantifies how strongly a material opposes the flow of electric current. For example, saltwater will be less resistive (more conductive) to electrical flow because it contains more ions than more resistive (less conductive) freshwater. To be able to see the contrast in electrical properties between freshwater and saltwater, software for geoelectrical modeling, ResIPy, was used to image the subsurface (Blanchy *et al.*, 2020). The software incorporates codes from R2 (Binley, 2019) to make importing, filtering, and error modeling possible for the geoelectrical dataset Pre-modeling flow chart can be seen in Figure 17.

4.4.1. Forward model and inversion model survey design

A modeling study was carried out to determine the optimum electrode spacings and configurations (e.g. Wenner and dipole-dipole) for resolving freshwater lenses. By solving the forward problem, we go from a conceptual model to the true resistivity, generating data on apparent resistivities. Apparent resistivity refers to the value determined by the field measurements of potential difference, multiplied by the geometric factor, which depends on the electrode configuration. The forward model takes a synthetic model and generates theoretical data. Theoretical data can be represented as resistances or apparent resistivity. Inversion modeling takes the measured apparent resistivity and tries to find the true resistivity of the subsurface. The Geneva lens substructure, Figure 19, was modeled to generate synthetic measures from the electrode spacing and the type of configuration. Data derived from the forward model was used to develop the inversion model. Here, the inversion model does a comparative analysis between known models starting with the assumption that the earth is homogeneous. I used the forward model to generate the type of data that would be measured in the field.

In carrying out the modeling it was assumed that the geology is uniform, i.e., the matrix material is homogenous, has a high porosity, and the only variation is in the nature of the pore fluids. The solution to forward modeling is found through discretization, the process of assigning each cell on a grid spatial coordinates and values of conductivity or resistivity. Every survey must be planned accordingly before going into the fields. Geophysical methods locate boundaries with contrast in physical properties. It is important to note which method yields an anomaly and/or give a geophysical signature response. Two electrode configurations (dipole-dipole and Wenner) were used in the modeling. Spatial resolution and depth of investigation are determined by electrode spacings and the length of the transect line. Input features were entered into the graphical user interface (GUI) through a series of steps:

a. Select 2D, forward model in the importing tab.
b. Determine electrode spacings (e.g. 50 electrodes, 5-meter spacings)
c. Define a mesh (e.g. unstructured triangular or structured quadrilateral). Quadrilateral mesh is good for condition with infinite boundaries. Triangular meshes are for more complicated geometry (Blanchy *et al.*, 2020). Because of its versatility, for this study the triangular mesh was chosen to infer water properties.
d. Assign resistivity values for each elemental shape based on values discussed in the literature.
e. Select configuration: Wenner (e.g. a=5) and dipole-dipole (e.g. a=5, n=8; 40, etc).
f. Generate synthetic data predicted by the model
g. Add noise to simulate the natural settings of an urban area (e.g. 5%). There are various sources of noise ranging from man-made (e.g. electrical cables, pipes, drains, etc.) to natural sources (e.g. wind, rain, etc.) (Reynolds, 1997).
h. Invert the data and interpret the model.

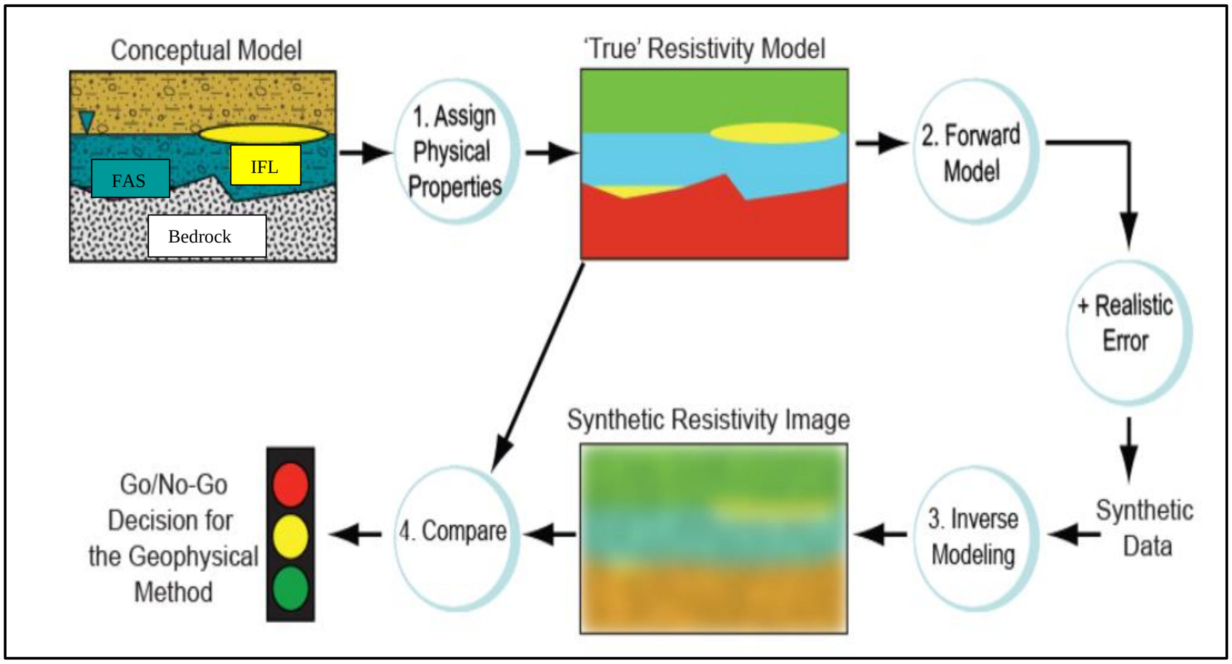

Figure 18. Pre-modeling workflow done within ResIPy software.

Image source: Day-Lewis et al., 2017.

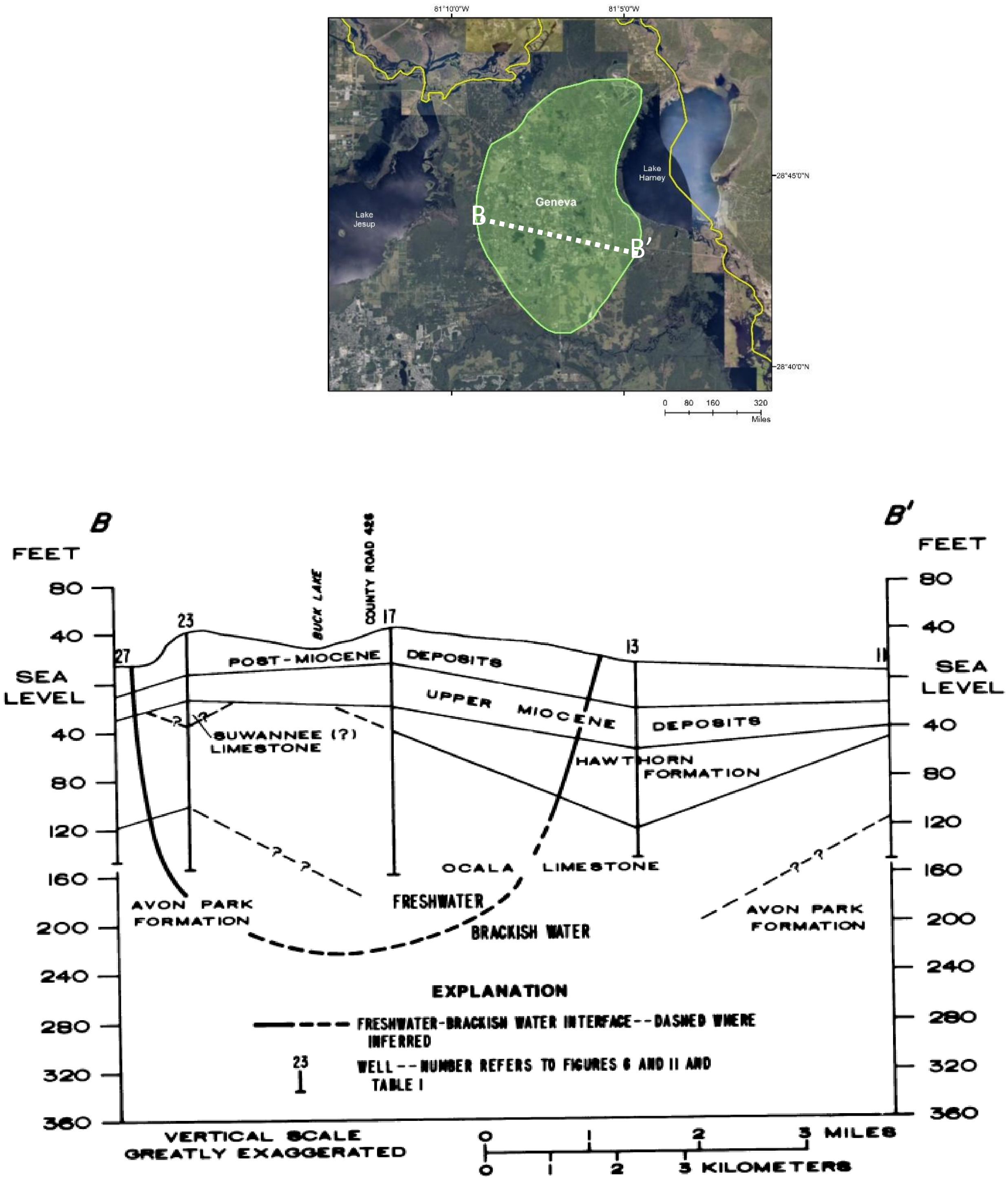

Figure 19. Geological cross section of the study area along line B to B'.
Image source: Phelps & Rohrer, 1987

CHAPTER 5

RESULTS AND DISCUSSION

5.1. WIOA GIS data-driven model

The GIS model classifies Geneva IFL as an area with post-Miocene surficial sediments. In the center of the recharge zone, the surface elevation is 24 feet and varies spatially throughout the area. The confining unit is thinnest in the center of the lens. The LULC consists predominantly of woody wetlands, pasture/hay, evergreen forest, and developed areas where impervious surfaces account for less than 20% of the total land cover. The 30-year mean precipitation is 50 inches/year. Transmissivity ranges from 4080.12 - 22,733.5 ft^2/day; lower transmissivity is in the center of the lens and increases towards the outer perimeter of the lens. The initial model ranks and recognizes the Geneva lens as a moderate potential zone (rank = 4). The lower classification into moderate instead of high IFL potentiality zone could be a result of how each layer was weighted in the WIOA. The determination of the weights for each class is an important part of the integrated analysis, yet there is no standard scale for a simple weighted overlay. Relative importance between the parameters leads to a better representation of the ground surfaces (Samson and Elangovan, 2015).

The GIS-data driven model has a ranking system from 1-6. For this study, rankings are as follows: ranking 1-2 are classified as low potential zones, 3-4 moderate zones, and 5-6 high potential zones. Of the 662 ranked locations, 32 areas were classified as low potential zones (green to yellow), 449 considered moderate potential zones (orange to brown), and the remaining 181 is classified as high potential zones (white). Visually, the highest ranking IFL zones are clustered in the southern part of the study area (Figure 20). A global linear regression, OLS, and a local linear regression, GWR models were used to analyze the spatial patterns of IFL potentiality.

OLS and GWR IFLPM geostatistical models were assessed for overall goodness of fit between observed and predicted using the adjusted coefficient of determination (adjusted R^2). OLS accounts for 47% IFLPM variability, whereas 58% variability was explained by GWR (Table 3). The Akaike information criterion (AIC_c) that helps in comparing the difference in regression models are 1,396 for OLS and 2,156 for GWR models. Even though the linear regression on the explanatory variable explains close to 60% of the variation in the dependent variable within the GWR model, the lower AIC_c is preferred as a means of comparing models. Considering all the IFL rankings, the OLS model shows a significantly clustered distribution pattern ($p < 0.00$, $z = 10.03$). The diagnostic statistics showed that there was no redundancy among explanatory variables (VIF < 7.5), the relationship modeled are not consistent (Koenker (BP): p-value < 0.01)), and lastly, the model predictions are biased and the residuals are not normally distributed (Jarque-Bera Statistic: p-value < 0.01).

Although both models are visually represented, for this study, the OLS model is considered the better model mapped to compare the observed and predicted IFL potential zone. Within this model elevation, the IAS thickness, precipitation, and lithology exhibited statistically significant (p value < .01) for IFL formation within the watershed, followed by land use and land cover, recharge, salinity, and transmissivity. In order of parameter importance listed above, the parameter weights in the WIOA were 20, 20, 10, 10, 10, 20, 5, 5. Thereby, rejecting the null hypothesis that there is no difference in IFL formation factor to predict IFL potentiality. P-values less than .01 means that there is less than a 1 percent chance of seeing these results (or more extreme results), in the world where the null hypothesis is true.

The comparison between observed GIS and predicted OLS and GWR rankings are present in Figures 21-24. The OLS geostatistical prediction, Figure 21, identified fewer high potential zones in the south and identifies the north of the watershed mostly as a low potential zone. The GWR predictions, Figure 23, highlight similar high and medium potential zones throughout the watershed. The model differs in the center where low potentials are inferred as more moderate potential zones. Overall, the slope and y-intercept of both models presents slight over predictions of observed IFL rankings (m= 0.5188, b = 1.86) and (m= 0.5571, b= 1.80) for OLS and GWR, respectively. The coefficient of determination describes a weak relationship between observed and predicted for OLS (R^2 = 0.42) and GWR (R^2 = 0.46).

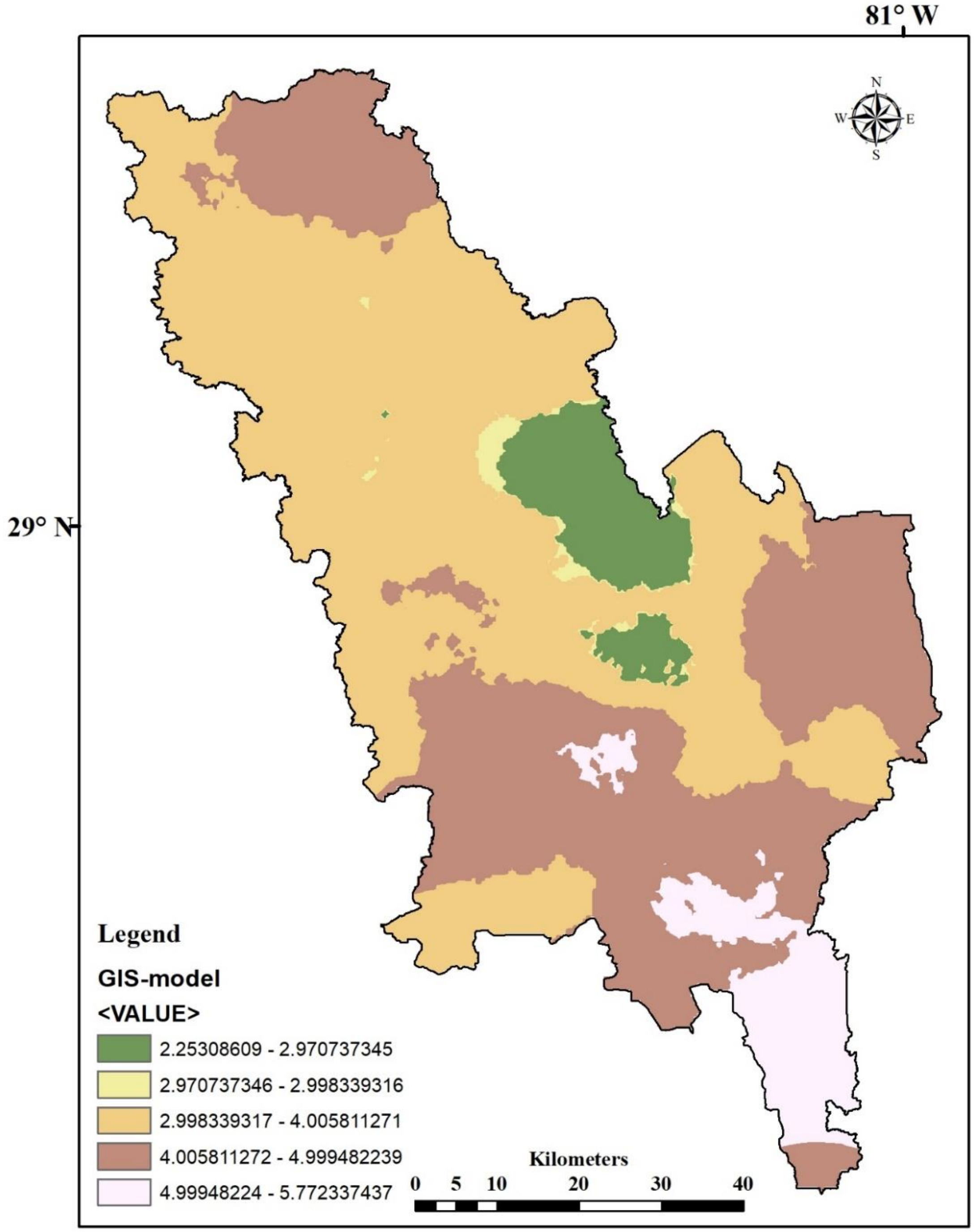

Figure 20. GIS-data driven model output for IFL potential zones within the Middle St. Johns watershed. Areas in green marks areas for low, orange for moderate, and pink for high potential zones.

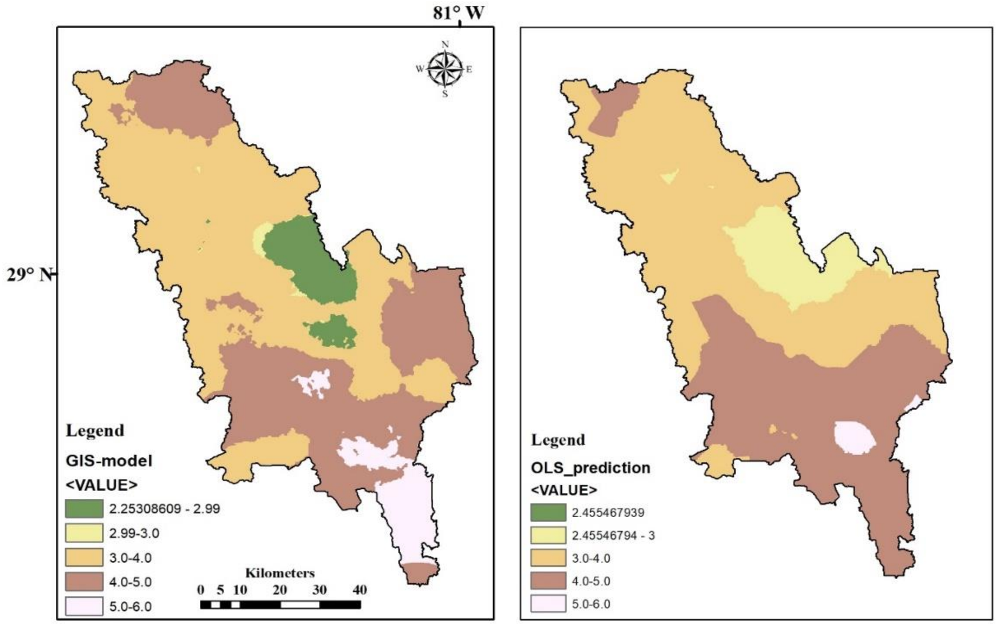

Figure 21. Left to right: GIS model observation and OLS geostatistical predictions.

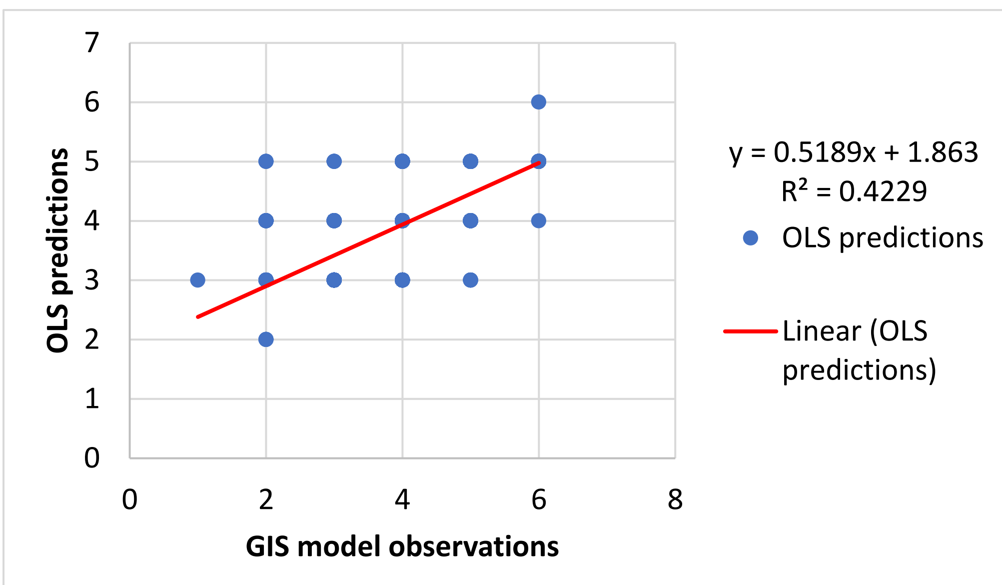

Figure 22. Relationship between GIS model observations and OLS geostatistical predictions.

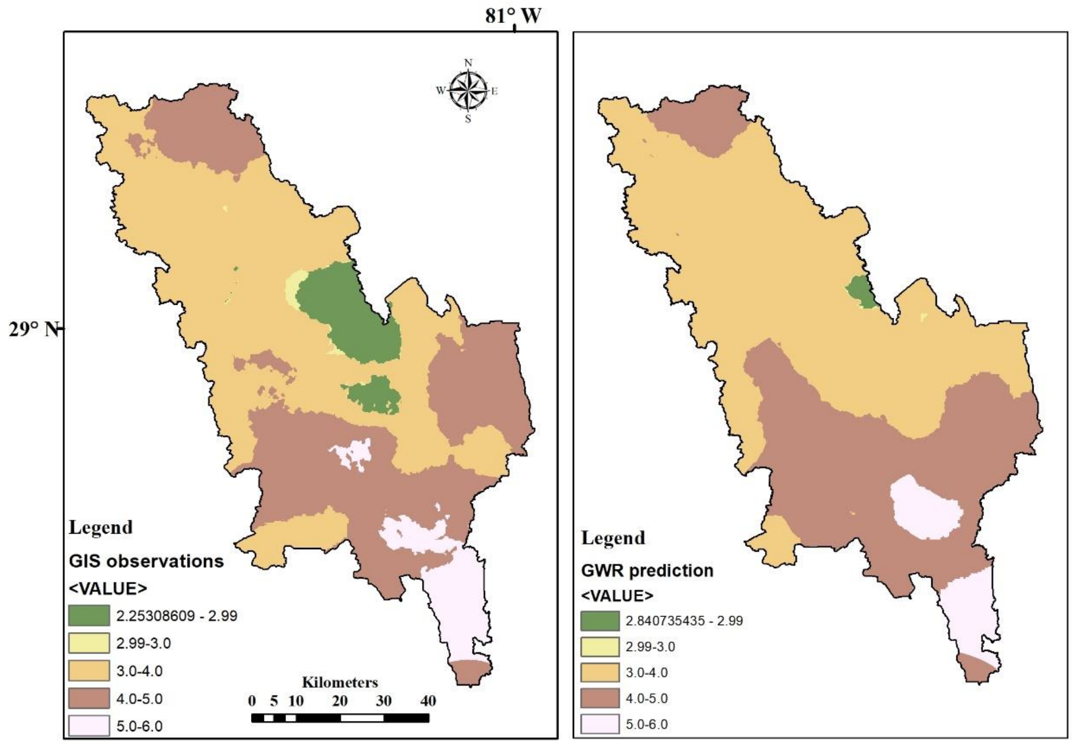

Figure 23. Left to right: GIS model observations and GWR geostatistical predictions.

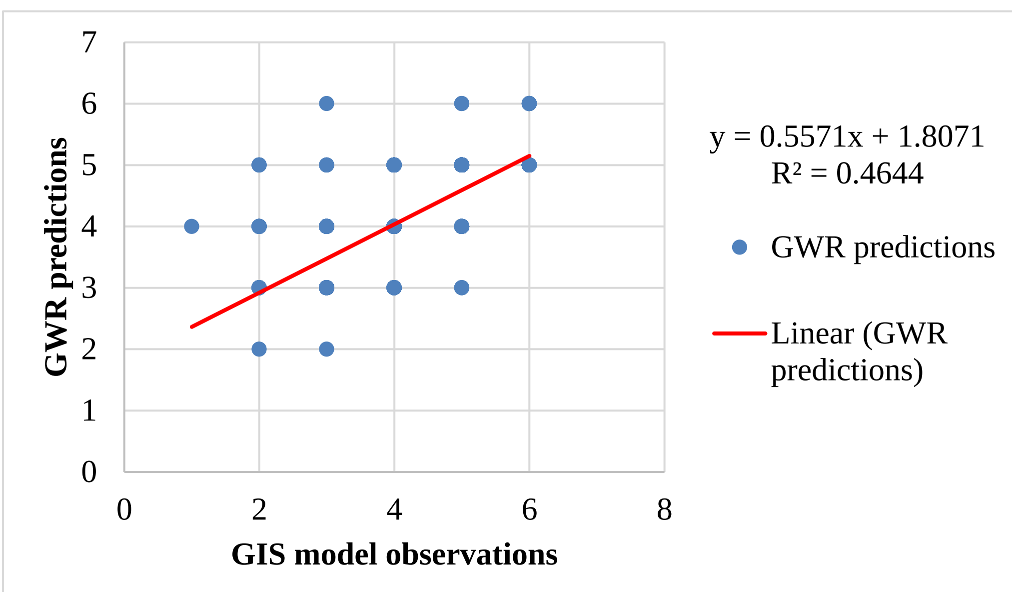

Figure 24. Relationship between GIS model observation and GWR geostatistical prediction.

Table 3. Result summary of the OLS and GWR models. The results from the IFL potential zone with the lowest Akaike Information Criterion (AIC_c) were used to produce the predictive maps.

	IFLPM GIS model
OLS Adjusted R^2	0.47
OLS AIC_C	1396.50
GWR Adjusted R^2	0.58
GWR AIC_C	2156.72

Table 4. OLS independent variable p-value results from the IFL potential map datasets.

Explanatory variables	p-value GIS model	VIF
Elevation	0.00*	2.53
IAS confining layer thickness	0.00*	2.25
Land Use and Land Cover	0.16	1.18
Salinity	0.32	1.13
Precipitation	0.00*	1.49
Transmissivity	0.67	2.04
Lithology	0.00*	1.04
Recharge	0.21	2.24

*An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$)

5.2. Random Forest model

As described by Breiman (2001), the out-of-bag (OOB) rate estimates the general error depending on the number of trees. As seen in Figure 25, the OOB error is a function of trees and reduces as the number of trees are added to the random forest algorithm. The red line shows the error rate when misclassifying low potential zones; gold and green masks moderate potential zones; blue and purple for high potential zone misclassification; pink shows the overall OOB error rate. Based on this analysis, OOB equal to .2124, *mtry* and *ntree* were obtained 3 and 1000, respectively. We also see that with this model, the error rates stabilize and can be truncated at 750 trees for all IFL potential classification.

OOB estimate of the error of 21% means that 79% of the OOB samples were correctly classified by the random forest. The confusion matrix shows how the random forest classified IFL ranking and the error estimates made with each wrong prediction using the training dataset. Of the 466 trained data, 11/14 ranking low potential zones were correctly classified, 270/340 moderate potential zones were correctly classified, and 86/112 high potential zone were correctly classified. The model was validated using the 30% testing dataset; 43/52 were correctly classified as high potential zones, 8/11 for low and 128/135 were classified correctly for moderate potential zone.

Figure 26 shows the formation factors by importance. As depicted, some categorial layers were strong and others were weak. The most influencing conditioning factors on IFL potentiality within the Middle St. Johns watershed were estimated to be confining layer thickness, recharge, elevation, and precipitation. The other variable in decreasing order of importance is land use and land cover, lithology, transmissivity, and salinity. The mean decrease in accuracy for each parameter is as follows: 85.95, 85.56, 72.29, 44.14, 37.06, 35.75, 36.44, 28.68 in order of parameter importance. The parameter weights in the WIOA were 20, 20, 20, 10, 10, 10, 5, 5 in order of decreasing order of importance. This suggest there might be a link between IFL important formation mechanism and how each parameter was weighted in the model.

To examine the efficiency and reliability of the IFLPM using random forest, both the success-rate and prediction-rate curves were calculated. The success-rate curve uses the training dataset to determine how well the resulting IFLPM has classified its ranking. The predictive-rate uses the validated IFLPM dataset to determine how well the model and formation factor to forecast IFL development (Rahmati, Pourghasemi and Melesse, 2016). Figure 27 shows the success-rate curve and the predictive-rate.

For quantitative comparison, the area under the prediction-rate curve was considered. As shown in Figure 27b, the AUC for the prediction-rate of the IFLPM produced by RF was 98.1%. Based on the classification described by Yesilnacar (2005), (90-100%) showed very good accuracy in predicting IFL potential. The results from the RF model can be seen in Figure 28-29. The RF model compared to the initial GIS-model shows similar IFL potentiality ranking throughout the watershed, except in the south and southwestern part of the watershed where IFL rankings are slightly different. Although there are some over predictions in the model (m= 0.8438, b= 0.59) the coefficient of determination describes a strong relationship between observed and predicted (R^2 = 0.77).

Figure 25. Optimization number of trees bases on OOB estimates of the error rate in the RF model.

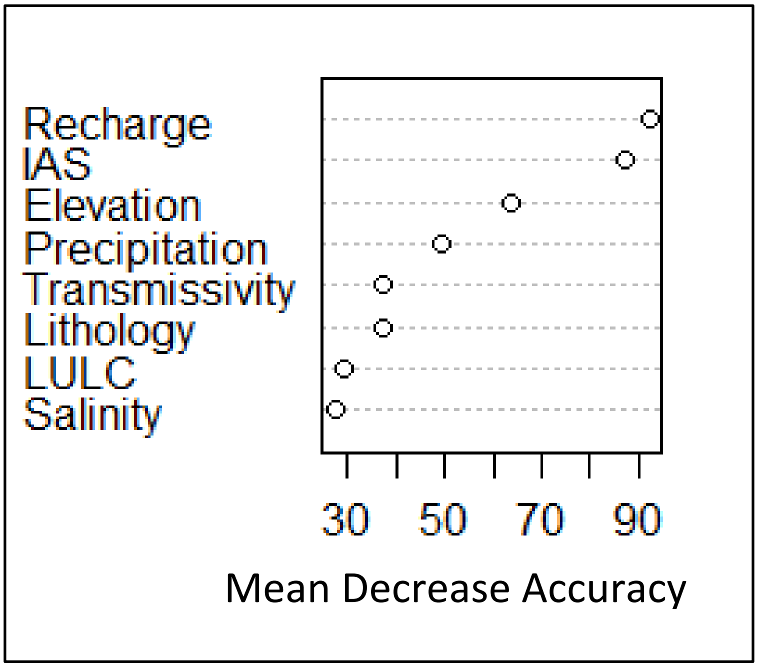

Figure 26. Variable importance derived from the RF model.

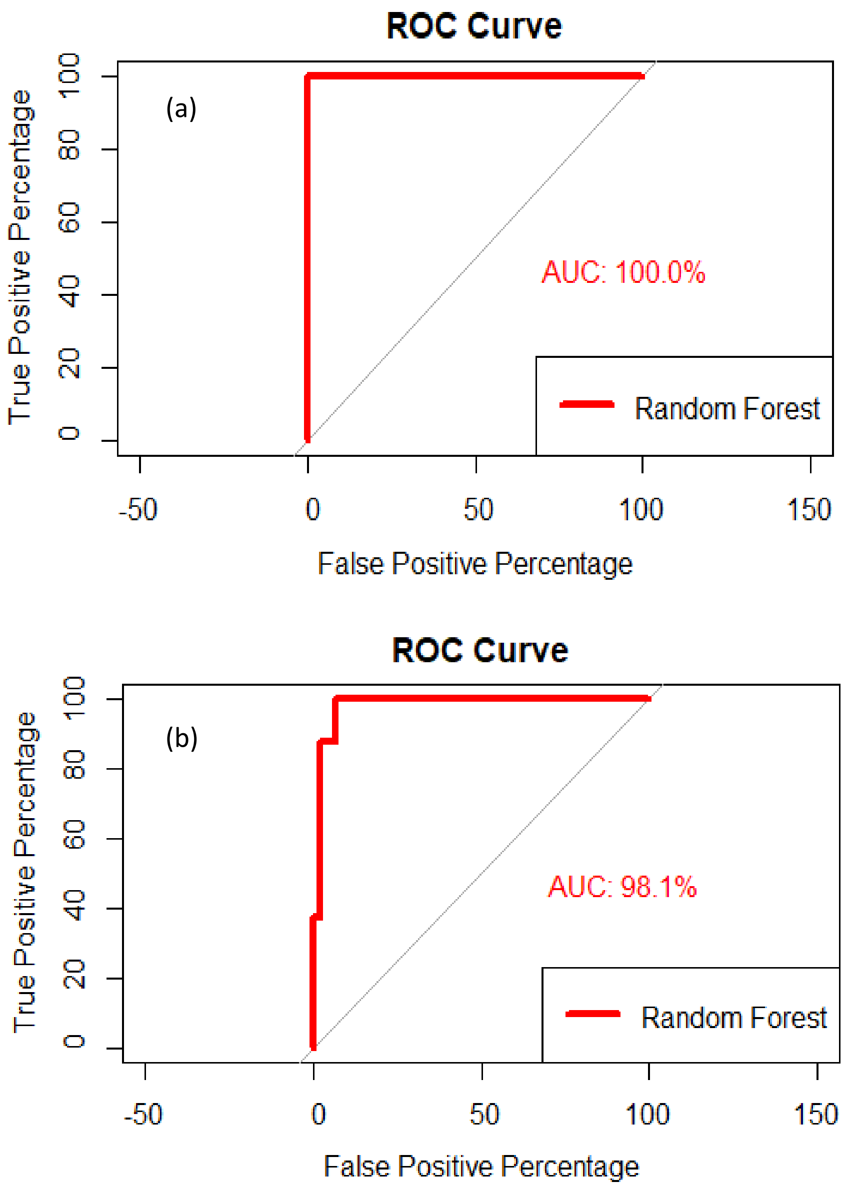

Figure 27. ROC curve: (a) successive rate (b) predictive rate for RF model.

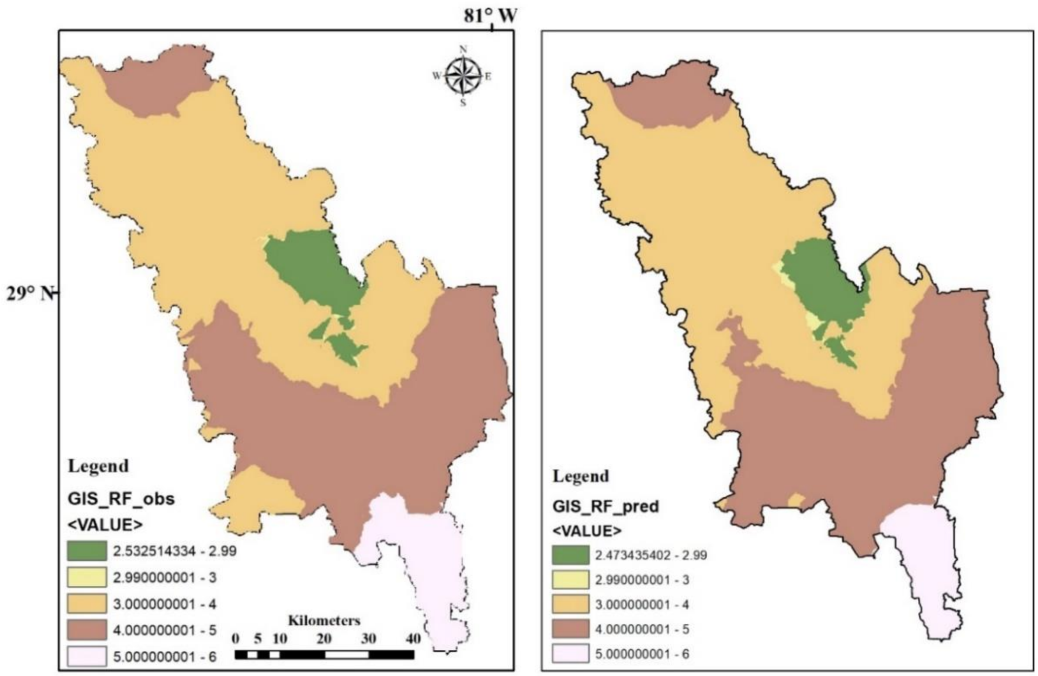

Figure 28. Inland freshwater lens potential map produced by Random Forest model using 30% testing dataset to make predictions.

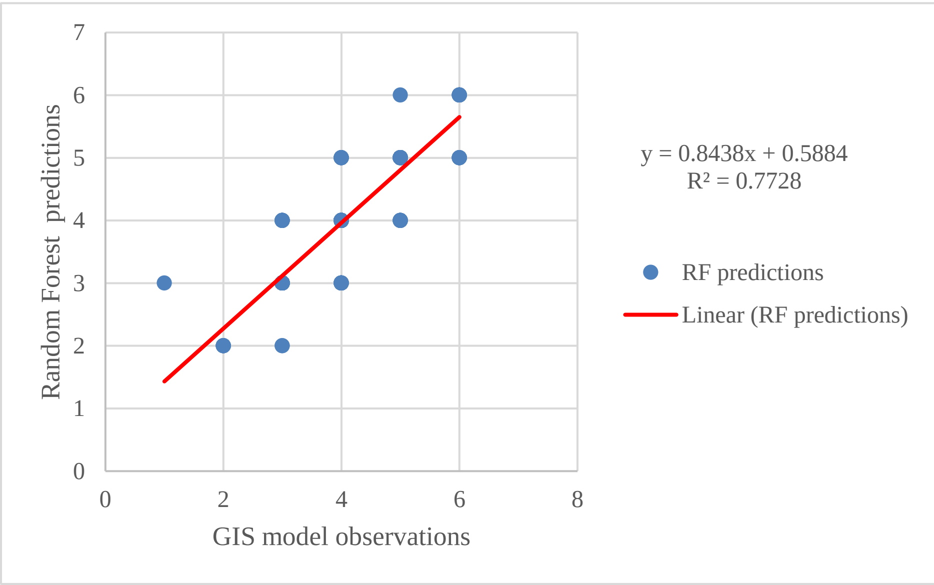

Figure 29. Relationship between GIS model observation and machine learning algorithm, RF predictions.

5.3. Electrical Resistivity

Simple ERT inversion tests were carried out for the four water class boundary layers that were defined within the ERT forward model. These layers are freshwater, brackish water, moderately saline (transitional zone), and saline water. Region 1 is the saline water, region 2 is freshwater, region 3 is brackish water, and region 4 is the transitional zone (Figure 30). Model resistivities of 0.29 Ω-m, 14.29 Ω-m, 5 Ω-m, and 1 Ω-m, were assigned to regions 1, 2, 3, and 4, respectively (Figure 31). The transect line extends to 250 meters, and the depth of investigation for dipole-dipole is 50 meters, whereas the depth of investigation for Wenner is 6 meters below the surface. The difference in maximum depth was controlled by the spacings used for each electrode configuration. In real situations, the resistivity is determined by different lithologies and geological structures. As noted earlier, it was assumed for this study that the geology of the subsurface was uniform, i.e., the matrix material was homogeneous, and the only variation was in the nature of the pore fluid. Apparent resistivity is controlled in part by the electrode configuration (dipole-dipole and Wenner). Also, another important thing to consider concerning the result and the variability in apparent resistivity is the noise recorded by the instrument that alters the quality of the data. The data presented is at a 5% noise level, accounting for the fact that this water class is in an urban area. The results of the inversion can be seen in Figure 32. Freshwater (yellow), is the more resistive material, and resistivity decreases across the water classes, reaching its lowest value within the saline class (purple). The transition zone is distinguishable in the inverted data at 50-meter depth. Similar results are displayed for Wenner, however, the water classes are stretched to better resolve horizontal features. For the Wenner array results, only one value of electrode spacing (5 meters) was used, therefore, variations in depth could not be resolved. Moreover, larger values of spacing would allow greater depths of penetration. Scaled to 100-meter length (25 electrodes, a = 4), a partial area of the Geneva lens was inverted. The inversion model can be seen in Figure 33. Resistivity values of 20 Ω-m, 10 Ω-m, and 1 Ω-m were assigned to the surficial sandy clay deposit, freshwater lens, and saltwater, respectively. Resistivity values were adopted from (Febriani *et al.*, 2019) and assigned to the 3 layer model. The unsaturated zone was not inferred into the model only the saturated zone was inferred in the model.

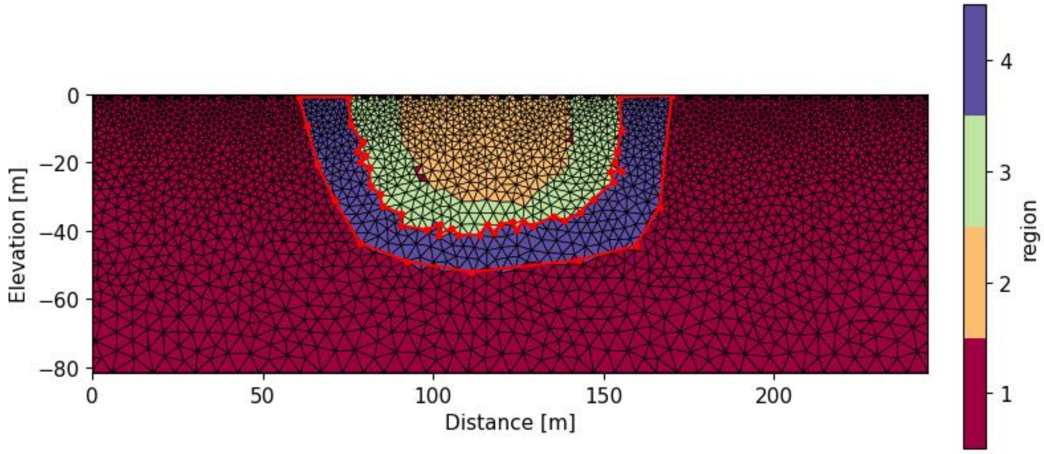

Figure 30. Water class boundaries defined within the geoelectric model. Unsaturated zone not included (water table very close to the surface).

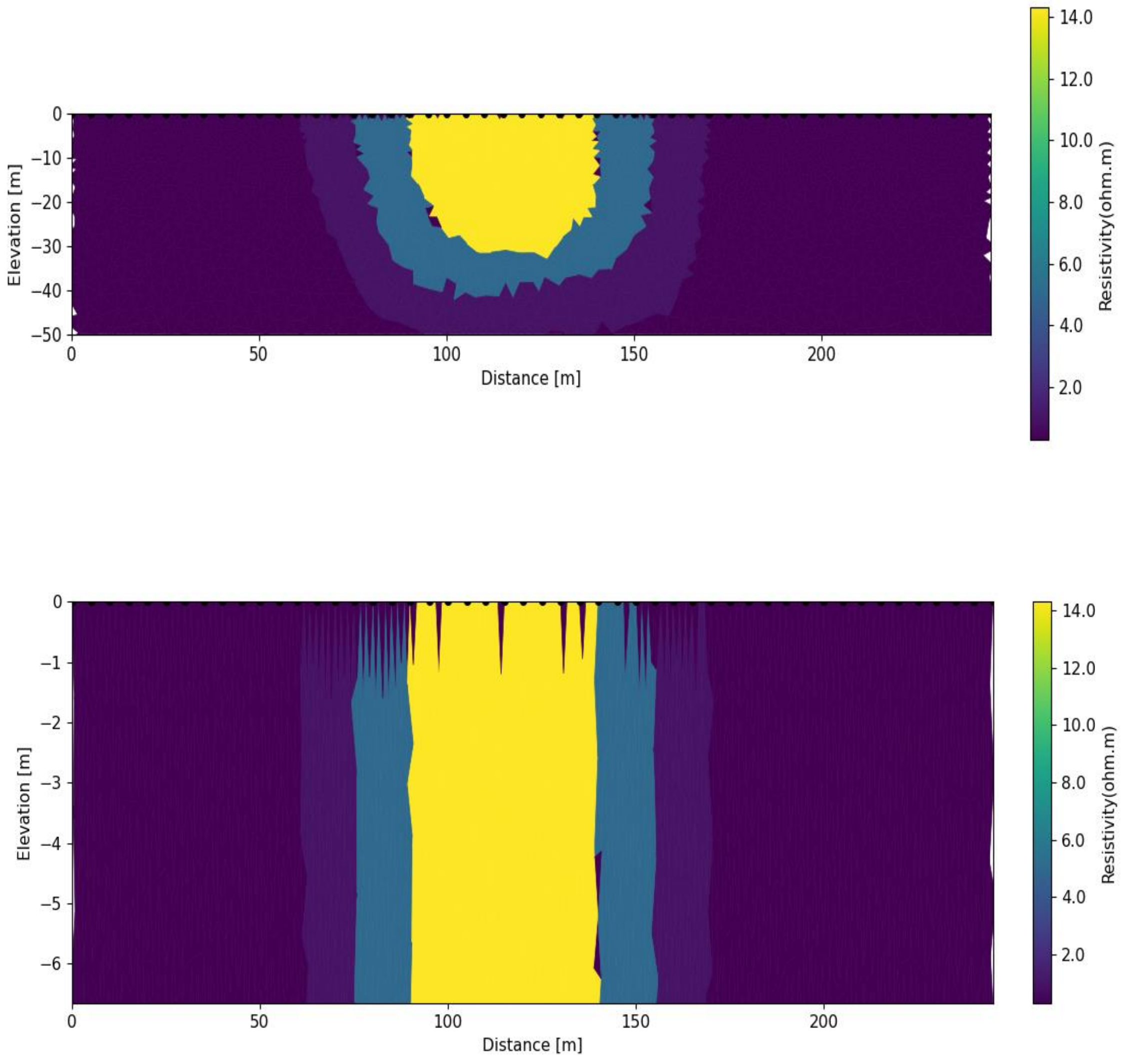

Figure 31. Modeled resistivity values assigned to water classes as they relate to a one-dimensional and homogenous Earth. The top image shows the model used for the dipole-dipole array and the bottom image shows the model used for the Wenner array.

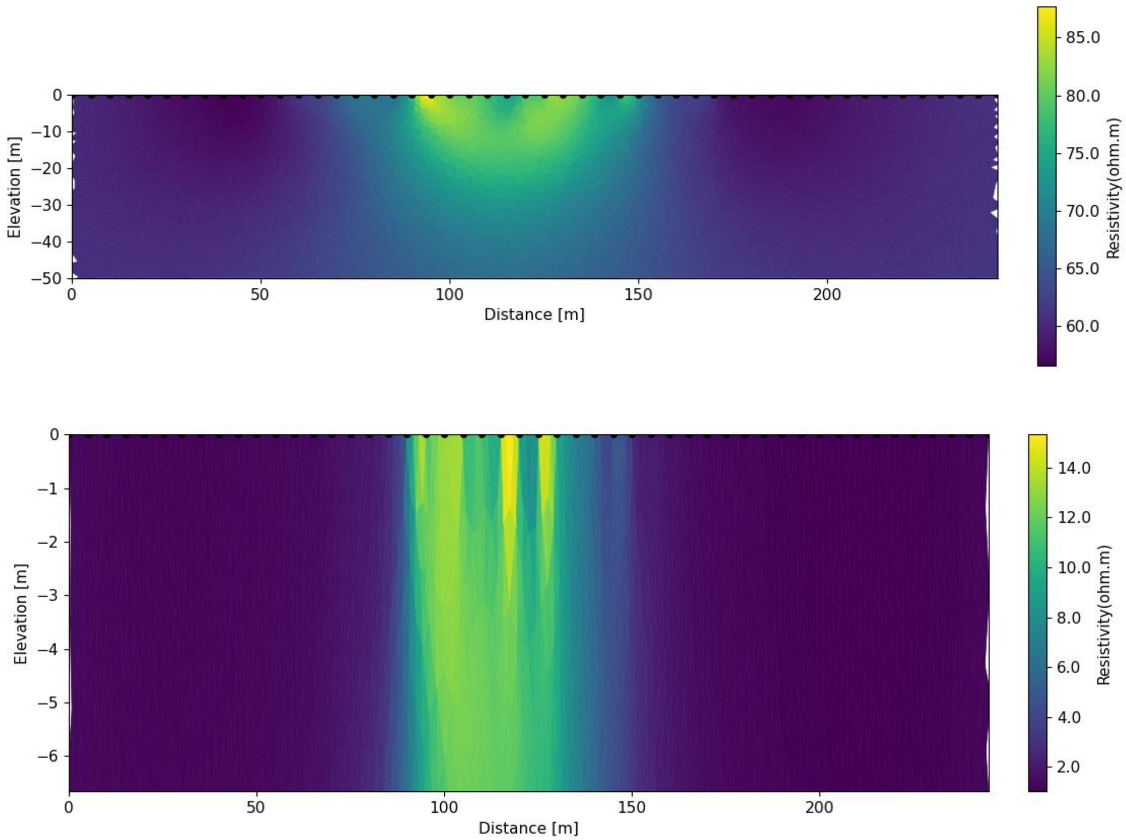

Figure 32. Inverted true resistivity values for water classes for dipole-dipole (top) and Wenner array (bottom).

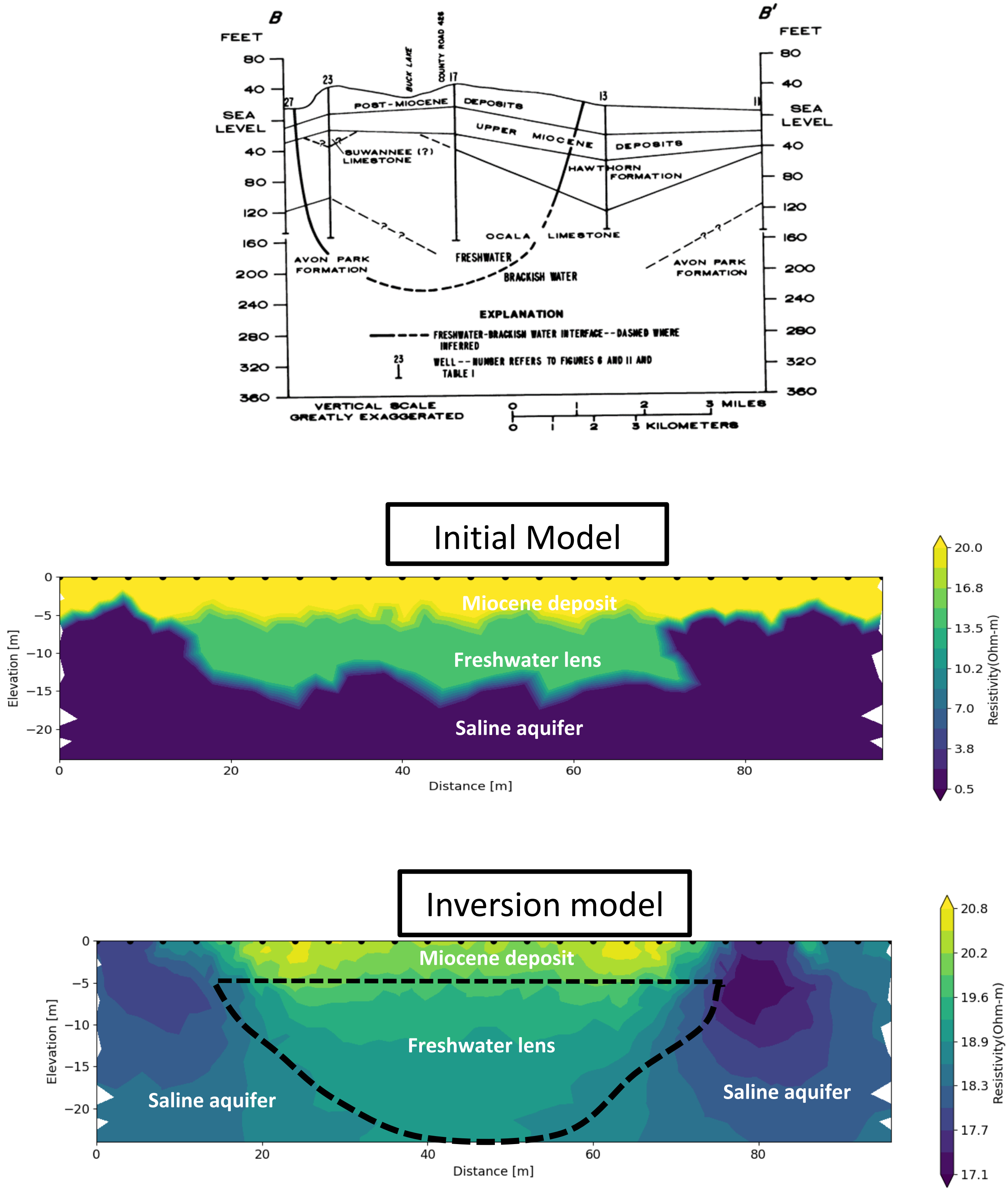

Figure 33. Wenner configuration on part of the Geneva lens (model is not drawn to scale).

5.4. Future work

From the results, the higher weighted parameters were identified by the model as the most contributing factor to IFL development within a coastal setting. Future work should be done to evaluate if the model would still identify these same parameters as predominant if equal weights were assigned.

Regarding geophysical modeling, more values of *n* for the dipole-dipole configuration should be included in the model to resolve both depth and lateral variation. The goal would be to determine the minimum thickness that can be resolved over various depth range. Due to the complexity and variability of pore networks, carbonate rocks of similar porosity may display a wide range of electrical resistivity. Therefore, it’s suggested to assign different resistivity values for the different boundary layer using existing literature. Lastly, additional work should be conducted to determine the better model and experimental design to capture the Geneva lens and a high IFL potentiality thickness in the field.

CHAPTER 6

CONCLUSION

Over the past 100 years, global water demand has increased by 600%, and expect to increase another 20-30% by 2050 (Boretti and Rosa, 2019). To address regions with water-related sustainability problems one of the key elements is to characterize and quantify renewable water resources for better water budget management. Freshwater development above brackish or saline groundwater is a valuable resource that can meet current and future water demands. Groundwater potentiality mapping is an area of research that has seen a growing interest over the past decades. With the recent increasing interest in water resources, survey projects relating to groundwater problems must be carried out. For the first time, this research mapped groundwater potential in the form of IFL potentiality or occurrences within a complex coastal aquifer system, determined the most important settings conducive for IFL development, and demonstrated geophysical capabilities to confirm and quantify the amount of water available for local use. Various approaches have been adopted for this research that has been used by numerous researchers in groundwater potential mapping and geophysical data acquisition. This thesis integrated a GIS-based model and machine learning random forest for inland freshwater lens potential mapping (IFLPM) within the Middle St. Johns watershed, Seminole County, Florida. To produce IFLPM, the first step was the selection and preparation of IFL conditioning factor data sets (e.g. recharge zones, elevation, IAS thickness, precipitation, land use and land cover, lithology, total dissolved solids, transmissivity) that affects IFL potential and the formation of an existing IFL within the same watershed. Then, the IFL data were randomly split into a training dataset 70% (e.g. 466) for training the model the remaining 30% (e.g. 196) was used for validating purposes. Using the mentioned conditioning factors, IFLPM was analyzed using GIS-predictive regression models (OLS and GWR) and RF model. All results were plotted in the GIS-environment. From the initial GIS observations recollected from remote sensing data, predictive models OLS, GWR, and RF identified the highest potential zone for IFL potential is in the southern part of the watershed. The sensitivity analysis identified by the OLS model is elevation, IAS thickness, precipitation, and lithology as statistically significant, whereas, the RF model identified IAS thickness, recharge zones, elevation, and salinity. For RF, percent relative to the decrease of AUC values are 65.47, 54.12, 56.24, 41.52, respectively. It is shown that the RF model showed better predictive performance than both OLS and GWR models. Although there were no redundancy among explanatory variables in both models, the models exhibit weak relationships between observed and predicted rankings, whereas RF, a nonparametric test, showed a strong correlation. Additionally, RF is easy to code and compute. To run a success RF model, it is important to first identify the two parameters (e.g. n_{tree} and m_{try}) for model optimization.

The results of this current study to determine how potential and susceptible groundwater mapping in conjunction with a minimally invasive technique can be of value to water resource managers to meet current and future water demands. Postulate untapped groundwater given certain hydrogeological parameters to promote stability and growth through localized water resources can be done using applied science. Additionally, sensing water remotely, cuts down the cost it takes to drill wells in order locate groundwater. Understanding favorable conditions for inland freshwater lens development is vital for all communities worldwide, especially in areas where the groundwater is or to ever become saline. Insights gained from this research can show applied science being used to postulate untapped groundwater given certain hydrogeological parameters to promote stability and growth through localized water resources.

REFERENCES

AlAli, E. H. (2008) 'Groundwater history and trends in Kuwait', in Sustainable Irrigation Management, Technologies and Policies II. Southampton, UK: WIT Press (WIT Transactions on Ecology and the Environment), pp. 153–164. doi: 10.2495/SI080161.

Allen, G. M. and Main, M. B. (2015) Florida's Geological History 1. Available at: http://edis.ifas.ufl.edu. (Accessed: 23 May 2020).

Archie, G. E., 1942, The electrical resistivity log as an aid in determining some reservoir characteristics: Journal of Petroleum Technology, v. 5, p. 1–8

Asghar, M.N., Prathapar, S., Shafique, M., 2002. Extracting relatively-fresh groundwater from aquifers underlain by salty groundwater. Agric. Water Manag. 52 (2), 119–137. http://dx.doi.org/10.1016/S0378-3774(01)00130-5.

Barraclough, J.T., 1962. Ground-water resources of Seminole County, Florida. Fla. Geol. Surv. Rep. Invest., No. 27, 91 pp. and 10 sheets.

Barrett, B. et al. (2002) 'Geophysical methods in saline groundwater studies: Locating perched water tables and fresh-water lenses', Exploration Geophysics. Taylor & Francis, 33(2), pp. 115–121. doi: 10.1071/EG02115.

Bauer, P. et al. (2006) 'Coupled flow and salinity transport modelling in semi-arid environments: The Shashe River Valley, Botswana', Journal of Hydrology. Elsevier, 316(1–4), pp. 163–183. doi: 10.1016/j.jhydrol.2005.04.018.

Bear J, Cheng A-HD, Sorek D, Ouazar DS, Herrera I (2010) Seawater intrusion in coastal aquifers concepts, methods and practices. Springer, Berlin, 625 pp

Binley, A. 2019. R2 (Version 3.3). Lancaster University website: www.es.lancs.ac.uk/people/amb/Freeware/R2/R2.htm.

Blanchy, G. et al. (2020) 'ResIPy, an intuitive open source software for complex geoelectrical inversion/modeling', Computers and Geosciences. Elsevier Ltd, 137, p. 104423. doi: 10.1016/j.cageo.2020.104423.

Boretti, A. and Rosa, L. (2019) 'PERSPECTIVE Reassessing the projections of the World Water Development Report'. doi: 10.1038/s41545-019-0039-9.

Breiman, L. (2001) Random Forests.

Breiman L, Freidman J, Olshen R, Stone C. 1984. Classification and regression trees. Belmont (CA): Wadsworth, 358 p.

Bugg, S. F. and Lloyd, J. W. (1976) A STUDY OF FRESH WATER LENS CONFIGURATION IN THE CAYMAN ISLANDS USING RESISTIVITY METHODS, Jl Enffng Geol. Available at: https://gsw.silverchair-cdn.com/gsw/Content_public/Journal/qjegh/9/4/10.1144_GSL.QJEG.1976.009.04.02/2/qjegh_009_004_0291.pdf?Expires=2147483647&Signature=arP3svafoKH5Ss74VounRQJQybXoKPZwuekNOXXBYUwmM~AIlnTQRsMAjd3Mlw9WrkxUcbwKHJC4HJxD6a1UoK5JGKefJWqF0HuH (Accessed: 22 August 2018).

Cellone, F., Tosi, L. and Carol, E. (2018) 'Estimating the freshwater-lens reserve in the coastal plain of the middle Río de la Plata Estuary (Argentina)', Science of the Total Environment. Elsevier B.V., 630, pp. 357–366. doi: 10.1016/j.scitotenv.2018.02.236.

Chafetz, H. S., McIntosh, A. G. and Rush, P. F. (1988) 'Freshwater phreatic diagenesis in the marine realm of recent Arabian Gulf carbonates', Journal of Sedimentary Petrology, 58(3), pp. 433–440. doi: 10.1306/212F8DAF-2B24-11D7-8648000102C1865D.

Chang-Jo, F., Fabbri, A.G., 2003. Validation of spatial prediction models for landslide hazard mapping. Nat. Hazards 30 (3), 451–472

Chongo, M. et al. (2015) 'Mapping localised freshwater anomalies in the brackish paleo-lake sediments of the Machile-Zambezi Basin with transient electromagnetic sounding, geoelectrical imaging and induced polarisation'. doi: 10.1016/j.jappgeo.2015.10.002.

Day-Lewis, F. D. et al. (2017) 'An overview of geophysical technologies appropriate for characterization and monitoring at fractured-rock sites', Journal of Environmental Management. Academic Press, 204(Pt 2), pp. 709–720. doi: 10.1016/j.jenvman.2017.04.033.

Elbeih, S. F. (2015) 'An overview of integrated remote sensing and GIS for groundwater mapping in Egypt', Ain Shams Engineering Journal. Elsevier, 6(1), pp. 1–15. doi: 10.1016/J.ASEJ.2014.08.008.

Elewa, H. H. and Qaddah, A. A. (2011) 'Groundwater potentiality mapping in the Sinai Peninsula, Egypt, using remote sensing and GIS-watershed-based modeling', Hydrogeology Journal. Springer-Verlag, 19(3), pp. 613–628. doi: 10.1007/s10040-011-0703-8.

Fallatah, O. A. et al. (2019) 'Assessment of modern recharge to arid region aquifers using an integrated geophysical, geochemical, and remote sensing approach', Journal of Hydrology. Elsevier B.V., 569, pp. 600–611. doi: 10.1016/j.jhydrol.2018.09.061.

Febriani, Y. et al. (2019) 'Determination of Groundwater Using Geoelectric Methods: Schlumberger Configuration in Rokan Hulu Regency', Jurnal Ilmiah Pendidikan Fisika Al-

Biruni. Raden Intan State Islamic University of Lampung, 8(2), pp. 141–152. doi: 10.24042/jipfalbiruni.v0i0.3918.

Fetter, C.W., 1972. Position of the saline water interface beneath oceanic islands. Water Resour. Res. 8, 1307-1315.

Fetter, C. W. (Charles W. (2001) Applied hydrogeology. Prentice Hall.

Fratesi, B. (2013) 'Hydrology and geochemistry of the freshwater lens in coastal karst', in Coastal Research Library. Springer, pp. 59–75. doi: 10.1007/978-94-007-5016-6_3.

Garing, C. et al. (2014) 'Electrical and flow properties of highly heterogeneous carbonate rocks', AAPG Bulletin. American Association of Petroleum Geologists, 98(1), pp. 49–66. doi: 10.1306/05221312134.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T., 2006. Methods and uncertainties in bioclimatic envelope modeling under climate change. Prog. Phys. Geogr. 30, 751–777.

Hine, A. C. (2008) Geology of Florida Summary of Content. Available at: www.cengagecustom.com (Accessed: 23 May 2020).

Hinton, J. C. (1996) 'GIS and remote sensing integration for environmental applications', International Journal of Geographical Information Systems. Taylor and Francis Ltd, 10(7), pp. 877–890. doi: 10.1080/02693799608902114.

Houben, G. et al. (no date) 'The freshwater lens of Benjamín Aceval, Chaco, Paraguay: a terrestrial analogue of an oceanic island lens'. doi: 10.1007/s10040-014-1169-2.

Kim, J. R. and McCuen, R. H. (1979) ‘FACTORS FOR PREDICTING COMMERCIAL WATER USE’, Journal of the American Water Resources Association, 15(4), pp. 1073–1080. doi: 10.1111/j.1752-1688.1979.tb01085.x.

Kuniansky, E.L., Bellino, J.C., and Dixon, J.F., 2012, Transmissivity of the Upper Floridan aquifer in Florida and parts of Georgia, South Carolina, and Alabama: U.S. Geological Survey Scientific Investigations Map 3204, 1 sheet, scale 1:100,000, available only at https://pubs.usgs.gov/sim/3204.

Kwarteng, A. Y. et al. (2000) ‘Formation of fresh ground-water lenses in northern Kuwait’, Journal of Arid Environments, 46(2), pp. 137–155. doi: 10.1006/jare.2000.0666.

Laattoe, T. et al. (2017) ‘Terrestrial freshwater lenses: Unexplored subterranean oases’, Journal of Hydrology, 553, pp. 501–507. doi: 10.1016/j.jhydrol.2017.08.014.

Langford, R. P., Rose, J. M. and White, D. E. (2009) ‘Groundwater salinity as a control on development of eolian landscape: An example from the White Sands of New Mexico’, Geomorphology. Elsevier, 105(1–2), pp. 39–49. doi: 10.1016/j.geomorph.2008.01.020.

Madani, A. and Niyazi, B. (2015) ‘Groundwater potential mapping using remote sensing techniques and weights of evidence GIS model: a case study from Wadi Yalamlam basin, Makkah Province, Western Saudi Arabia’, Environmental Earth Sciences. Springer Verlag, 74(6), pp. 5129–5142. doi: 10.1007/s12665-015-4524-2.

Maupin, M. A. and Barber, N. L. (2015) Estimated Withdrawals from Principal Aquifers in the United States, 2000. Available at: http://www.usgs.gov/ (Accessed: 21 May 2020).

Mekonnen, M. M. and Hoekstra, A. Y. (2016) ‘Sustainability: Four billion people facing severe water scarcity’, Science Advances. American Association for the Advancement of Science, 2(2). doi: 10.1126/sciadv.1500323.

Meyer, R., Engesgaard, P. and Sonnenborg, T. O. (2019) ‘Origin and Dynamics of Saltwater Intrusion in a Regional Aquifer: Combining 3-D Saltwater Modeling With Geophysical and Geochemical Data’, Water Resources Research. Blackwell Publishing Ltd, 55(3), pp. 1792–1813. doi: 10.1029/2018WR023624.

Milewski, A. et al. (2014) ‘Geologic and hydrologic settings for development of freshwater lenses in arid lands’, Hydrological Processes, 28(7), pp. 3185–3194. doi: 10.1002/hyp.9823.

Naghibi, S. A., Ahmadi, K. and Daneshi, A. (2017) ‘Application of Support Vector Machine, Random Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping’, Water Resources Management. Springer Netherlands, 31(9), pp. 2761–2775. doi: 10.1007/s11269-017-1660-3.

Naghibi, S. A. and Pourghasemi, H. R. (2015) ‘A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping’, Water Resources Management. Kluwer Academic Publishers, 29(14), pp. 5217–5236. doi: 10.1007/s11269-015-1114-8.

Nampak, H., Pradhan, B. and Manap, M. A. (2014) ‘Application of GIS based data driven evidential belief function model to predict groundwater potential zonation’, Journal of Hydrology. Elsevier, 513, pp. 283–300. doi: 10.1016/j.jhydrol.2014.02.053.

Oliveira, S. et al. (2012) ‘Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest’, Forest Ecology and Management. Elsevier, 275, pp. 117–129. doi: 10.1016/j.foreco.2012.03.003.

Panday, S. et al. (1993) ‘A density-dependent flow and transport analysis of the effects of groundwater development in a freshwater lens of limited areal extent: The Geneva area (Florida, U.S.A.) case study’, Journal of Contaminant Hydrology. Elsevier, 12(4), pp. 329–354. doi: 10.1016/0169-7722(93)90004-C.

Parson’s Corporation (1961) Fresh ground-water resources, Rawdatain area, Kuwait: final report. Parson’s, Los Angeles.

Phelps G, Rohrer K (1987) Hydrogeology in the area of a freshwater lens in the Floridan aquifer system, northeast Seminole County, Florida. US Geol Surv Water Resour Invest Rep 96-4181

Prasad, A. M., Iverson, L. R. and Liaw, A. (2006) ‘Newer classification and regression tree techniques: Bagging and random forests for ecological prediction’, Ecosystems. Springer, 9(2), pp. 181–199. doi: 10.1007/s10021-005-0054-1.

Rahmati, O., Pourghasemi, H. R. and Melesse, A. M. (2016) ‘Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran’, Catena, 137, pp. 360–372. doi: 10.1016/j.catena.2015.10.010.

Raji WO (2014) Review of electrical and gravity methods of near-surface exploration for groundwater. Niger J Technol Development 11(2):31–38

Razandi, Y. et al. (2015) ‘Application of analytical hierarchy process, frequency ratio, and certainty factor models for groundwater potential mapping using GIS’, Earth Science Informatics. Springer Verlag, 8(4), pp. 867–883. doi: 10.1007/s12145-015-0220-8.

Reese, R.S., and Wacker, M.A., 2009, Hydrogeologic and Hydraulic Characterization of the Surficial Aquifer System, and Origin of High Salinity Groundwater, Palm Beach County, Florida: U.S. Geological Survey Scientific Investigations Report 2009–5113, 83 p.

Reynolds, J. M. (1997) An introduction to applied and environmental geophysics, An introduction to applied and environmental geophysics. doi: 10.1071/pvv2011n155other.

Rodriguez-Galiano, V. F. et al. (2012) 'Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture', Remote Sensing of Environment. Elsevier, 121, pp. 93–107. doi: 10.1016/j.rse.2011.12.003.

Rotz, R., Milewski, A. and Rasmussen, T. C. (2020) 'Transient Evolution of Inland Freshwater Lenses: Comparison of Numerical and Physical Experiments', Water. MDPI AG, 12(4), p. 1154. doi: 10.3390/w12041154.

Rotz, R. R. and Milewski, A. M. (2019) 'Physical modeling of inland freshwater lens formation and evolution in drylands', Hydrogeology Journal. Springer Berlin Heidelberg, pp. 1–14. doi: 10.1007/s10040-019-01940-1.

Salem, H. S. and Chilingarian, G. V (1999) The cementation factor of Archie's equation for shaly sandstone reservoirs, Journal of Petroleum Science and Engineering. Available at: www.elsevier.comrlocaterjpetscieng (Accessed: 19 July 2020).

Sameen, M. I., Pradhan, B. and Lee, S. (2019) 'Self-Learning Random Forests Model for Mapping Groundwater Yield in Data-Scarce Areas', Natural Resources Research. Springer New York LLC, 28(3), pp. 757–775. doi: 10.1007/s11053-018-9416-1.

Samson, S. and Elangovan, K. (2015) 'Delineation of Groundwater Recharge Potential Zones in Namakkal District, Tamilnadu, India Using Remote Sensing and GIS', Journal of the Indian

Society of Remote Sensing. Springer India, 43(4), pp. 769–778. doi: 10.1007/s12524-014-0442-0.

Schneider, J. C. and Kruse, S. E. (2003) ‘A comparison of controls on freshwater lens morphology of small carbonate and siliciclastic islands: Examples from barrier islands in Florida, USA’, Journal of Hydrology. Elsevier, 284(1–4), pp. 253–269. doi: 10.1016/j.jhydrol.2003.08.002.

Smith, S. K. (no date) Florida Population Growth: Past, Present and Future. Available at: https://www.researchgate.net/publication/228869969 (Accessed: 20 July 2020).

Stoeckl, L. and Houben, G. (2012) ‘Flow dynamics and age stratification of freshwater lenses: Experiments and modeling’, Journal of Hydrology. Elsevier, 458–459, pp. 9–15. doi: 10.1016/j.jhydrol.2012.05.070.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240 (4857), 1285–1293

Tang, Q. et al. (2009) ‘Remote sensing: hydrology’, Progress in Physical Geography: Earth and Environment. SAGE PublicationsSage UK: London, England, 33(4), pp. 490–509. doi: 10.1177/0309133309346650.

Temesgen, B., Mohammed, M. U. and Korme, T. (2001) ‘Natural hazard assessment using GIS and remote sensing methods, with particular reference to the landslides in the Wondogenet area, Ethiopia’, Physics and Chemistry of the Earth, Part C: Solar, Terrestrial and Planetary Science. Pergamon, 26(9), pp. 665–675. doi: 10.1016/S1464-1917(01)00065-4.

Tibbals, C.H., 1977. Availability of ground water in Seminole County and vicinity, Florida. U.S. Geol. Surv. Water-Resour. Invest. Rep. 76-97, 15 pp.

Tobin, B. D. and Weary, D. J. (2004) Digital engineering aspects of Karst map: a GIS version of Davies, W.E., Simpson, J.H., Ohlmacher, G.C., Kirk, W.S., and Newton, E.G., 1984, Engineering aspects of Karst: U.S. Geological Survey, National Atlas of the United States of America, Scale 1:7,500,000, Open-File Report. doi: 10.3133/OFR20041352.

Underwood, M. R., Peterson, F. L. and Voss, C. I. (1992) ‘Groundwater lens dynamics of Atoll Islands’, Water Resources Research. John Wiley & Sons, Ltd, 28(11), pp. 2889–2902. doi: 10.1029/92WR01723.

VACHER, H. L. (1988) ‘Dupuit-Ghyben-Herzberg analysis of strip-island lenses’, GSA Bulletin. GeoScienceWorld, 100(4), pp. 580–591. doi: 10.1130/0016-7606(1988)100<0580:dghaos>2.3.co;2.

Viezzoli, A., Auken, E. and Munday, T. (2009) ‘Spatially constrained inversion for quasi 3D modelling of airborne electromagnetic data – an application for environmental assessment in the Lower Murray Region of South Australia’, Exploration Geophysics. Taylor & Francis, 40(2), pp. 173–183. doi: 10.1071/EG08027.

Vörösmarty, C. J. et al. (2010) ‘Global threats to human water security and river biodiversity’, Nature. Nature Publishing Group, 467(7315), pp. 555–561. doi: 10.1038/nature09440.

Wallis, T. N., Vacher, H. L. and Stewart, M. T. (1991) ‘Hydrogeology of freshwater lens beneath a Holocene strandplain, Great Exuma, Bahamas’, Journal of Hydrology. Elsevier, 125(1–2), pp. 93–109. doi: 10.1016/0022-1694(91)90085-V.

Van Weert, F. (2009) Global Overview of Saline Groundwater Occurrence and Genesis (Report number: GP 2009-1) Water Resources Assessment Yemen View project World Water

Assessment Programme (WWAP) View project. Available at: https://www.researchgate.net/publication/304348604 (Accessed: 21 June 2020).

Werner, A. D. and Laattoe, T. (2016) 'Terrestrial freshwater lenses in stable riverine settings: Occurrence and controlling factors'. doi: 10.1002/2015WR018346.

Werner AD, Bakker M, Post VEA, Vandenbohede A, Lu C, Ataie- Ahstiani B, Simmons CT, Barry DA (2012) Seawater intrusion processes, investigation and management: recent advances and future challenges. Adv Water Resour 51:3–26

Willard, D. A. et al. (2007) 'Deglacial climate variability in central Florida, USA', Palaeogeography, Palaeoclimatology, Palaeoecology. Elsevier, 251(3–4), pp. 366–382. doi: 10.1016/j.palaeo.2007.04.016.

Williams, L.J., and Kuniansky, E.L., 2016, Revised hydrogeologic framework of the Floridan aquifer system in Florida and parts of Georgia, Alabama, and South Carolina (ver. 1.1, March 2016): U.S. Geological Survey Professional Paper 1807, 140 p., 23 pls., http://dx.doi.org/10.3133/pp1807.

Yao, Y. et al. (2019) 'Development of fresh groundwater lens in coastal reclaimed islands', Journal of Hydrology. Elsevier B.V., 573, pp. 365–375. doi: 10.1016/j.jhydrol.2019.03.062.

Yesilnacar EK (2005) The application of computational intelligence to landslide susceptibility mapping in Turkey. Ph.D Thesis Department of Geomatics the University of Melbourne, p 423

Young, M. E. et al. (2004) 'Electromagnetic detection of deep freshwater lenses in a hyper-arid limestone terrain', Journal of Applied Geophysics. Elsevier, 57(1), pp. 43–61. doi: 10.1016/J.JAPPGEO.2004.09.002.