5-2018

Geostatistical Analysis of Potential Sinkhole Risk: Examining Spatial and Temporal Climate Relationships in Tennessee and Florida

Kimberly Blazzard
East Tennessee State University

Geostatistical Analysis of Potential Sinkhole Risk: Examining Spatial and Temporal Climate Relationships in Tennessee and Florida

---

A thesis

presented to

the faculty of the Department of Geosciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Geosciences

---

by

Kimberly Blazzard

May 2018

Andrew Joyner, Ph.D., Co-Chair

Ingrid Luffman, Ph.D., Co-Chair

Arpita Nandi, Ph.D.

Keywords: Sinkhole, Doline, GIS, Climate, Predictive Modeling, Karst, Tennessee, Florida, Logistic Regression, MaxEnt, R

ABSTRACT

Geostatistical Analysis of Potential Sinkhole Risk: Examining Spatial and Temporal Climate Relationships in Tennessee and Florida

by

Kimberly Blazzard

Sinkholes are a significant hazard for the southeastern United States. Although differences in climate are known to affect karst environments differently, quantitative analyses correlating sinkhole formation with climate variables is lacking. A temporal linear regression for Florida sinkholes and two modeled regressions for Tennessee sinkholes were produced: a general linearized logistic regression and a MaxEnt derived species distribution model. Temporal results showed highly significant correlations with precipitation, teleconnection patterns, temperature, and CO_2, while spatial results showed highly significant correlations with precipitation, wind speed, solar radiation, and maximum temperature. Regression results indicated that some sinkhole formation variability could be explained by these climatological patterns and could possibly be used to help predict when/where sinkholes may form in the future.

ACKNOWLEDGEMENTS

I would like to share my thanks to my wonderful advisors- Drs. Andrew Joyner, Ingrid Luffman, and Arpita Nandi- who have not only meticulously reviewed my work but have legitimately been concerned with my overall success. I would also like to give my special thanks to fellow graduate students, Laura Rooney, Kingsley Fasesin, Reagan Cornett, and Michael Shoop for supporting and encouraging me through the frustrations of the past two years, thesis-related or not. Lastly, I would like to thank my parents, Gale and Raymond Keeler, who have always pushed me to get an education and to be the best I can be; to them, I dedicate this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Sinkhole Problems

Some of the most sinkhole susceptible areas coincide with areas of low socio-economic status; between the 1970s and 1980s, 38 lives were lost in sinkholes in the Gualeng Province of South Africa (Buttrick et al. 2001). Cooper et al. (2011) outlined additional sinkhole-related issues. Storm runoff can exacerbate sinkhole formation and increase pollution into karstic aquifers, improperly installed heat pumps can allow more pollution into the karst aquifers, bridge spans may collapse in them, drainage ditches commonly cause sinkholes along roads, and leaking pipes, failed sewer and water systems, and poorly designed drainages also contribute to sinkhole development (Veni et al. 2001; Waltham et al. 2005). A study in Pretoria, South Africa determined that 96% of nearly 400 sinkholes were at least influenced by anthropogenic activities (Veni et al. 2001).

Weary and Doctor (2014) approximated that 18% of the United States is underlain by soluble rocks with the potential for karstic formations; these karstic rock types are found within all 50 states. From 2000-2015, the cost of damages due to karst-related incidents in the US averaged $300,000,000 per year (Weary 2015).

Cooper et al. (2011) also indicated that many counties and states within the United States have varying regulations concerning what is required for sinkholes in hazard avoidance and land development. In Florida, the influence of urban and agricultural land use is altering and reducing natural drainage and recharge, increasing the susceptibility of karst aquifer contamination (Tihansky and Knochenmus 2001). More than 100 million dollars in sinkhole-related structural damage is reported every year in Florida (Florea 2008). Numerous insurers either ceased writing

homeowner's policies or tripled insurance premiums by 2006 due to increasing sinkhole claims (Florea 2008). As of 2017, Florida and Tennessee were the only U.S. states to require property insurance companies to offer sinkhole coverage. The 2017 Florida Statute 627.706 requires property insurance to include damages to "Catastrophic ground cover collapse". Tennessee laws required similar until a 2014 amendment of the 2014 Tennessee Code 56-7-130 changed the law to only "make available" sinkhole coverage. In Florida, residential property insurance deductibles may range from 1 – 10% of sinkhole losses within the policy dwelling limits. These amendments and high property deductibles help transfer the cost of damages back on the homeowner saving the insurance company on costs. For example, if a house sustains $100,000 of damages due to a sinkhole, the homeowner could be responsible for a $10,000 deductible. Preventative measures could reduce the financial burden laid on both homeowners and insurance companies.

Formation

Sinkholes naturally form in five main soluble rock types: limestone, dolostone, gypsum, halite (salt), and chalk (Cooper et al. 2011). The focus of this research is on sinkholes in the carbonate (limestone and dolostone) regions of Tennessee and Florida. Karst consists of unique landforms that form from weathering and erosion of soluble rock (Ritter et al. 2011). Karst is typically associated with landforms formed in limestone and dolomite such as caves, sinkholes, and sinking rivers. Limestone is made of the mineral calcite ($CaCO_3$) and dolostone is made of the mineral dolomite ($CaMg(CO_3)_2$). (Note: the terms dolostone and dolomite are often used interchangeably.) As water passes through the atmosphere and through soil, carbon dioxide (CO_2) is dissolved into the water creating a weak carbonic acid. The carbonic acid then dissolves the $CaCO_3$ creating a void (Borsato et al. 2015).

Another term occasionally used for a sinkhole is “doline.” Although Ritter et al. (2011) claim that in North America they are used interchangeably (Ritter et al. 2011), in other locations the terms have different meanings. In South Africa, sinkholes are landforms that form quickly without warning and generally have a diameter less than 100 m, while dolines form slowly over weeks to years and are typically large depressions of 30 m to 1 km in diameter (Buttrick et al. 2001).

Six main types of sinkholes are recognized: solution sinkholes, collapse sinkholes, caprock sinkholes, dropout sinkholes, suffosion sinkholes, and buried sinkholes. Solution sinkholes undergo dissolution at the surface, whereas collapse, caprock, dropout, and buried sinkholes undergo dissolution under the surface of the earth creating a void. Collapse sinkholes form when the surface collapses into the void below. Caprock sinkholes have a cap of insoluble rock that eventually collapses into the void; these sinkholes slowly evolve over more than 10,000 years. Dropout sinkholes form when soil collapses into a soil void. Buried sinkholes have been buried with soil (Ritter et al. 2011). Suffosion sinkholes form with the down-washing of soil into bedrock features.

Sinkhole formation is accelerated by increases in: water, exposed rock surface, CO_2 partial pressure, and temperature. Although calcite and dolomite decrease in solubility with increasing temperature, the solubility increases with increasing CO_2 partial pressure (Ritter et al. 2011) (Fig 1). For example, increasing vegetation will increase the CO_2 levels in water resulting in an increase in soil CO_2 partial pressure. The soil CO_2 partial pressure (pCO_2) is the main determinant in groundwater pCO_2 (Zeng et al. 2016). Increasing the outside air temperature will allow for greater vegetative growth assuming there are sufficient water and nutrients and assuming the plant is not overburdened by excessive heat. Since the temperature of groundwater

is generally more stable than the temperature of the outside air (Winter et al. 1998), an increase in temperature for outside air should ultimately allow for an increase in CO_2 levels dissolved in groundwater.

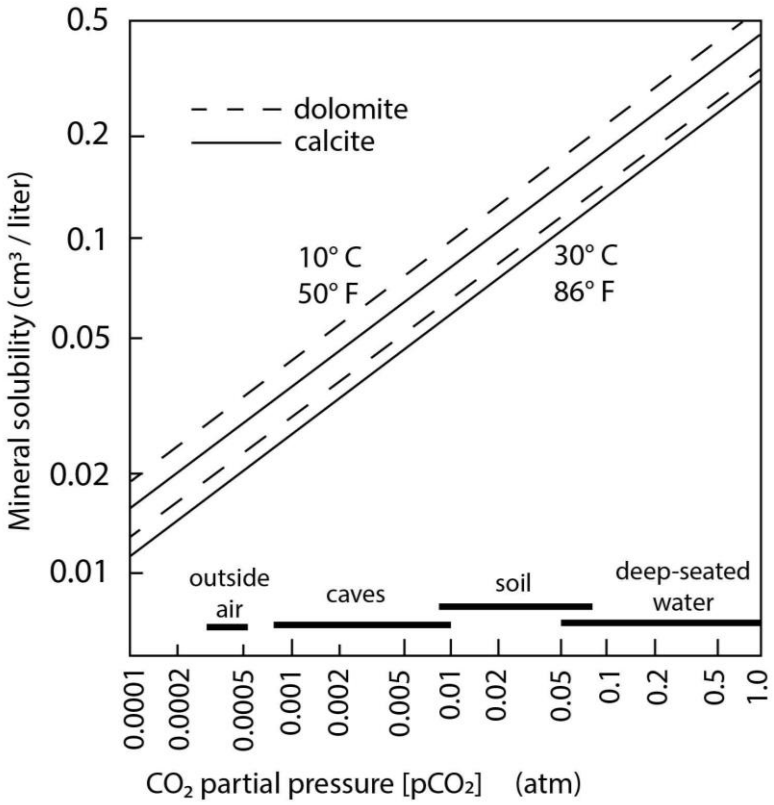

Figure 1: Mineral solubility vs CO_2 partial pressure. (Adapted from Ritter et al., 2011)

Other features affecting sinkhole formation in particular include slope and geologic structures such as bedding-plane partings and fractures (Sweeting 2017). Greater slopes restrict pooling, decreasing the chance for water to dissolve the underlying rocks. In this way, solution sinkhole frequency is inversely proportional to the surface slope (Ritter et al. 2011). Jones and Banner (2003) support this idea with their work on a Barbados aquifer. They found that regions of high sinkhole densities have characteristics of low relief where dry valleys do not have well

developed karst landforms. Geologic structures, in particular fractures and porosity, affect how water infiltrates through the rock. In Barbados, soil permeabilities are much lower than the permeabilities of the underlying Pleistocene limestone. As sinkholes are filled with low-permeability/high porosity clay-rich soils, infiltration rates are slowed, ponding is minimized, and losses due to evapotranspiration are minimized. This leads to better aquifer recharge through karst shafts than through larger interfluvial sinkholes (Tullstrom 1964; Smart and Ketterling 1997; Jones and Banner 2003). In turn, this creates more sinkholes. In addition, sinkholes are more likely to form in dense limestones that are well-jointed; as the grain size increases from micritic to sparitic, the dissolution process is negatively influenced as less grain surface area is exposed to dissolution (Sweeting 2017). The fractures allow for selective dissolution over uniform surficial dissolution.

Collapse dolines are more than just a collapse of sediment into a cave, as dolines can have volumes a hundred times larger than the largest cave systems. As collapse happens, it blocks the routes of water systems allowing for an increased hydraulic gradient and flow velocity. This in turn accelerates the dissolution process in new areas allowing for continued collapse (Gabrovšek and Stepišnik 2011). Two main types of collapse infilling may occur: continuous and discontinuous (Fig 2). Continuous consists of fractures (not at the surface) that maintain a certain limiting aperture by collapsing material into the void at the same rate as is dissolved. Discontinuous consists of fractures that grow in size until a collapse shrinks the aperture back down to a smaller size then restarts growth (Gabrovšek and Stepišnik 2011).

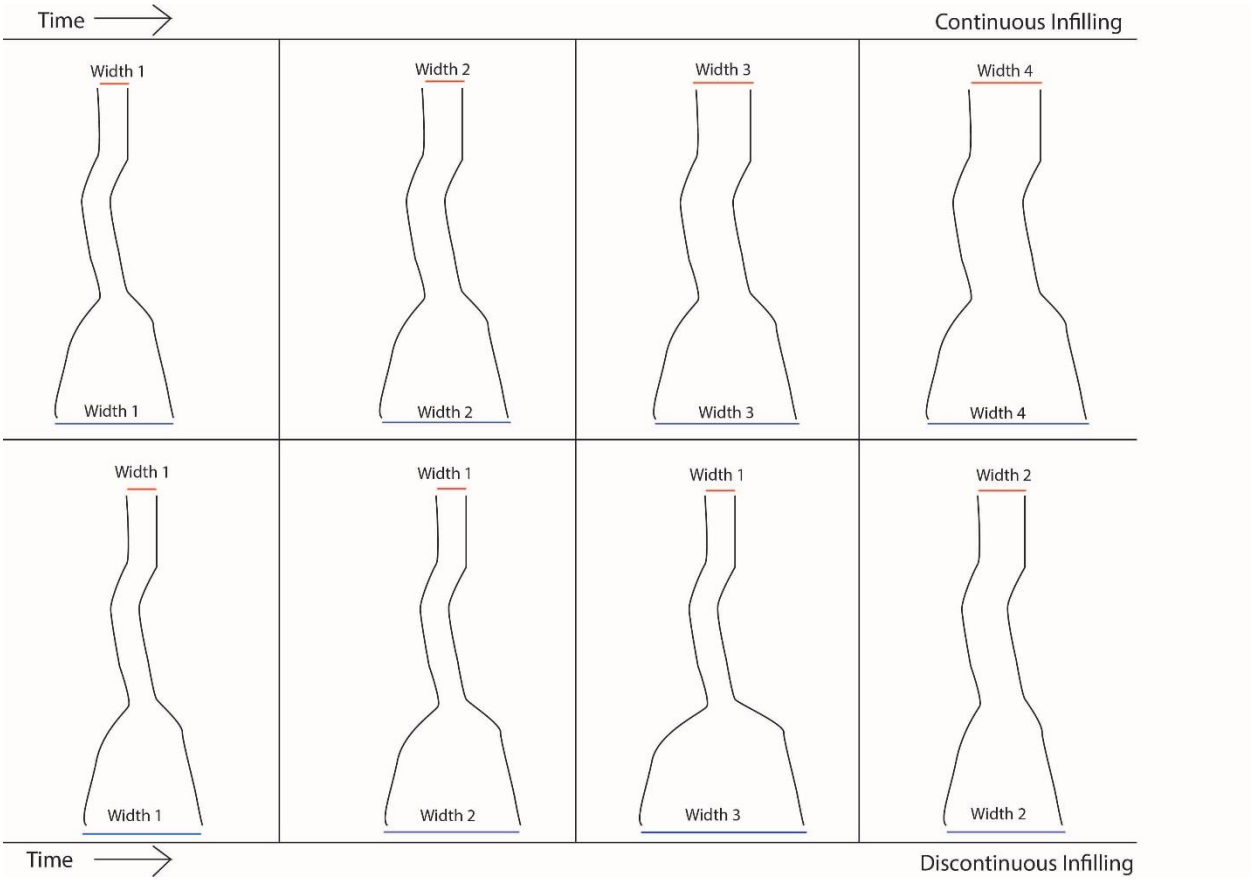

Figure 2: Continuous vs Discontinuous sinkhole infilling over time

When water percolates through caves or if the humidity in the chambers is high, weathering allows for cracks to form (Parise and Lollino 2011). This reduces the rock tensile strength allowing shearing along fractures and joints, floor heaving and roof lowering within the underground cavern (seen as ground surface subsidence), and "rock noises" before total collapse (Parise and Lollino 2011). Building structures on top of these at-risk areas increases the force and lowers the stability of the rock even more.

Since water seeps through fractures, and sinkholes form around these fractures, sinkholes can be used to display fault zones and groundwater paths. Florea (2005) used a geographic information system (GIS) to map all sinkholes found on USGS 1:24000 scale Kentucky

topographic maps. On a macro scale, Florea was able to trace the Rough River Fault zone with sinkhole sites and find karst boundaries along fault offsets. On a meso-scale, sinkholes were found tracing along NW-trending fractures and surrounding the Versailles impact structure. It was noted that not all fractures have associated sinkholes and many sinkhole alignments currently follow along unknown faults and fractures. One hypothesis for this suggests that karstic regions with faults but no sinkholes may be a sign that mylonitization is preventing groundwater flow (Taylor 1992; Florea 2005). Correlation between sinkhole and fault locations in Kentucky infer that similar correlations may exist within other states, such as Tennessee and Florida.

Climate Impacts

Rising levels of CO_2 are amplifying changes in climate around the globe, manifesting in warmer temperatures overall and increasingly erratic and extreme rainfall. Polar amplification entails that higher latitudes are more affected by this warming (Moran 2009). As temperature rises, polar ice caps melt and seawater experiences thermal expansion. According to the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report, global mean sea level is expected to rise approximately 0.35 - 0.54 m by 2080 (IPCC 2014). Coastal and terrestrial water sources are influenced by sea level. As sea levels rise, so does the water table. Cowling (2016) researched this correlation with respect to glacial-interglacial cycling and found that a cooling, dry climate associated with glaciation lowered sea level and water table. In contrast, a shallower water table would concentrate dissolution at lesser depths, which could allow for an increase in sinkhole formation over time. Examples of this type of occurrence are occasionally visible around dams (Flores-Berrones et al. 2011).

As CO_2 levels increase in the atmosphere, CO_2 returns to the ground by means of acid rain. Increased atmospheric CO_2 also increases the efficiency of plant water usage (Gedney et al.

2006; Macpherson et al. 2008); which increases CO_2. The natural groundwater pH at present is ~6 in the majority of the world (Knutsson 1994). In areas with weathering-resistant minerals (non-carbonates) in bedrock and soils, Knutsson (1994) found that groundwater acidification is increasing. In carbonate systems, the opposite effect has been documented. Zeng et al. (2016) identified examples attributed to increased CO_2. In the past half century, the Mississippi River increased in alkalinity approximately 20% (Raymond and Cole 2003; Raymond et al. 2008; Zeng et al. 2016). The Konza Prairie, USA increased in alkalinity 13% between 1990 and 2005 (Macpherson et al. 2008; Zeng et al. 2016). Macpherson et al. (2008) found an increase in dissolved calcium and magnesium in these karstic waters. Together, the two minerals made up 82-94% (by weight) of the total dissolved material. They proposed that the increase in CO_2 was the cause for the increased chemical weathering and resulted in increasing alkalinity in karstic waters (Fig 3). As CO_2 levels increase, the probability of sinkhole formation may therefore increase but to what extent is unknown. Understanding how these levels influence sinkholes for future protective measures is needed.

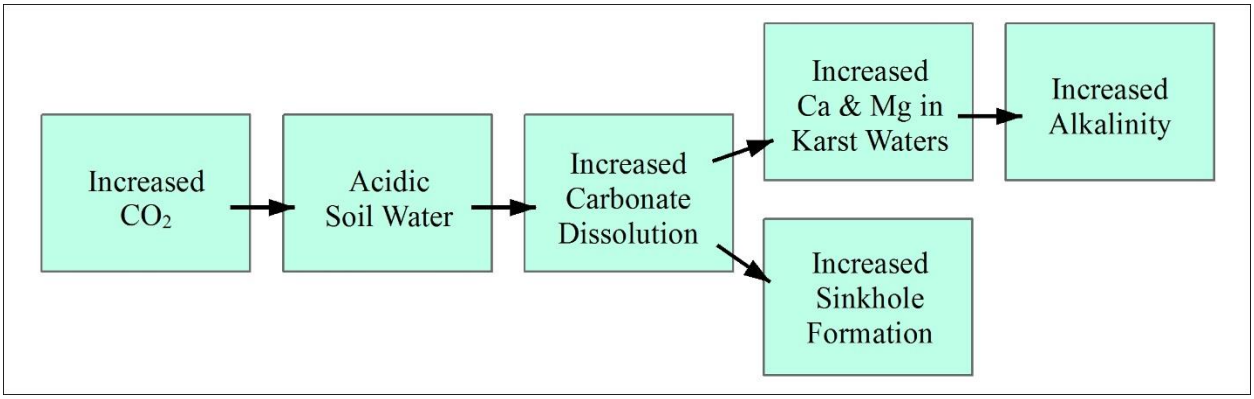

Figure 3: Diagram of CO_2 affecting sinkhole formation

Another influence on weather-related factors includes teleconnections. Teleconnections are recurring deviations of standard pressure heights above sea level, impacting weather patterns

at a regional scale (Wallace and Gutzler 1981; Barnston and Livezey 1987; Duffy et al. 2005). The most well-known teleconnection impacting the United States is the El Niño/Southern Oscillation (ENSO). The National Oceanic and Atmospheric Administration (NOAA) monitors the teleconnection monthly indices for the northern hemisphere as they are the main drivers of climatic patterns for the United States. These teleconnections directly influence precipitation, temperature, significant weather events, etc.

Modeling Methods

Buttrick et al. (2001) developed a method for hazard and risk assessments for the dolomitic lands of South Africa to morphometrically determine sinkhole risk including variables such as sinkhole size, depth, and whether dewatering had occurred. The risk assessment also considered measures suggested by the Joint Structural Division (2000) to prevent catastrophes, such as building raft foundations underneath housing units to span a sinkhole; this would prevent the house from collapsing, allow an escape for the occupants and limit structural damage to the house (as cited in Buttrick et al. 2001).

MacGillivray (1997) reviewed different quantitative techniques for measuring karst terrain. He found that before the 1970s, most sinkholes were categorized qualitatively by shape and origin. By the 1980s, morphometric approaches were embraced. These approaches include indices for surface roughness, bedrock properties (including porosity, purity, permeability, and strength), and a double Fourier series analysis of topographic variance. Morphometric techniques are the current main facilitators for the research in karst geohazards and anthropogenic-related karst problems.

Study Objectives and Research Questions

Climate factors are known to influence sinkholes but few studies link the two. Geospatial (GIS) models are increasingly employed to model sinkhole hazards, however, these models generally do not include climatic variables. Generally, sinkholes are mapped at a local scale, negating the change in climatic variables over space. For example, variables used to model sinkhole risk in a temperate environment are different than those used in a tropical environment, so the two environments are modeled separately. Gaining climate-related information from multiple locations, or at a regional scale, may help to link these differing environments. Also, as climate changes over time, these answers may expose potential shifts in sinkhole hazard locations.

This thesis is composed of two main studies: a temporal study of Florida sinkholes and a two-part spatial study of Tennessee sinkholes. Due to differing data, conducting a spatial study of Florida sinkholes and a temporal study of Tennessee sinkholes was deemed inappropriate. The thesis will seek to answer the following questions.

Study One:

1. How are specific weather patterns associated with sinkhole formation in Florida?
2. To what extent does climate impact sinkhole-forming regions in Florida?
3. Where are hotspots for sinkhole formation in Florida?

Study Two:

1. Which climate variables are the most influential in predicting sinkhole formation?
2. Where are sinkholes expected to form in Tennessee based on climate-sinkhole relationships?

CHAPTER 2

TELECONNECTION AND CLIMATE EFFECTS ON FLORIDA SINKHOLES – A TEMPORAL STEPWISE MULTIPLE LINEAR REGRESSION APPROACH

By:

Kimberly Blazzard, Ingrid Luffman, T. Andrew Joyner

Abstract

Little is known about the temporal correlation between Florida sinkhole formation and climatic patterns. This study utilized a stepwise linear regression methodology to examine relationships between Florida sinkhole and subsidence events, teleconnection phases, and other climatological patterns. Significant (northern hemisphere) monthly and seasonal teleconnection phases (National Weather Service Climate Prediction Center), statewide precipitation and temperature averages (Florida Climate Center), and average CO_2 levels (NOAA Earth System Research Laboratory) were the covariates. Event records were also offset by month/multi-month and season/multi-season time periods to examine lagged relationships. Results showed a statistically significant ($p < 0.01$) positive correlation between sinkhole event formation and the East Atlantic Pattern, precipitation, temperature, and CO_2. Regression results indicated that as much as 23% of sinkhole formation variability could be explained by these teleconnection phases and climatological patterns and may be useful in identifying periods at higher risk for sinkhole formation.

Keywords: Sinkhole, Doline, GIS, Climate, Predictive Modeling, Karst, Florida, Linear Regression

Introduction

Environmental Background

Florida has a humid subtropical climate north of Lake Okeechobee and a tropical climate south of Lake Okeechobee. Average annual precipitation is 150.4 cm (59.21 inches). Average low temperature ranges from 15° C (39° F) in January to 22° C (72° F) in July and August. Average high temperature ranges from 18° C (64° F) in January to 33° C (92° F) in July and August (USClimateData, 2017).

Florida covers 170,304 km^2 (65,755 mi^2) of land which is almost entirely underlain by carbonates (Lane, 1986). Although Florida's basement rock consists of Paleozoic igneous and metamorphic rocks, it is overlain by varying thicknesses of carbonates with interbedded sandstones and shales (Fig 4); this is the basis for the Florida Platform which makes up the Florida peninsula (Scott, 2001). The siliciclastics (the sandstones and shales) are Late Miocene to Pliocene and are underlain by Middle Miocene carbonates and overlain by Quaternary carbonates. The Florida Platform is considered to be one of the largest carbonate platform systems created during Earth's geologic history (Cunningham et al., 2003; Poag, 1991). Since all regions of Florida contain carbonate layers, all regions are at risk of sinkhole formation; locations not labeled as carbonates may still have dissolution below the top layer of rock allowing for the surface layers of rock and soil to suddenly collapse forming a sinkhole (Fig 5).

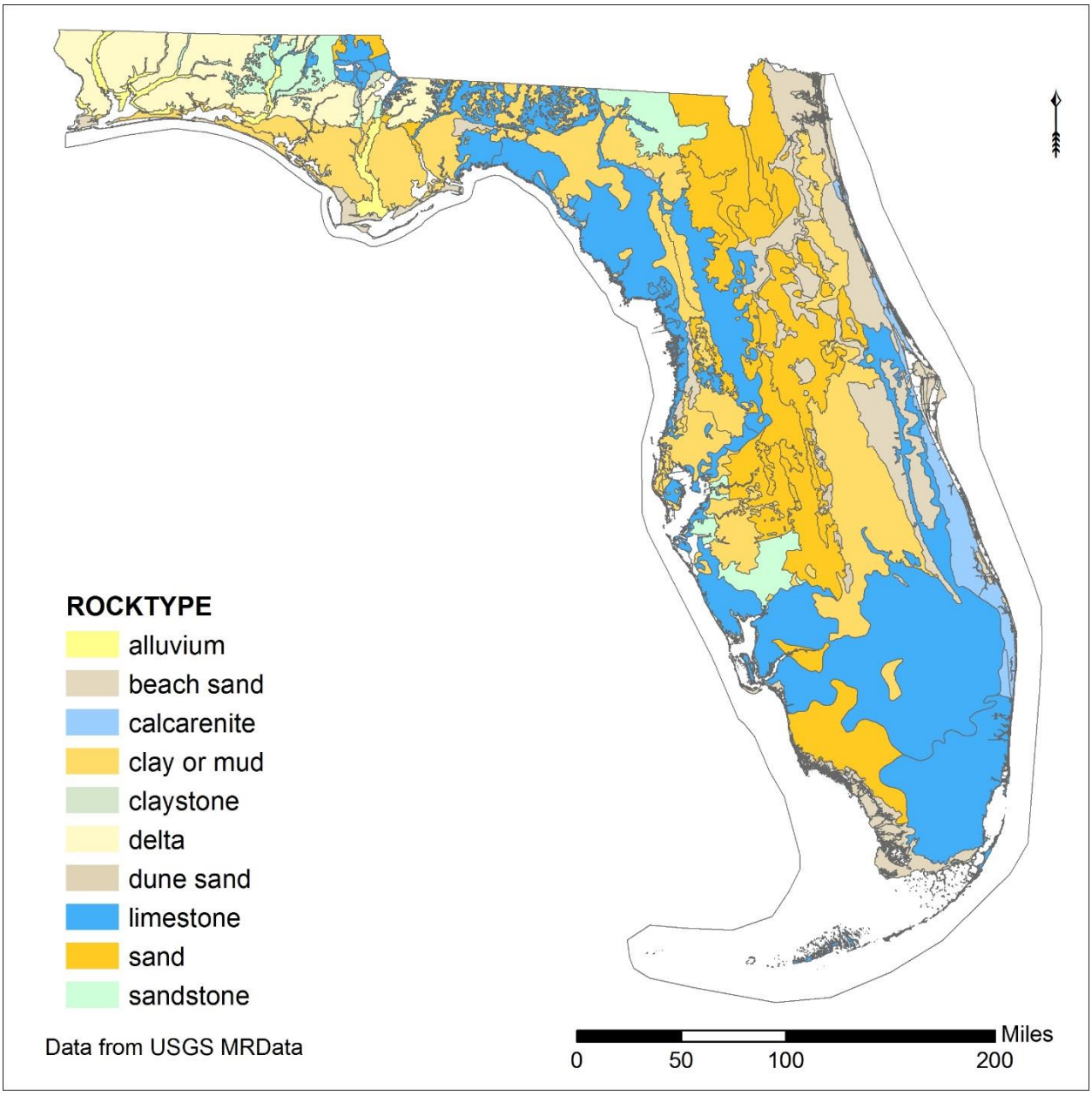

Figure 4: Florida geologic map created from USGS MRData.

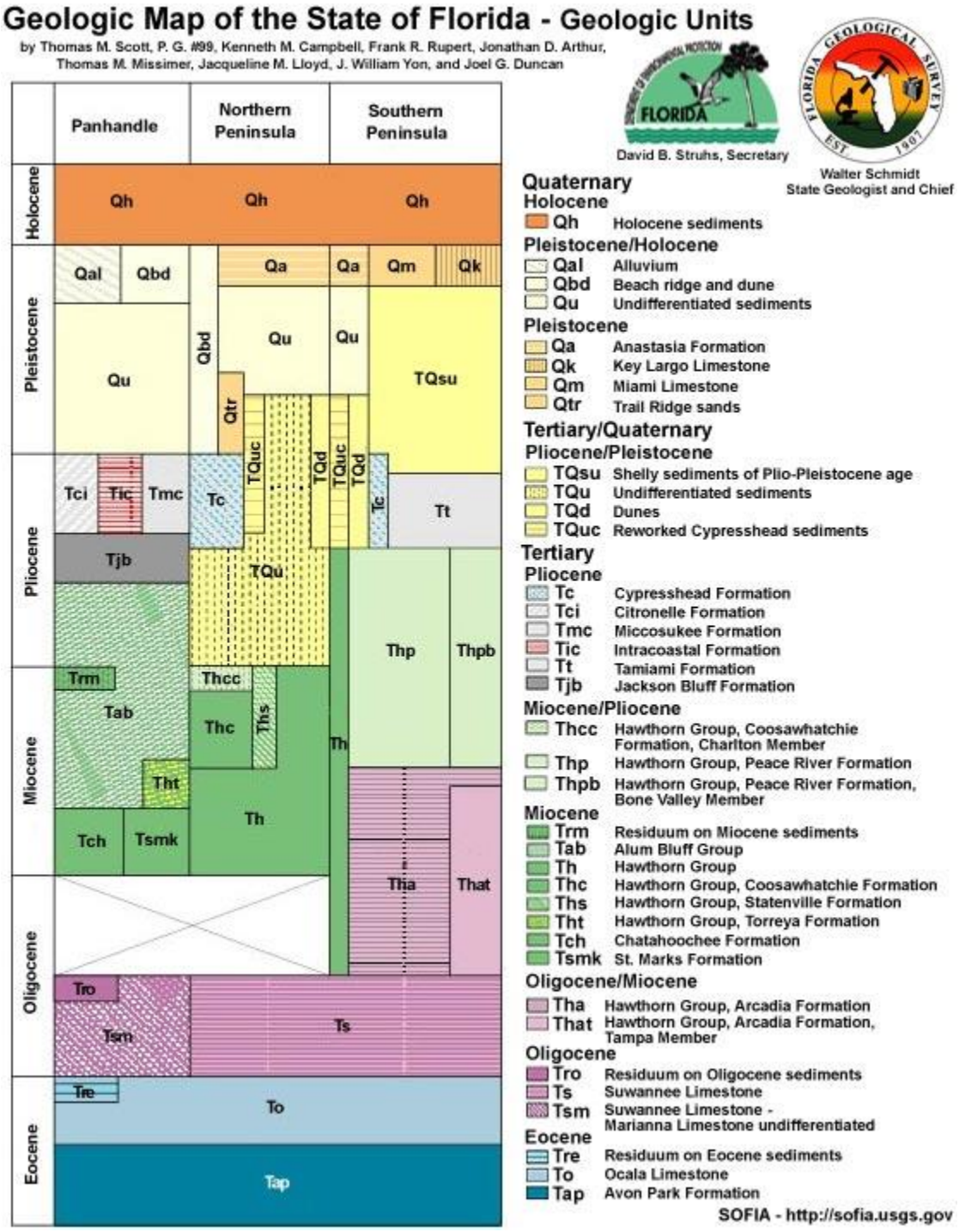

Figure 5: Florida geologic timescale produced as part of the Florida Geological Survey, Open File Report No. 80

According to the National Land Cover Database 2011 (NLCD2011), Florida contains approximately 35% wetlands, 17% planted/cultivated land, 17% forest, and 15% developed land. Of the sinkholes mapped since 1954, 56% of the sinkholes reside within land of minimal development (open space developed – low intensity developed), 19% within medium to high developed lands, 7.7% within forests, and 6.5% within planted/cultivated lands (with the remaining 10.8% distributed among the remaining land types) (Fig 6).

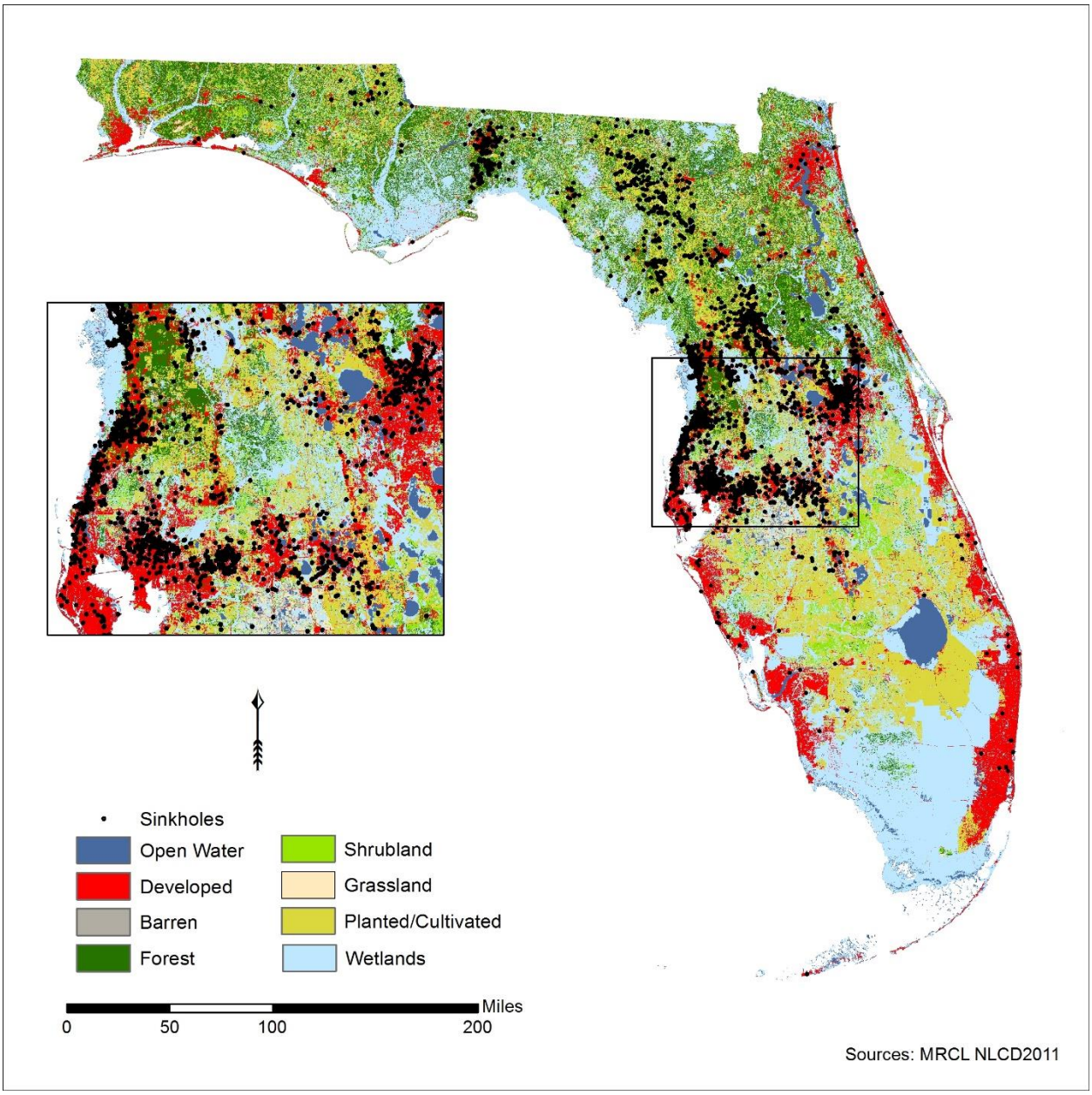

Figure 6: Florida land cover map with sinkhole locations.

Statistical Background

Linear regression is a statistical technique used to model a linear relationship between two variables by creating a linear equation. The dependent variable (the variable to be predicted) is assumed to be continuous and normally distributed (at least for smaller samples) (Sainani,

2013). Multivariate linear regression includes multiple variables as predictors (Sainani, 2013), where the intercept (*A*) is a constant and the coefficients (b_i) are multiplied to each independent variable (X_i):

$$Y = A + b_1 * X_1 + b_2 * X_2 + \ldots + b_n * X_n$$

When working with multiple independent variables, not all variables may combine together to affect the dependent variable. A stepwise regression is useful for reducing the number of independent variables (Parinet et al., 2015).

After a regression model is developed, two main diagnostic statistics must be explored, the R^2 value and the *p*-value. The R^2 value determines how much variability in the dependent variable is explained by the independent variables. For example, an R^2 value of 0.123 would explain 12.3% of the variability. The *p*-value determines how significant the different predictors are and the overall significance of the model. A *p*-value below .05 is considered significant while a *p*-value below .01 is considered highly significant.

Research Questions

1. How are specific weather patterns associated with sinkhole formation in Florida?
2. To what extent does climate impact sinkhole-forming regions in Florida?
3. Where are hotspots for sinkhole formation in Florida?

Data and Methods

Spatial Distribution of FL sinkholes

Sinkhole and subsidence event records, available from 1948 to 2016 from the Florida Department of Environmental Protection Geospatial Open Data, were obtained and converted

into a point shapefile using ArcMap (Fig 7). The records include 3,516 data points with accompanying latitude and longitude. Additional information including the report source, repair status, soil type, event date, date reviewed, county, township, width, depth, slope, and extra notes were attached for each sinkhole/subsidence report if known. Florida roads were downloaded as a 2013 TIGER/line shapefile from Data.Gov. All data were projected into the NAD 1983 HARN Florida GDL Albers (meters) projection.

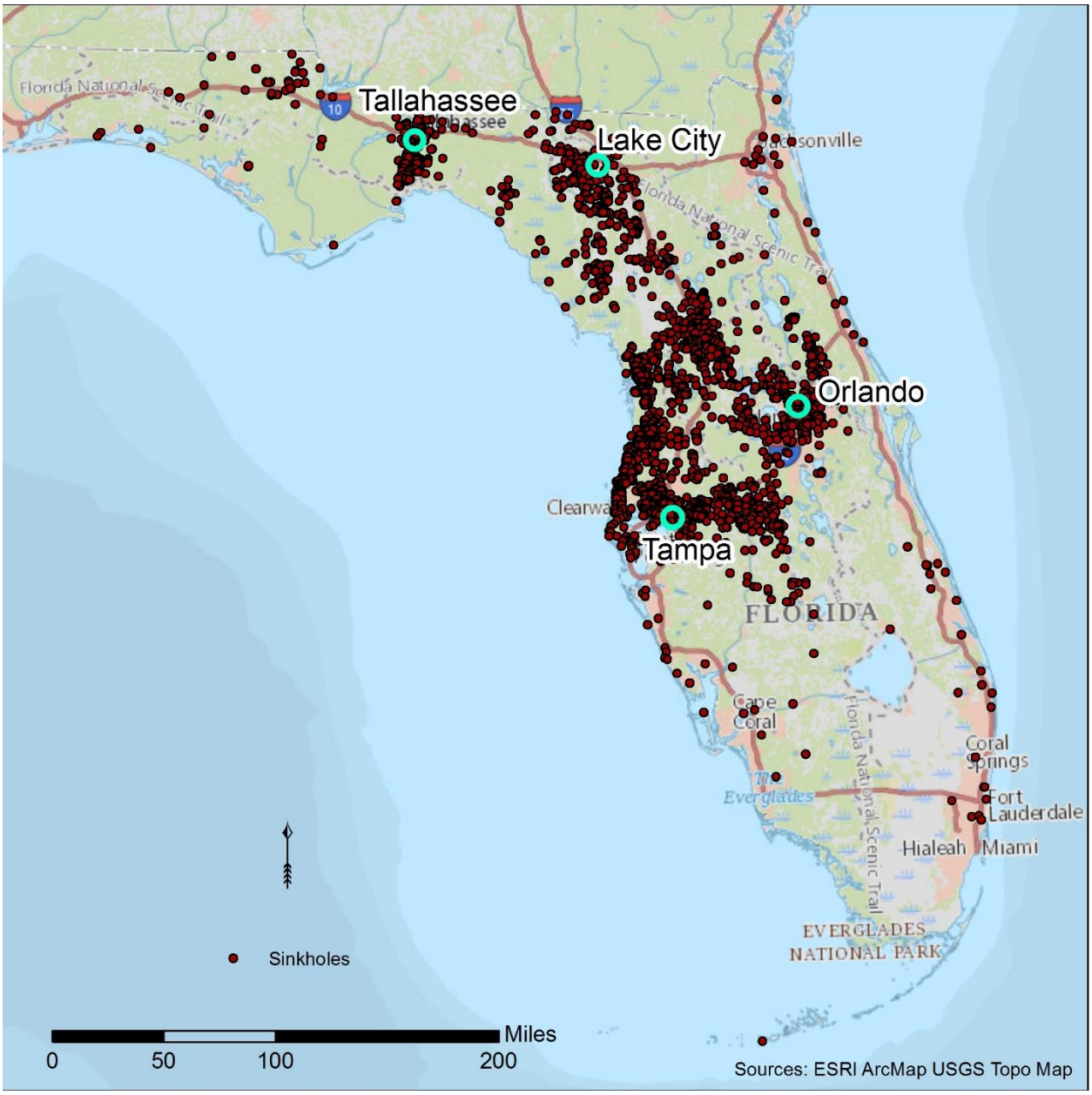

Figure 7: Florida sinkholes recorded since 1948.

Crimestat IV was used to compute spatial statistics for the Florida sinkholes (using the road network as a distance reference), while ESRI ArcMap 10.5 was used to visualize the sinkholes, roads, and resulting statistics. First, a spatial distribution analysis was conducted to find the average location of the sinkholes and to find where the majority of the sinkholes are

forming. This was done by calculating the mean center, the geometric mean, the harmonic mean, the median center and the mean center of minimum distance.

Next, a Nearest Neighbor Analysis (NNA) was conducted to quantitatively describe the dispersion of the sinkholes. Results of the NNA are calculated on the Nearest Neighbor Index scale (NNI). An NNI score less than one implies that the incidents are clustered, equal to one indicates they are randomly dispersed in space, and more than one implies they are regularly dispersed. A rectangular border correction was applied to compensate for the fact that the nearest neighbor to some of the sinkholes may actually be north of the Florida boundary.

Then, a hot spot analysis was conducted to identify where the sinkholes concentrate. This was computed using a Nearest-Neighbor Hierarchical Spatial Clustering analysis. A cluster was defined as a minimum of 100 sinkhole sites within a 10-mile radius. A 1x standard deviation was utilized to show the standard deviation of distances of the sinkhole locations from the mean center.

Lastly, three kernel density estimates (KDE) were developed to model locations with the highest risk for sinkhole formation. Since sinkholes require certain features (such as a karst lithology), it is possible that the area of influence can decline rapidly near an incident; three different kernel distribution bandwidth combinations were selected for this reason. KDE#1 included a negative exponential distribution, an adaptive bandwidth, and a minimum sample size of 100. KDE#2 included a normal distribution, an adaptive bandwidth, and a minimum sample size of 100. KDE#3 included a normal distribution and a fixed interval of h_0 = 13.63 km (8.5 miles) determined from the Fotheringham Rule, where N is the total number of occurrences/points and σ is the standard distance deviation.

$$h_0 = [2/(3N)]^{1/4}\sigma$$

Sinkholes were also graphed displaying the number of sinkholes formed for each month and season, and mapped to show the sinkhole locations. The general location map and cluster analysis were used to determine which cities in Florida are most vulnerable to sinkholes and to select focal areas for linear regression modeling.

Florida Temporal Linear Regression

Karst sinkholes often develop shortly after periods of heavy rain and may be connected to larger, macro-climatic patterns. Recent research indicated that a 200-m-long collapse zone, in Guanxi, China, was preceded by a year-long drought followed by a heavy, single day rain event (469.8 mm total) (Gao, 2013). To determine empirical linear trends in long-range forecasting, linear regression is historically the base method (Blender, Luksch, Fraedrich, and Raible, 2003).

This study primarily utilized stepwise linear regression to examine relationships between Florida sinkhole collapse/formation and other subsidence events, teleconnection phases, and climatological patterns. Monthly and seasonal teleconnection phases, available from the National Weather Service Climate Prediction Center, were used for regression analysis and include the following teleconnections: North Atlantic Oscillation (NAO), East Atlantic Pattern (EA), West Pacific Pattern (WP), East Pacific/North Pacific Pattern (EP/NP), Pacific/North American Pattern (PNA), East Atlantic/West Russia Pattern (EA/WR), Scandinavia Pattern (SCA), and El Niño-Southern Oscillation (ENSO) (utilizing the Niño 3.4 SST Index). These teleconnections include the National Oceanic and Atmospheric Administration (NOAA) prioritized monthly indices. NOAA monthly precipitation totals and monthly average temperatures for the Tampa, Lake City, Tallahassee, and Orlando, Florida regions were downloaded. Average global CO_2 levels were obtained from the NOAA Earth System Research Laboratory (measurements from the Mauna Loa Observatory in Hawaii). Combined sinkhole and subsidence event records were

available from the Florida Department of Environmental Protection Geospatial Open Data portal. Although subsidence events may form differently from sinkholes (for example, from broken water lines), the dataset did not differentiate between the two event types. Sinkhole records were subdivided into regions within 50 mile and 100 mile radii around the Tampa, Lake City, Tallahassee, and Orlando weather stations. Sinkholes were totaled per month and totaled per season. These totals were also log-transformed for analyses. In addition, the event records were offset by month/multi-month (1-3 months) and season/multi-season time periods (1-2 seasons), compared to the independent variables (total monthly sinkholes, total season sinkholes, etc), to examine possible lagged relationships. All statistical calculations were processed in SPSS Statistics. The time range included January 1954 to June 2016. For comparative analysis, a Poisson loglinear regression with the zero count data, a Poisson loglinear regression without the zero count data, and a negative binomial with log link, and an OLS enter-method (vs. stepwise) regression models were developed using the same variables on the Tampa 50 mile, monthly total dataset.

The carbon dioxide (CO_2) levels oscillated higher and lower along an overall increasing trend, ranged from 327.20 ppm to 403.95 ppm (Fig 8). The dataset contains recorded values from 1974 to present. The teleconnections oscillate between a positive and negative around a value of 0 (showing normal climate conditions) (Fig 9). Visually, the ENSO data shows an overall oscillatory pattern ranging from 24.52° C to 29.14° C (measured from sea surface temperatures) (Fig 10).

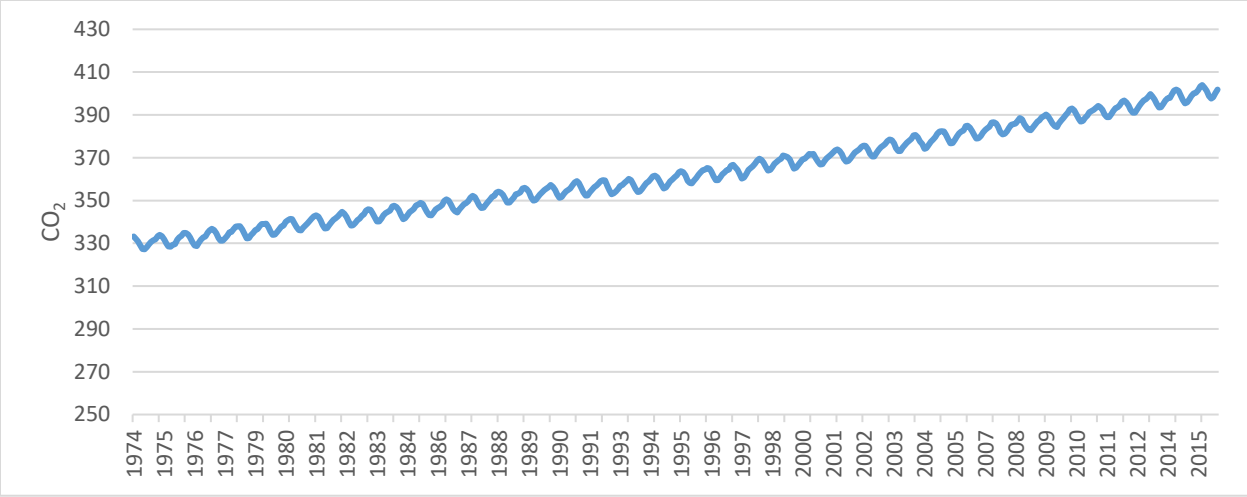

Figure 8: Change of CO_2 levels (ppm) since 1974

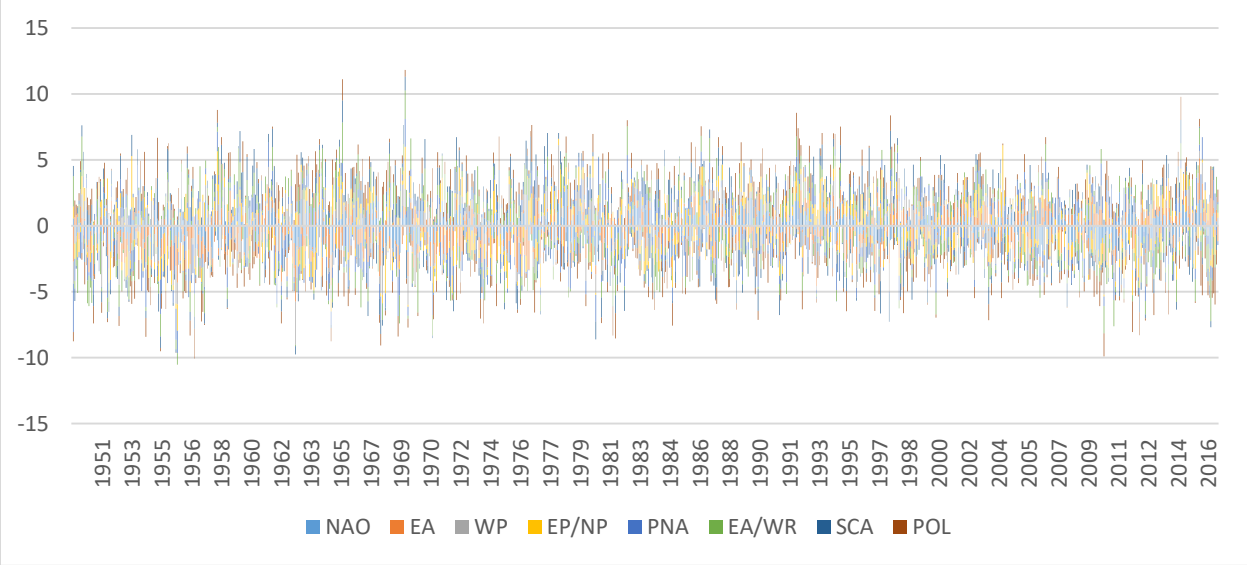

Figure 9: Monthly teleconnection trends showing general oscillation of indices

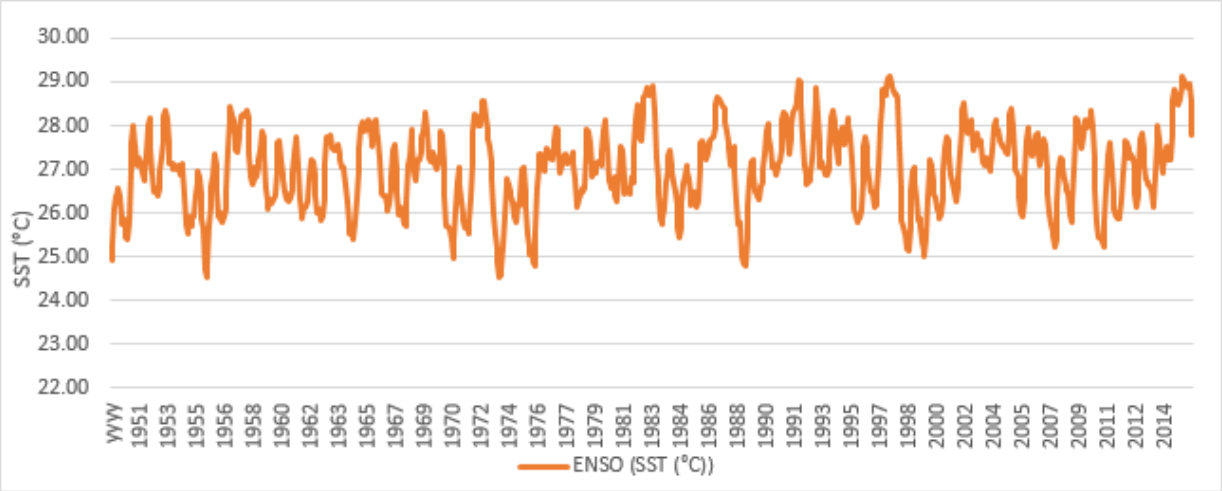

Figure 10: Monthly ENSO trend since 1950

Results

Spatial Distribution of FL sinkholes

The geometric mean and mean center of minimum distance are approximately 13 km apart. The mean center is located at -82.40 W, 28.81 N or approximately 9 km west of Fort Cooper State Park (Fig 11). The mean center is approximately 20 km NE of the harmonic mean, 9.45 km NE of the geometric mean, 13 km NNE of the median center, and 16.7 km NNW of the mean center of minimum distance. The five points range across an expanse of 139.9 km^2.

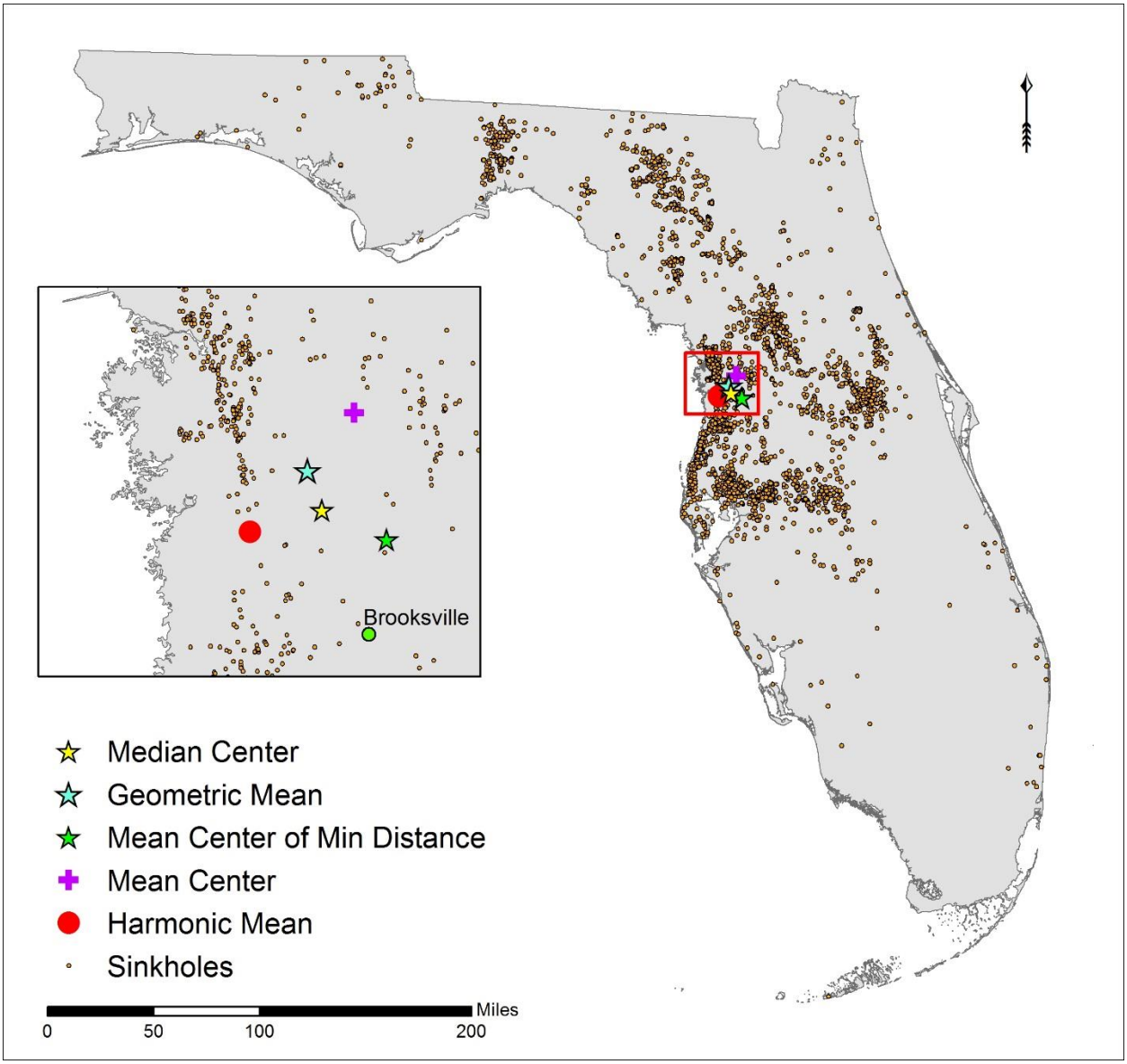

Figure 11: Centrographic statistics of Florida sinkholes displaying the mean center, harmonic mean, geometric mean, median center, and mean center of the minimum distance. The overall trend shows the mean point north of Brooksville.

The mean nearest neighbor distance was 1,232.90 m with a minimum distance of 0.00 m. The Nearest Neighbor Index equaled 0.202 showing that sinkholes are spatially clustered (p = 0.0001).

The hot spot analysis found nine clusters primarily located around major cities (Fig 12). These major cities include Tallahassee, Lake City, Tampa, and Orlando. The majority of the clusters are found within the Tampa region.

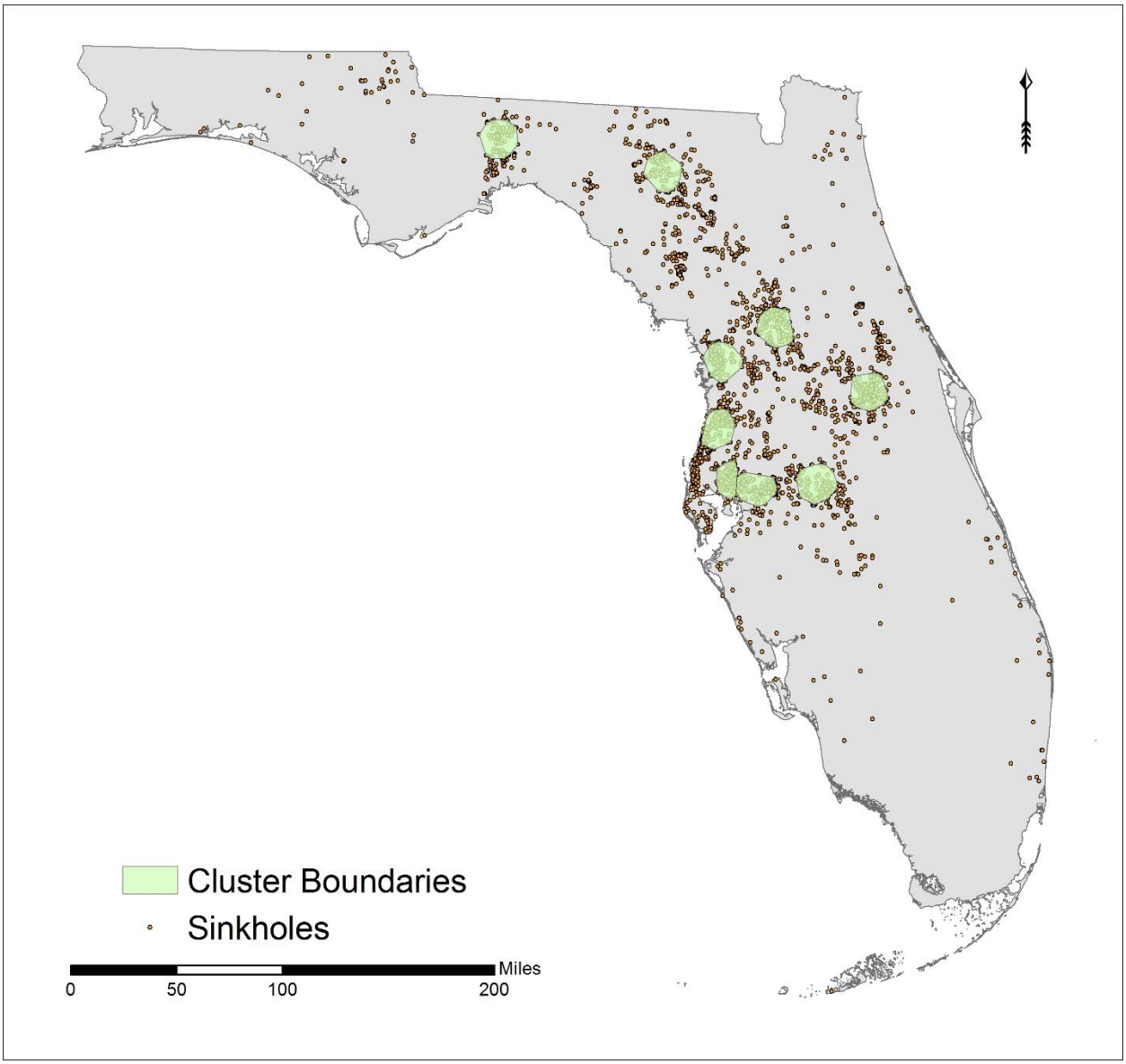

Figure 12: Florida cluster analysis showing the regions where there are at least 100 sinkholes within a 10 mile radius. Nine clusters were found around Tallahassee, Lake City, Tampa, and Orlando.

KDE#1 showed no areas of sinkhole density, so it was discarded. KDE#2 (Fig 13) and KDE#3 (Fig 14) showed better interpolated surfaces, however, KDE#3 best fit the actual locations of the recorded sinkholes. KDE#2 included large sections of low sinkhole density as being susceptible. KDE#3 shows which areas, within the clustered regions, have the highest density of sinkholes.

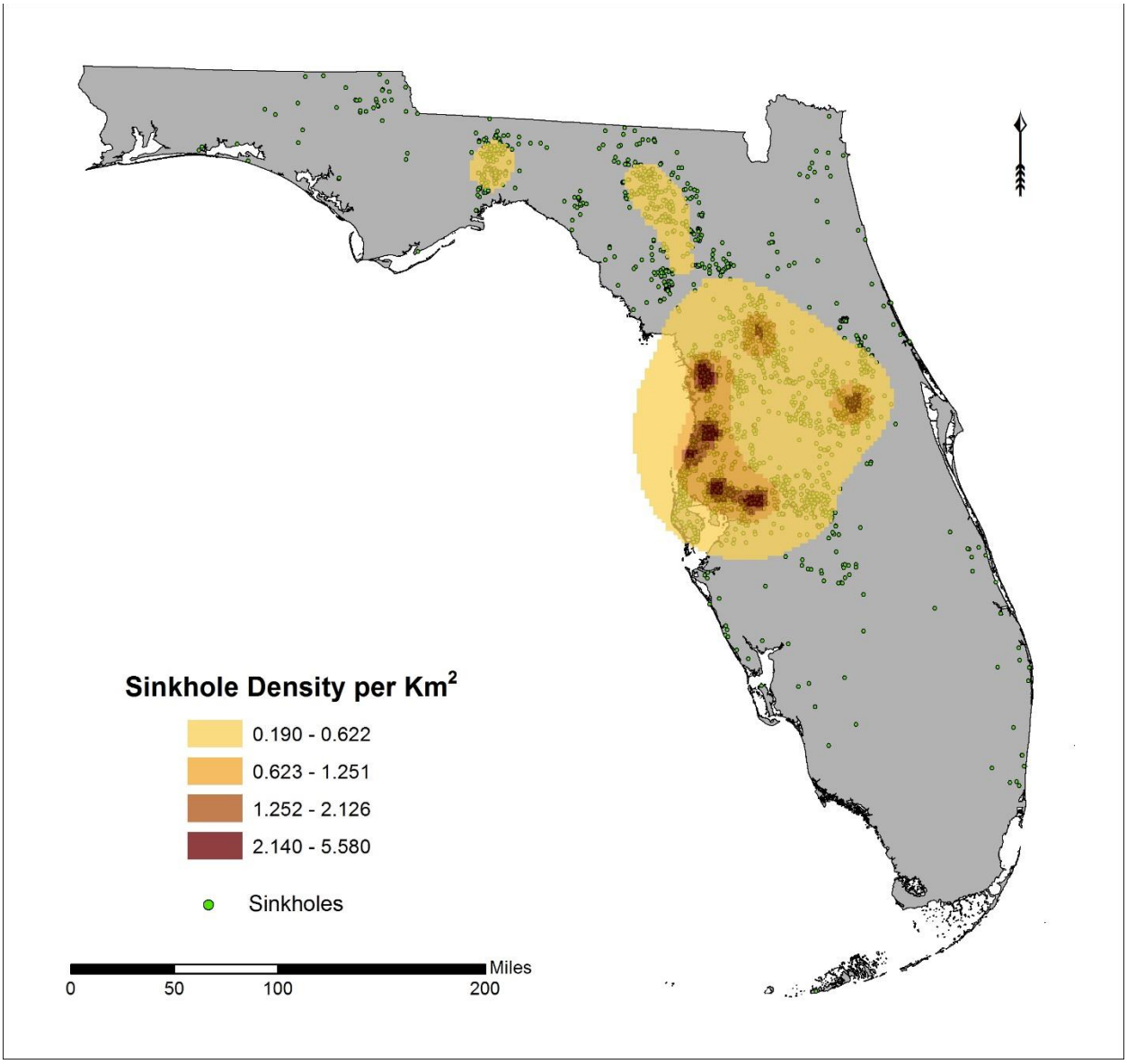

Figure 13: Florida KDE#2 shows an interpolated density surface utilizing a normal distribution, adaptive bandwidth, and a minimum sample size of 100.

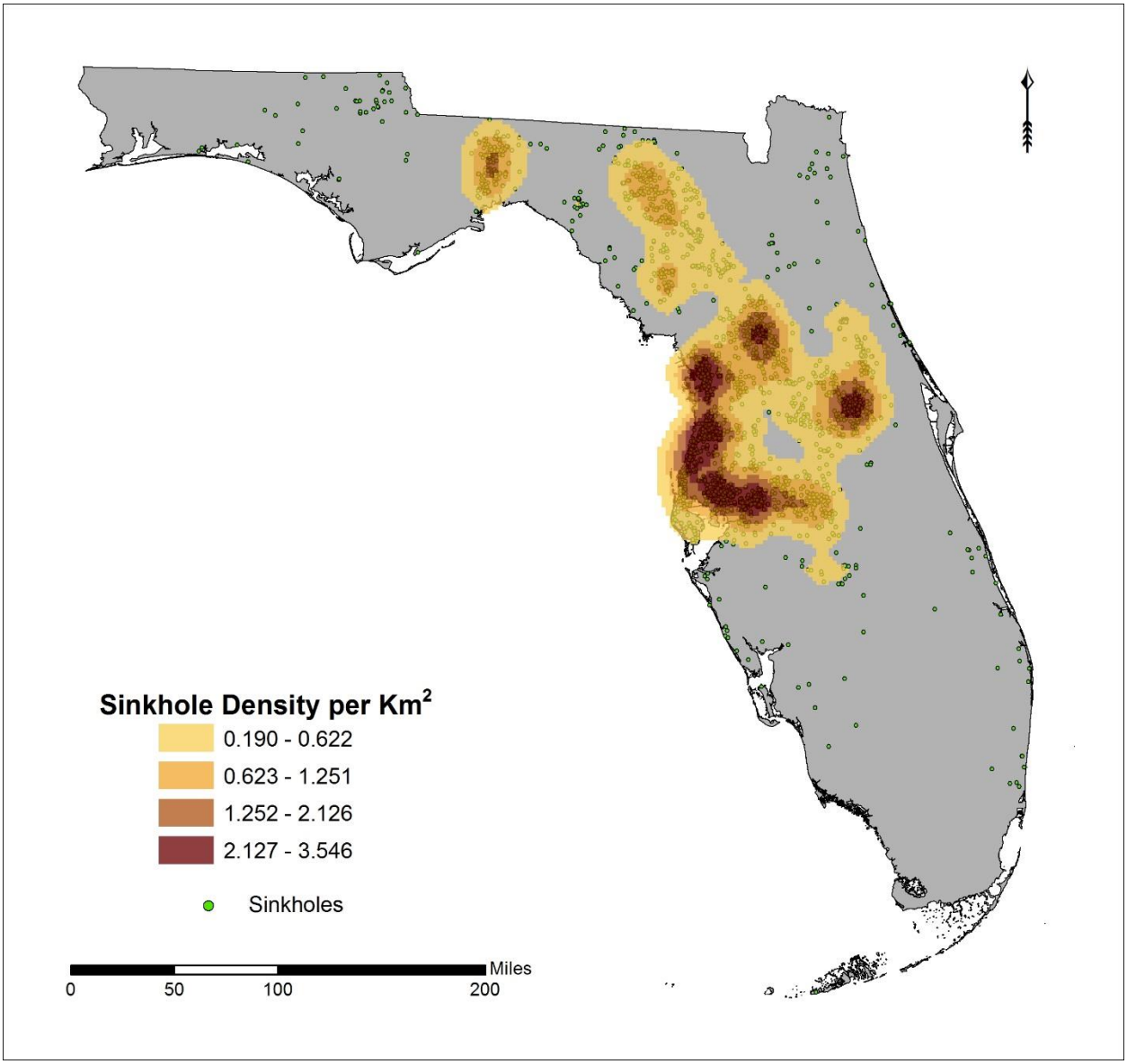

Figure 14: Florida KDE#3 shows an interpolated density surface utilizing a normal distribution and a fixed interval of 13.63 km.

Results from Linear Regression of Florida

Since the largest clusters were centered around Tampa, Lake City, Tallahassee, and Orlando, these cities were selected for linear regression modeling.

Approximately 24.5% ($R^2 = 0.245$, $p = 0.000$) of sinkhole variability within 100 miles of Orlando was explained using precipitation, ENSO, CO_2, and the PNA (Table 1). Within 50 miles

of Lake City, 24.0% ($R^2 = 0.240$, $p = 0.000$) of sinkhole variability could be explained by the precipitation, NAO, and SCA. Only 11.2% ($R^2 = 0.112$, $p = 0.000$) of sinkhole variability could be explained in the region of Tallahassee by precipitation. The Tampa region had minimal sinkhole variability explained by precipitation. The most significant climate variable affecting the sinkholes in all four cities was precipitation. The included explanatory variables were determined as statistically significant ($p < 0.05$). The sinkhole monthly lag and seasonal lag times showed a lower R^2 with each variable (if at all) compared to the month/season in which the sinkhole actually formed. Seasonal data, compared to monthly data, showed an increase in R^2 values by 0.6% for the Lake City region, but decrease in R^2 for all other regions. Explanatory teleconnection patterns for the differing regions included ENSO, PNA, NAO, EP/NP, SCA, and WP.

Table 1: Florida sinkhole regression results based on differing climate variables.

Time frame	Conditions	Dependent Variable	Independent Variables Retained	Adjusted R^2
Monthly	100 mile (160 km) radius from Orlando	Log sinkholes	Precipitation, ENSO, CO2, PNA	0.227
		Log of 1 month time lag	ENSO, Precipitation, CO2	0.123
		Sinkholes	NAO	0.020
		1 month time lag	EP/NP, ENSO, NAO	0.070
	50 mile (80 km) radius from Tallahassee	Log sinkholes	Precipitation	0.105
		Log of 1 month time lag	NAO, EP/NP, CO2	0.076
		Sinkholes	Precipitation, CO2, ENSO	0.098
	50 mile (80 km) radius from Tampa	Log sinkholes	Precipitation, EP/NP	0.038
		Log of 1 month time lag	None	-
		Sinkholes	NAO	0.011
	50 mile (80 km) radius from Lake City	Log sinkholes	Precipitation, NAO, SCA	0.219
		Log of 1 month time lag	Precipitation	0.057
		Sinkholes	Precipitation	0.081
		1 month time lag	Precipitation	0.010
	50 mile (80 km) radius from Orlando	Log sinkholes	CO2, WP, ENSO	0.177
		Log of 1 month time lag	CO2	0.086
		Sinkholes	Precipitation, CO2, EA	0.080
		1 month time lag	CO2, Temperature	0.035
	All Florida	Log of Sinkholes	Precipitation	0.036
		Log of 1 month time lag	ENSO, EA	0.038
		Sinkholes	Precipitation, NAO	0.018
		1 month time lag	Precipitation	0.010
Seasonal	100 mile (160 km) radius from Orlando	Log of Sinkholes	Precipitation	0.153
		Log of season time lag	Precipitation	0.105
		Sinkholes	NAO	0.127
		Log of 1 season time lag	None	-
	50 mile (80 km) radius from Tampa	Log of Sinkholes	Precipitation, NAO	0.228
		Sinkholes	Precipitation, NAO	0.127
		Log of 1 season time lag	None	-
		1 season time lag	None	-

Precipitation, average temperature, and the EA significantly influenced sinkhole formation while the PNA, NAO, SCA, and WP negatively influenced sinkhole formation. ENSO, CO_2, and the EP/NP displayed both positive and negative influences on sinkhole formation depending on the location.

The largest spikes in sinkhole formation occurred in January 2010 with 160 sinkholes and June 2012 with 153 sinkholes (Fig 15). The next largest spikes occurred in May 1964 and September 1988 with 47 sinkholes each. No significant rain events occurred during or directly before any of the four spikes (Fig 16). This shows that although there is a general correlation between precipitation and sinkhole formation within this area, it is not the primary factor affecting the most significant events.

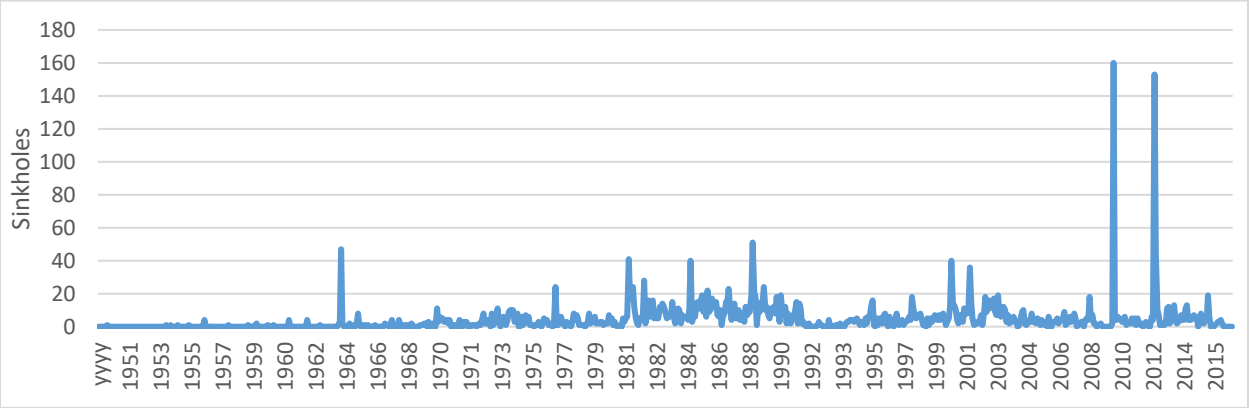

Figure 15: Florida Sinkhole Formation over time graph displaying the total sinkholes formed per month since 1954.

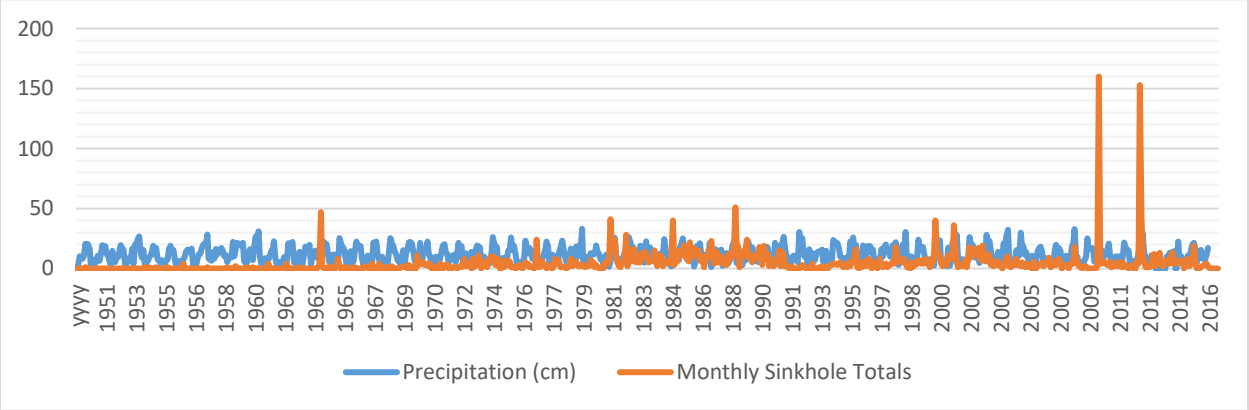

Figure 16: Total Florida sinkholes and precipitation each month from 1950-2016

The seasons ranked from largest to smallest total sinkhole formations included summer (776 sinkholes), spring (757 total sinkholes), winter (632), and fall (506) (Fig 17). The average sinkhole total per season ranged from 2.72 in fall to 4.15 in summer (Fig 18). The months with the largest sinkhole totals included January (368) and May (324). The months with the lowest sinkhole totals included November (112) and December (116) (Fig 19). The average sinkhole formed per month ranged from January (5.84) to November (1.81) (Fig 20). Since the data do not include the months of July – Dec 2016, the data totals are slightly skewed, yet the small averages for each of these months show that the missing data would have negligible impact on overall monthly totals. For example, adding an additional ~2 sinkholes to the month of November (bringing the total to 114) would still leave November with the lowest total of sinkholes formed. An independent-samples Kruskal-Wallis test determined no significant differences between seasons or between months. However, January 2010 recorded 160 sinkholes and June 2012 recorded 153 sinkholes; this accounts for over 1/3 of the total sinkholes recorded for January and over ½ of the total sinkholes for June. When these events are treated as outliers and not included within analysis, the Kruskal-Wallis tests determined a significant difference between months of sinkhole formation ($p = 0.021$) but still not between seasons of sinkhole formation ($p = 0.088$).

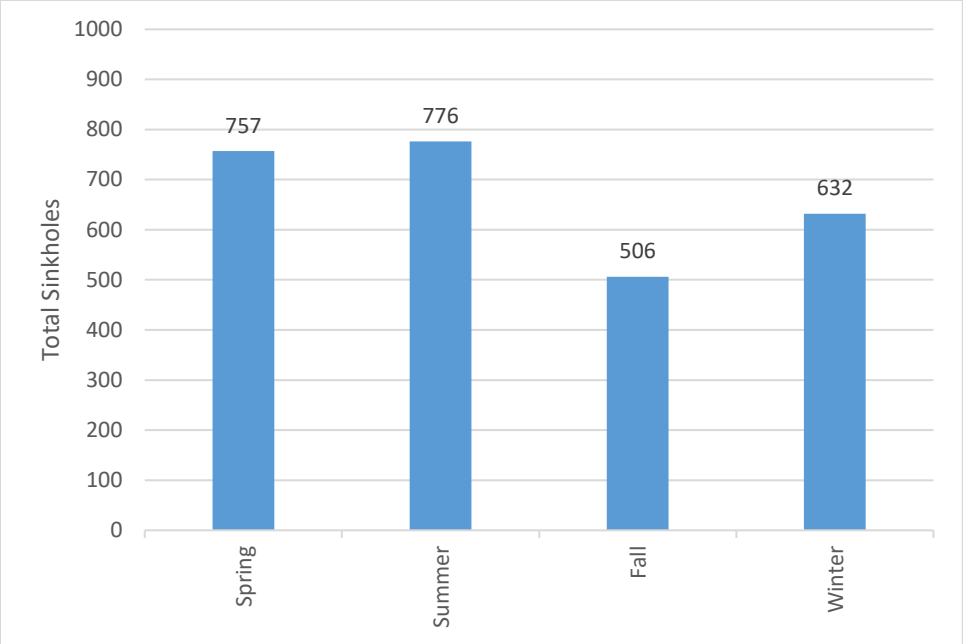

Figure 17: Total Florida sinkholes per season from 1954 to 2016.

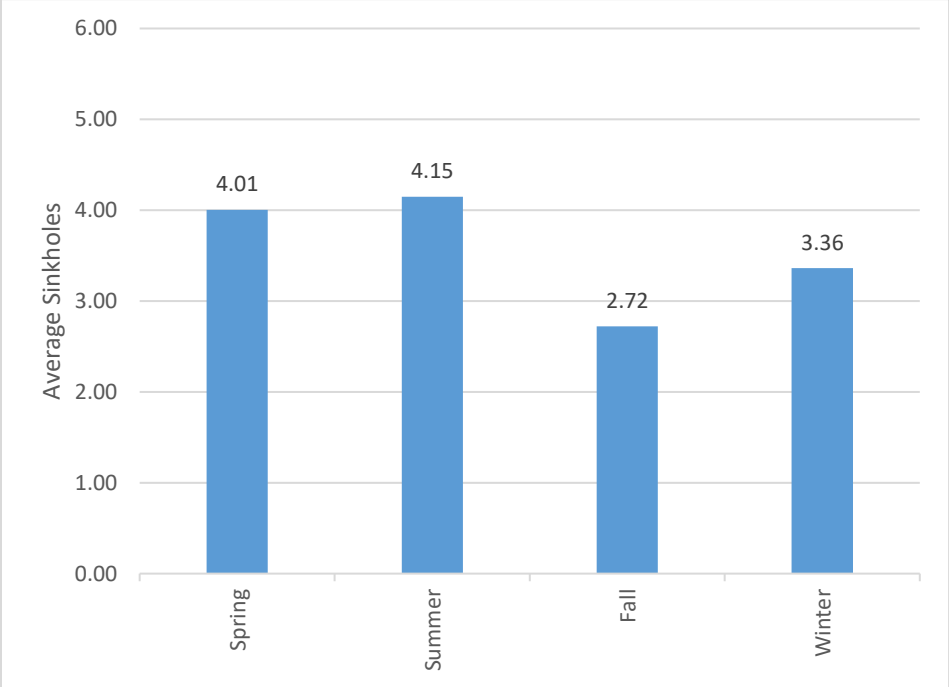

Figure 18: Average total sinkholes formed in Florida for each season from Jan 1954 - June 2016.

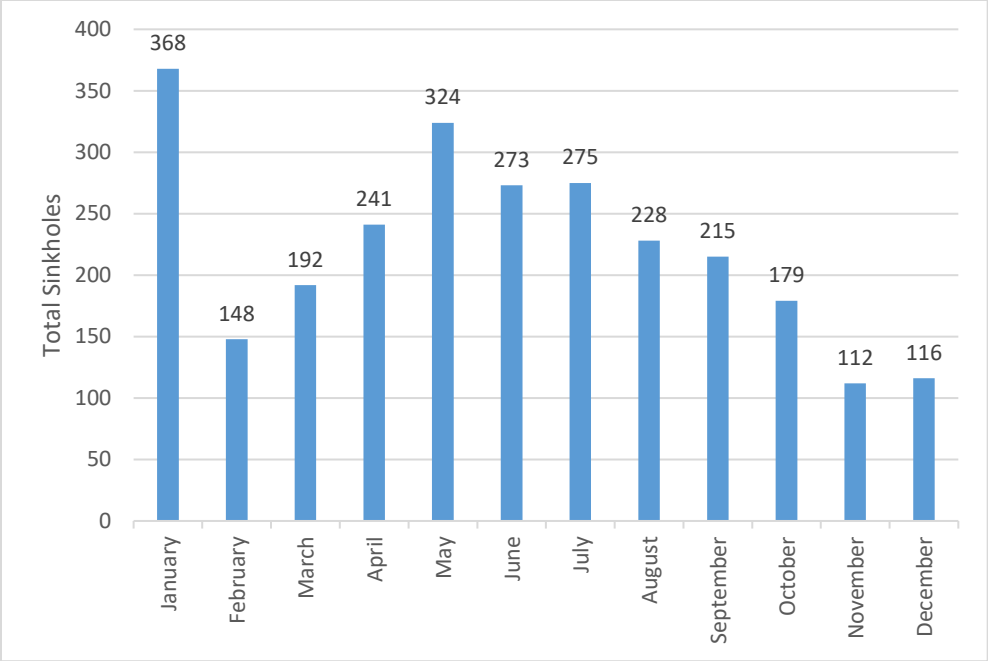

Figure 19: Total sinkholes formed in Florida per month between Jan 1954 - June 2016.

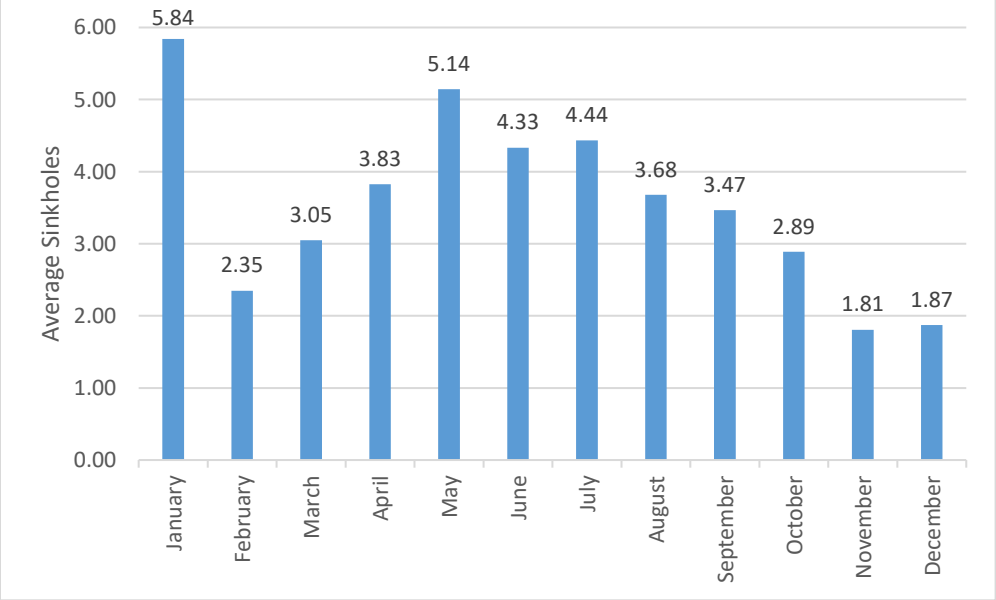

Figure 20: Average total sinkholes formed in Florida from Jan 1954 - June 2016.

Of the three Poisson models, the "negative binomial with log link" model was determined to be the best fit (AIC = 1965) compared to the "Poisson loglinear" models (zero count data removed AIC = 2206; zero count data included AIC = 3182) as it had a better AIC. The negative binomial with log link model (including the zero count months) included precipitation, average temperature, CO_2, NAO, EA, and EP/NP as significant variables in sinkhole formation ($p < 0.05$).

Discussion

Regression Results and Covariates

During the dates of increased sinkhole formation, significant spells of freezing temperatures occurred. Aurit, Peterson, and Blanford (2013) found that when Florida experiences a frost-freeze event, the farmers utilize a larger supply of groundwater to protect crops by the use of spray-freezing. In particular, this practice is often used on strawberry crops as Florida is the primary producer of strawberries in the US during the winter. Although other methods to protect the crops exist, this method is particularly effective and affordable. An excess of groundwater withdrawal causes the water table to drop allowing for the sinkholes to quickly form, which are generally clustered near groundwater pumps. This could explain why average temperature was not found to be a significant factor. Although dissolution should be greater with warmer temperatures, the lowering of the groundwater table is more significant. Having greater rates for both high and low temperatures is a non-linear relationship that is not properly captured in linear models. It also shows that although climate factors do have an impact on sinkholes, anthropogenic influences (such as groundwater use) can have a more pronounced effect.

Precipitation was the most highly correlated variable; this was expected as sinkholes require water for dissolution. However, the reaction of the karst environments to water table

fluctuations is more telling than the regular influx of runoff into the system (Cahalan, 2015) making the correlation harder to detect.

Of the teleconnections, ENSO and NAO were the most commonly incorporated variables within the models. These variables match the expected result. ENSO is known to significantly affect precipitation, temperature, and CO_2 plant uptake in Florida (Malone et al., 2014).

The monthly regression determined better results for the Orlando, Tallahassee, and Lake City regions while the seasonal regression had better results for the Tampa region. The log-transformed sinkhole datasets showed the best results compared to the total sinkhole count, time lag counts, and log-transformed time lag counts.

Limitations

The dataset contains a mixture of sinkholes and subsidence events. If a small subsidence event occurs due to an anthropogenic event, such as a broken pipe line, occurs, it should not be counted as a sinkhole. Since no differentiation could be made for certain throughout the dataset, all sinkhole/subsidence occurrences were treated equally.

The dataset contains the dates when the sinkholes occurred; this means that people were around at the time of the occurrence to notice the sinkhole appear or form. More sinkholes may form under the same conditions, but if they are in more remote locations, than they may not be recorded skewing the data. Saying that the sinkholes are only clustered around the cities would be a poor assumption as more sinkholes may be in more remote locations away from observance.

The NOAA Global Monitoring Division collects CO_2 data at four different observatories (Mauna Loa, Hawaii; American Samoa; Barrow, Alaska; and South Pole, Antarctica). Since none of these observatories are located in Florida, the Mauna Loa, Hawaii data were utilized. Since data were not taken directly from Florida, the actual values may differ; Florida and Hawaii

have differing weather patterns, differing total vegetative areas and densities (for absorbing CO_2), and differing distances to CO_2 emitting volcanoes.

Future Work, Implications, and Conclusions

Future work includes replicating the models with daily temperature and precipitation data compared to the used monthly averages. With this finer temporal scale, it may be easier to observe specific freeze/frost events and/or large precipitation events. For example, if farmers know that a frost/freeze event is about to happen, then groundwater may significantly drop right before the event. These frost/freeze days may then show a correlation to a negative time lag of sinkhole formation.

Another consideration is to add a binomial frost/freeze event variable determining whether or not a frost/freeze event has occurred. This could help account for specific dewatering events while keeping the average temperature variable to account for normal dissolution processes.

Carbon Dioxide levels have generally increased over time while maintaining their natural fluctuations. De-trending the CO_2 may reveal a better correlation between the natural CO_2 fluctuations and sinkhole formation.

The Poisson regression models for Tampa had very different results than the OLS stepwise regression. The OLS regression only included the NAO, whereas the Poisson negative binomial with log link model included precipitation, average temperature, CO_2, NAO, EA, and EP/NP to be significant variables in sinkhole formation ($p < 0.05$). The Poisson regression does not include an adjusted R^2 value for the model but AIC values could be calculated for both models. The AIC for the OLS regression was 4836 showing the OLS regression as the worst

method of all the utilized methods. Future research includes re-running all the models with the Poisson negative binomial with log link analyses.

For having such a high risk of sinkhole formation, Florida should have a more complete catalogue of sinkholes within the state than what is currently available. The creation of such a dataset is highly recommended. Knowing where the sinkholes are and how different factors influence sinkhole formation should be a high priority for Florida for the safety of the people. This study showed that increased precipitation is the most influential (climate-related) variable on sinkhole formation; with this information more widely published, the public can be better prepared for when sinkholes are more likely to form.

In conclusion, sinkhole formation can minimally be modelled using precipitation, temperature, CO_2, and teleconnection data with a linear regression model. Other models, such as the Poisson negative binomial regression may show better results, but is currently part of ongoing investigation. Using these variables to help predict when sinkholes might form may allow for a forewarning to the public within high risk zones allowing time for safety measures to be implemented.

References

Aurit, M.D., Peterson, R.O., and Blanford, J.I., 2013, A GIS analysis of the relationship between sinkholes, dry-well complaints and groundwater pumping for frost-freeze protection of winter strawberry production in Florida: PLoS ONE, v. 8, doi: 10.1371/journal.pone.0053832.

Blender, R., Luksch, U., Fraedrich, K., and Raible, C.C., 2003, Predictability study of the observed and simulated European climate using linear regression: Quarterly Journal of the Royal Meteorological Society, v. 129, p. 2299–2313, doi: 10.1256/qj.02.103.

Cahalan, M.D., 2015, Sinkhole formation dynamics and geostatistical-based prediction analysis in a mantled karst terrain [M.S. thesis]: Athens, University of Georgia, 64 p.

Cunningham, K.J., Locker, S.D., Hine, A.C., Bukry, D., Barron, J.A., and Guertin, L.A., 2003, Interplay of Late Cenozoic Siliciclastic Supply and Carbonate Response on the Southeast Florida Platform: Journal of Sedimentary Research, v. 73, p. 31–46, doi: 10.1306/062402730031.

Florida Center for Instructional Technology (FCIT): University of South Florida, 2002, Florida's land then and now: Exploring Florida: http://fcit.usf.edu/FLorIDA/lessons/land/land.htm (accessed January 2018).

Gao, Y., Luo, W., Jiang, X., Lei, M., and Dai, J., 2013, Investigations of Large Scale Sinkhole, in Proceedings, 13th Sinkhole Conference, NCKRI Symposium 2: Carlsbad NM, U.S., National Cave and Karst Research Institute, p. 327-332.

Lane, E., 1986, Karst in Florida: Florida Geological Survey Special Publication, n. 29, 100 p.

Malone, S.L., Staudhammer, C.L., Oberbauer, S.F., Olivas, P., Ryan, M.G., Schedlbauer, J.L., Loescher, H.W., and Starr, G., 2014, El Nino Southern Oscillation (ENSO) enhances CO_2

exchange rates in freshwater marsh ecosystems in the Florida Everglades: PLoS ONE, v. 9, 30 p., doi:10.1371-journal.pone.0115058.

Parinet, J., Julien, M., Nun, P., Robins, R.J., Remaud, G., and Hohener, P., 2015, Predicting equilibrium vapour pressure isotope effects by using artificial neural networks or multi-linear regression - A quantitative structure property relationship approach: Chemosphere, v. 134, p. 521-527, doi: 10.1016/j.chemosphere.2014.10.079.

Poag, C.W., 1991, Rise and demise of the Bahama-Grand Banks gigaplatform, northern margin of the Jurassic proto-Atlantic seaway: Marine Geology, v. 102, i. 1-4, p. 63-130, https://doi.org/10.1016/0025-3227(91)90006-P.

Sainani, K.L., 2013, Understanding linear regression: American Academy of Physical Medicine and Rehabilitation, v. 5, p. 1063-1068, doi: 10.1016/j.pmrj.2013.10.002

Scott, T.M., 2001, Text to accompany the geologic map of Florida: Florida Geological Survey Open-File Report 80, 28 p., https://sofia.er.usgs.gov/publications/maps/florida_geology/OFR80.pdf (accessed March 2017).

U.S. climate data, 2016, Climate Florida: Your Weather Service: https://www.usclimatedata.com/climate/florida/united-states/3179 (accessed Jan 2017).

CHAPTER 3

CLIMATE EFFECTS ON TENNESSEE SINKHOLES: GLM LOGISTIC REGRESSION AND MAXENT DERIVED APPROACHES

By:

Kimberly Blazzard, T. Andrew Joyner, Ingrid Luffman

Abstract

In Tennessee, sinkholes are prevalent in the center and eastern portions of the state, and 18,081 sinkholes have been recorded from topographic maps, indicating that sinkholes are a serious risk. For this study, two models were used to model the probability of sinkhole formation: a general linearized (GLM) logistic regression approach and a MaxEnt derived species distribution model. WorldClim climate normals were used for analysis and combined with known significant non-climate variables. Results showed a highly significant ($p < 0.001$) correlation between sinkhole events and precipitation, maximum temperature, solar radiation, wind speed, slope, distance to rivers, carbonate bedrock, and distance to faults and were retained as variables for both models. The final logistic regression model had a pseudo-R^2 value of 0.329 and correctly identified 87.6% of the validation data within very high and high risk zones. Areas of highest sinkhole risk were found in the Valley and Ridge Province and the Nashville Basin Province.

Keywords: Sinkhole, Doline, GIS, Climate, Predictive Modeling, Karst, Tennessee, Logistic Regression, MaxEnt, R

Introduction

Environmental Background

Tennessee covers 109,849km^2 (42,143 mi^2). It ranks first among all US states for the number of caves (Veni et al., 2001). With over 18,000 sinkholes currently mapped and numerous others still forming, sinkholes are a significant problematic feature of the Tennessee landscape. These sinkholes include both dissolution and collapse types within carbonate lithologies.

Most of Tennessee experiences a temperate climate, with mild winters and warm summers (UTIA, 2017). Normal precipitation ranges from 109 to 203 cm (43 to 80 inches) per year with the lowest values within the northeastern portion of the state and the highest values along the southeastern half of the North Carolina border. The normal average temperature ranges from 6° to 18° C (43° to 64° F) with a general trend of cooler temperatures in the northeast and warmer temperatures in the southwest; the coolest temperatures do not follow this trend as they are located at the highest elevations within the Appalachian Mountain range (Fig 21). Normal maximum temperature ranges from as high as 35° C (95° F) in July to as low as 1° C (34° F) in January, and normal minimum temperatures range from as low as -9° C (16° F) in January to as high as 24° F (75° F) in July (Tennessee Climate Office).

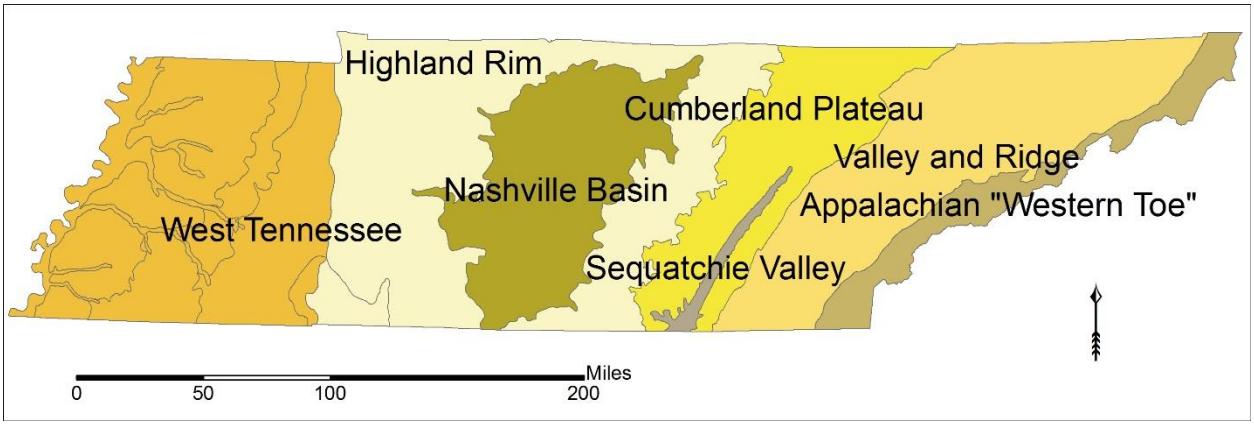

Figure 21: Tennessee Physiographic Provinces; USGS data.gov

The National Land Cover Database 2011 (NLCD) indicates that land use in the state consists of forests (including deciduous, evergreen, and mixed) (39.7%), pasture/hay/cultivated crops (29.0%), developed landscape (9.8%), shrub/grasslands (6.2%), wetlands (2.9%), open water (2.1%), and barren landscape (0.2%).

The western half of Tennessee is relatively flat with gently rolling plains. This region ranges from 61-76 m (200-250 ft) near the Mississippi River to ~183m (~600 ft) above sea level near the Tennessee River (UTIA, 2017; NCDC). This region is primarily underlain by Tertiary-Quaternary silts, sand, and mudrocks (Fig 22) (Table 2). The easternmost portion of the state is encompassed by the Appalachian Mountain Range. This region is primarily underlain by Cambrian-Ordovician carbonates but with a mixture of greywackes, siltstones, shale, sandstone, carbonates, and quartzites on the easternmost boundary. The central portion of the state is an eroded dome (Tennessee Central Basin) exposing an Ordovician limestone surrounded by Mississippian cherts, mudstones, and claystones (USGS MRData). The carbonate regions of interest, which are most susceptible to sinkhole formation, are within the Mississippian, Ordovician, and Cambrian age rocks.

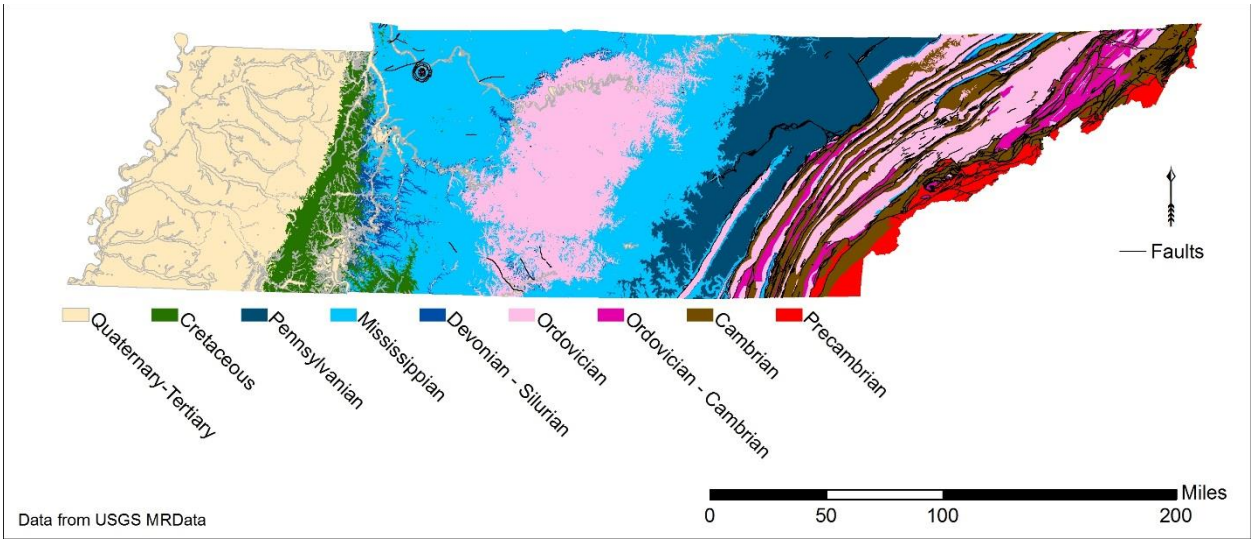

Figure 22: Tennessee Geologic Map from USGS MRData

Table 2: Tennessee rock types per geologic unit. Data obtained from the TN Department of Environment & Conservation.

Quaternary – Tertiary	Sand, silt, clay, gravel, and loess
Cretaceous	Sand, clay, silt, and gravel
Pennsylvanian	Sandstone, shale, conglomerate, siltstone, and coal
Mississippian	Limestone, chert, shale, siltstone, sandstone, and dolostone
Devonian – Silurian	Limestone, chert, shale, and sandstone
Ordovician	Limestone, shale, dolostone, siltstone, sandstone, and claystone
Ordovician – Cambrian	Dolostone, limestone, shale, chert, siltstone, and sandstone
Cambrian	Shale, dolostone, limestone, sandstone, conglomerate, quartzite, arkose, greywacke, and siltstone
Precambrian	Sandstone, conglomerate, siltstone, arkose, greywacke, quartzite, phyllite, slate, schist, metamorphosed lavas and tuffs, metagabbro, rhyolites, diorite, granite, granitic gneisses, monzonite, quartz latites, anorthosite, and diabase

Statistical Background

Menard (2010) explained that if the regression was trying to show a linear relationship (as with an ordinary least squares (OLS) linear regression,) the intercept would represent the value of the dependent variable when the value of the independent variables equal zero. The independent variable coefficients represent the change in the value of the independent variable with a one-unit increase in the value of the dependent variable. The logistic regression does not estimate a linear relationship but produces an s-shaped curve model predicting the probability of an occurrence happening (Fig 23). Logistic modeling has been successfully used to model

sinkhole probability for varying locations including Konya, Turkey (Ozdemir, 2015); Guangxi, China (Zhou, Yan, Chen, and Zhang, 2016), and the Apulia Region, Italy (Pellicani, Spilotro, and Gutiérrez, 2016).

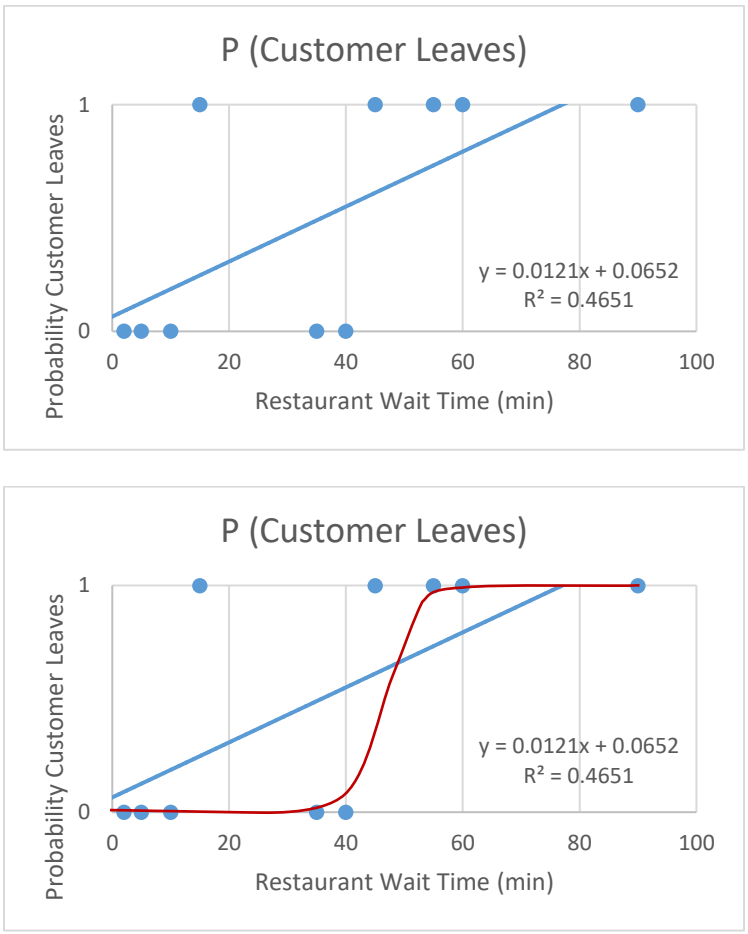

Figure 23: An OLS regression (top) creates a linear relationship while a logistic regression (bottom) creates an s-shaped relationship showing the probability between the dependent variable occurring vs. not occurring. This theoretical example shows the probability a customer would wait to be seated at a restaurant based on the approximate wait time.

Logistic regression results produce regression coefficients for each variable expressing their log-odds (or logit) of occurrence. Other ways to interpret the results include calculating the odds:

$$Odds = e^{logit(Y)}$$

or determining the probability:

$$Probability = odds / [1 + odds]$$

The odds describe the relative chance of the dependent variable occurring (a sinkhole forming); it is bounded below by 0, but not bounded above (Fox, 2008). For example, an odds-ratio of 3 can be read as: a 3 to 1 chance that a sinkhole will form. The probability is the chance of something happening ranging from 0 (not likely to happen) to 1 (most likely to happen). For example, a 0.5 probability can be read as: a fifty percent chance that a sinkhole will form. Note: the probability surface is created within a logistic regression. The logit is the natural logarithm of the odds; this creates symmetry around 0 allowing for infinitely large and small values (Fox, 2008; Menard, 2010). Of the three result interpretation styles, the odds and the probability are generally easier for people to comprehend while the logit is considered the best mathematical form for analyzing dichotomous dependent variables (Menard, 2010).

A variable with an odds less than 1 implies a negative relationship, while an odds ratio more than 1 implies a positive relationship (King, 2017). For positive relationships, the higher the odds ratio, the higher likelihood that a sinkhole will form; for negative relationships, the lower the odds ratio, the lower the likelihood that the dependent variable (in this case, a sinkhole) will form as the independent variable increases (King, 2017).

Area under the curve (AUC) is calculated by a receiver operating characteristic (ROC) analysis in MaxEnt (Phillips, Aneja, Kang, and Arya, 2006; Anderson, Lew, and Peterson, 2003). An AUC of 0.5 indicates that the model's performance is no better than random with better values closer to 1.0 (Young, Carter, and Evangelista, 2011). If the AUC value, or the probability of success, is greater than 0.5, than the model is qualitatively considered a success, while a value less than 0.5 is qualitatively considered a failure (Menard, 2010).

The MaxEnt model output includes a series of response curves for each variable. Each plot models one independent variable while holding the other independent variables constant at their mean (BCCVL, 2016), and displays the change of probability (from 0 to 1) over the range of values for the particular variable. So, if a variable has a changing probability from one side of the graph to the other, than the area of higher probability shows the optimal values for "occurrence" of the point variable. The threshold applied to the response curves indicates the "cut-off point" for the optimal value ranges to be considered. If the response curve is relatively flat, then no optimal range is found.

Research Questions

1. Which climate variables are the most influential in predicting sinkhole formation?
2. Where are sinkholes expected to form in Tennessee based on climate-sinkhole relationships?

Data and Methods

Spatial Distribution of TN sinkholes

A sinkhole shapefile was created from a database of sinkhole locations recorded from USGS TN topographic quadrangles and obtained from the University of Tennessee-Knoxville (tnlandforms.us). It includes 18,081 sinkhole records with the following attribute data: latitude, longitude, perimeter, area, depth, volume, and elevations (Fig 24). Lithologic data, including faults systems (line shapefiles) and rock types (polygon shapefiles), were collected from USGS Mineral Resources. Tennessee roads were downloaded as a line shapefile from Data.Gov. All

data were projected into the NAD 1983 StatePlane Tennessee FIPS 4100 (meters) projection and mapped using ESRI ArcMap 10.5.1.

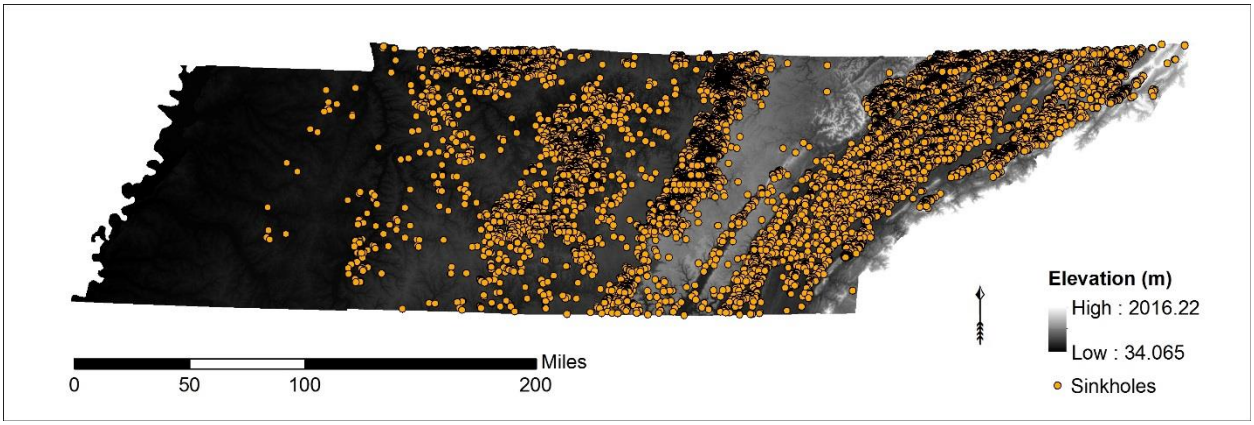

Figure 24: Mapped Tennessee sinkholes and elevation.

Spatial distribution analysis was conducted to locate the geographic center of the sinkhole dataset and to identify sinkhole clusters. This was done by calculating the mean center, the geometric mean, the harmonic mean, the median center, and the mean center of minimum distance using Crimestat IV.

Second, a Nearest Neighbor Analysis (NNA) was conducted to quantitatively describe dispersion of the sinkholes using the Nearest Neighbor Index (NNI). A rectangular border correction was applied as some of the sinkholes' nearest neighbors' may actually lie outside of the Tennessee boundaries.

Third, a hot spot analysis was conducted to identify where the sinkholes concentrate. This was computed using a Nearest-Neighbor Hierarchical Spatial Clustering analysis. A cluster was determined if there was a minimum of 100 sinkhole sites within a 10 mile radius. A 1x standard deviation was utilized.

Lastly, three kernel density estimates (KDE) were conducted to model hotspots of highest risk for sinkhole formation. Since sinkholes require certain features (such as a karst lithology), it is possible that the area of influence can decline rapidly near an incident; three different kernel distribution and bandwidth combinations were chosen for this reason. KDE#1 used a negative exponential distribution, an adaptive bandwidth, and a minimum sample size of 100. KDE#2 used a normal distribution, an adaptive bandwidth, and a minimum sample size of 100. KDE#3 used a normal distribution and a fixed interval of 11.26 km (7 miles) calculated using the Fotheringham Bandwidth Formula with N equaled the total number of occurrences/points and σ was the standard distance deviation.

$$h_0 = [2/(3N)]^{1/4}\sigma$$

The concluding KDE surface was then visually compared to faults and lithology within Tennessee.

Logistic Regression Model

This study utilized a logistic regression model to examine relationships between Tennessee sinkholes and climatological patterns and other long-term variables known to impact sinkholes. These other variables include: slope (derived from ⅓ arc-second (10 m^2) National Elevation Dataset), bedrock lithology (USGS MRData), distance to major faults (derived from USGS MRData), and distance to rivers/streams (derived from USGS HydroSHEDS). Climate normals data, available from WorldClim at a resolution of 30 arc seconds (1 km^2) included annual minimum temperature, annual maximum temperature, average range of temperature, and annual precipitation interpolated from 1950-2000 data and annual average water vapor pressure (kPa), annual average wind speed (m s^{-1}), and annual average solar radiation (kJ m^{-2} day^{-1}) interpolated from 1970-2000 data since data were not available prior to 1970. The bedrock

lithology was converted into a binary variable representing presence of carbonate rock; if the lithology included a carbonate rock (limestone, dolostone, or calcarenite), it was assigned a 1, while all other lithologies were assigned a 0. This file was then converted into a raster file.

The normal wind speed ranged from 1.87 to 6.2 m s^{-1} (Fig 25). Generally, the highest wind speeds were found in areas of high and low elevation where there was higher exposure. The highest values were found within the Appalachian Mountain Range and within the western third of the state. The lowest values were found within the Valley and Ridge Province and the Sequatchie Valley Province (Fig 21 and 25).

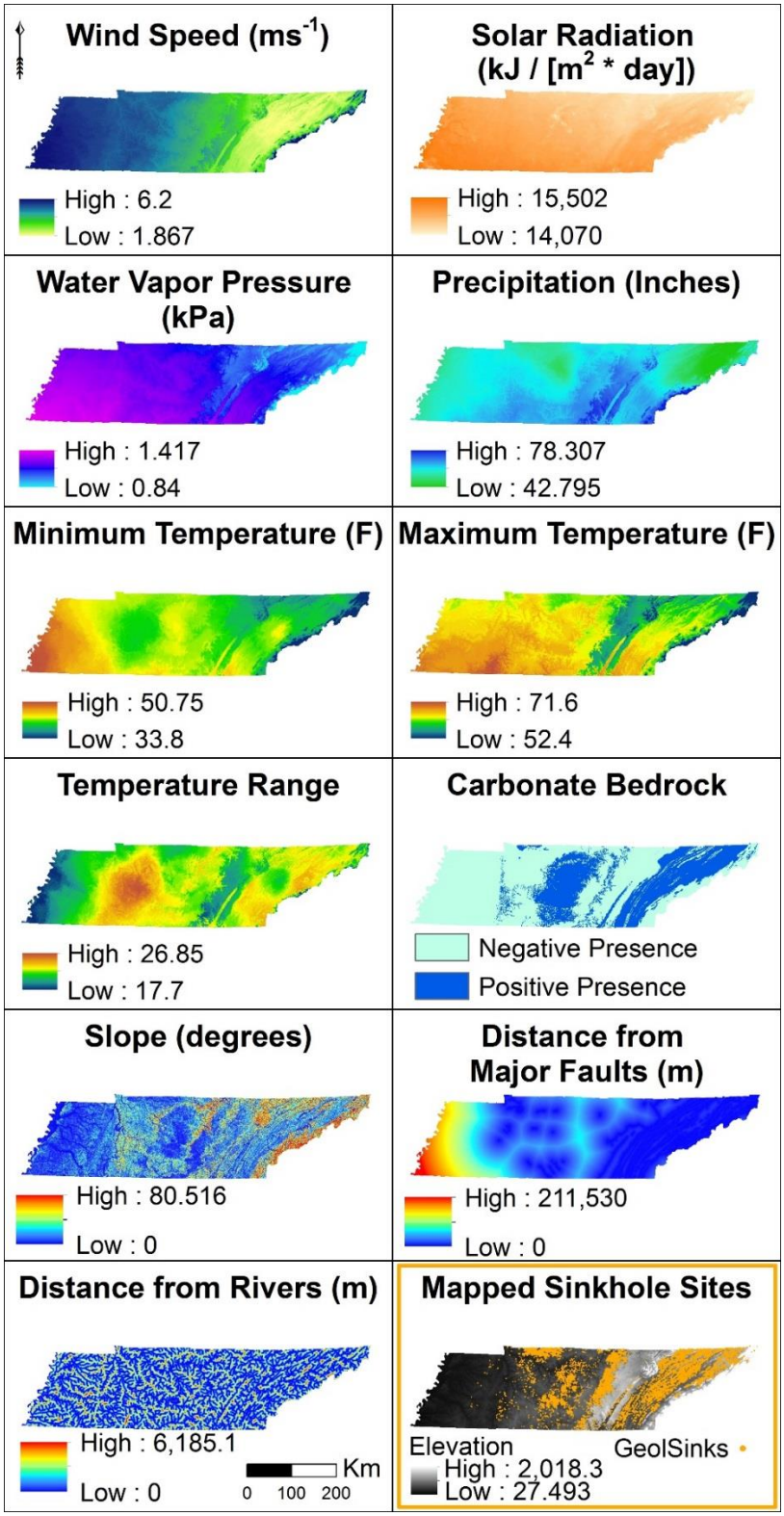

Figure 25: Candidate variables relevant to sinkhole formation (independent variables) compared to mapped sinkhole sites (dependent variable). Sources: National Elevation Dataset, USGS MRData, USGS HydroSHEDS, WorldClim, University of Tennessee

Normal solar radiation ranges from 14,070 to 15,502 kJ m^{-2} day^{-1}. The lowest values are primarily located around the regions of higher population. This is distinctly seen around Nashville, Memphis, Chattanooga, and the Tri-Cities region. The lowest values are found within the south-western portion of the state, not including the Memphis suburban region. Note: this variable is typically measured as the irradiance, or the amount of light energy from a light source (such as the sun) that hits a square meter of an object (such as the ground) over a second (Garner 2008).

The normal water vapor pressure ranges from 0.84 to 1.42 kPa. The lowest values are found along the eastern boundary of the state, particularly within the northeastern corner of the Appalachians. The highest values are found within the southwestern corner of the state near Memphis. There is a general gradient of values ranging from these low to high areas.

The normal precipitation ranges from 108.7 to 198.9 cm (42.8 to 78.3 inches). The highest values are found along the southeastern border and surrounding the Sequatchie Valley. The lowest values are found within the northern portion of the Valley and Ridge Province.

The normal minimum temperature ranges from 1° C (33.8° F) to 10.4° C (50.75° F). The lowest of these temperatures follow along the eastern boundary of the state at the highest elevations within the Appalachians, while the highest of these temperatures are found in the south-western portion of the state. There is a general gradient between the low and high values.

The normal maximum temperature ranges from 11.3° C (52.4° F) to 22° C (71.6° F). There is a general gradient from lower temperatures in the north to higher temperatures in the south. However, the lowest of these temperatures are found within the Appalachian Mountains and within the Cumberland Plateau Province.

The normal average monthly temperature range spans from 9° - 14° C (17.7° to 26.85° F). The smallest temperature range is located along the western boundary of the state and within the Cumberland Plateau Province. The region with the largest normal temperature range is within the Central Basin Province and the Western Highland Rim Province.

The carbonate bedrock primarily underlies the Valley and Ridge, Sequatchie Valley, and Nashville Basin Provinces. The carbonate layers overlap into the other provinces, but only the top bedrock layer is mapped during geologic mapping (not including cross sections). Therefore, an area may have a differing bedrock top layer, but may still have an underlying carbonate layer allowing for collapse sinkholes to form. This variable was created from a general bedrock lithology shapefile delineated into a bivariate layer of positive (1) or negative (0) carbonate. The carbonate bedrocks included limestone, dolostone, and calcarenite (a carbonate with sand-sized clasts (Orme 1982)).

The slope ranges from 0 to 80.9°. The areas of highest slope are within the Appalachian Mountain Range. There are several areas of low slope, with the largest areas including the Nashville Basin Province and the western quarter of the state.

The majority of the mapped faults are along the Appalachian Mountains and throughout the Valley and Ridge Province. Fewer faults are mapped within the western portion of the state, forming a gradient from small distances to faults in the east to greater distances to faults in the west. The distances range from 0 meters to 211,530 meters.

The distance from rivers ranges from 0 meters to 6185.1 meters. There is no particular pattern differentiating areas of short/long distance. The rivers are generally evenly-distributed throughout the state.

Absence data points were also created to compare actual sites to areas lacking in sinkholes. These points were created in ESRI ArcMap 10.5 using a systematic sampling method for the same rectangular extent as the Tennessee boundary. The method created a grid with points located within each grid. These points did not overlap with actual sinkhole locations, and points outside of the Tennessee boundary were discarded. This created a total 1,308 absence sites periodically distributed.

Data from the independent variables were extracted at presence and absence points. From the presence sites, the data were partitioned randomly into training and testing data. For this purpose, 70% of records (12,657 sinkholes) were used to model the regression and 30% (5424 sinkholes) were used for model validation. The 70% data point set and the absence points, with their accompanying attribute data, were joined and a new binary variable was created assigning the true sinkhole sites a value of 1 and the absence sites a value of 0. Spearman correlation coefficients were calculated in SPSS 24 and models were developed in R.

Within R, a general linearized model (GLM) was utilized for the logistic regression.

```
> log_data1 = read.table("TNSinks9.csv", sep=",", header = TRUE)
> names(log_data1); dim (log_data1)
> model2 <- glm(ID ~ precinchpr + RiverDist2 + FaultDist2 + WindCLip_P + avesolar_P +
tngeolrast + tnslp2 + tmax_proj, data = log_data1, family = binomial ())
> summary (model2)
```

The R^2 value was calculated by the following equation:

$$1 - (Null\ Deviance\ /\ Residual\ Deviance)$$

The first regression model included all variables. After determining which variables were not significant, the regression was reproduced without the non-significant variables. The model was tested several times with differing individual variables to determine if certain variables were not

included due to possible multicollinearity as no multicollinearity diagnostics were produced with the model.

The logistic regression coefficients were then applied back to the independent variables within ArcMap to create a new map. First, independent variables were resampled to the same spatial resolution (30 arc-seconds) and extent. Second, all independent variables were multiplied by their coefficients (calculated in R). Next, the rasters were combined into one mapped surface using the equation for logistic regression to display the probability with the intercept (α), the slopes of each independent variable (β_i), and each independent variable (x_i):

$$Yi = (e^{\alpha + \beta_i x_i}) / (1 + e^{\alpha + \beta_i x_i})$$

The hazard areas were classified into five different groups using a geometrical interval classification to determine zones of very low risk, low risk, medium risk, high risk, and very high risk. Probability at both the 70 percent sinkhole training sites and the 30 percent sinkhole testing sites was extracted, the percentages of each risk category were compared for the testing and training data to determine model accuracy. An accurate model will have similar percentages of sinkholes within each category for both training and testing data. If the model does not project the validation sites onto higher risk zones, then the percentages would not be similar and the model would be poor.

Finally, the determined risk zones were compared to Tennessee land cover classifications (National Land Cover Database, 2011) to determine the types of land cover most at risk for sinkhole development and to determine how much developed land area is at risk. This developed area includes open space, low intensity, medium intensity, and high intensity. The National Land Cover Database explains the difference between the upper and lower ranges as:

Developed, Open Space: areas of mixed vegetation and constructed materials with impervious surfaces accounting for less than 20% of total land cover.

and,

Developed, high intensity: areas of large development and population with impervious surfaces accounting for 80% to 100% of total land cover.

MaxEnt Model

A second model was created utilizing MaxEnt software Version 3.4.1 - a platform originally developed to produce maximum entropy species distribution models. Although MaxEnt is most commonly used for ecological modelling, it is useful for non-species presence data. Successful examples include landslide spatial modeling (Chen, Pourghasemi, Kornejady, and Zhang, 2017) and rhythmite spatial modeling (Shunk, 2009). The MaxEnt model is considered a spatial logistic regression while the first model used a general linearized logistic regression that has been spatially displayed. The spatial logistic regression attempts to accommodate for spatial autocorrelation. Spatial autocorrelation is when an observed value is dependent on the values of neighboring locations (Diao, 2015). A comparison of the two models should determine if the MaxEnt program is a good alternative to traditional spatial modelling methods. For the MaxEnt model, the same dependent and independent variables were used as with the first model (GLM).

One way to determine the validity of the model is to determine how many validation points are accurately predicted above a determined threshold. MaxEnt produces various thresholds specific to the model. The recommended threshold method for large datasets of presence/absence data, where random points are used instead of true absences, is one that maximizes the sum of the sensitivity and specificity (Liu, Newell, and White, 2016). Liu et al.

(2016) explains that the sensitivity is the proportion of correctly predicted presences from all presences and the specificity is the proportion of correctly predicted absences from all absences. Menard (2010) further explains that when the sensitivity is 100%, the specificity is 0% and when the specificity is 100%, the sensitivity is 0%. Two thresholds were used for this model: the "Equal training sensitivity and specificity" threshold and the "Maximum training sensitivity plus specificity" threshold. The maximum training threshold was also used to quantitatively analyze the response curves produced for each of the variables. These response curves show how each variable influences the MaxEnt model by displaying how the predicted probability of presence changes as each independent variable changes, while isolating all other variables at their mean sample value.

Results

Spatial Distribution of TN sinkholes

The harmonic mean, geometric mean, and the mean center converged in similar locations, approximately 27.4 km (17 miles) northeast of Crossville (Fig 26). The median center is approximately 21 km (13 miles) WNW from the mean center and the mean center of the minimum distance is approximately 22.5 km (14 miles) WSW from the mean center.

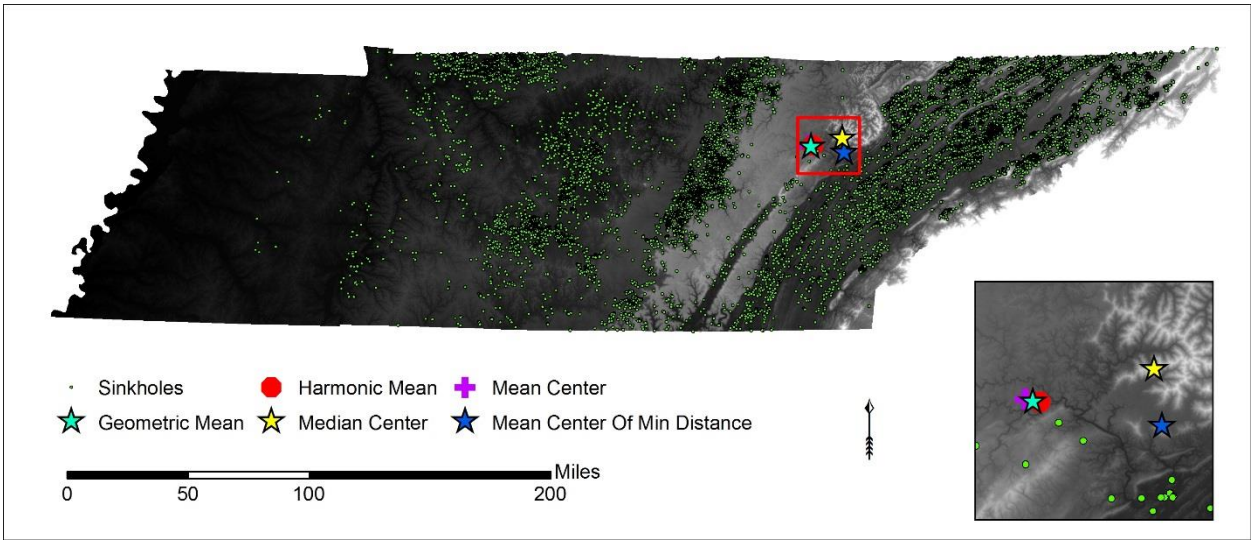

Figure 26: Crimestat statistical analysis of Tennessee sinkholes displaying the mean center, harmonic mean, geometric mean, median mean and mean center of the minimum distance. The overall trend shows the mean point in the northeastern portion of the state.

The mean nearest neighbor distance was 513.64 meters with a minimum distance of 38.98 meters. The Nearest Neighbor Index equaled 0.348, indicating that sinkholes are clustered. The results were highly significant with a *p*-value = 0.0001.

The hot spot analysis found 41 clusters, of which 23 were within the Valley and Ridge Province along the eastern border of Tennessee (Fig 27). The other clusters were found within the Eastern Highland Rim Province, the northernmost portion of the Western Highland Rim Province, and the center of the Nashville Basin Province of Tennessee.

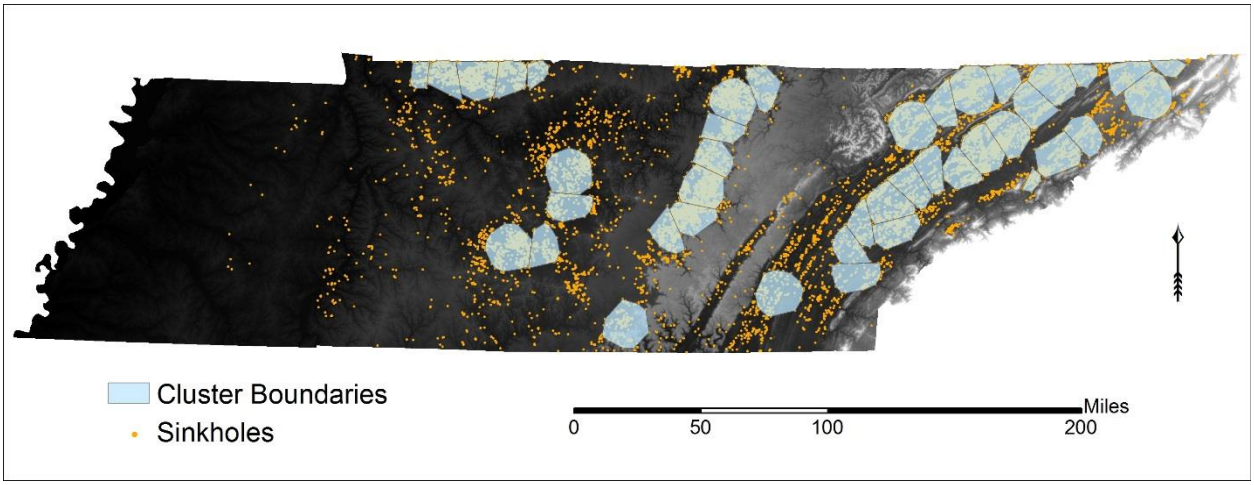

Figure 27: Tennessee cluster analysis showing the regions where there are at least 100 sinkholes within a 10 mile radius. 41 clusters were found, mostly within the Appalachian Mountains.

KDE#1 showed hundreds of tiny, sporadic areas of sinkhole density; this is not helpful for estimating which regions have higher susceptibility, so it was discarded. The visual assessment of KDE#2 (Fig 28) and KDE#3 (Fig 29) showed better interpolated surfaces, however, KDE#3 best fit the actual locations of the recorded sinkholes. KDE#2 overestimates the density in areas with fewer sinkholes and underestimates the density in areas with higher sinkhole occurrences. KDE#3 shows which areas, within the clustered regions, have the highest density of sinkholes. Since this KDE surface is the best of the three models, it is utilized for the rest of the comparisons.

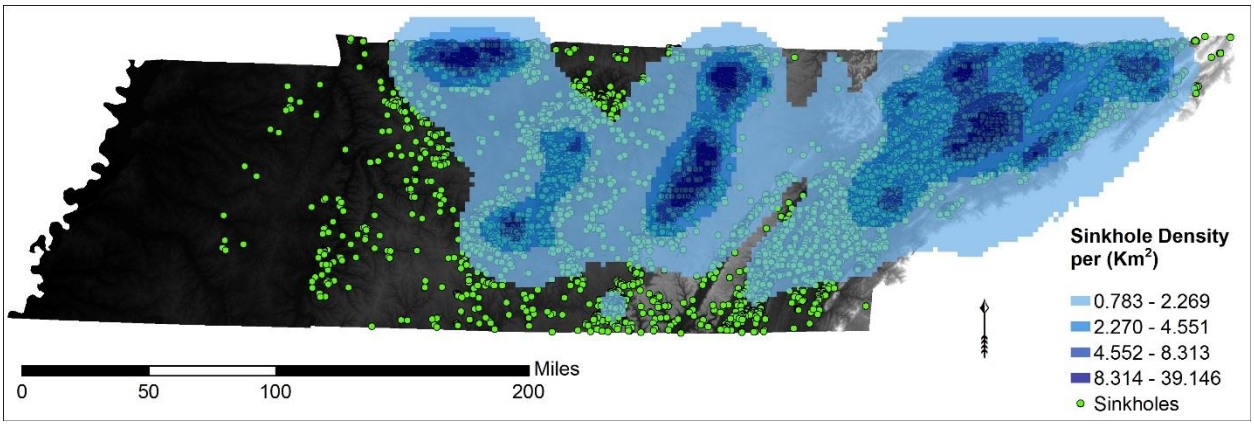

Figure 28: Tennessee KDE#2 shows an interpolated density surface utilizing a normal distribution, adaptive bandwidth and a minimum sample size of 100.

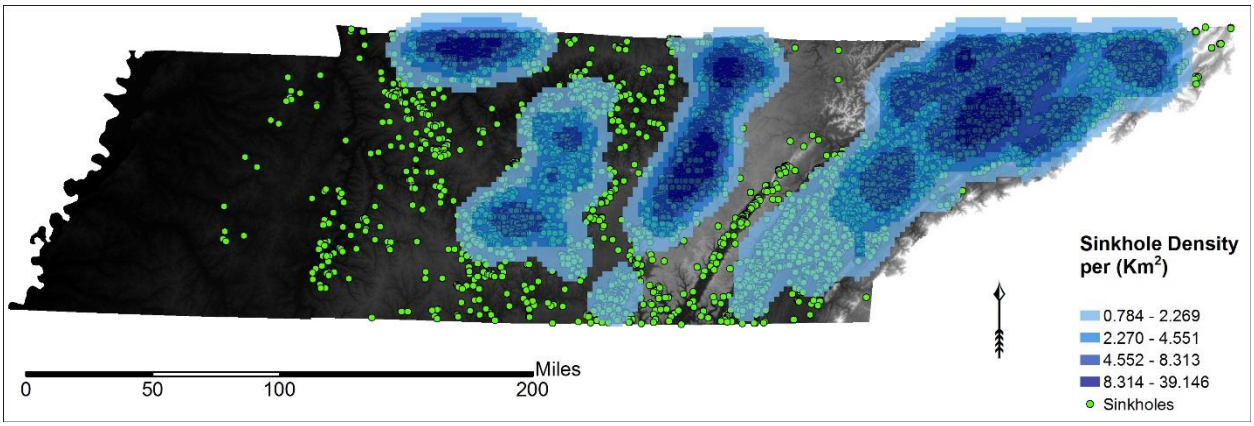

Figure 29: Tennessee KDE#3 shows an interpolated density surface utilizing a normal distribution and a fixed interval of seven miles.

When compared to a lithologic map, all regions of highest sinkhole density generally matched with the carbonate areas (Fig 30).

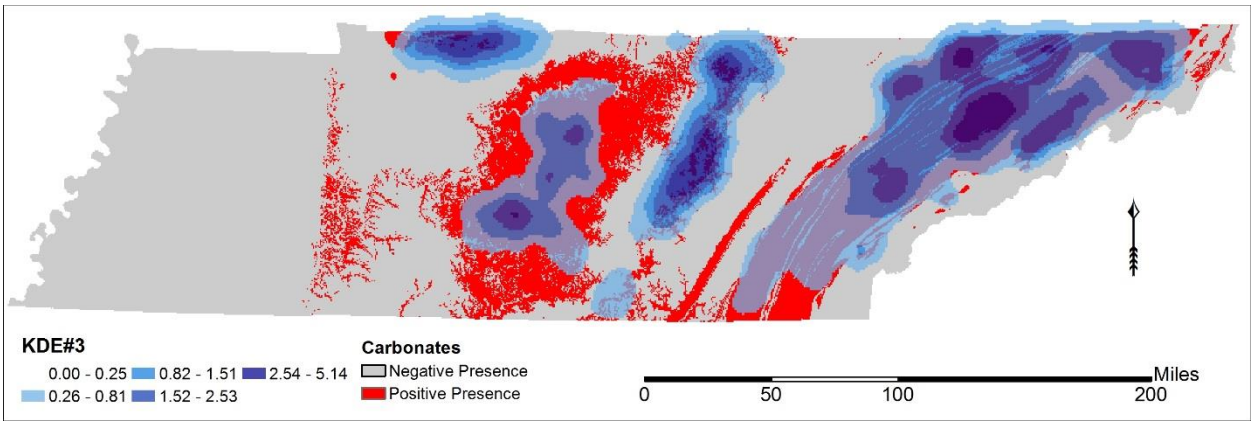

Figure 30: KDE#3 overlaid on a Tennessee carbonate map. The areas of highest density overlap with the carbonate layers.

The visual qualitative assessment of the fault comparison showed that the highest density areas matched with fault zones in the Valley and Ridge Province, but not with the faults in the other regions (Fig 31).

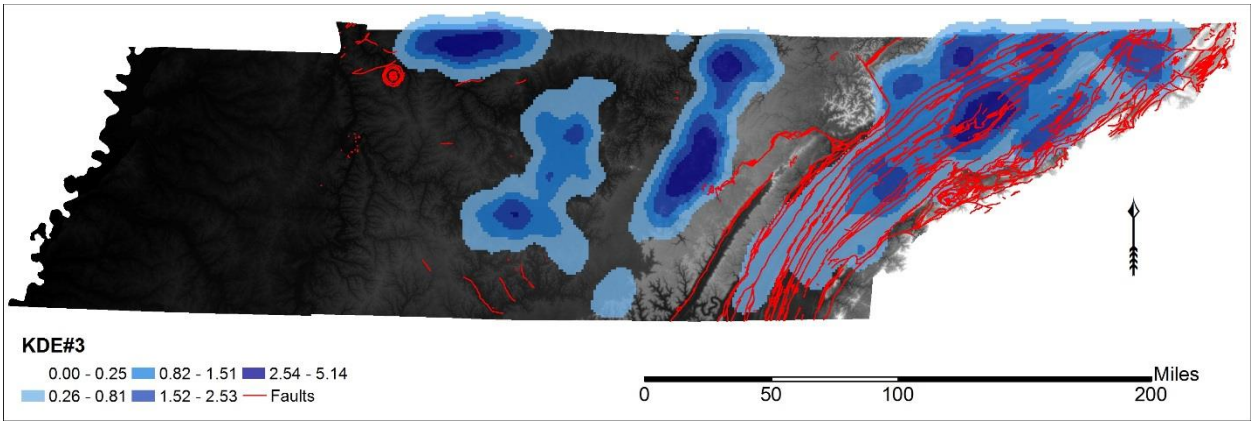

Figure 31: KDE#3 overlaid with the Tennessee major faults show a visual correlation in the eastern portion of the state (the Appalachian Mountains), but not in the central or western portions of the state.

Logistic Regression Model

The calculated linear equation (log odds) inputted into the logistic equation for the mapped model (Fig 32) with accompanying correlation coefficients was:

$$z = \alpha + \beta_i x_i = 81.25 + (0.04906 * precipitation) + (0.0001221 * rivers) + (-0.00001424 * faults) + (-1.31 * wind) + (-0.008339 * solar\ radiation) + (1.171 * carbonates) + (0.000006737 * \max temperature) + (-0.1506 * slope)$$

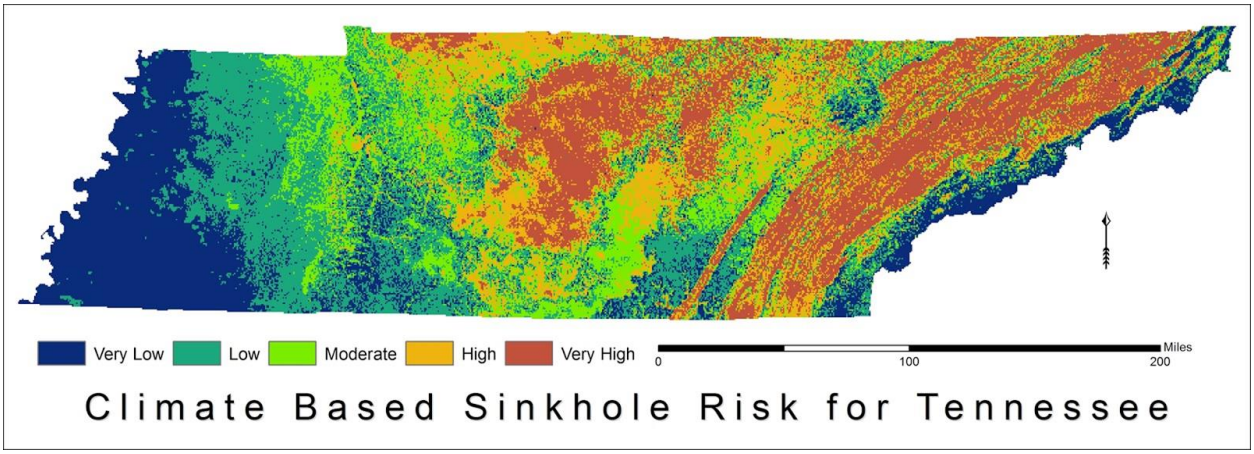

Figure 32: Tennessee sinkhole risk zones predicted using precipitation, high temperature, wind speed, solar radiation, distance to major faults, distance to rivers, lithology, and slope.

The overall pseudo-R^2 of 0.329 indicates that the model accounts for approximately 33% of the variability. An AUC of the model calculated as 0.838. The water vapor pressure, average annual temperature, and average temperature range were rejected from the model, either for non-significance or for multi-collinearity. The average minimum temperature showed a possible correlation with the sinkholes, however, when both minimum and maximum temperatures are added to the model, the minimum temperature was rejected. When the solar radiation and wind were excluded, the overall pseudo-R^2 dropped to 0.176.

When the independent climate variables retained by the model were separately compared to the sinkhole locations, the order of highest correlation to least correlation was: solar radiation

(pseudo-R^2 = 0.119), wind (pseudo-R^2 = 0.115), carbonates (pseudo-R^2 = 0.103), distance from faults (pseudo-R^2 = 0.102), precipitation (pseudo-R^2 = 0.0517), slope (pseudo-R^2 = 0.0496), and maximum temperature (pseudo-R^2 = 0.0236) (Fig 33). The individual model for distance from rivers (pseudo-R^2 = 0.0002) became non-significant.

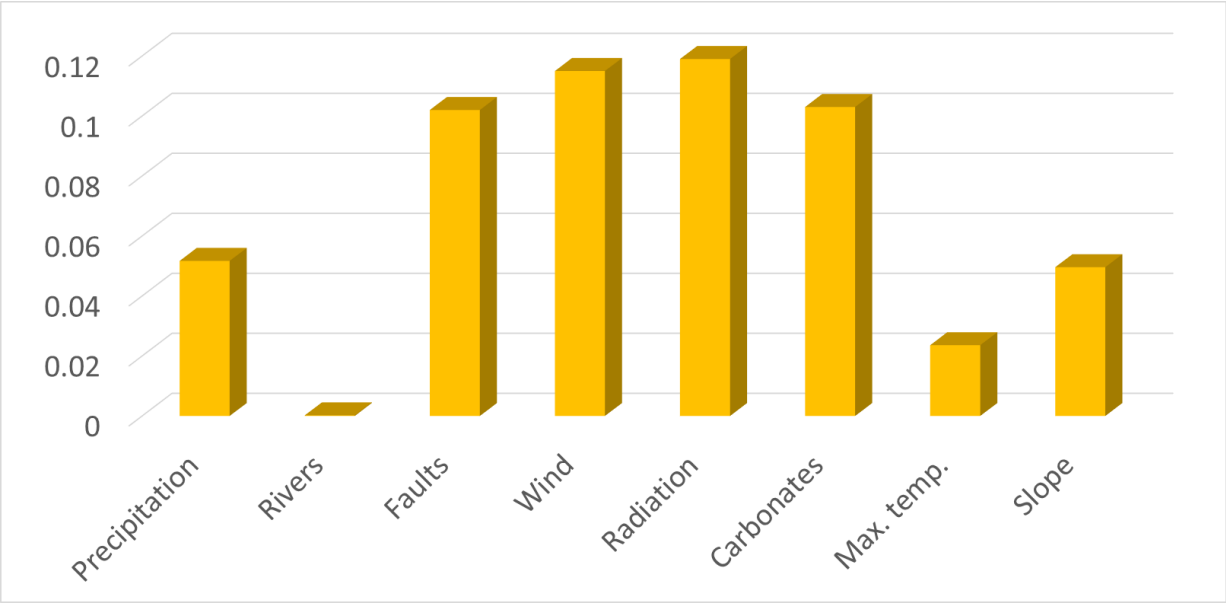

Figure 33: Pseudo-R^2 for univariate spatial logistic regression models on sinkhole presence.

The regression coefficient, for a particular variable, may display the amount of influence it has on the model. This means that for a one-unit increase in precipitation, a 0.049 increase in the log-odds for sinkhole formation is expected. Therefore, the odds of sinkhole formation are increased with a one-unit increase in precipitation by 1.05 and the probability by 0.512. The carbonate bedrock had the largest positive relationship (3.22 odds ratio); this shows that the odds of sinkhole formation are increased by a factor of 3.22 in areas of carbonate bedrock. The largest negative relationship was with wind (0.27 odds ratio); this shows that the odds of a sinkhole are increased by a factor of 3.7 as wind decreases (Table 3).

Table 3: Significant independent variables with regression coefficients, significance, odds, and probability. All individual models were significant at $p < 0.01$.

Variable	*Regression Coefficient*	*Odds*	*Probability*
(Intercept)	8.125e+01		
Precipitation	4.906e-02	1.05033	0.512
Distance to Rivers	1.221e-04	1.00012	0.500
Distance to Faults	-1.424e-05	0.99999	0.500
Wind	-1.310e+00	0.26982	0.212
Solar Radiation	-8.339e-03	0.99169	0.498
Carbonate Bedrock	1.171e+00	3.22199	0.763
Maximum Temperature	6.737e-01	1.96207	0.662
Slope	-1.506e-01	0.86019	0.462

All variables, except for distance to rivers, were found to have a correlation with sinkhole formation (Table 4). Water vapor pressure had a high correlation (above 0.5) with precipitation, average temperature, distance to faults, wind speed, solar radiation (0.928), maximum temperature, and minimum temperature; this shows high multicollinearity.

Comparing the pseudo-R^2 with the correlations chart shows that the wind and solar radiation both have the largest (negative) correlations and largest pseudo-R^2 when compared to sinkhole formation. While the

Table 4: SPSS significant Spearman correlation coefficients to $p < 0.01$ (**).

	Sinkholes	Precipitation	Average temperature	Distance to faults	Wind speed	Solar radiation	Water vapor pressure	Maximum temperature	Temperature range	Minimum temperature	Slope
Sinkholes	1	-.165**	-.138**	-.147**	-.248**	-.217**	-.194**	-.143**	.069**	-.147**	-.074**
Precipitation	-.165**	1	.405**	.456**	.380**	.442**	.564**	.385**	-.245**	.474**	-.071**
Average temperature	-.138**	.405**	1	.392**	.405**	.694**	.876**	.893**	-.121**	.878**	-.226**
Distance to faults	-.147**	.456**	.392**	1	.746**	.077**	.528**	.357**	-.333**	.493**	-.143**
Wind speed	-.248**	.380**	.405**	.746**	1	.147**	.581**	.423**	-.202**	.481**	-.145**
Solar radiation	-.217**	.442**	.694**	.077**	.147**	1	.750**	.762**	.106**	.664**	-.112**
Water vapor	-.194**	.564**	.876**	.528**	.581**	.750**	1	.928**	-.093**	.908**	-.232**
Maximum temperature	-.143**	.385**	.893**	.357**	.423**	.762**	.928**	1	.107**	.853**	-.223**
Temperature range	.069**	-.245**	-.121**	-.333**	-.202**	.106**	-.093**	.107**	1	-.350**	.057**
Minimum temperature	-.147**	.474**	.878**	.493**	.481**	.664**	.908**	.853**	-.350**	1	-.236**
Slope	-.074**	-.071**	-.226**	-.143**	-.145**	-.112**	-.232**	-.223**	.057**	-.236**	1

The overall sinkhole hazard ranged in probability from 0.000 – 0.999 (Table 5). From the training data, 117 sinkholes were not assigned a risk zone as they were located on the Tennessee border. Of the training data assigned a risk zone, 57.6% were projected into the very high risk category, 30.0% in the high risk category, 8.4% in the medium risk category, 3.4% in the low risk category, and 0.63% in the very low risk category.

Table 5: Logistic Regression sinkhole risk zones based on range of probability

Sinkhole Risk	***Range of Probability***
Very High	0.945-0.999
High	0.837-0.945
Medium	0.626-0.837
Low	0.211-0.626
Very Low	0.000-0.211

From the 30% validation group, 45 sinkholes were not assigned a risk zone as they were located on the Tennessee border. No probability could be calculated for sinkholes directly on the border. Of the sinkholes that were assigned a risk zone, 58.2% were projected into the very high risk category, 29.6% into the high risk category, 8.4% into the medium risk category, 3.2% into the low risk category, and 0.6% into the very low risk category (Figs 34-35; Table 5). This shows that at least 96.2% of the sinkholes were accurately predicted within a medium to very high risk zone.

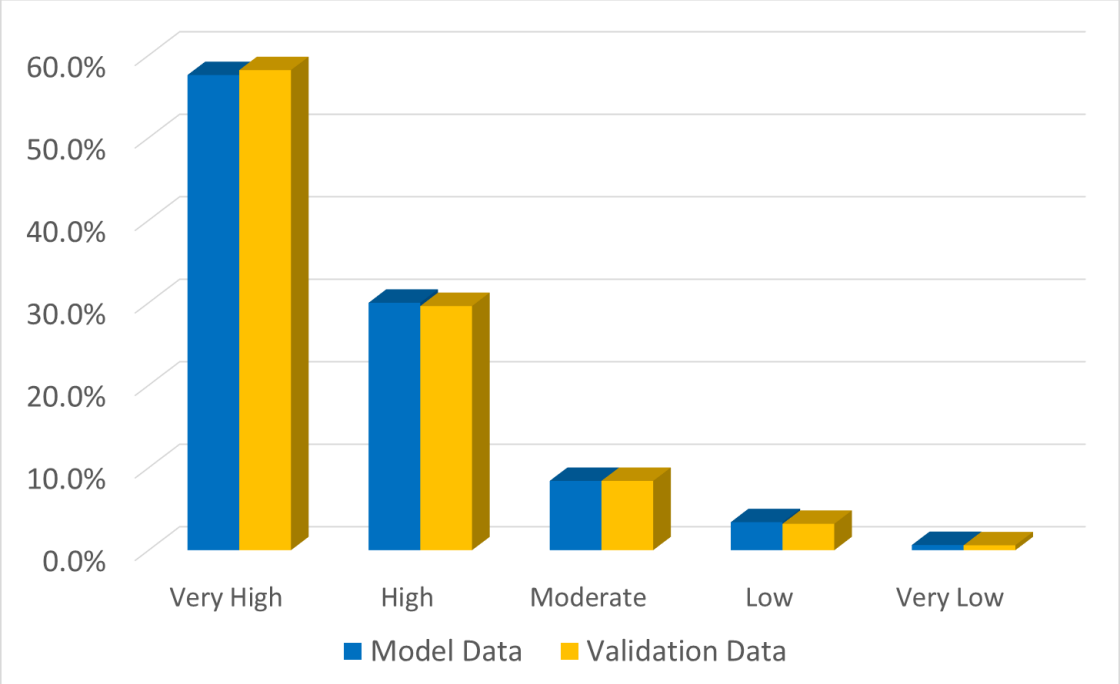

Figure 34: Comparison of sinkhole probabilities of model data vs. validation data indicates good model fit. Distribution of validation data with 87.6% of points in very high and high risk zones also indicates good model fit.

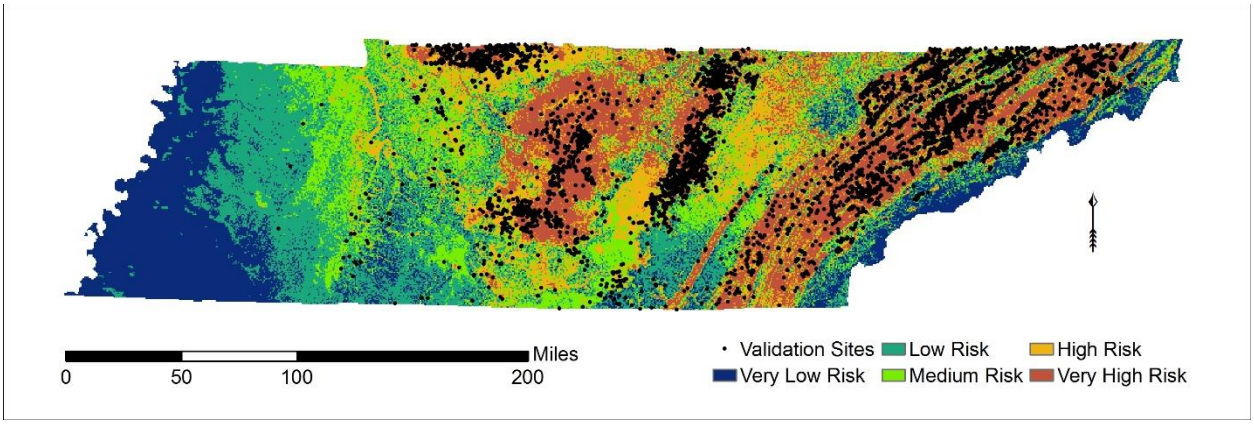

Figure 35: Sinkhole probability map with overlay of sinkhole validation dataset.

The highest hazard zones were located in the Valley and Ridge Province and the Nashville Basin Province, and overlap with the mapped sinkhole clusters/regions of highest sinkhole density. The lowest risk zones were located within the West Tennessee Province.

When compared to National Land Cover Data, 26846.96 km^2 of land (24.4% of Tennessee) falls within the very high risk zone. Within this zone, the highest affected areas include the Pasture/Hay and Cultivated Crops classification (9552.56 km^2) followed by the Deciduous Forest classification (7021.53 km^2). Within the combined Developed classifications (open space, low intensity, medium intensity, and high intensity), a total of 5115.68 km^2 is included (Fig 36).

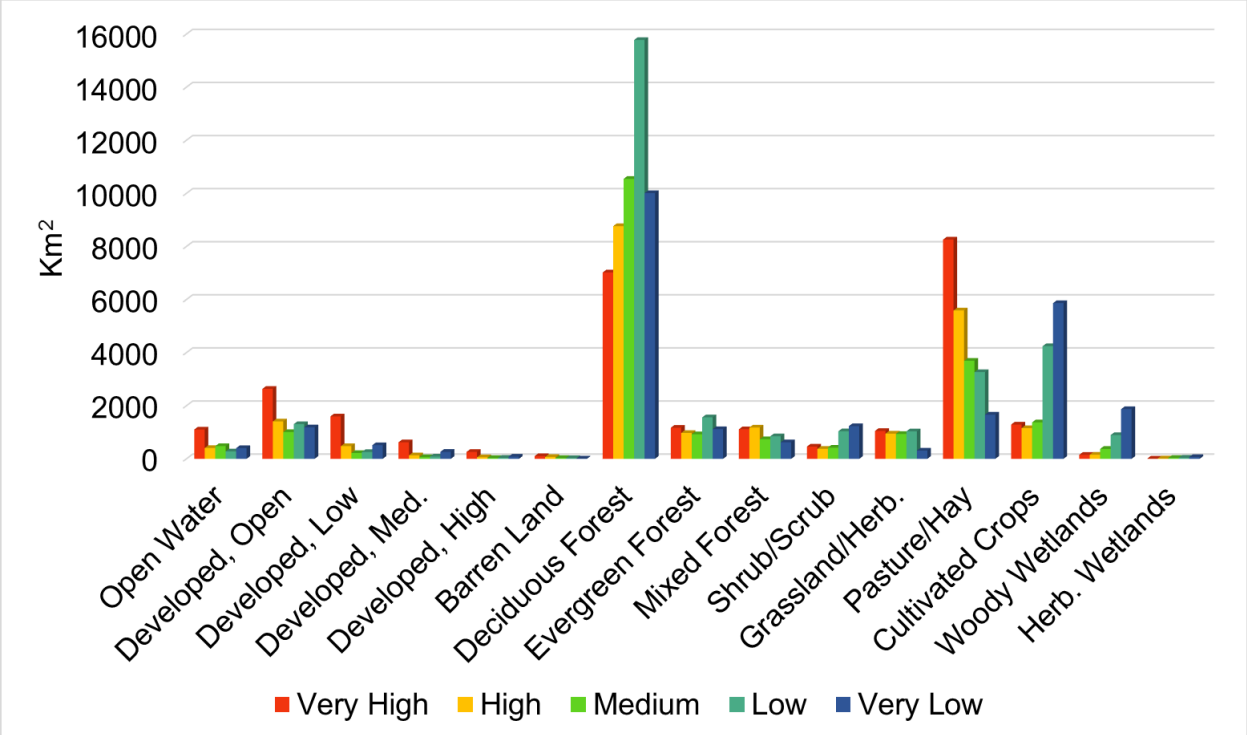

Figure 36: Graph displaying the total amount of differing land covers (Km^2) within each risk zone.

For the high risk zone, 21,713.12 km^2 was included (20.0% of Tennessee). Of this, the majority of the land falls within the Deciduous Forest classification with 8764.8 km^2 followed by

the Pasture/Hay and Cultivated Crops classification (6744 km^2). Of the combined Developed classifications, approximately 2078.79 km^2 of land are included.

Of all risk zones, the developed classifications primarily overlie the very high risk zone. The land covers with the highest percentages within the very high risk zone include: open water, barren land, grassland/herbaceous pasture/hay, and all developed classifications. The land covers with the highest percentages within the very low risk zone include: herbaceous wetlands, woody wetlands, cultivated crops, and shrub/scrub.

Overall, the Deciduous Forest classification covered the largest area of land within each risk zone. The largest percentage of Deciduous Forest lies within the low risk zone; the evergreen forest classification had the same result.

MaxEnt Model

The AUC value for the MaxEnt model is 0.684 (Fig 37). Yackulic et al. (2012) explains that the MaxEnt AUC does not follow the traditional definition for an AUC which classifies presences vs. absences; MaxEnt calculates AUC from the presence vs. background points (pseudo absences created by the model in lieu of true absences). This difference precludes comparison with AUC values from other models. However, the MaxEnt AUC is generally accepted otherwise.

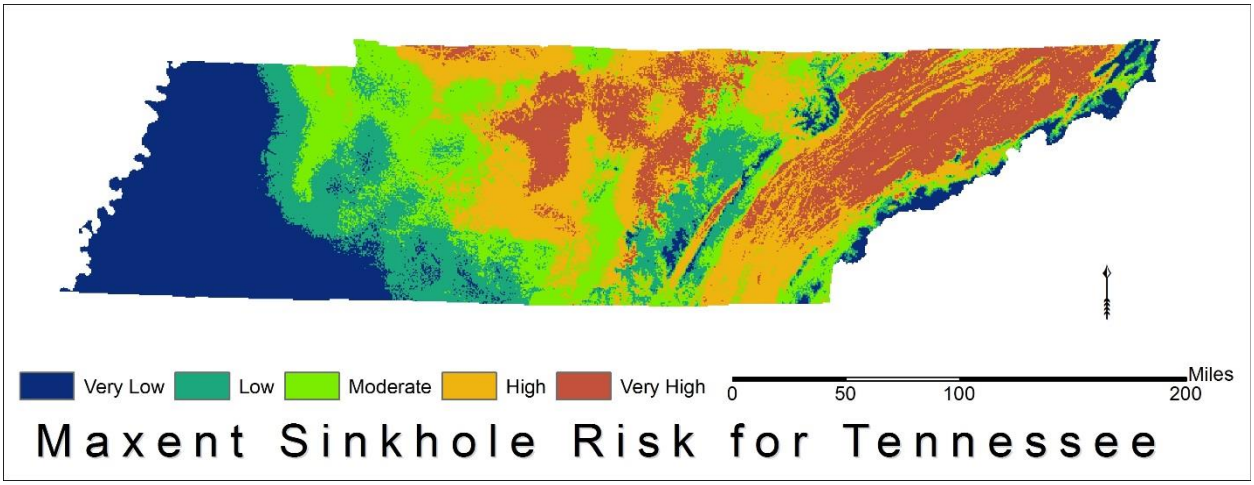

Figure 37: Sinkhole Risk map created using MaxEnt.

The "equal training sensitivity and specificity" threshold yielded a value of 0.505 and the "maximum training sensitivity plus specificity" threshold yielded a value of 0.419 (Fig 38). For the equal training threshold, 65% of the test sinkhole sites were included and for the maximum training threshold, 89% of the test sinkhole sites were included.

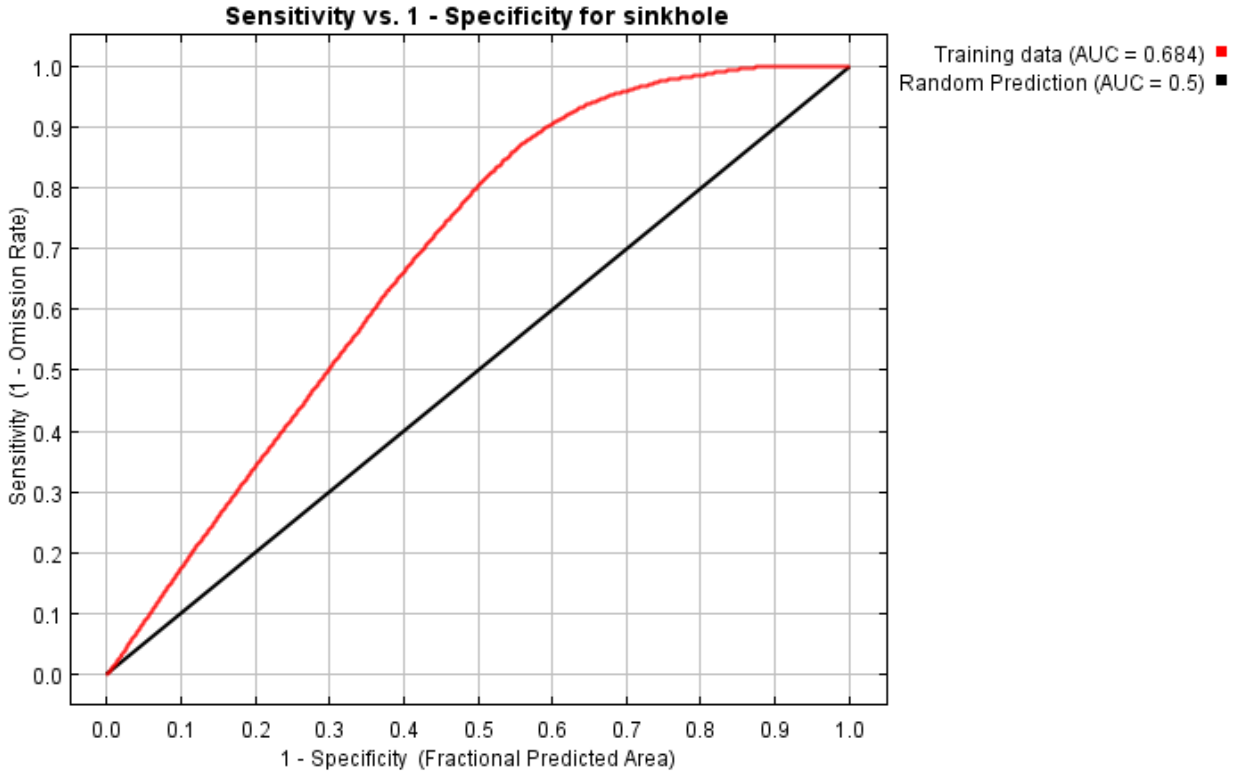

Figure 38: Sensitivity vs. 1-Specificity for sinkhole graph displaying the MaxEnt derived AUC.

A jackknife of regularized training gain (Fig 39) displays the gain of each variable for the model if ran in isolation, the gain of the model without the variable, and the training gain with all variables. This model included carbonate bedrock, distance from faults, precipitation, distance from rivers, solar radiation, maximum temperature, slope, wind, and water vapor pressure as significant, although, no *p*-values were produced for each variable.

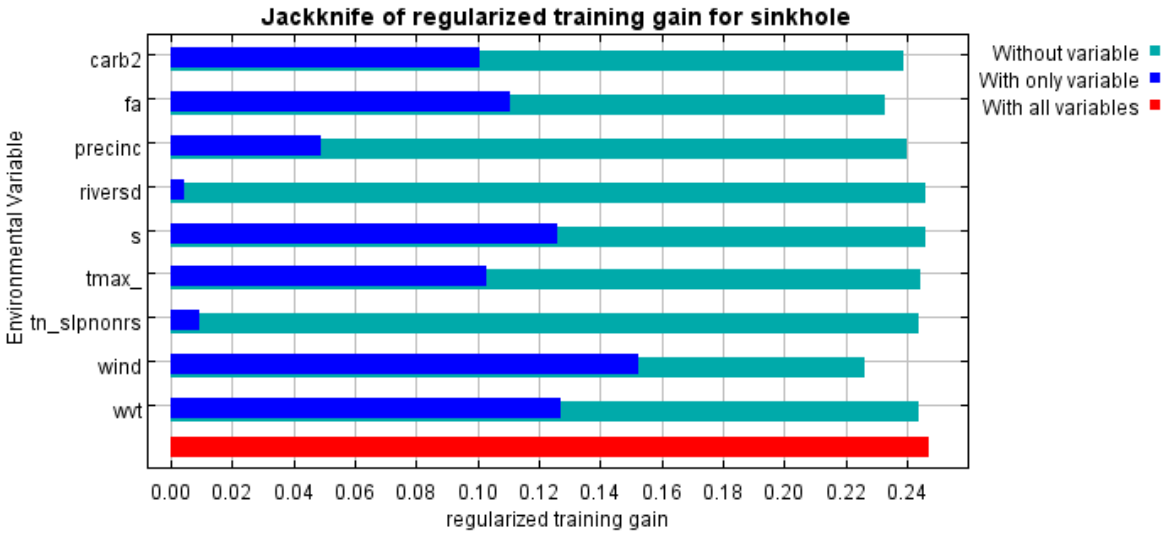

Figure 39: MaxEnt-derived jackknife of regularized training gain, which indicates how the independent variables influenced sinkhole formation under differing circumstances.

The thresholds applied to the response curves determined the optimal ranges for each of the continuous variables (Table 6). Each of the response curves are included below (Figs. 40 – 47).

Table 6: MaxEnt-derived the ranges for each of the independent variables for optimal sinkhole formation.

Variable	***Optimal Range for sinkhole formation***
Distance from faults	< 0.6 x 10^5 m
Precipitation	< 147 cm (58 in)
Solar radiation	< 15,200 kJ/m^2*day
Maximum Temperature	> 18° C (64.5° F)
Slope	< 48°
Wind	< 2.9 m/s
Distance to rivers	N/A
Water Vapor Pressure	N/A

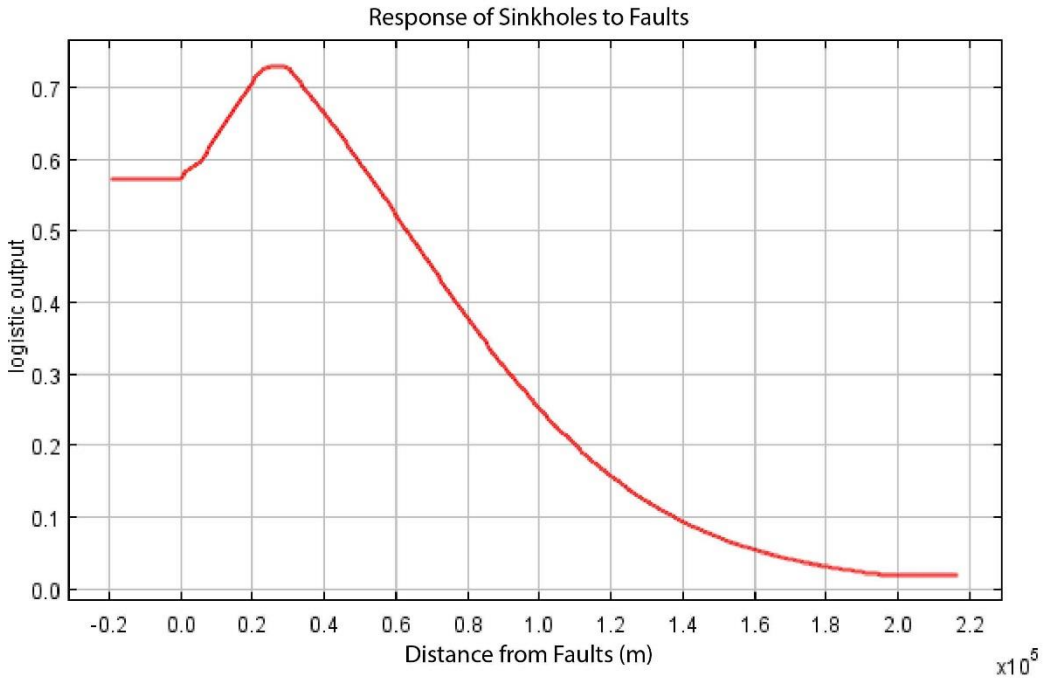

Figure 40: MaxEnt derived optimal distance to fault range for sinkhole formation.

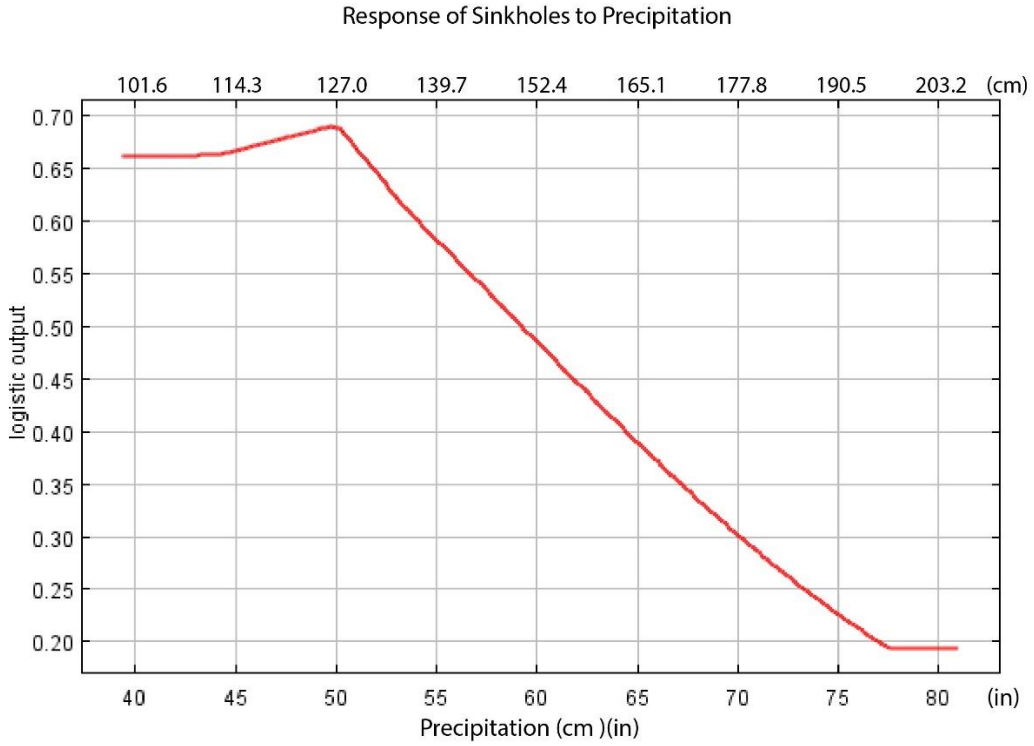

Figure 41: MaxEnt derived optimal precipitation range for sinkhole formation.

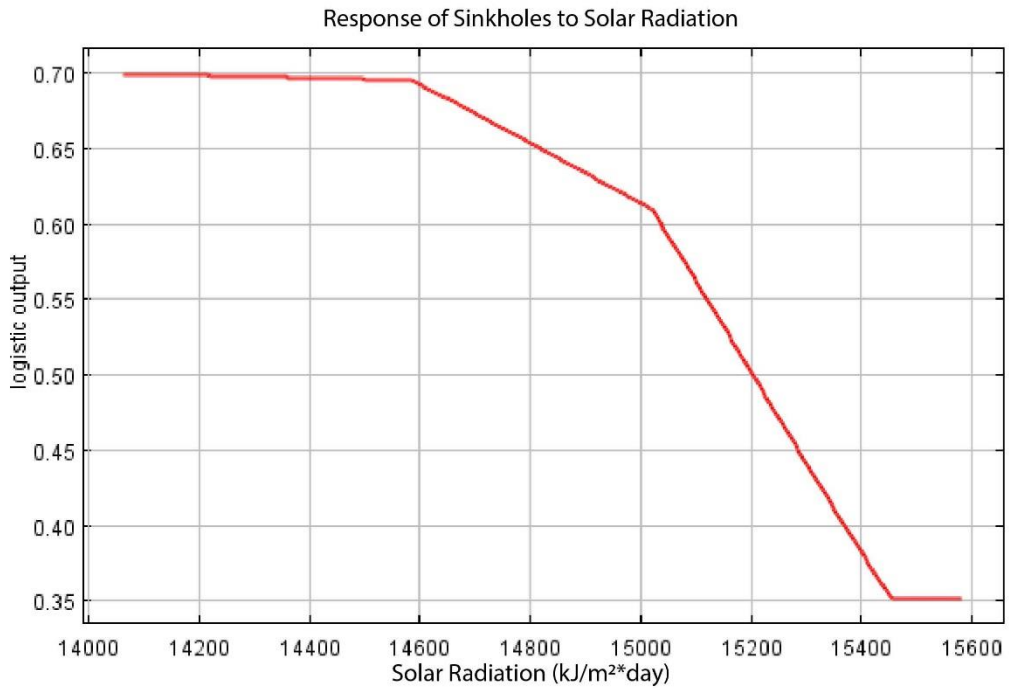

Figure 42: MaxEnt derived optimal solar radiation range for sinkhole formation.

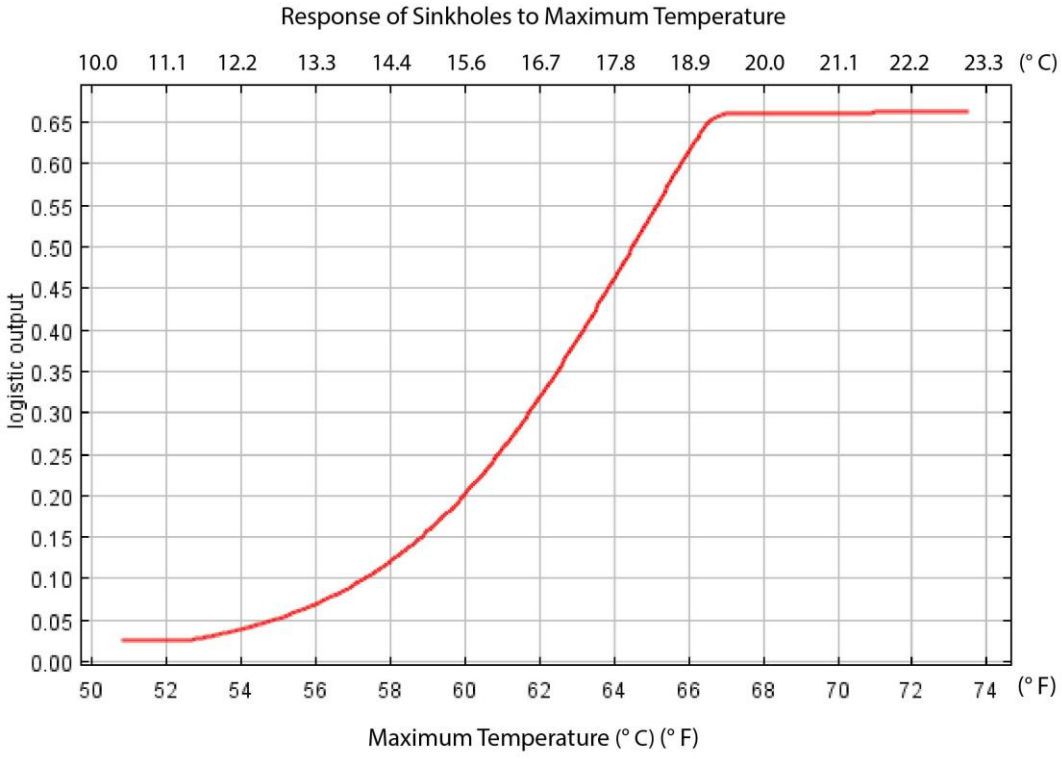

Figure 43: MaxEnt derived optimal maximum temperature range for sinkhole formation.

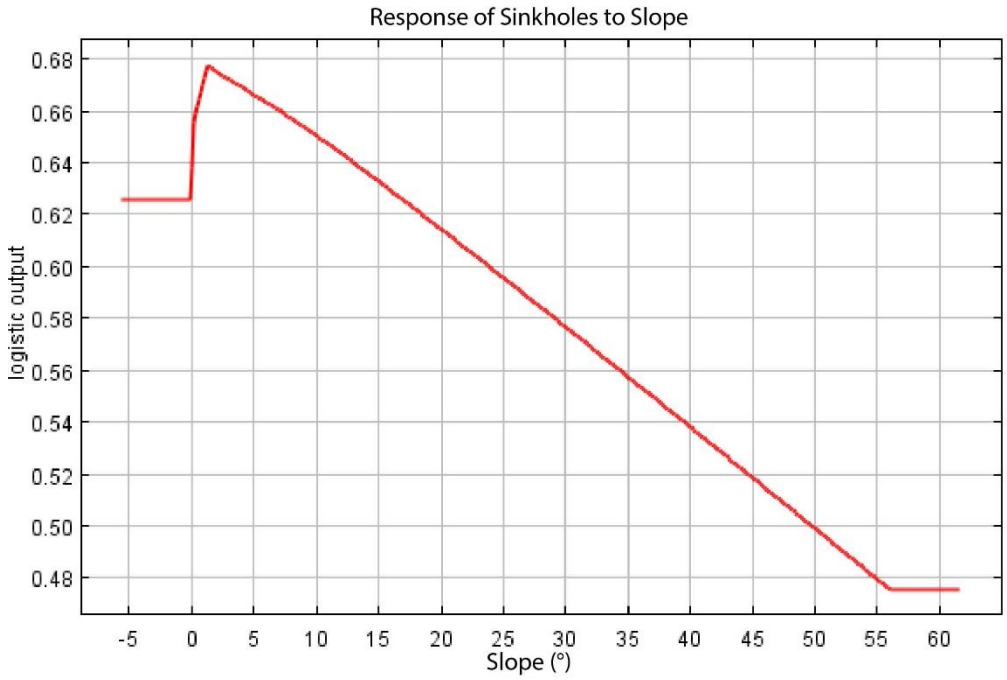

Figure 44: MaxEnt derived optimal slope range for sinkhole formation.

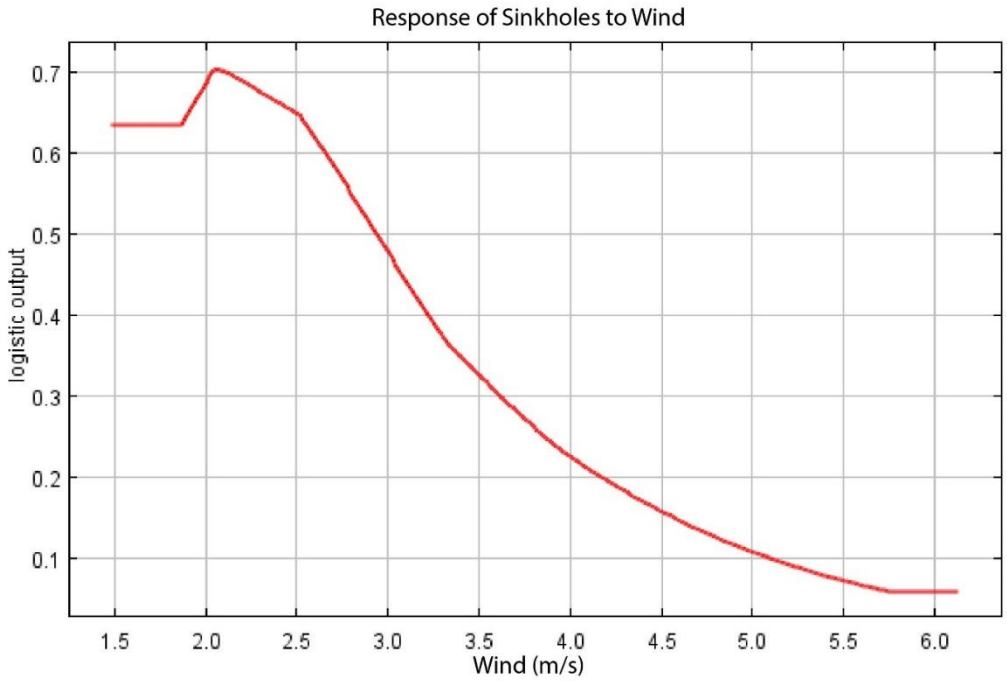

Figure 45: MaxEnt derived optimal wind speed range for sinkhole formation.

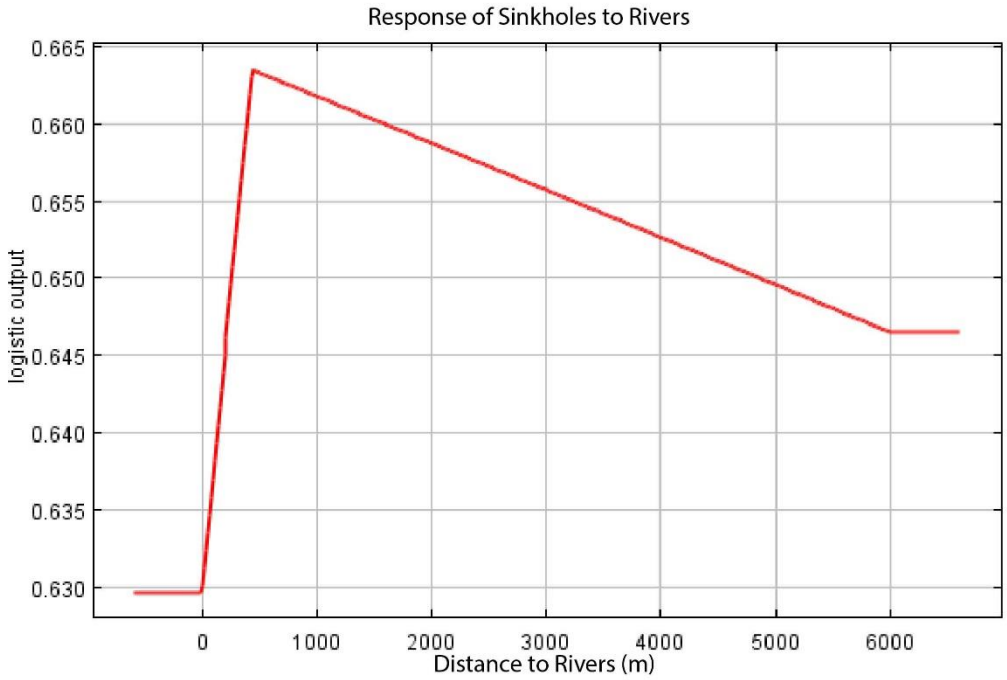

Figure 46: MaxEnt derived optimal distance to rivers range for sinkhole formation.

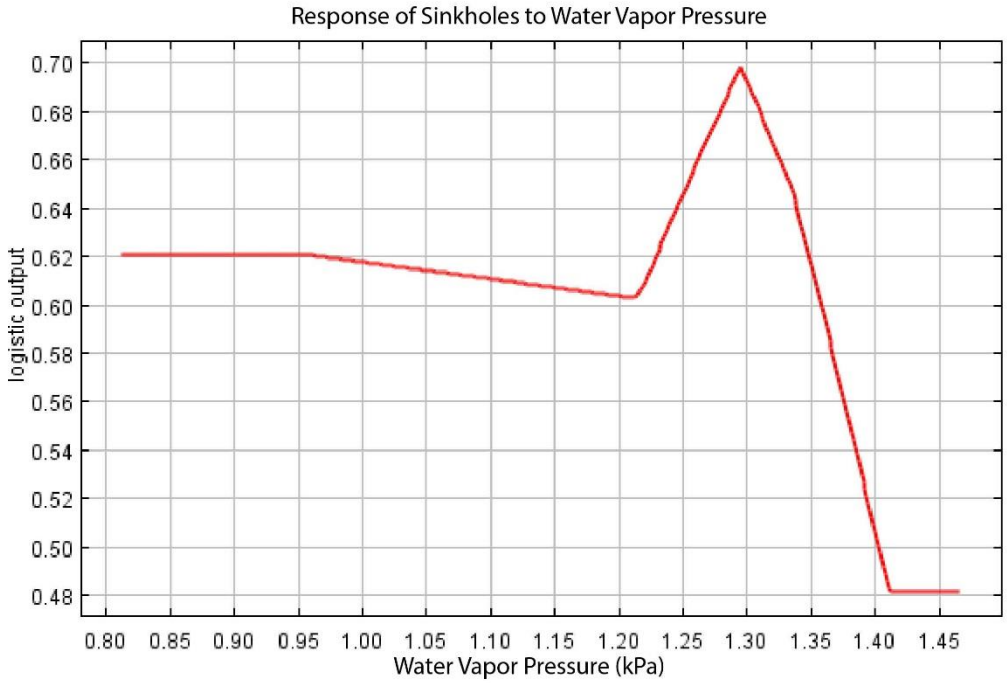

Figure 47: MaxEnt derived optimal water vapor pressure range for sinkhole formation.

A comparison of the overall mapped surfaces shows that the MaxEnt model predicts a lower probability for each risk zone than the logistic regression model, which varies spatially. The northern portions of the state had relatively equal areas within their prospective risk zones. The southern portions of the state displayed an overall lower risk within the MaxEnt model.

For the MaxEnt model, the overall sinkhole risk probability ranged from 0.000 to 0.742 (Table 7). From the 30% test site group, 38 sinkholes were not assigned a hazard zone as they were located on the Tennessee border. No probability could be calculated for sinkholes directly on the border. Of the sinkholes that were assigned a hazard zone, 64.0% were projected into the very high risk category, 31.6% into the high risk category, 2.8% into the medium risk category, 1.4% into the low risk category, and 0.20% into the very low risk category. This shows that at least 98.4% of the sinkholes were accurately predicted within a medium to very high risk zone.

Table 7: MaxEnt model sinkhole risk probability ranges

Sinkhole Risk	***Range of Probability***
Very High	0.509-0.742
High	0.339-0.509
Medium	0.215-0.339
Low	0.124-0.215
Very Low	0.000-0.124

When the MaxEnt model's risk zones are ranged with the same probability ranges as the logistic regression, the risk zones are dramatically reduced in area. No areas fall within the high or very high risk zones and small patches of medium risk zones are scattered within the northern portions of the Nashville Basin, Valley and Ridge, and Eastern Highland Rim. The medium risk

zone contains 3.4% of sinkholes from the 30% test group, while the low risk zone contains 95.1%, and the very low risk zone contains 1.5%.

Discussion

Model Comparison

Both models projected the highest risk zones within the following provinces: the Valley and Ridge, the Nashville Basin, the northwestern-most portion of the Highland Rim, the northern half of the eastern side of the Highland Rim, and the Sequatchie Valley. Both models displayed the areas of lowest risk within the West Tennessee Province and the Appalachian Province. However, the MaxEnt model incorporates less of the Appalachian Province than the original model, particularly along the western portion of the province.

Both models included maximum temperature, distance to faults, solar radiation, slope, wind, rivers, and precipitation. The difference in probability between the logistic regression and MaxEnt models is partially attributed to the inclusion of water vapor pressure in the MaxEnt model. If the water vapor pressure is truly non-significant (or at least causing collinearity), as the logistic regression suggests, then the MaxEnt model should not have included this variable allowing for distorted results.

Another difference between the probabilities of the two models is due to varying spatial autocorrelation differences. The cluster analysis determined that there is spatial autocorrelation. The MaxEnt model treats the spatial autocorrelation as one occurrence directly influencing another occurrence. For clarification, the MaxEnt program is generally used to map species, for example, red pandas. Where one red panda is located, there are likely to be more in the same location because animals need each other to survive. Also, one red panda may create another red panda hence more pandas in the same area. Sinkholes are not necessarily clustered because they

are creating each other and need each other to survive; sinkholes cluster only because they have similar necessary environmental conditions. Both models had very similar results and complimented each other. Since the first study obtained slightly better results, it shows that incorporating measures for spatial autocorrelation may have limited benefits.

This jackknife display, created by MaxEnt, agrees with the logistic regression model in that the most correlated variable is wind and the least correlated variable is the distance to rivers.

Covariates

The negative "distance from faults" correlation shows how groundwater often follows faults and joints. Since the faults form a path of least resistance (compared to going through the rock), the continuous flow of water through the faults allows more dissolution. This creates greater voids and thus a higher sinkhole susceptibility. The optimal range created from the threshold determined that after approximately 0.6×10^5 m, the distance to a fault is not as influential. This determination is reasonable as fault-related fractures are not continuous; the fractures related to faults become smaller and fewer with distance from the fault reducing the likelihood of a sinkhole.

The "positive" precipitation correlation not only matches the expected result, but also agrees with the Florida results. Although the two regressions' results are interpreted differently and to different degrees, they both indicate that increased precipitation influences sinkhole formation. Increased precipitation increases groundwater flowing through the karstic system allowing for increased dissolution. The MaxEnt model shows a slight positive correlation between the precipitation and sinkhole formation, but to a point (with a maximum probability around 50 inches); overall, it has a negative correlation. It is unknown why this model has an overall trend opposite the other models. Two possible theories include: 1) when the groundwater

system is saturated, i.e., as the threshold is reached, the infiltration reduces and collects as surface water runoff for streams and 2) like in Florida, as the groundwater system is filled, the hydrostatic pressure is able to support the ground from collapsing and requires a sudden drop in water level (example: excessive water pumping) before the ground can collapse forming a sinkhole. These reasons also explain why the optimal precipitation range includes values less than 58 inches.

For solar radiation, no significant correlation was expected; however, a negative correlation between solar radiation and sinkhole risk was observed. One explanation for the observed negative correlation is that areas of lower solar radiation coincide with areas of higher population and development, such as around Nashville, Memphis, and Knoxville (determined qualitatively). Anthropogenic influences, such as leaking pipes have been found to form "artificial" sinkholes (Lee, Shin, Ko, and Chang, 2016). This anthropogenic influence may have increased sinkhole occurrence in these areas, and may be the factor related to sinkhole formation, with solar radiation having a spurious relationship. Lower solar radiation values around cities can be indicative of smog. This could explain why the optimal solar radiation range is for values less than 15,200 kJ/m^2*day. To test this hypothesis, SPSS was used to compare solar radiation values at sinkhole sites with a bivariate variable of the National Land Cover Data's (MLCC, 2017) land cover designated as "Developed" (including all the developed classifications) vs. undeveloped. This comparison found a highly significant but small correlation between the two ($r = -0.088$, $p = 0.000$). Although this result supports the hypothesis (less radiation where there is development), the small correlation indicates there may be a better, unknown explanation for the observed correlation between solar radiation and sinkhole formation.

The positive “temperature” correlation matched the expected results. Although carbonate solubility decreases as temperature increases, warmer water can hold more dissolved material than colder water (Fig 1). The more material is able to dissolve, the higher the sinkhole likelihood. Also, warmer outside air temperatures, with enough nutrients and water, increase plant photosynthesis and soil CO_2 levels. This increase in CO_2 has a greater impact on mineral solubility than groundwater temperature. The warmer temperatures match the optimal range for values above 64.5 degrees Fahrenheit.

A negative “slope correlation” suggests that infiltration and dissolution are more likely to happen where water can puddle. If water is quickly moved downhill from a source, there is not enough time for dissolution to occur. This idea is reflected well within the MaxEnt model’s optimal range of slope values below 48 degrees, with an increase in sinkhole probability as the slope decreases.

Wind had a significantly negative correlation. This shows that as wind decreases, the likelihood of a sinkhole forming increases. It is possible that the decrease in wind would allow more aeration of atmospheric CO_2 into the underlying soil and plants. It is also possible that the areas of higher wind speeds are also at higher elevations. Since karst limestone is easily dissolved, it is less likely to be found at these same elevations. Elevation was not used as a determining factor in this study as it was not a climate variable, to avoid model over-specification, and to minimize an excess of variables. When wind data were compared to elevation in SPSS Statistics 23, a significantly strong negative correlation ($r = -0.515$, $p = 0.000$) was found. When comparing elevation with wind data confined within the optimal range suggested for sinkhole formation, only a weak correlation was found ($r = -0.077$, $p = 0.000$). When compared at a more local level, opposing results were found; the Valley and Ridge

Province = 0.578 (p = 0.000) and the Highland Rim Province = -0.512 (p = 0.000) showing opposing wind/elevation correlations within carbonate provinces. Also, the areas of highest wind speed are located within the highest and lowest elevation locations. This determines that no real determination can be made concerning correlation between wind and elevation within the carbonate regions. Although elevation could be considered for future studies, it does not explain the correlation between wind and sinkhole sites.

The lower wind speeds are not necessarily found at the lowest of elevations but within valleys, particularly in the Valley and Range Province and in the Sequatchie Valley Province. These valleys not only block wind but also allow the channeling of water. It is suggested that the wind variable acts as a surrogate variable for the topography. Although slope was used to try to show one aspect of topography, where areas of puddling would be highest, the wind variable was more effective. For example, a valley, like a cup, can hold a greater amount of water than a plane than is flat like a plate. It is currently unknown if the optimal range for wind speed would be helpful for future studies, but should definitely be considered within future sinkhole studies as the optimal range of lower values may indicate possible valleys.

A positive "distance from rivers" correlation is also to be expected. In regions of high groundwater infiltration, there is less surface water flow and vice versa. This is due to the lithology of the system. For example, in a highly insoluble rock, the water cannot infiltrate into the ground, so it travels above the surface; whereas, carbonates are easily soluble, so instead of the water flowing on the surface, it infiltrates into the ground. It is expected that although the sinkholes are further from streams, they should be closer to groundwater conduits. However, as there is no groundwater conduit network dataset for the whole of Tennessee, it could not be utilized or proven within this case study. No optimal range for the distance of rivers was found

due to all values being above the threshold; according to the threshold, all the values are within the optimal range for sinkhole formation. The graph appears to show that the optimal distance from rivers is at ~450 meters, but the value's probability range is less than 3.5% from lowest to highest probability making the optimal probability relatively negligible and not useful for this particular model.

Logistic Regression Model Validation

For the sample sites and test sites comparison, the test sites should be categorized with similar percentages to each risk category. If the percentages were very different, then the model would not be accurate for predicting where the sinkholes would form. Since all of the test site percentages for each of the hazard categories were within 1% of the sample site percentages, this comparison determines the model to be good.

The overall pseudo-R^2 value may seem low compared to other methods for predicting sinkhole formation sites. Other significant variables not included in this model include: rock hardness, soil depth, depth to groundwater, and proximity to pipes and wells. Since all of these variables are very significant factors in sinkhole formation, they are the primary focus of sinkhole prediction and may outweigh the climatic variables in significance. However, the climatic variables should not be ignored. Since this study focused on the impact of climatic variables (even with their smaller significance) on sinkholes and not known physical factors, this model is determined to be useful in identifying the additional influence of climate on sinkhole formation.

Implications, Limitations, and Future Work

As the Tennessee sinkholes did not include the dates when the sinkholes formed, the climate data may not match the exact conditions under which each sinkhole formed. Without

these data, future climate values cannot be substituted for current climate conditions to predict possible future sinkhole risk zones. However, as the landscape changes slowly over time, the overall change in climate values over space should have similar correlations. For example, even if average temperatures increase, the ratio of differing temperatures from one region to another should remain relatively the same. This allows for this study to still be valid. A database of newly forming sinkholes with their formation dates should be started for the state.

The spatial statistics showed the mean center in an area lacking in sinkholes. A better representation can be made by splitting the dataset into two groups representing the regions with the most clusters, particularly the regions to the east and west of the Cumberland Plateau Province.

Assuming the wind variable is a surrogate variable for topography distinguishing where the valleys are located, there may be a better variable to use in place of wind. One such possibility includes using a curvature index. Determining this or other possible substitutes for correlation is needed for future work.

The work should also be replicated at a finer scale for more precise results. Light detection and ranging (LiDAR) would allow finer scaled data for elevation and slope. A finer scaled slope should show more accurate locations for puddling than the current 1/3 arc second data used for this study. However, obtaining climate data at the same scale as LiDAR data may be difficult to obtain without sufficient weather station and may be unobtainable for larger regions.

The majority of Tennessee developed land within the very high risk sinkhole zones shows that this risk is a definite concern for the public and not just for buildings within the

forests or grasslands. Making sure that the public and developed lands are protected from sinkholes should be considered during future development.

Conclusions

In conclusion, sinkhole formation can be modelled using precipitation, maximum temperature, solar radiation, wind speed, slope, distance to rivers, carbonate bedrock, and distance to faults. While, spatial and non-spatial logistic regression models are complementary, a spatial model may not best map sinkhole hazard. Using these variables to help better map the sinkhole hazard for Tennessee may help homebuyers, construction companies, and insurance companies make wiser decisions when deciding how and where to buy and how to best protect a building from potential sinkhole collapse.

References

Anderson, R.P., Lew, D., and Townsend Peterson A., 2003, Evaluating predictive models of species' distributions: criteria for selecting optimal models: Ecological Modelling, v. 162, p. 211-232.

BCCVL, 2016, Interpretation of Model Outputs: Biodiversity and Climate Change Virtual Laboratory: https://support.bccvl.org.au/support/solutions/articles/6000127046-interpretation-of-model-outputs (accessed December 2017).

Chen, W., Pourghasemi, H.R., Kornejady, A., and Zhang, N., 2017, Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques: Geoderma, v. 305, p. 314-327, doi: 10.1016/j.geoderma.2017.06.020.

Diao, M., 2015, Selectivity, spatial autocorrelation and the valuation of transit accessibility: Urban Studies, v. 52(1), p. 159-177, doi: 10.1177/0042098014523686.

Fox, J., 2008, Applied regression analysis and generalized linear models, Edition 2: Los Angeles, Sage, 688 p.

Garner, R., 2008, Solar Irradiance: National Aeronautics and Space Administration: https://www.nasa.gov/mission_pages/sdo/science/solar-irradiance.html (accessed December 2017).

King, E., 2017, There's nothing odd about the odds ratio: Interpreting binary logistic regression: Statistics Solutions: http://www.statisticssolutions.com/theres-nothing-odd-about-the-odds-ratio-interpreting-binary-logistic-regression/ (accessed November 2017).

Liu, C., Newell, G., and White, M., 2016, On the selection of thresholds for predicting species occurrence with presence-only data: Ecology and Evolution, v. 6, p. 337–348, doi: 10.1002/ece3.1878.

Marella, R.L., 2015, Water withdrawals in Florida, 2012: U.S. Geological Survey Open-File Report 2015–1156, 10 p., http://dx.doi.org/10.3133/ofr20151156.

Menard, S., 2010, Logistic regression: from introductory to advanced concepts and applications, 4-14 p.

Multi-Resolution Land Characteristics Consortium (MLCC), 2017, National Land Cover Database 2011 (NLCD2011): https://www.mrlc.gov/nlcd11_leg.php (accessed January 2018).

Orme, G.R., 1982, Calcarenite, in Beaches and Coastal Geology: Encyclopedia of Earth Science, Springer, doi:10.1007/0-387-30843-1_86.

Ozdemir, A., 2015, Sinkhole susceptibility mapping using logistic regression in Karapinar (Konya, Turkey): Bulletin of Engineering Geology and the Environment, v. 75, i. 2, p. 681-707, doi: 10.1007/s10064-015-0778-x.

Pellicani, R., Spiltro, G., and Gutiérrez, F., 2017, Susceptibility mapping of instability related to shallow mining cavities in a built-up environment: Engineering Geology, v. 2017, p. 81-88, doi: 10.1016/j.enggeo.2016.12.011.

Phillips, S.B., Aneja, V.P., Kang, D., and Arya, S.P., 2006, Modelling and analysis of the atmospheric nitrogen deposition in North Carolina: International Journal of Global Environmental Issues, v. 6, p. 231–252, doi: 10.1016/j.ecolmodel.2005.03.026.

Shunk, A.J., 2009, Late Tertiary paleoclimate and stratigraphy of the Gray Fossil Site (eastern TN) and Pipe Creek sinkhole (northcentral IN) [Ph.D. thesis]: Baylor University, 105 p.

UTIA Institute of Agriculture, 2017, Tennessee Climatological Service: https://ag.tennessee.edu/climate/Pages/climatedataTN.aspx (accessed December 2017).

Veni, G., DuChene, H., Crawford, N.C., Groves, C.G., Huppert, G.N., Kastning, E.H., Olson, R., and Wheeler, B.J., 2001, Living with karst: AGI Environmental Awareness Series, v. 4, 69 p.

Webbers, A., 2003, Groundwater use by public water-supply systems in Tennessee, 2000: U.S. Geological Survey Open-File Report 03-47.

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., and Veran, S., 2013, Presence-only modelling using MAXENT: When can we trust the inferences? Methods in Ecology and Evolution, v. 4, p. 236–243, doi: 10.1111/2041-210x.12004.

Young, N., Carter, L., and Evangelista, P., 2011, A MaxEnt model v3.3.3e tutorial (ArcGIS v10): Natural Resource Ecology Laboratory at Colorado State University and the National Institute of Invasive Species Science.

Zhou, G., Yan, H., Chen, K., and Zhang, R., 2016, Spatial analysis for susceptibility of second-time karst sinkholes: A case study of Jili Village in Guangxi, China: Computers and Geosciences, v. 89, p. 144–160, doi: 10.1016/j.cageo.2016.02.001.

CHAPTER 4

CONCLUSION

The objectives of this thesis were to quantitatively determine climate correlations to sinkhole formations at a regional scale to form regression models for predictive sinkhole formation zones and times.

While Florida's primary water source is groundwater, Tennessee has a greater variability of groundwater and surface water intake sources. The USGS estimated that in 2000, Tennessee consumed approximately 321 million gallons per day of public groundwater (Webbers 2003), while in Florida during 2012, approximately 2.03 billion gallons per day of public groundwater were used (Marella 2015). More surface water intake should reduce the amount of anthropogenic-related sinkhole appearances. Although the risk is still there for certain areas, Tennessee should be a better study site for these climatic studies than Florida.

Study one found that although sinkholes have a higher likelihood of forming after large precipitation events, they are more likely to form during frost-freeze events due to excessive groundwater withdrawal. As sinkholes are more likely to form with warmer temperatures (when frost-freeze events are absent), no correlation could be made between sinkhole formation and temperature. Some teleconnection indices, including the NAO, PNA, ENSO, EP/NP, and SCA, could also be linked to sinkhole formation but with limited effect. The effect of these parameters on sinkhole formation may be as much 25% in some locations.

Study two found that sinkholes are correlated with precipitation, temperature, distance to faults, and slope. It also found highly significant correlations with solar radiation and wind speed, which had greater regression coefficients and R^2 values than the other climate variables. Wind speed was found to be a surrogate variable for topography, better expressing the areas of

puddling than slope. Solar radiation has a small correlation with the type of landcover, but a better explanation is still needed; this should be considered for future research. Of the different temperature variables (maximum temperature, minimum temperature, average temperature, and temperature range), the maximum temperature was found to have the highest correlation with sinkholes.

Of the two Tennessee models, the MaxEnt model was found to be slightly less accurate than the logistic regression model. Although the MaxEnt model accounts for spatial autocorrelation, unlike the logistic regression model, the MaxEnt model included water vapor pressure, which was deemed to have high multi-collinearity with the other variables. It also presented lower probabilities for each hazard level. Overall, both models were accurate at predicting sinkholes but the logistic regression is recommended model of choice due to its better accuracy and higher AUC.

The Florida study showed that CO_2 was significantly correlated with sinkhole formation, but long term CO_2 datasets are spatially limited. Results for the spatial analysis could be improved with spatial CO_2 levels. The Tennessee correlations may also be more accurately predicted with the addition of sinkhole occurrence dates. Obtaining these data could significantly and positively influence future studies; it would determine the exact weather conditions in which sinkholes form. With this data, future predicted climate levels could be applied to help determine if the sinkhole formation zones are expanding and/or intensifying. Future research includes using these same techniques on different regions for comparative analysis if the data are available.

The two states' sinkhole databases could be greatly improved. For Tennessee, currently forming sinkholes should be catalogued with their formation dates like the Florida database. Determining the date of formation may be difficult for dissolution sinkholes. One suggestion for

dealing with this is to set a standard depth a sinkhole must attain before being included within the database. When the sink has obtained this minimum depth, then the approximate date should be recorded. For Florida, a more complete sinkhole database should be created including all the mapped sinkholes in the state like the Tennessee sinkhole catalogue. With this information, a spatial study could be made for the Florida sinkholes and a temporal study for the Tennessee sinkholes. Then the two states could be compared to determine possible climate links between the two differing climatic regions.

Sinkhole catalogues should not include subsidence events as subsidence events are formed differently. Specific governmental requirements should be enforced to determine which catalogue an event should be included in; such requirements should include depth, roundness, perimeter, surface area/volume, and probable cause of occurrence including whether or not it is natural or anthropogenic in nature (for example: is it caused from a broken water line or a precipitation event?).

With this information, the public could become more informed of the possible sinkhole risk in their location. Better regulations could be made to either determine sinkhole insurance rates or to include sinkhole prevention methods during land development. Knowing the locations of the highest risk zones could help save both insurers and insureds money. Knowing the optimal weather patterns that influence sinkhole formation could keep both the public and the government on alert when optimal sinkholes conditions are met; this could possibly save lives.

The Florida study complimented other studies showing that groundwater withdrawal may have a distinct influence on sinkhole formation. The amount of water exiting through groundwater pumps should be regulated. If sinkholes are forming on private property due to water overuse, then insurance companies may not want to cover these particular claims. If

farmers or other groundwater pump owners are causing sinkholes and/or sinkhole related damages to land not of their own, then they could be held responsible.

REFERENCES

1. Agriculture UI of. Climate data for Tennessee. 2017.
2. Anderson RP, Lew D, Peterson AT. Evaluating predictive models of species' distributions: criteria for selecting optimal models. 2003;162(3):211–232. doi:10.1016/S0304-3800(02)00349-6
3. Aurit, M. D., Peterson, R. O., & Blanford, J.I. (2013). A GIS analysis of the relationship between sinkholes, dry-well complaints and groundwater pumping for frost-freeze protection of winter strawberry production in Florida. PLoS ONE, 8. doi: 10.1371/journal.pone.0053832
4. Barnston AG, Livezey RE. Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns. 1987;115(6):1083–1126. doi:10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2
5. BCCVL. Interpretation of model outputs. 2016.
6. Blender R, Luksch U, Fraedrich K, Raible CC. Predictability study of the observed and simulated European climate using linear regression. 2003;129(592 PART A):2299–2313. doi:10.1256/qj.02.103
7. Borsato A, Frisia S, Miorandi R. Carbon dioxide concentration in temperate climate caves and parent soils over an altitudinal gradient and its influence on speleothem growth and fabrics. 2015;40(9):1158–1170. doi:10.1002/esp.3706
8. Buttrick DB, Van Schalkwyk A, Kleywegt RJ, Watermeyer RB. Proposed method for dolomite land hazard and risk assessment in South Africa. 2001;43(2):27–36.

9. Cahalan MD, Milewski AM. Sinkhole formation mechanisms and geostatistical-based prediction analysis in a mantled karst terrain. 2018;165(February 2017):333–344. doi:10.1016/j.catena.2018.02.010
10. Chen W, Pourghasemi HR, Kornejady A, Zhang N. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. 2017;305(May):314–327. doi:10.1016/j.geoderma.2017.06.020
11. Cooper AH, Farrant AR, Price SJ. The use of karst geomorphology for planning, hazard avoidance and development in Great Britain. 2011;134(1–2):118–131. doi:10.1016/j.geomorph.2011.06.004
12. Cowling SA. Sea level and ground water table depth (WTD): A biogeochemical pacemaker for glacial-interglacial cycling. 2016;151:309–314. doi:10.1016/j.quascirev.2016.09.009
13. Cunningham KJ, Locker SD, Hine AC, Bukry D, Barron JA, Guertin LA. Interplay of Late Cenozoic Siliciclastic Supply and Carbonate Response on the Southeast Florida Platform. 2003;73(1):31–46. doi:10.1306/062402730031
14. Diao M. Selectivity, spatial autocorrelation and the valuation of transit accessibility. 2015;52(1):159–177. doi:10.1177/0042098014523686
15. Duffy PA, Walsh JE, Graham JM, Mann DH, Duffy PA, Walsh JE, Graham JM, Mann DH, Rupp TS. Impacts of Large-Scale Atmospheric-Ocean Variability on Alaskan Fire Season Severity. 2005;15(4):1317–1330.1. Florea LJ. Geology and Hydrology of Karst in West-Central and North-Central Florida. 2008;(August):225–239.

16. Florea LJ. Using State-wide GIS data to identify the coincidence between sinkholes and geologic structure USING STATE-WIDE GIS DATA TO IDENTIFY THE. 2005;(August):120–124.
17. Flores-Berrones R, Asce F, Ramírez-Reynaga M, Macari EJ. Internal Erosion and Rehabilitation of an Earth-Rock Dam. 2011;137(February):150–160. doi:10.1061/ASCÊGT.1943-5606.0000371
18. Fox J. Applied regression analysis and generalized linear models. 2nd ed. Los Angeles: Sage; 2008.
19. Gabrovšek F, Stepišnik U. On the formation of collapse dolines: A modelling perspective. 2011;134(1–2):23–31. doi:10.1016/j.geomorph.2011.06.007
20. Gao Y, Luo W, Jiang X, Lei M, Dai J. Investigations of Large Scale Sinkhole.
21. Garner R. Solar irradiance. 2008.
22. IPCC. Climate Change 2014: Synthesis report: 5th Assessment Report. 2014.
23. Jones IC, Banner JL. Hydrogeologic and climatic influences on spatial and interannual variation of recharge to a tropical karst island aquifer. 2003;39(9):1–10. doi:10.1029/2002WR001543
24. King E. There's nothing odd about the odds ratio: Interpreting binary logistic regression. 2017.
25. Knutsson G. Acidification effects on groundwater - prognosis of the risks for the future. 1994;(222).
26. Lane E. Karst in Florida: Florida Geological Survey Special Publication. 1986.
27. Lee EJ, Shin SY, Ko BC, Chang C. Early sinkhole detection using a drone-based thermal camera and image processing. 2016;78:223–232. doi:10.1016/j.infrared.2016.08.009

28. Liu C, Newell G, White M. On the selection of thresholds for predicting species occurrence with presence-only data. 2016;6(1):337–348. doi:10.1002/ece3.1878
29. MacGillivray C. A review of quantitative techniques for measuring karst terrain, with examples from the Caribbean. 1997;8(2):81–95.
30. Macpherson GL, Roberts JA, Blair JM, Townsend MA, Fowle DA, Beisner KR. Increasing shallow groundwater CO2and limestone weathering, Konza Prairie, USA. 2008;72(23):5581–5599. doi:10.1016/j.gca.2008.09.004
31. Malone SL, Staudhammer CL, Oberbauer SF, Olivas P, Ryan MG, Schedlbauer JL, Loescher HW, Starr G. El Niño Southern Oscillation (ENSO) enhances CO2 exchange rates in freshwater marsh ecosystems in the Florida Everglades. 2014;9(12):1–30. doi:10.1371/journal.pone.0115058
32. Marella RL. Water withdrawals in Florida, 2012. 2015;(2015–1156):10. doi:10.3133/ofr20151156
33. Menard S. Logistic regression: from introductory to advanced concepts and applications. Los Angeles: Sage; 2010.
34. Moran JM. Weather studies: Introduction to atmospheric science. 4th ed. Boston: the American Meteorological Society; 2009.
35. MRLC. National land cover database 2011. 2017.
36. Orme GR. Calcarenite. In: Beaches and Coastal Geology. Dordrecht: Kluwer Academic Publishers; 1982. p. 186–186. doi:10.1007/0-387-30843-1_861.
37. Ozdemir A. Sinkhole susceptibility mapping using logistic regression in Karapınar (Konya, Turkey). 2016;75(2):681–707. doi:10.1007/s10064-015-0778-x

38. Parinet J, Julien M, Nun P, Robins RJ, Remaud G, Höhener P. Predicting equilibrium vapour pressure isotope effects by using artificial neural networks or multi-linear regression - A quantitative structure property relationship approach. 2015;134:521–527. doi:10.1016/j.chemosphere.2014.10.079

39. Parise M, Lollino P. A preliminary analysis of failure mechanisms in karst and man-made underground caves in Southern Italy. 2011;134(1–2):132–143. doi:10.1016/j.geomorph.2011.06.008

40. Pellicani R, Spilotro G, Gutiérrez F. Susceptibility mapping of instability related to shallow mining cavities in a built-up environment. 2017;217:81–88. doi:10.1016/j.enggeo.2016.12.011

41. Phillips SB, Aneja VP, Kang D, Arya SP. Modelling and analysis of the atmospheric nitrogen deposition in North Carolina. 2006;6(2–3):231–252. doi:10.1016/j.ecolmodel.2005.03.026

42. Poag WC. Rise and demise of the Bahama-Grand Banks gigaplatform, northern margin of the Jurassic proto-Atlantic seaway. 1991;102(1–4):63–130. doi:10.1016/0025-3227(91)90006-P

43. Ritter DF, Kochel RC, Miller JR. Process Geomorphology. 5th ed. Long Grove: Waveland Press, Inc.; 2011.

44. Sainani KL. Understanding linear regression. 2013;5(12):1063–1068. doi:10.1016/j.pmrj.2013.10.002

45. Scott TM. Text to accompany the geologic map of Florida. 2001;80:28.

46. Shunk AJ. Late Tertiary paleoclimate and stratigraphy of the Gray Fossil Site (eastern TN) and Pipe Creek sinkhole (northcentral IN). Baylor University; 2009.

47. Sweeting M. Some aspects of karst hydrology. 1973;139(2):280–28. doi:10.1007/s11206-004-0694-5

48. Tihansky AB, Knochenmus L a. Karst features and hydrogeology in west-central Florida — A field perspective. 2001:198–211.

49. U.S. climate data. Climate Florida. 2016.

50. University of South Florida / Florida Center for Instructional Technology (FCIT). Florida's land then and now. 2002.

51. USGS. Ground-water use by public water-supply systems in Tennessee, 2000. 2003:2003.

52. Veni G, DuChene H, Crawford NC, Groves CG, Huppert GN, Kastning EH, Olson R, Wheeler BJ. Living with karst: a fragile foundation. 2001. doi:PNR61

53. Wallace JM, Gutzler DS. Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. 1981;109(4):784–812. doi:10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2

54. Waltham T, Bell F, Culshaw M. Sinkholes and Subsidence. 2005. doi:10.1007/b138363

55. Weary D. the Cost of Karst Subsidence and Sinkhole Collapse in the United States Compared With Other Natural Hazards. 2015:433–446.

56. Weary DJ, Doctor DH. Karst in the United States: a digital map compilation and database. 2014;1156:23. doi:http://dx.doi.org/10.3133/ofr20141156

57. Webbers A. Ground-water use by public water-supply systems in Tennessee, 2000: U.S: Geological Survey Open-File Report 03-47. 2003.

58. Winter T, Harvey J, Franke O, Alley W. Ground water and surface water a single resource: U.S. Geological Survey Circular 1139. 1998.

59. Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Campbell Grant EH, Veran S. Presence-only modelling using MAXENT: When can we trust the inferences? 2013;4(3):236–243. doi:10.1111/2041-210x.12004

60. Young N, Carter L, Evangelista P. A MaxEnt model v3.3.3e tutorial (ArcGIS v10). 2011:1–30.

61. Zeng S, Jiang Y, Liu Z. Assessment of climate impacts on the karst-related carbon sink in SW China using MPD and GIS. 2016;144:171–181. doi:10.1016/j.gloplacha.2016.07.015

62. Zhou G, Yan H, Chen K, Zhang R. Spatial analysis for susceptibility of second-time karst sinkholes: A case study of Jili Village in Guangxi, China. 2016;89:144–160. doi:10.1016/j.cageo.2016.02.001

63. Sinkhole losses: TN Code 56-7-130. 2014.

64. Sinkhole insurance; catastrophic ground cover collapse: definitions. 2017.

VITA

KIMBERLY BLAZZARD

Education:	Las Vegas Academy of International Studies, Performing and Visual Arts, Las Vegas, NV, Honors, 2006
	A.S. Geological Science, College of Southern Nevada, Las Vegas, NV, High Honors, 2014
	A.S. Earth Science, College of Southern Nevada, Las Vegas, NV, High Honors, 2014
	B.A. Art, Photography Concentration, University Nevada Las Vegas, Las Vegas, NV 2009
	B.S. Geology, University of Nevada Las Vegas, Las Vegas, NV 2016
	M.S. Geosciences, Geospatial Analysis Concentration, East Tennessee State University, Johnson City, TN 2018
Professional Societies:	Association of American Geographers
	Southeastern Division Association of American Geographers
	The National Scholars Honor Society
	National Society of Collegiate Scholars
Professional Experience:	Research Assistant, ETSU, Tennessee Climate Office, Fall 2016-Fall 2017
	Teaching Assistant, ETSU, Spring 2017-Spring 2018
	GIS Help Desk Manager and Computer Lab Coordinator, ETSU, College of Arts and Sciences, Fall 2017
	Teaching Assistant, Northside Elementary School, TN, Fall 2017-Spring 2018
Presentations:	SEDAAG Conference, Starkville, MS 2017

	Friends of Karst Mini-Symposium, Morgantown, WV 2017
	AAG Conference, Boston, MA 2017
	AEG Southern Nevada Chapter Student Banquet 2016
	University of Nevada Las Vegas Annual Geosymposium 2016
Honors and Awards:	AEG-Southern NV Conference first prize poster winner 2016
	Nomination for Golden Key International Honour Society 2017
	Dean's Honor List 2006, 2008, 2009, 2015
	President's List 2010, 2011, 2012
	Bernada French Scholarship
	Millennium Scholarship
	Certificate of Special Congressional Recognition 2005
Professional Service:	Steele Creek Park, TN, GIS based trail reroute, Spring 2017
	ETSU, guest lecture for building stratigraphic columns, 2017
	ETSU, GPR based utility mapping for utility services, 2017