STARS

Electronic Theses and Dissertations, 2004-2019

2018

Sinkhole Detection and Quantification Using LiDAR Data

Amirarsalan Rajabi
University of Central Florida

STARS Citation

SINKHOLE DETECTION AND QUANTIFICATION USING LIDAR DATA

by

AMIRARSALAN RAJABI

B.S. SHARIF UNIVERSITY OF TECHNOLOGY, 2015

A thesis submitted in partial fulfillment of the requirements

for the degree of Master of Science

in the Department of Civil, Environmental and Construction Engineering

in the College of Engineering and Computer Science

at the University of Central Florida

Orlando, Florida

Spring Term

2018

ABSTRACT

The state of Florida is highly prone to sinkhole incident and formation, mainly because of the soluble carbonate bedrock which is susceptible to dissolution and groundwater recharge that causes internal soil erosions. Numerous sinkholes, particularly in Central Florida, have occurred. Florida Subsidence Incident Report (FSIR) database contains verified sinkholes with Global Positioning System (GPS) information. In addition to existing detection methods such as subsurface exploration and geophysical methods, a remote sensing method can be an alternative and efficient means to detect and characterize sinkholes with a wide coverage.

the first part of this study is aimed at developing a method to detect sinkholes in Missouri by using Light Detection and Ranging (LiDAR) data. Morphometrical parameters such as TPI (Topographic Position Index), CI (Convergence Index), SI (Slope Index), and DEM (Digital Elevation Model) have a high potential to help detect sinkholes, based on local ground conditions and study area. The GLM (General Linear Model) built in R software is used to obtain morphometrical indices of the study terrain to be trained and build a logistic regression model to detect sinkholes. In the second part of the study, a semi-automated model in ArcMap is then developed to detect sinkholes and also to estimate geometric characteristics of sinkholes (e.g. depth, length, circularity, area, and volume). This remote sensing technique has a potential to detect unreported sinkholes in rural and/or inaccessible areas.

This work is dedicated to my parents and my wife

Without their support, this work would have not been possible.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor, Professor Boo Hyun Nam, for his friendship, support, and mentorship. Appreciation is also extended to my committee members, Professor Dingbao Wang, and Professor Arvind Singh for their valuable comments.

Special thanks to Dennis Filler, Yong Je Kim, Adam Lane Perez, Ryan Shamet, Moataz Soliman, for their friendship and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Problem statement

Sinkholes are geologic features and naturally occurring in karst terrain. Sinkholes play an important role in public safety and health because they cause structural damages, property losses, and dramatic impacts on public life. For instance, the economic loss to housing is estimated to be five-million dollars per year for the city of Tampa itself and is expected to increase in future (Lerche, 2006). The most destructive sinkhole ever reported occurred on May 8-10, 1981, in Winter Park, and caused over four-million dollars of damages.

By creating pathways between surface water and underlying aquifers, sinkholes threaten water and environmental resources. Surface contaminants can be transmitted into underlying aquifers causing degrade of ground water resources (Tihansky, 1999). In one incident, on September 16th, 2016, a sinkhole occurred over a gypsum stack. "It drained millions of gallons of acidic water laced with sulfate and sodium from a pool atop a 120-foot gypsum stack. An unknown amount of gypsum, a fertilizer byproduct with low levels of radiation, also fell into the sinkhole, which is believed be at least 300 feet deep", reported *Tampa Bay Times* (Figure 1).

Figure 1. Aerial image of sinkhole occured in Mulberry, FL (Tampa Bay Times, 2016)

<div align="center">1.2 Objectives</div>

Due to heavy impact that each sinkhole incident may have to public life and property,

sinkhole detection and quantification of sinkholes in specific areas is of great interest and may

help researchers predict and reduce the risks. Many studies have aimed to develop method to

detect and quantify sinkholes, although the topic is in early stages of development.

This study is an attempt to develop a methodology to automate the sinkhole detection

process and also quantify the geometric characteristics of sinkholes, by using LiDAR data.

<div align="center">1.3 Outline</div>

Chapter 1 provides the objectives and problem statement of this study. In addition, the

organization of thesis chapter is presented.

Chapter 2 is a literature review on the process of sinkhole development and different types of

sinkholes, and a review of studies being conducted on sinkhole risk assessment and mapping

Chapter 3 discusses a logistic regression model being developed to detect sinkholes based on the morphometrical characteristics of the terrain

Chapter 4 presents the details of a semi-automated method which is being developed to quantify the geometric characteristics of sinkholes in Central Florida.

Chapter 5 is the summary and conclusions. Limitations and recommendations are also presented.

CHAPTER 2: LITERATURE REVIEW

2.1 Sinkholes in Florida

Karst topography is formed by the geomorphic process involving dissolution of soluble carbonate bedrock, resulting in an underground network of drainage with high hydraulic conductivity. Karst topography is mostly not regular due to the presence of acidic water in the region and dissolution of carbonate rock forming cavities and resulting in sediments subside or collapse. Sinkholes (or dolines) are a feature of all karst terrains (Waltham, Bell, & Culshaw, 2005). There are six types of sinkholes: collapse, buried, solution, caprock, suffusion, and dropout (Lowe, Waltham, & British Cave Research Association., 2002). Each of these sinkhole types have some equivalent names.

Figure 2. Cover-collapse sinkhole occurrences in Florida: (a) Winter Park, FL (1981), (b) Orlando, FL (2013), and (c) Pasco County, FL (2014)

Sinkholes in Florida, by the form of either cover-collapse or cover-subsidence, are formed by erosion of subsurface soils caused by dissolution of soluble bedrock in karst landscapes. Cover-collapse sinkholes are sudden sinkhole collapse and cover-subsidence sinkholes are gradual ground subsidence phenomenon. Both sinkhole types have caused severe damage to infrastructure/buildings, and also affect water quality in underlying carbonate acquirers (Shaban & Darwich, 2011). Studies on Florida sinkholes associated with sinkhole

mechanism, sinkhole hazard assessment, and numerical analyses have been conducted (Perez et al., 2017) (Xiao et al., 2017.) (Shamet, Perez, & Nam, 2017) (Kim & Nam, 2017) (Nam and Kim, 2017)

Carbonates are a group of minerals which all contain (CO3) in their molecular formation. The most important carbonates are calcite, dolomite, and aragonite. Calcite is the most prevalent carbonate mineral. Limestones in Florida are either calcite or dolomite with calcite being more predominant. The sinkhole formation in Florida is caused by dissolution of limestone or other soluble carbonate rocks by groundwater flow. As acidic water from rainfall infiltrates into the groundwater system and encounters soluble limestones on top of the carbonate bedrock within the confined Floridan Aquifer System (FAS), the rocks naturally and very slowly begin to dissolve away and physically erode along the fractures, creating small cavities and voids.
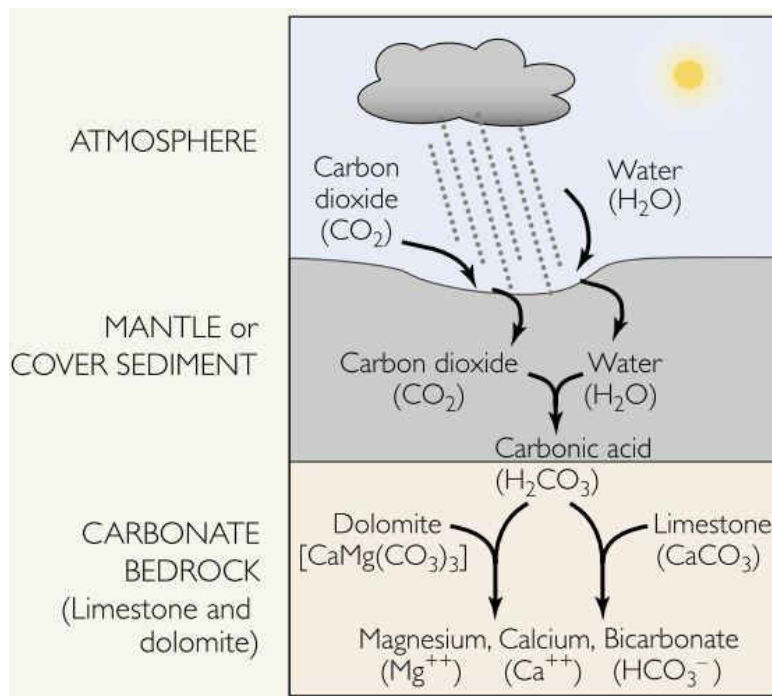
Figure 3. Dissolution process (Tihansky, 1999)

As they grow larger with time, the overlying surficial soils keep moving downward to fill into the cavities and voids, resulting in upward raveling due to continuous dissolution and soil erosion. As a result, sinkholes occur when overburden sediments either abruptly collapse or slowly subside. It is noted that the growth of individual cavities and voids can coalesce and lead to hydraulic interconnection, and thus, to increase groundwater flow and to accelerate dissolution and erosion rates.

Sinkholes in Florida are commonly classified as dissolution sinkholes, cover-subsidence sinkholes, and cover-collapse sinkholes depending on the thickness and composition of overburden materials and the local hydrologic conditions (Figures 4, 5, and 6).
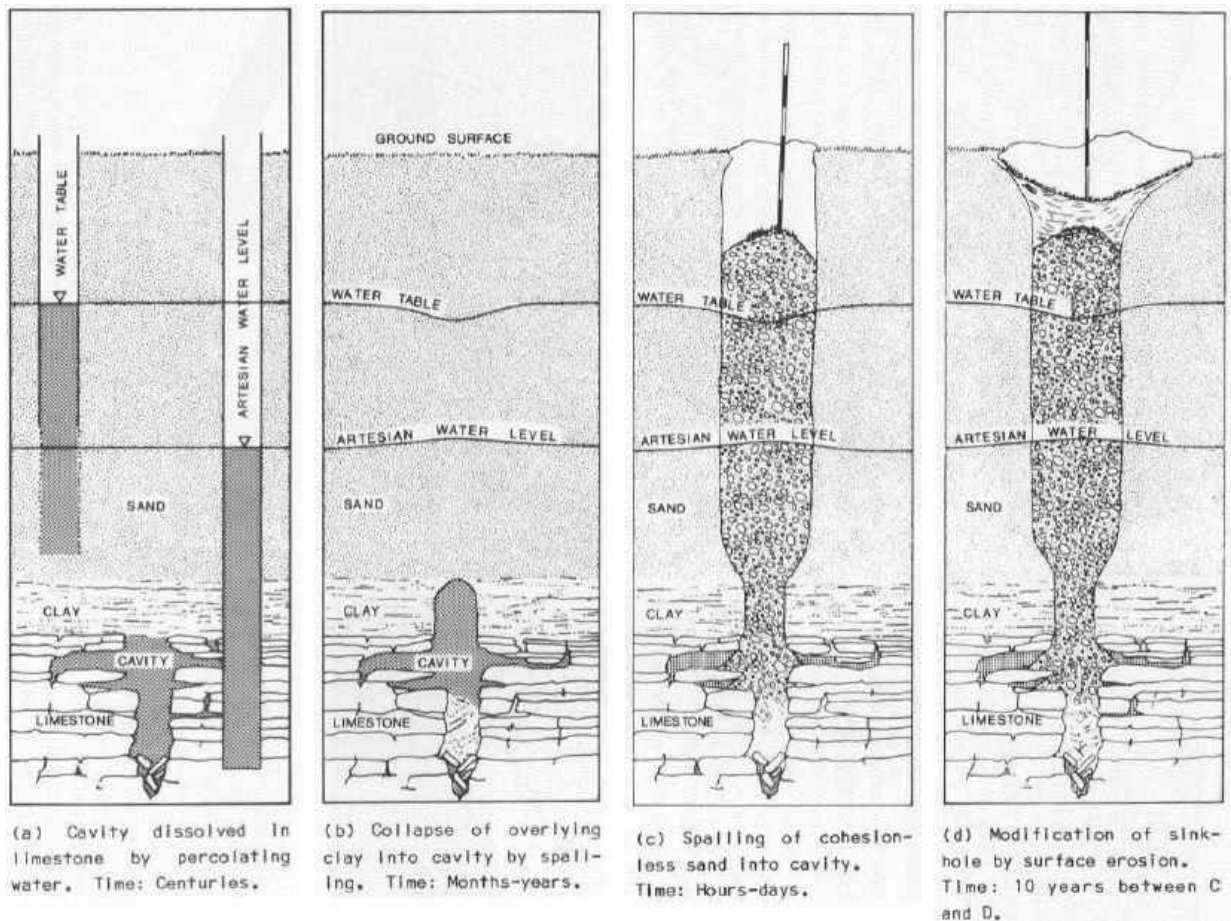
Figure 4. Cover-collapse sinkhole (Beck, 1986)

Dissolution sinkholes occur where carbonate bedrocks are slowly dissolved and carried away from the surface due to weakly acidic rail water resulting in small cavities and voids. As the surficial materials fail and move downward to infill the cavities and voids, a gradual depression on land surface is formed. These sinkholes are prominent in areas where the overburden deposits are thin and highly permeable. However, due to the slow formation process of dissolution sinkholes, damages to human lives and properties could be minor.

Compared to dissolution sinkholes, cover-subsidence and cover-collapse sinkholes are likely to occur where the overburden deposits are thicker and less permeable. Cover-subsidence sinkholes are developed in areas where covering sediments are relatively permeable non-cohesive sands and gradually settling into cavities and voids to form slow depressions in land surface. On the other hand, cover-collapse sinkholes are developed due to erosion and upward raveling of soil structures, where covering sediments contains a significant of cohesive and impermeable clays. The process of cover-subsidence sinkholes take place for a very long time, and these sinkholes may be undetectable for long periods in areas where there are thicker cover materials. Cover-collapse sinkholes, however, can develop abruptly and cause catastrophic damages.

U.S. Geological Survey (USGS) has categorized sinkhole of Florida into 4 areas (Figure 7). The first category Area I the covering material is very permeable and thin. Area I is reportedly very less susceptible to cover-collapse sinkholes. Solution sinkholes dominate this area. Area II consists of incohesive material and permeable sand. In this area, sinkholes are low in number and the thickness of cover material is between 30 feet to 200 feet. Cover-subsidence sinkholes are more prevalent in this area. Area III consists of cohesive clayey sediments with low

permeability and like area II, cover material thickness is in range of 30 to 200 feet. Sinkholes are a lot in number, the diameter of sinkholes differ, and they are developed precipitously. Cover-collapse sinkholes are more prevalent in this area. Clay component provides degree of cohesiveness to overlay material that allows bridge between surficial aquafer and develops sinkhole in carbonate aquafer (Tu, 2016). Areas II and III have similar sinkhole occurrence mechanism. The difference is that developing cavity in Area II is will be filled by incohesive soil, while the cavity in Area III will be supported by cohesive soil on the top, and in the last stage of this process, failure of the cohesive layer result in cover-collapse sinkhole. The difference in water levels between sand aquafer and its underlain carbonate aquafer is a lot. Due to this water head difference, the clay layer between these two will be stressed under hydrostatic pressure. The other pressure which the clay layer suffers are cover layers and its weight. An increase in water head difference, often by flooding or decline of water level in carbonate aquafer, will then result in a collapse. Also, pumping the water form low aquafers might the most important man-made causes of sinkholes. Area IV consists of cohesive sediments, and the thickness of cover layer is more than 200 feet. Due to the high amount of thickness, sinkholes rarely occur and the rare occurred sinkholes are very large and deep, mostly cover-collapse.
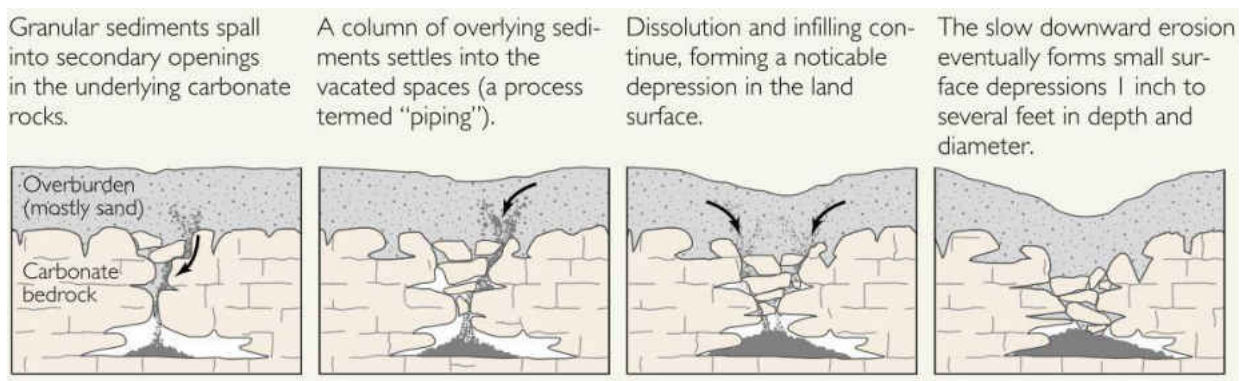
Figure 5. Cover-subsidence sinkholes in Florida (Tihansky, 1999)
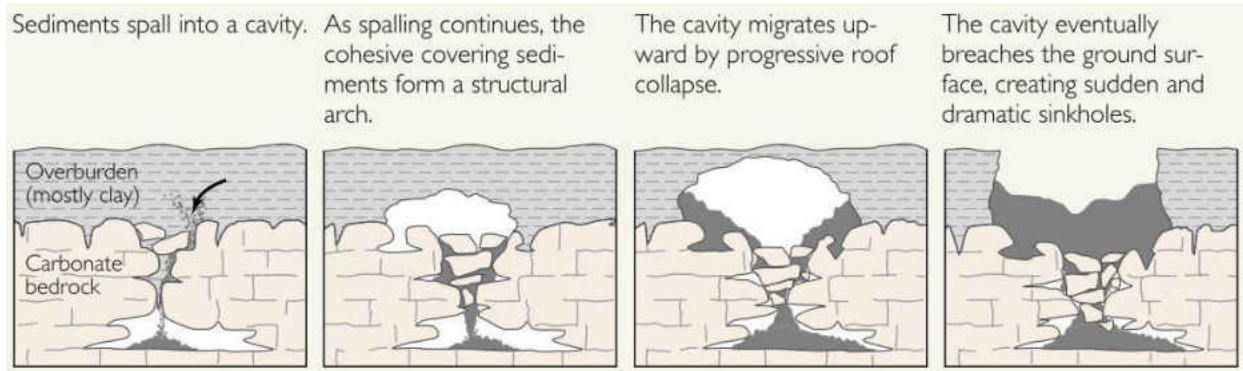
Figure 6. Cover-collapse sinkholes in Florida (Tihansky, 1999)

AREA I. Thinly covered Limestone. Sinkholes are few and shallow. Solution sinkholes dominate

AREA II. Cover 30 to 200 feet thick. Sinkholes are few, shallow, small diameter. Cover-subsidence sinkholes dominate

AREA III. Cover 30 to 200 feet thick. Sinkholes numerous of varying size, and develop abruptly. Cover-collapse sinkholes dominate

AREA IV. Cover more than 200 feet thick. Sinkholes are very few, with large diameters. Seep sinkholes occur. Cover-collapse sinkholes dominate

Figure 7. Four sinkhole category areas in Florida (U.S. Geological Survey, 1985)

2.2 LiDAR technology and sinkhole mapping techniques

Mapping sinkholes is critical for the success of public safety and infrastructure/building management. Past methods for the sinkhole mapping were mainly relied on visual interpretation

9

with low-resolution topographic maps and aerial photographs that require field verification. These ways are manually done, which are labor-intensive and time-consuming environment in the analysis. Field check of individual sinkhole is most times not practical; as a result, manually digitized sinkhole data may not be reliable (Doctor & Young, 2013). Mapping sinkholes by the use of visual topographic maps and aerial imagery is time consuming and mostly inaccurate. Also sinkholes under forested areas are impossible to be detected by aerial imagery and topographic maps. manual interpretation of karst features have shown that subjectivity in the methodology can result in false positive and false negative identification of karst features (Doctor & Young, 2013). Some previous studies (Rahimi & Alexander, 2013); (Zhu, Taylor, Currens, & Crawford, n.d.); (Wu, Deng, & Chen, 2016a) found that sinkholes may be changing fast because of natural causes or human activity.
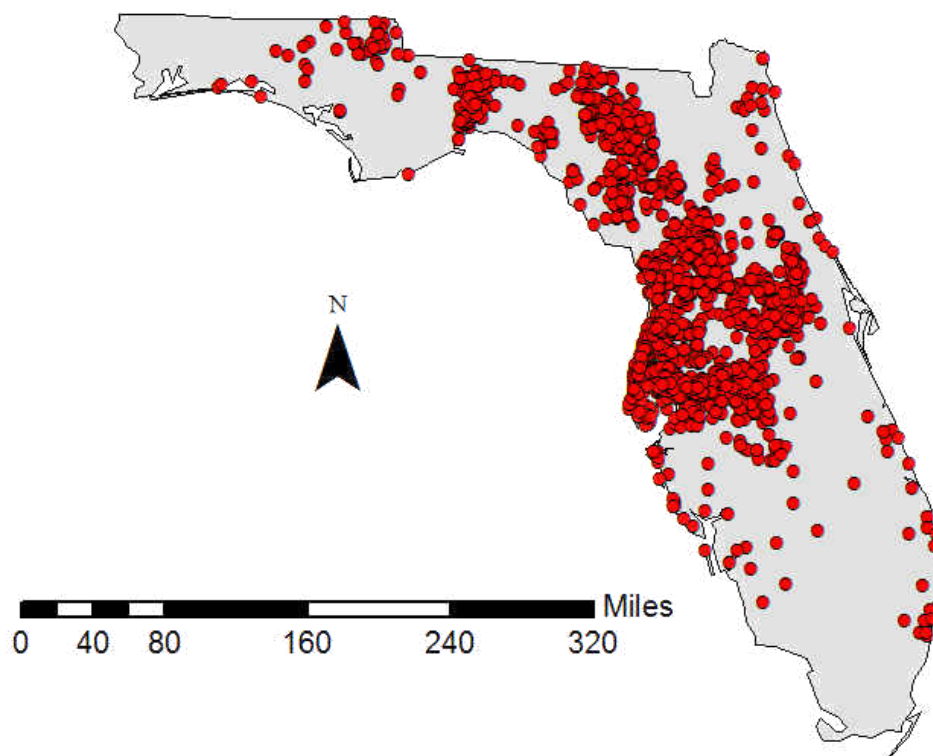
Figure 8. Map showing reported sinkholes throughout the state of Florida (Florida Geological Survey, 2017)

LiDAR (Light Detection and Ranging) technology has made possible more accurate study of karst terrain features, e.g. sinkholes. LiDAR is relatively a recent remote sensing technology that produce large volumes of accurate and high spatial-resolution topographical measurements. In this technology, LiDAR data is collected using a low-flying aircraft flying over a designated area, illuminating laser pulses and receiving them back by sensors embedded inside the aircraft, and finally calculating distance to target and finally making 3d-representation of the designated area using the wavelength and time of laser pulses (Figures 9 and 10). Airborne lidar data are collected and used by remote sensing companies, in which they can be used to create Digital Terrain Models (DTM) and Digital Elevation Models (DEM). High-resolution digitized elevation data from LiDAR enables more accurate delineation and small-scale analyses on geomorphological features and landscapes(Galve, Lucha, Castañeda, Bonachea, & Guerrero, 2011) (Wu et al., 2016a).

Figure 9. Left: Airborne LiDAR (Deepreef Explorer). Right: LiDAR 3d-representation (Rise Media Productions)
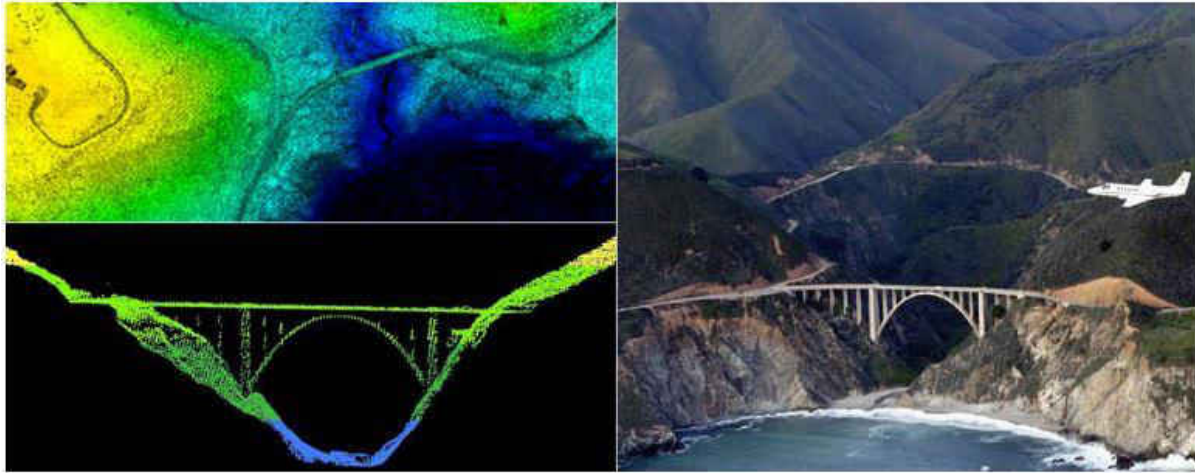
Figure 10. An application of LiDAR technology (US Department of Commerce, n.d.)

Moreover, several studies have used LiDAR data to identify and characterize sinkholes. Mukherjee (2012) employed a sink-filling method on LiDAR to determine the depression by subtracting the depressionless/filled digital elevation model (DEM) from the original DEM. Different thresholds were applied to the subtracted layer, thus sinkholes are identified. Some have used a similar sink-filling method and reported that potential sinkholes are four times the existing database in the same area (Wu, Deng, & Chen, 2016b). Some studies have used image-processing techniques to detect and delineate sinkhole boundaries. (Obu & Podobnikar, 2015) implemented kernel windows using focal functions. (Rahimi & Alexander, 2013) used active-contour approach to detect sinkhole boundaries based on elevation gradient in the surrounding region around the seed point.

CHAPTER 3: SINKHOLE DETECTION USING A LOGISTIC REGRESSION MODEL AND LIDAR DATA

3.1 Study area and data

In this chapter, a pixel-based method is proposed to detect sinkholes of a karst terrain. A free 1 x 1m LiDAR-derived DEM (Digital Elevation Model) was acquired from Missouri Spatial Data Information Service ("MSDIS LiDAR DEM File Download Tool," n.d.). The data is a high-resolution DEM, covering an area of approximately 14 km² in Greene County, MO. The region is underlain by thick, carbonate rock units that host a wide variety of karst features. The sinkholes in the region are formed by a process, similar to sinkholes in Florida.The region is located between -93.383° and -93.342° west-east longitudes and 37.326° and 37.293° north-south latitudes (Figure 11). The Geological Survey Program of Missouri Department of Natural Resources has identified 15,981 sinkholes in the state and hundreds non-reported sinkholes also exist in the region.
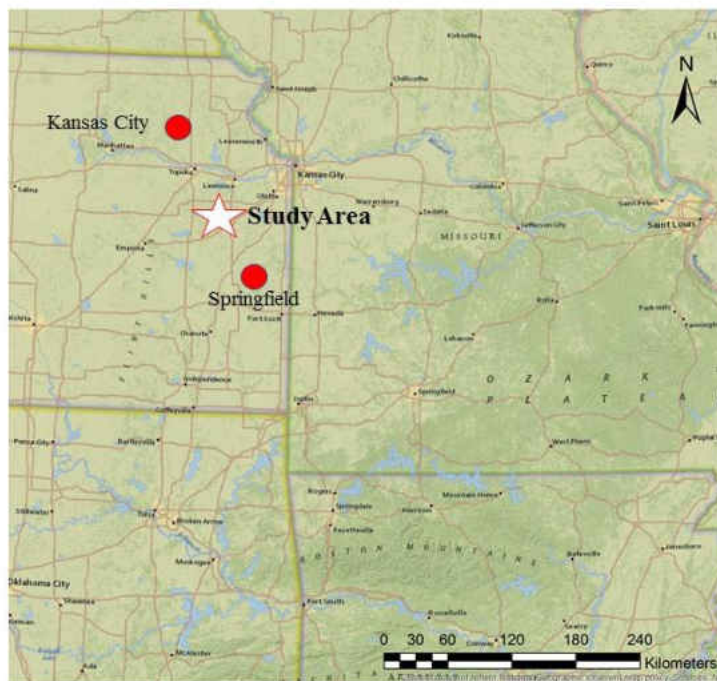
Figure 11. Location map of study area, Missouri

A 1 m x 1 m spatial resolution DEM obtained from the study area is shown in Figure 12. The

green polygons illustrate sinkhole boundaries which are provided by the city of Springfield's

Governmental Open Data ("Sinkhole Boundaries," n.d.). The data is presented as a shapefile

(vector data storage format for storing the location, shape, and attributes of geographic features

compatible with GIS softwares). This shapefile is a set of polygons delineating the boundaries of

reported sinkholes in the region. This shapefile was used to build the logistic regression model.
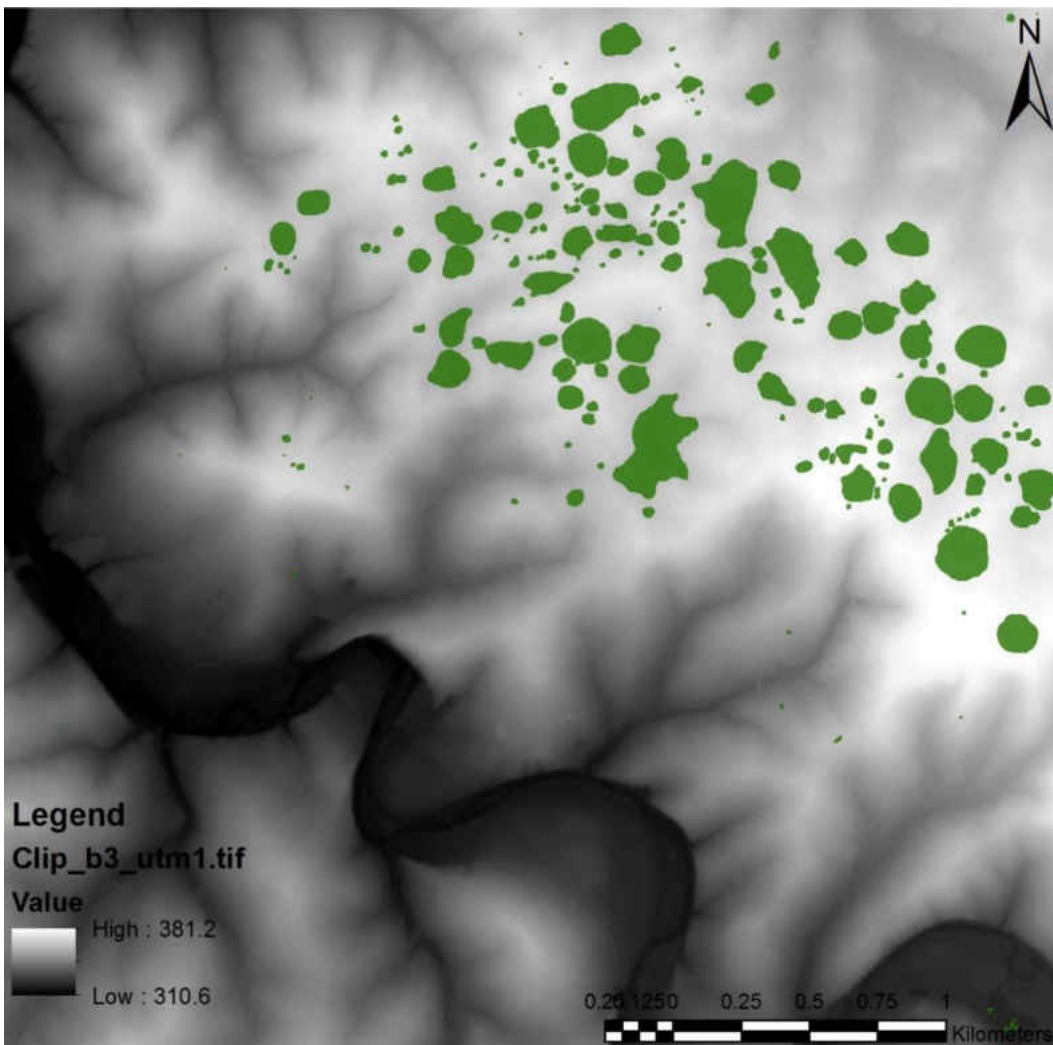
Figure 12. Digital Eleveation Model (DEM) and sinkhole boundaries in the study area

In the first step, a new raster was produced in the same extents of the original DEM, using "rasterize" function in R software. In this new raster, the cells inside the reported sinkholes of the region (polygons) were reclassified to have the value of one and the cells outside any polygon were reclassified to have the value of zero. The produced raster was used as the response variable in model building process. It represented a binomial response variable, with each cell having the value of one as "success" (or sinkhole existing) and the remaining cells with the zero value as "failure" (or sinkhole not existing) (see Figure 13).
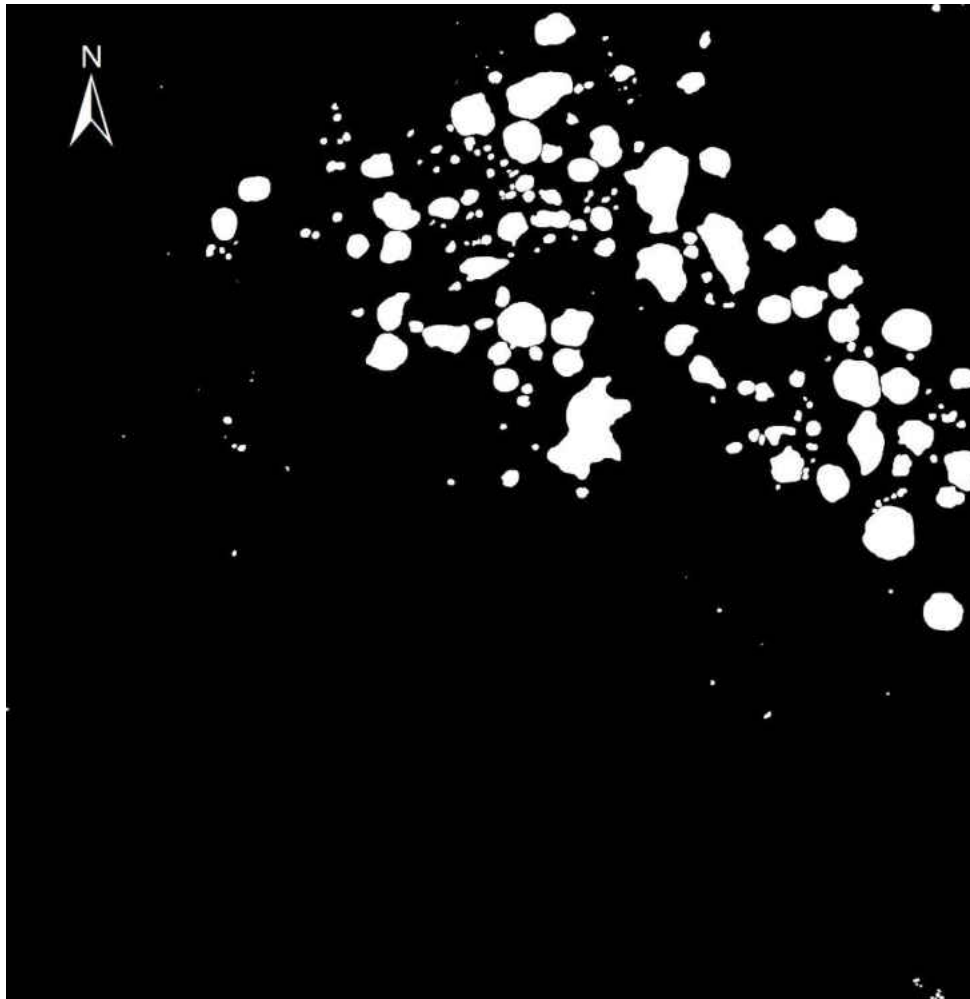
Figure 13. Reclassified study area

3.2.1 Logistic regression

Researchers are often trying to make models to analyze the relationship among some predictor variables (i.e., independent variables) and response variables (i.e., dependent variable). Logistic regression (LR) is a regression analysis to conduct when the response variable is dichotomous (i.e., binary). The following is the general equation of logistic regression:

$$logit\big(p(Y = 1)\big) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \cdots b_k X_k$$

Where   p = probability

Y = response variable

X = predictor variable

3.2.2 Morphometrical indices

In the second step, morphometrical parameters (or indices) were derived from the DEM and then were used as predictor variables in the GLM model building process. Seven (7) different morphometrical indices were derived. Using the morphometrical indices as predictor values, a Logistic Regression (LR) model was built, and most important variables associated with the existence of sinkholes in the region were identified.

Seven different morphometrical indices were derived from the DEM raster. Each morphometric index was presented as a raster. Each cell in these rasters represents the morphometrical index of the corresponding cell in DEM. Topographic Position Index (TPI) compares the elevation of each cell in a DEM to the mean elevation of a specified neighborhood around that cell (Jenness, 2001; De Reu et al., 2013).The neighborhood radius of 60 meters was used to compute TPI. Convergence Index with a search radius of 50 meters was also derived (Kiss, 2004). Normalized height and standardized height were also computed. Normalized height allots value 1 to the highest and value 0 to the lowest position within a respective reference area.

Standardized height is the product of normalized height multiplied with absolute height

(Dietrich, Physischen, & 2008, 2008). Topographic Wetness Index (TWI) which is a steady-state

wetness index was also derived (Sørensen, Zinko, & Seibert, 2006). Downslope Distance

Gradient is an index used to quantify downslope controls on local drainage (Hjerdt, McDonnell,

Seibert, & Rodhe, 2004). DEM was also considered and used as a morphometrical index (see

Figure 14).
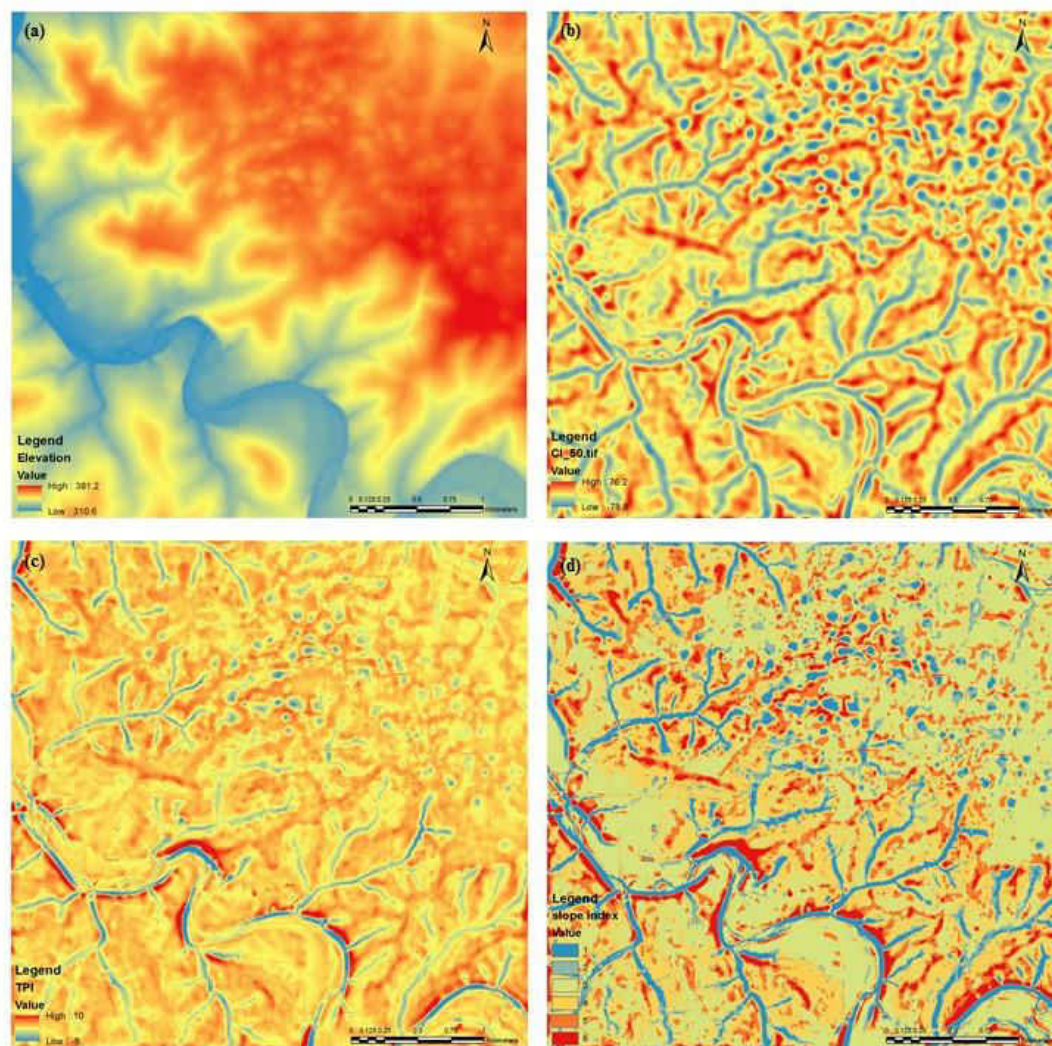
Figure 14. Morphometrical indices for the study area. **a)** Digital Elevation Model (DEM), **b)** Convergence index (CI), **c)** Topographic Position Index (TPI), and **d)** Slope index (SI)

3.2.3 GLM model selection

Prior to building GLM model and to avoid multicollinearity issue, a correlation analysis was conducted on predictor variables (morphometrical indices) and variables with a high correlation were removed. Next step, different combinations of predictor variables were used to make logistic regression models and the variable significance was checked for each model.

In order to reduce overfitting issue, we tried to use fewer predictor variables in the model. The best model was the with Convergence Index (CI), Topographic Positioning Index (TPI), Digital Elevation Model (DEM), and Slope Index (SI) as predictor variables. In Table 1 and Table 2, show two examples of models being built and tested, and model 3 (Table 3) is the best model.

Table 1. Model using CI, SH, DEM, and SI as predictor variables

Variable	Estimate	Std. Error	Z value	Pr(>|z|)
Intercept	-2.277e+03	2.286e+04	-0.100	0.921
CI	-6.609e-02	6.387e-05	-1034.852	<2e-16
SH	1.421e+01	1.464e+02	0.097	0.923
DEM	-6.919e+00	7.280e+01	-0.095	0.924
SI	2.001e-01	1.981e-03	101.01	<2e-16

$R^2 = 0.29$

Table 2. Model using CI, SH, TPI, and SI as predictor variables

Variable	Estimate	Std. Error	Z value	Pr(>|z|)
Intercept	-1.074e+02	1.235e-01	-869.61	<2e-16
CI	-5.771e-02	1.062e-04	-543.57	<2e-16
SH	3.113e-01	3.646e-04	853.85	<2e-16
TPI	-3.077e-01	3.164e-03	-97.24	<2e-16
SI	1.801e-01	2.201e-03	81.83	<2e-16

$R^2 = 0.55$

Table 3. Model using CI, TPI, DEM, and SI as predictor variables

Variable	Estimate	Std. Error	Z value	Pr(>|z|)
Intercept	-6.058e+01	7.021e-02	-862.82	<2e-16
CI	-5.762e-02	1.045e-04	-551.41	<2e-16
TPI	-5.361e-01	3.950e-03	-135.72	<2e-16
DEM	1.578e-01	1.862e-04	847.57	<2e-16
SI	1.961e-01	2.156e-03	90.96	<2e-16

$R^2 = 0.58$

3.2.4 Cutoff value

The result of a logistic regression model in this case, is a raster in which each cell shows the probability of that specific cell residing in a sinkhole. Therefore, a cutoff value should be chosen for this model. Figure 15 is the probability raster.

Sensitivity and Specificity are statistical measures of performance of a classification model (Bewick, Cheek, & Ball, 2004). The former quantifies the avoiding of false negatives detections and the latter quantifies the avoidance of false positives. With cutoff value increasing, specificity also increases, and sensitivity decreases. Based on sensitivity/specificity curve, the optimum cutoff value was chosen 0.63 which is the interception of the two curves (Figure 16).

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$
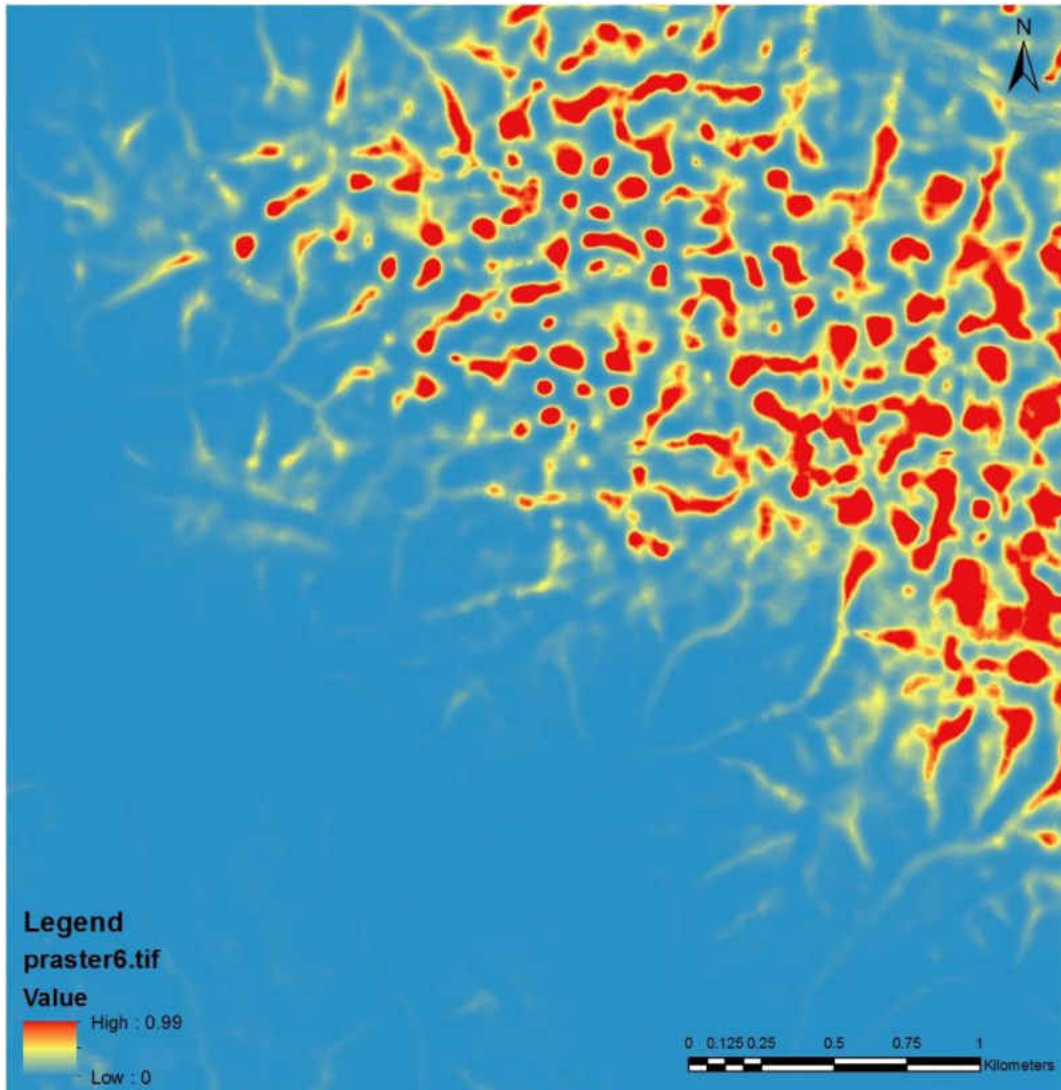
Figure 15. Sinkhole existence probability map

Based on chosen cutoff value, a new raster was produced in which a value of 1 was assigned to cells with probability of equal or more than 0.63 and 0 was assigned to the cells with a probability of less than 0.63. as Figure 17 is demonstrating, the method is ideal for locating sinkholes and not necessarily delineating the boundaries of sinkholes. By increasing the cutoff value, the boundaries generated by the model will approach the actual sinkhole boundaries,

although the chance of False Positive detections increases. Figure 18 shows the boundaries generated by the GLM model after implementing it on the study area.
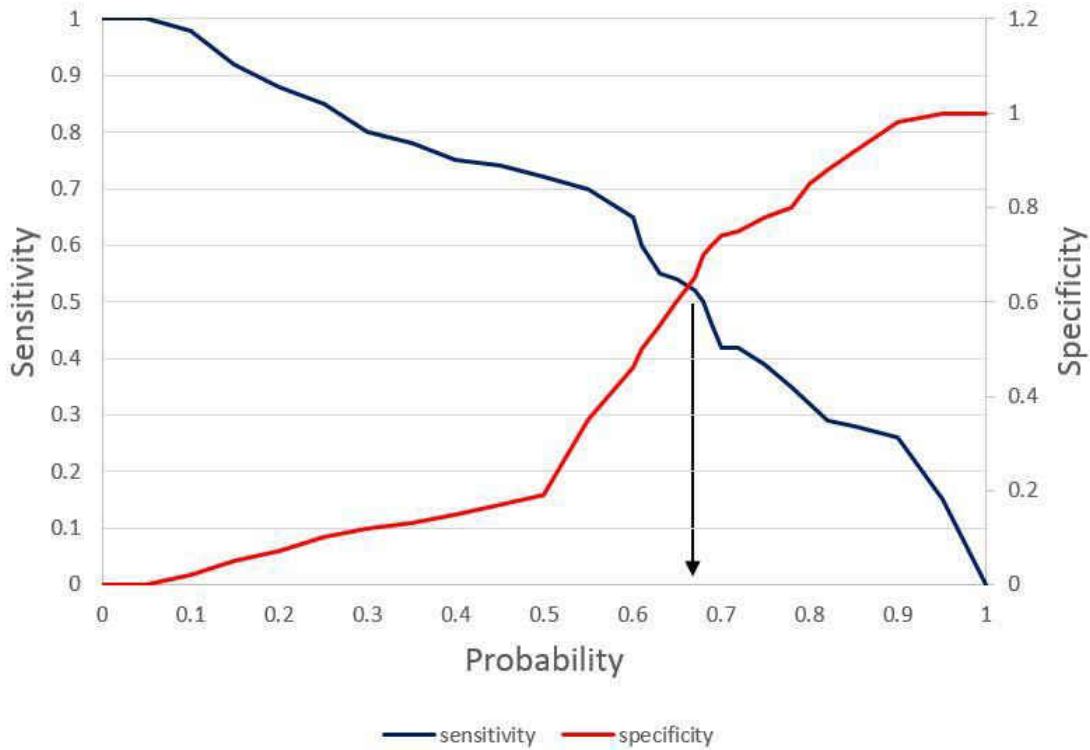
Figure 16. Sensitivity and specificity chart

All morphometrical indices were derived and computed and presented in QGIS® software. Model building was conducted in R software (RTeam, 2017) with glm2 package (Marschner, 2011). raster layers were read, modified, and produced using raster package (Hijmans, 2017). Cutoff value was determined using proc package (Xavier Robin, Natacha Turck, Alexandre Hainard & Frédérique Lisacek, 2011).

21
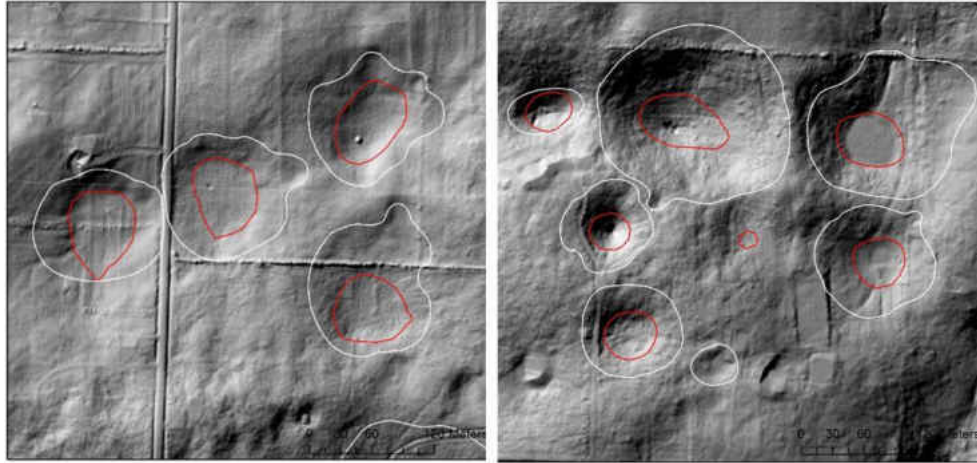
Figure 17. Validated sinkhole boundaries(white) and sinkhole boundaries detected by model(red)
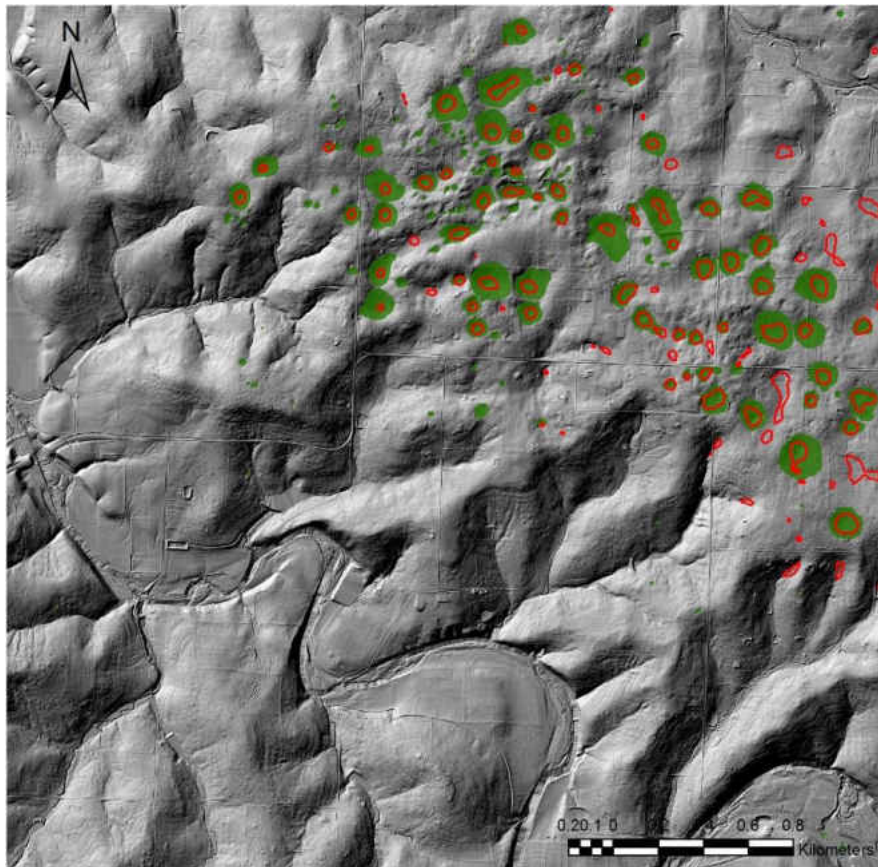
Figure 18. Sinkhole detected in study area (green polygons), and sinkholes detected by GLM model(red)

CHAPTER 4: SINKHOLE QUANTIFICATION USING LIDAR DATA [1]

4.1 Florida's sinkholes

The sinkhole formation in Florida is caused by dissolution of limestone or other soluble carbonate rocks by groundwater flow. As acidic water from rainfall infiltrates into the groundwater system and encounters soluble limestones on top of the carbonate bedrock within the confined Floridan Aquifer System (FAS), the rocks naturally and very slowly begin to dissolve away and physically erode along the fractures, creating small cavities and voids. As they grow larger with time, the overlying surficial soils keep moving downward to fill into the cavities and voids, resulting in upward raveling due to continuous dissolution and soil erosion. As a result, sinkholes occur when overburden sediments either abruptly collapse or slowly subside. It is noted that the growth of individual cavities and voids can coalesce and lead to hydraulic interconnection, and thus, to increase groundwater flow and to accelerate dissolution and erosion rates.

Sinkholes in Florida are commonly classified as dissolution sinkholes, cover-subsidence sinkholes, and cover-collapse sinkholes depending on the thickness and composition of overburden materials and the local hydrologic conditions. Dissolution sinkholes occur where carbonate bedrocks are slowly dissolved and carried away from the surface due to weakly acidic rail water resulting in small cavities and voids. As the surficial materials fail and move downward to infill the cavities and voids, a gradual depression on land surface is formed. These sinkholes are prominent in areas where the overburden deposits are thin and highly permeable.

---

[1] The content of this chapter also will appear in:
A. Rajabi, Y. Kim, S. Kim, B. Nam (2017). "*A preliminary study on use of LiDAR data to characterize sinkholes in Central Florida*". International Foundation Congress and Equipment Expo (IFCEE) 2018, Orlando.

However, due to the slow formation process of dissolution sinkholes, damages to human lives and properties could be minor.

Compared to dissolution sinkholes, cover-subsidence and cover-collapse sinkholes are likely to occur where the overburden deposits are thicker and less permeable. Cover-subsidence sinkholes are developed in areas where covering sediments are relatively permeable non-cohesive sands and gradually settling into cavities and voids to form slow depressions in land surface. On the other hand, cover-collapse sinkholes are developed due to erosion and upward raveling of soil structures, where covering sediments contains a significant of cohesive and impermeable clays. The process of cover-subsidence sinkholes take place for a very long time, and these sinkholes may be undetectable for long periods in areas where there are thicker cover materials. Cover-collapse sinkholes, however, can develop abruptly and cause catastrophic damages.

<div align="center">4.2 Methodology</div>

A procedure to identify sinkholes was developed. Once sinkholes are identified, the geometric characteristics can be determined. Figure 19 shows the flowchart for the methodology. GIS-based software, ArcGIS, was used. The procedure involves five steps: (1) creating Digital Elevation Model (DEM), (2) processing the DEM, extraction of depression DEM, (3) extraction of sinkhole boundary contours, (4) calculation of sinkhole geometric properties, and (5) eliminate non-sinkholes depressions based on threshold values. The threshold values are the criteria, with respect to geometric characteristics, to determine whether surface depressions are sinkhole or not. More details are described as below.
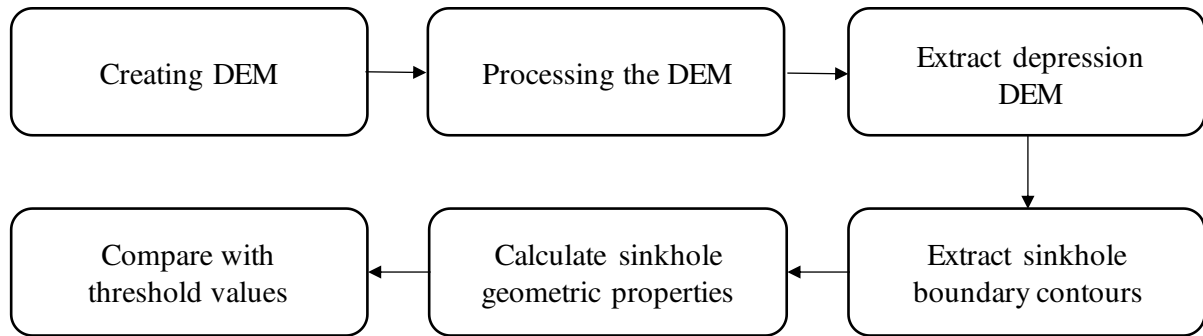
Figure 19. Sinkhole detection and quantification process flowchart.

The first step is to create 1 m x 1 m DEMs from the LiDAR point cloud using a triangulation interpolation method. (Figure 20). The second step is to identify and fill sinks in the DEM using the *Fill* tool in ArcGIS. The third step is to subtract the filled elevation raster from the original DEM (Figure 21). The result will be the difference raster where only the depressions have a value and all other pixels' values are zero. In the fourth step, *Reclassify* tool is used to classify pixels less than 20 cm (vertical accuracy of LiDAR data). The contours of the reclassified raster are delineated and converted into polygons. Polygons that do not meet the threshold requirements are then eliminated. The most outward contours, which remain after implementing threshold values, are depression, or in other word, sinkhole-candidates' boundaries. (Figure 22).
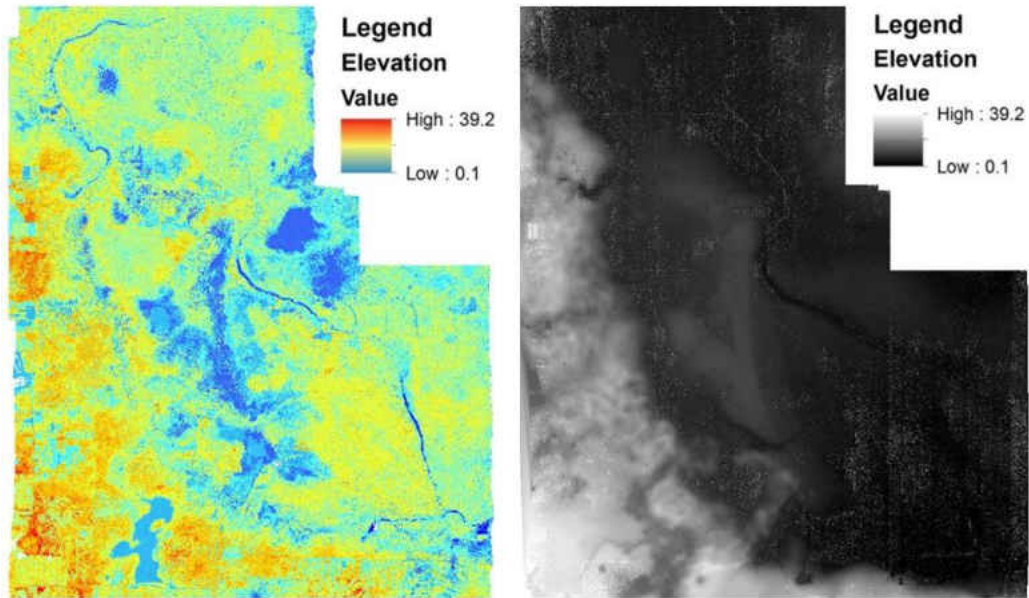
Figure 20. Left: raw LiDAR data. Right: DEM produced by LiDAR data
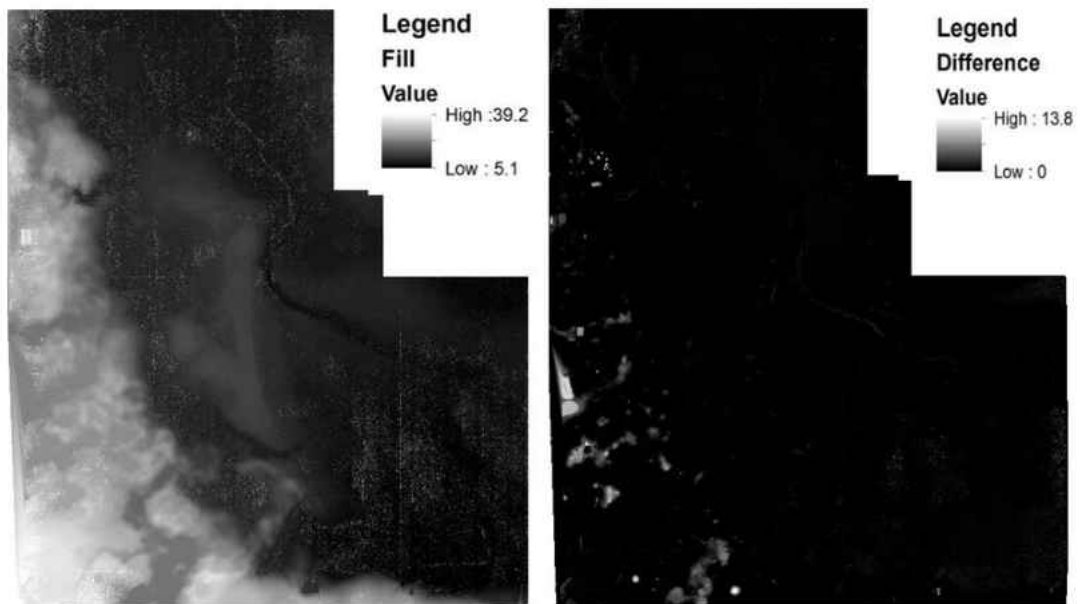
Figure 21. Left: Filled DEM raster. Right: Difference raster
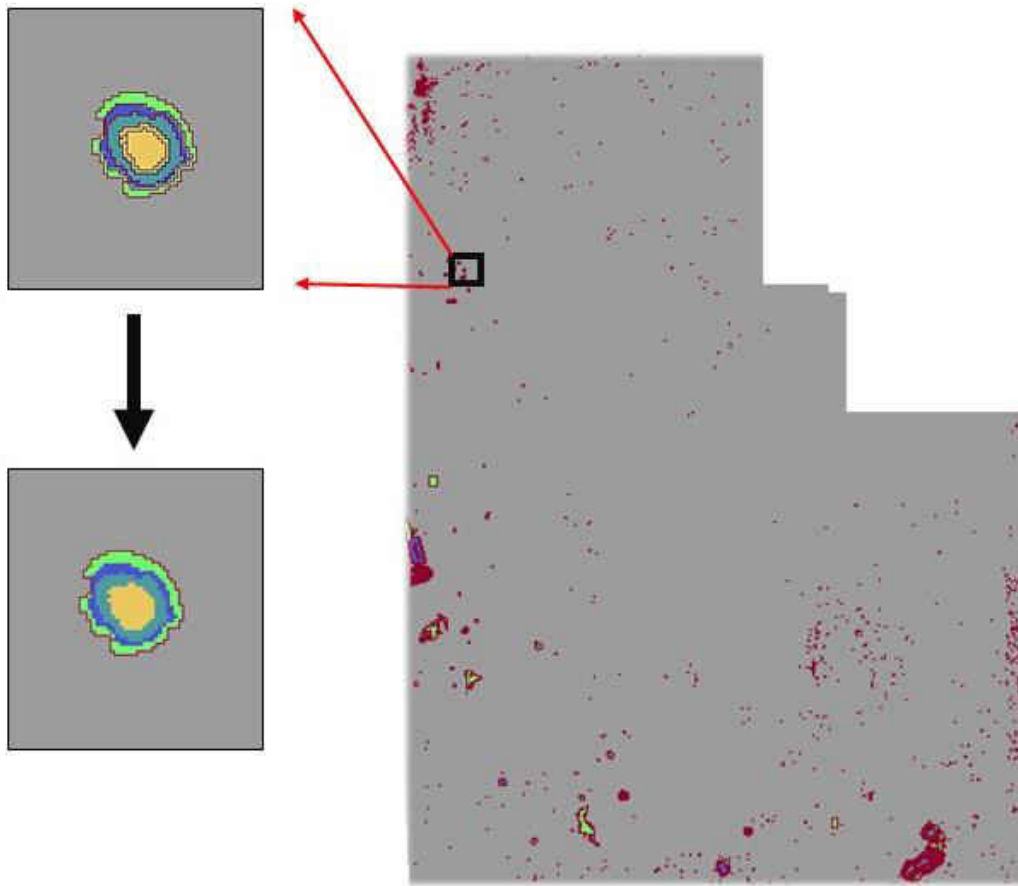
Figure 22. Contour drew over study area and the process of contour eliminating

4.3 Threshold values of sinkhole geometric characteristics

In this section, criteria to eliminate non-sinkholes from the identified polygons were established. Based on sinkholes in the selected areas where LiDAR data are analyzed, the upper and lower limits of area, perimeter, depth, and length are determined so that sinkholes outside the upper and lower limits are eliminated. First, basic geometric characteristics of reported sinkholes of Central Florida were evaluated. The area selected in constructing the thresholds include nine counties, including Marion, Sumter, Lake, Seminole, Orange, Osceola, Polk, Hardee, and Highlands Counties. The total number of sinkholes over the nine counties is 807. About 50

sinkholes do not contain geometric information; thus, they were not included in further analysis. A few "mega" sinkholes, for example greater than 8 m, are not included in the analysis because most sinkholes occurring in the area of central Florida are in the rage of 0.3 to 7 m. These mega sinkholes are considered as outlier in the sinkhole database. To eliminate those outliers, interquartile ranges (IQRs) method was used. For this purpose, first and third quartile of the data was calculated for each parameter. The IQR was then calculated by subtracting first quartile from third quartile. Any data which is more than 1.5*IQR below the first quartile or more than 1.5*IQR above the third quartile is regarded as an outlier, thus they are filtered out (Figure 24). The remaining data was used to calculate cumulative frequencies for length, area, perimeter, and depth. Figure 23 shows plotted cumulative frequencies for geometric parameters. Based on summary statistics of central Florida sinkholes, it appears that most sinkholes ranged from 5 and 95 percentiles in size, thus threshold values are determined by corresponding 5 and 95 percent of area, perimeter, depth, and length. The constructed thresholds are presented in Table 4. These thresholds will be used to eliminate non-sinkholes identified from LiDAR data.
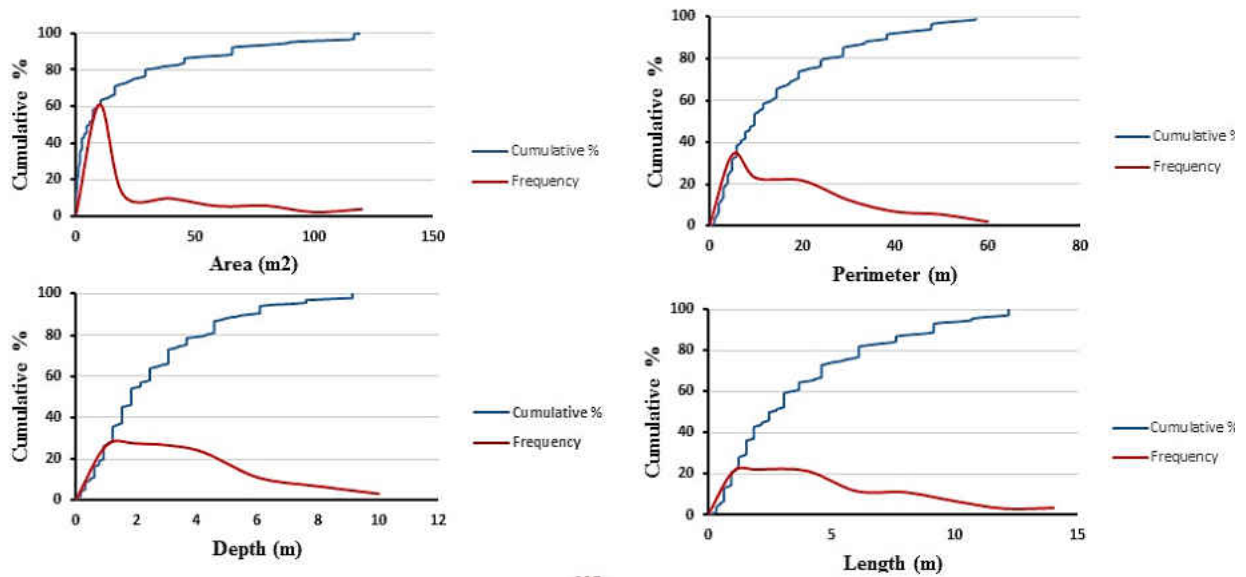
Figure 23. Cumulative frequency for (a) Area. (b) Perimeter. (c) Depth. (d) Length
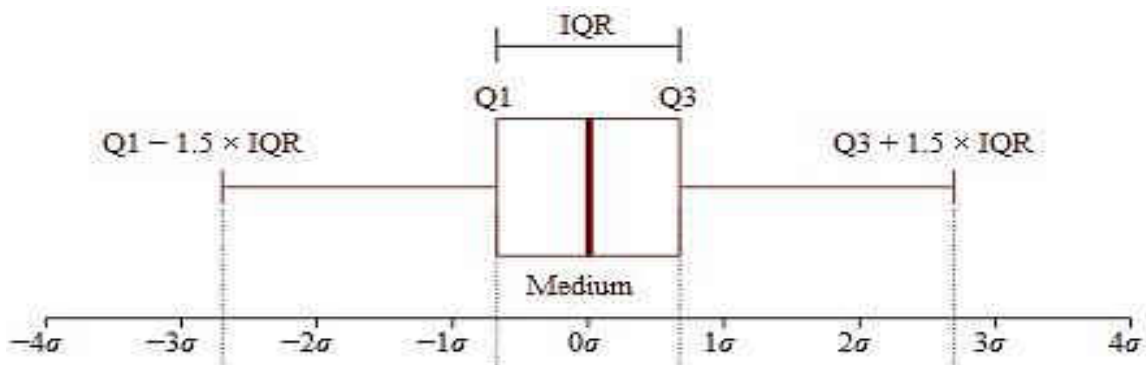
Figure 24. IQR (Interquartile Ranges Method)

Table 4. 5th and 95th percentile for Central Florida Reported Sinkholes.

Parameter	Area (m^2)	Perimeter (m)	Depth (m)	Length (m)
5th percentile	0.16	1.46	0.27	0.46
95th percentile	89.36	47.41	7.19	10.67

4.4 Determination of geometric characteristics

The detected depressions are then quantitatively characterized regarding their geometric characteristics, including area, depth and volume and circularity are calculated. *Zonal Statistics as Table* tool is used to calculate depth and standard deviation of elevation values inside the depressions. To determine the areas and volume of sinkholes in 2D and 3D, the elevation difference raster is first clipped by the sinkhole boundaries polygons and then the tool *Surface Volume* is used to calculate mentioned parameters.

Circularity is deviation of boundary of a geometric shape from a circle and is calculated using by the following equation. The circularity value is 1.0 for a circle and close to zero for a highly elongated shape.

$$C = \frac{4\pi A}{p^2}$$

where *C* is circularity, *A* is area of shape, *p* is perimeter of shape. To determine the circularity of depressions, *Minimum Bounding Geometry* tool is used and the length and area of the smallest convex polygon enclosing the depression is used.

4.5 Site description and LiDAR data

LiDAR data was acquired by the National Center for Airborne Laser Mapping (NCALM). Airborne LiDAR data were collected on June 25, 2011. The point density is 6.73 (pts/m^2) and covers approximately 49 km^2 in Orange County, FL and a portion of Seminole county. It is located between -81.5112° and -81.4429° west-east longitudes and 28.7827° and 28.7013° north-south latitudes (see Figure 25).

The area can be considered as highly vegetated and non-residential areas. Due to these accessibility issues, manual detection of depressions and aerial images may not be practical; thus, using LiDAR data for detection can be an effective approach.
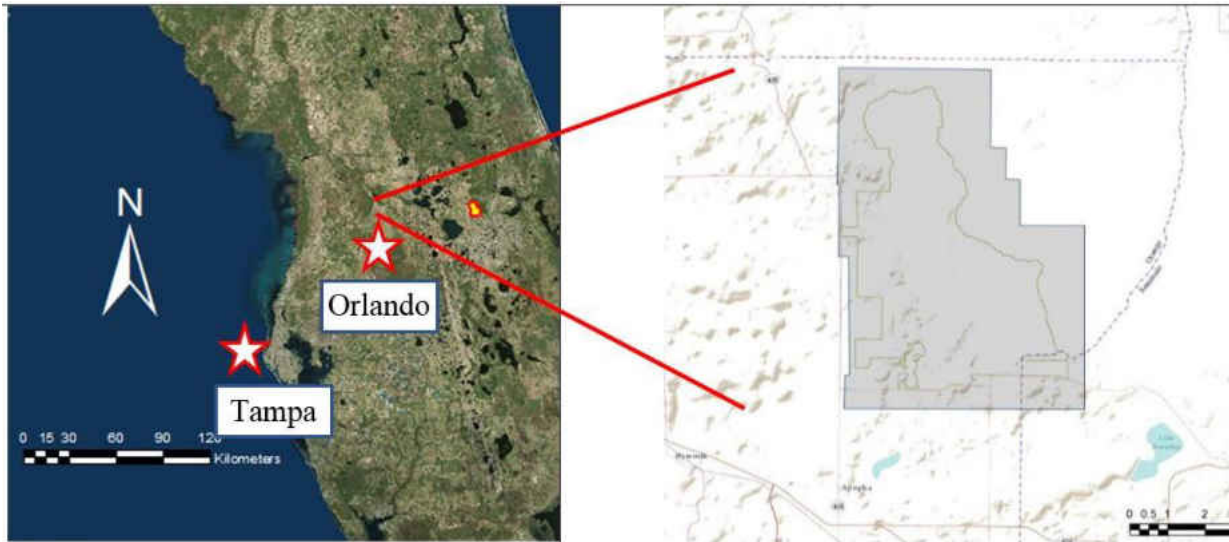
Figure 25. Location map of study area, Florida

4.6 Results and discussion

The methodology described in Figure 19 was applied to the study area shown in Figure 25. Figure 26 shows the aerial and hillshade images of the study area where potential sinkholes exist. Red closed circles are the boundaries of detected surface depressions. It is obvious that the aerial image does not show surface depressions/sinkholes due to high density of vegetation and presence of trees. On the other hand, the hillshade image (see Figure 26.b) shows two depressions (denoted as A and B in the figure). Depressions A and B were selected for further analysis to determine their morphometric characteristics. The 2D and 3D profiles of both depressions are shown in Figure 27. A LiDAR pulse can be reflected from many features and return more than one pulse and a set of filters can be used. The LAS dataset view includes all returns with no use of filters, therefore in addition to ground points, vegetation and trees are visible in profile view.

The threshold values established in Table 4 were applied to filter out non-sinkhole depressions. The geometric characteristics of Depression A and B are determined as shown in Table 5. Depression B is quite large having a depth of 8.67 m and an area of 383 m^2. According to the lower and upper limits in Table 4, 5th percentile and 95th percent respectively, this depression is not considered as a sinkhole because the area and depth exceeds the area and depth thresholds of 89.36 m^2 and 7.19 m. On the other hand, the depth, area, perimeter, and length of Depression A are within the lower and upper limits of area, depth, length, and perimeter shown in Table 4. Thus, there is a high potential that Depression A can be a naturally occurred sinkhole but had not been detected due to non-accessible area. We checked with Florida Subsidence Incident Reports (FSIR) and it is not included in the database.

LiDAR data accurately calculates the geometric characteristics of those depressions. These are valuable information with respect to damage assessment that help engineers to select optimum repair methods and decide the level of repair and/or reinforcement (e.g. volume of cement grouting). Extension of LiDAR-based sinkhole detection can enable more accurate sinkhole mapping that the existing FSIR database does not cover due to unreported sinkholes in non-residential areas.
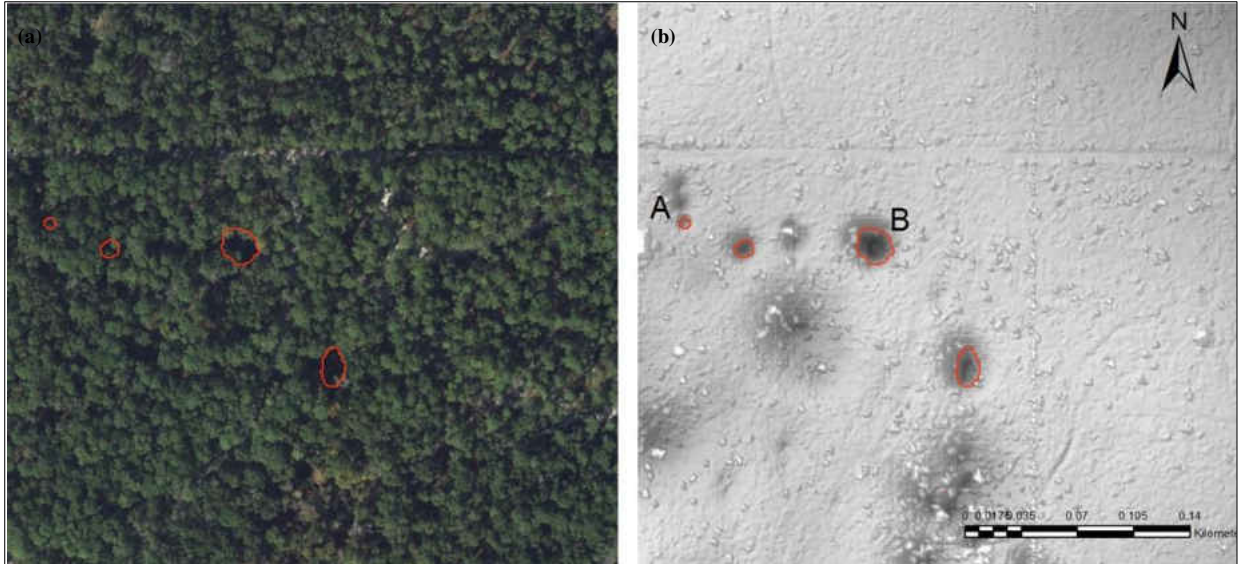
Figure 26. (a) Aerial image and depression boundaries. (b) Hillshade image and depression boundaries.
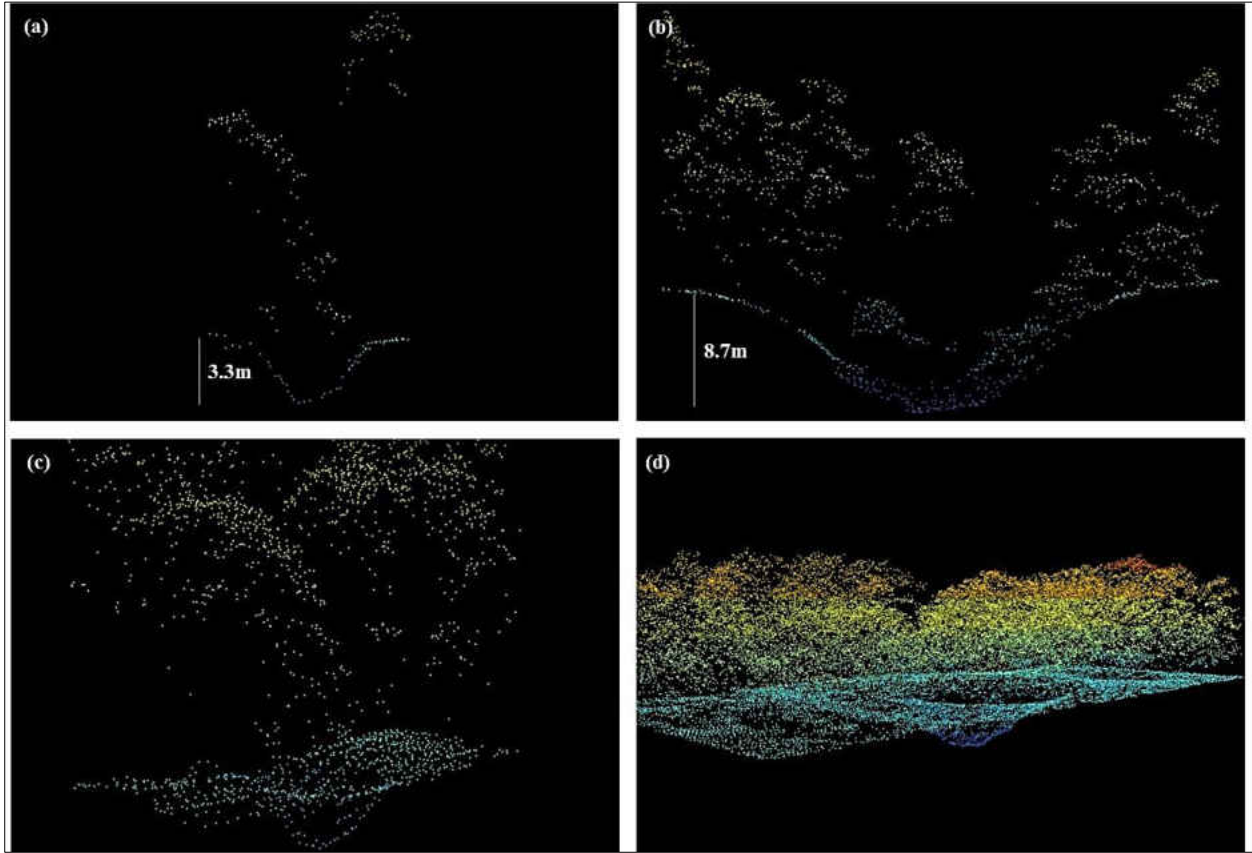
Figure 27. (a) and (b): Detailed 2D profile of depressions A and B. (c) and (d): Detailed 3D views of depressions A and B

Table 5. Results of sinkhole quantification.

Parameter	Depth (m)	Std. Error	Area (m²)	Volume (m³)	Perimeter (m)	Length (m)	Circularity
Intercept	3.31	0.92	37	50.7	23.31	7.71	0.95
CI	8.67	2.34	383	1715.67	73.76	24.8	0.96

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary

This study contains two achieved two important facts. In the second chapter, a method to detect sinkholes accurately was introduced. Although the method was very efficient for detecting and locating sinkholes, Due to the fact that the boundaries expand and shrink based on determined cutoff value, it is suggested for this method not to be used for quantification of sinkholes. Instead, the method proposed in chapter 3 is a reliable and accurate way to quantify sinkholes.

The method proposed in chapter 3 was applied to a selected area of Orange County, FL and noticeable surface depressions were identified. Two identified depressions were selected as case study and their morphometric characteristics were determined. By using the threshold values constructed, the authors believe that the method can distinguish between non-sinkhole depressions and sinkholes.

5.2 Conclusions

Based on the study results, following conclusions have been made:

1. The LiDAR-based remote sensing technique can be a potential means to effectively and accurately detect sinkholes.

2. Field validation and measurement of sinkholes are labor-intensive and often impossible due to large area coverage and inaccessibility.

3. geometric characteristics of detected sinkholes can be easily quantified by ArGIS tool.

4. These information is helpful in selecting effective engineering solutions and level of repair cost because dimension of sinkholes are provided.

This study is a preliminary study that investigates the potential of LiDAR remote sensing technique in sinkhole hazard assessment. Future research works are necessary to fully automate the process of sinkhole identification and quantification of geometric information. Additionally, advanced image processing and thresholds to filter out outliers will result in enhanced accuracy. The proposed methodology can be used as a tool not only for sinkhole detection in non-resident/rural areas but also for damage assessment to quantify location, distribution, and geometric information when natural and man-made events (e.g. hurricane and groundwater pumping after severe droughts) create many sinkholes in specific times and regions.

5.3 Limitations and Recommendations

Limitations in this study include:

1. Location and geometric characteristics of sinkholes provided in Florida Subsidence Incidents Report are not accurate. Therefore, the methodology explained in chapter three, was not possible to be tested for sinkholes of Florida.

2. Acquisition of free LiDAR data which is open to be used by public was challenging for this study and this factor determined the areas of study.

3. The LiDAR data which is used in chapter 4 of this study, is collected on 2011 and therefore might miss sinkholes which are formed after that time.

The following are recommended for future studies:

1. For method discussed in chapter 3 of this study, additional morphometrical indices should be tested as predictor variables.

2. The cut-off value which was chosen in this study was at the interception of specificity and sensitivity curves. Although this approach is based on optimization of the model, future research must be conducted on cut-off value and how to choose the best cutoff value based on objectives.

3. The method discussed in chapter 4 was semi-automated. Future study must be conducted to enhance and fully automate quantification of sinkholes.

APPENDIX: MULTIPLE REGRESSION MODELS TRIAL AND ERRORS

Table 6. Model using Convergence Index as predictor variable

```
Call:
glm(formula = sinks ~ CI, family = binomial, data = rdata)

Deviance Residuals:
   Min     1Q    Median    3Q      Max
-2.0065  -0.3801  -0.3325  -0.2921  2.8841

Coefficients:
           Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.7480042  0.0011742 -2340.4   <2e-16 ***
CI          -0.5678738  0.0009082  -625.2   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6834137  on 13743575  degrees of freedom
Residual deviance: 6464901  on 13743574  degrees of freedom
  (3663 observations deleted due to missingness)
AIC: 6508295

Number of Fisher Scoring iterations: 5
```

Table 7. Model using Topographic Position Index as predictor variable

call:
glm(formula = sinks ~ TPI, family = binomial, data = rdata)

Deviance Residuals:
   Min   1Q   Median   3Q   Max
-2.0065 -0.3801 -0.3325 -0.2921   2.8841

Coefficients:
          Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.7480042 0.0011742 -2340.4   <2e-16 ***
TPI   -0.5678738 0.0009082 -625.2   <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6834137 on 13743575 degrees of freedom
Residual deviance: 6464901 on 13743574 degrees of freedom
  (3663 observations deleted due to missingness)
AIC: 6508295

Number of Fisher Scoring iterations: 5

Table 8. Model using Slope Index as predictor variable

Call:
glm(formula = sinks ~ SI, family = binomial, data = rdata)

Deviance Residuals:
   Min   1Q   Median   3Q   Max
-0.6739 -0.3838 -0.3838 -0.2129   2.9632

Coefficients:
          Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.7647497 0.0027487 -278.2   <2e-16 ***
SI   -0.6021591 0.0009313 -646.5   <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6834137 on 13743575 degrees of freedom
Residual deviance: 6389291 on 13743574 degrees of freedom
 (3663 observations deleted due to missingness)
AIC: 6432608

Number of Fisher Scoring iterations: 6

Table 9. Model using Convergence Index and Slope Index as predictor variable

```
Call:
glm(formula = sinks ~ CI + SI, family = binomial, data = rdata)

Deviance Residuals:
   Min     1Q    Median    3Q     Max
-1.2585  -0.3959  -0.2696  -0.1834  3.3595

Coefficients:
          Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.266e+00  4.815e-03 -678.19   <2e-16 ***
CI      -4.737e-02  7.375e-05 -642.37   <2e-16 ***
SI       8.563e-02  1.332e-03   64.28   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6834137  on 13743575  degrees of freedom
Residual deviance: 5925739  on 13743573  degrees of freedom
  (3663 observations deleted due to missingness)
AIC: 5968958
```

Table 10. Model using Convergence Index and DEM as predictor variable

Call:
glm(formula = sinks ~ CI + DEM, family = binomial, data = rdata)

Deviance Residuals:
    Min   1Q   Median   3Q   Max
-2.3816 -0.2715 -0.1011 -0.0370   3.6800

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.728e+01 6.393e-02 -896.0   <2e-16 ***
CI   -6.609e-02 6.387e-05 -1034.9   <2e-16 ***
DEM   1.507e-01 1.742e-04   864.9   <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6834137 on 13743575 degrees of freedom
Residual deviance: 4095488 on 13743573 degrees of freedom
 (3663 observations deleted due to missingness)
AIC: 4138612

Number of Fisher Scoring iterations: 8

Table 11. Model using Slope Index and DEM as predictor variable

Call:
glm(formula = sinks ~ SI + DEM, family = binomial, data = rdata)

Deviance Residuals:
   Min   1Q   Median   3Q   Max
-1.7205 -0.3450 -0.1484 -0.0626 3.9244

Coefficients:
          Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.045e+01 4.695e-02 -861.6   <2e-16 ***
SI   -1.007e+00 1.292e-03 -779.2   <2e-16 ***
DEM   1.139e-01 1.317e-04   864.7   <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6834137 on 13743575 degrees of freedom
Residual deviance: 4929068 on 13743573 degrees of freedom
  (3663 observations deleted due to missingness)
AIC: 4972315

Number of Fisher Scoring iterations: 7

Table 12. Model using Convergence Index and DEM and Slope Index as predictor variable

Call:
glm(formula = sinks ~ SI + DEM + CI, family = binomial, data = rdata)

Deviance Residuals:
  Min   1Q   Median   3Q   Max
-2.3918 -0.2715 -0.1013 -0.0371   3.6760

Coefficients:
          Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.733e+01 6.414e-02 -893.854 <2e-16 ***
SI   1.612e-02 1.701e-03   9.475   <2e-16 ***
DEM   1.507e-01 1.743e-04 864.590   <2e-16 ***
CI   -6.660e-02 8.304e-05 -801.967   <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6834137 on 13743575 degrees of freedom
Residual deviance: 4095399 on 13743572 degrees of freedom
  (3663 observations deleted due to missingness)
AIC: 4138526

Number of Fisher Scoring iterations: 8

REFERENCES

Beck, B. F. (1986). *Sinkholes in Florida : an introduction*. Florida Sinkhole Research Institute, College of Engineering, University of Central Florida in cooperation with the U.S. Geological Survey.

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 13: receiver operating characteristic curves. *Critical Care (London, England)*, *8*(6), 508–12. https://doi.org/10.1186/cc3000

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., … Crombé, P. (2013). Application of the topographic position index to heterogeneous landscapes. *Geomorphology*, *186*, 39–49. https://doi.org/10.1016/J.GEOMORPH.2012.12.015

Dietrich, H., Physischen, J. B.-H. B. zur, & 2008, U. (2008). Cold air production and flow in a low mountain range landscape in Hessia (Germany). *Hg.mirror.ac.za*. Retrieved from ftp://hg.mirror.ac.za/sourceforge/s/sa/saga-gis/SAGA - Documentation/HBPL19/hbpl19_05.pdf

Doctor, D. H., & Young, J. a. (2013). An evaluation of Automated Gis Tools for Delineating Karst Sinkholes and Closed Depressions From 1-Meter Lidar-Derived Digital Elevation Data. *Thirteenth Multidisciplinary Conference on Sinkholes and the Engineering and Environmental Impacts of Karst*, (January), 449–458. https://doi.org/PNR61

Galve, J. P., Lucha, P., Castañeda, C., Bonachea, J., & Guerrero, J. (2011). Integrating geomorphological mapping, trenching, InSAR and GPR for the identification and

characterization of sinkholes: A review and application in the mantled evaporite karst of the Ebro Valley (NE Spain). *Geomorphology*, *134*(1–2), 144–156. https://doi.org/10.1016/J.GEOMORPH.2011.01.018

Hijmans, R. J. (2017). raster: Geographic Data Analysis and Modeling.

Hjerdt, K. N., McDonnell, J. J., Seibert, J., & Rodhe, A. (2004). A new topographic index to quantify downslope controls on local drainage. *Water Resources Research*, *40*(5). https://doi.org/10.1029/2004WR003130

Jenness, J. (2001). Topographic position and landforms analysis. *, A Weiss - Poster Conference, ESRI User Diego,*. Retrieved from http://www.jennessent.com/downloads/TPI-poster-TNC_18x22.pdf

Kim, Y. J., & Nam, B. H. (2017). Sinkhole Hazard Mapping Using Frequency Ratio and Logistic Regression Models for Central Florida. In *Geo-Risk 2017* (pp. 246–256). Reston, VA: American Society of Civil Engineers. https://doi.org/10.1061/9780784480717.023

Kiss, R. (n.d.). Determination of drainage network in digital elevation models, utilities and limitations. Retrieved from http://foldtan.hu/sites/default/files/Kiss_Richard_JHG.pdf

Lerche. (2006). Environmental and Economic Risks from Sinkholes in West-Central Florida. In *Environmental Risk Assessment* (pp. 67–79). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-29709-X_5

Lowe, D., Waltham, A. C. (Antony C., & British Cave Research Association. (2002). *A dictionary of karst and caves : a brief guide to the terminology and concepts of cave and karst science*. British Cave Research Association. Retrieved from

http://bcra.org.uk/pub/cs/index.html?j=10

Marschner, I. C. (2011). glm2: Fitting generalized linear models with convergence problems. *The R Journal*, *3*(2), 12–15.

MSDIS LiDAR DEM File Download Tool. (n.d.). Retrieved March 21, 2018, from http://www.msdis.missouri.edu/data/lidar/download/modem.html

Nam, B. (2017). Karst sinkhole detection, characterization, and engineering - A US case study. In *The 2017 KGS Spring Conference, Seoul, Korea,*.

Obu, J., & Podobnikar, T. (2015). Algorithm for karst depression recognition using digital terrain models. Retrieved from https://core.ac.uk/display/41057999

Perez, A. L., Nam, B. H., Alrowaimi, M., Chopra, M., Lee, S. J., & Youn, H. (2017). Experimental Study on Sinkholes: Soil–Groundwater Behaviors Under Varied Hydrogeological Conditions. *Journal of Testing and Evaluation*, *45*(1), 20160166. https://doi.org/10.1520/JTE20160166

Rahimi, M., & Alexander, E. C. (2013). Locating Sinkholes in Lidar Coverage Of A Glacio-Fluvial Karst, Winona County, MN. *Proceedings of the 13th Multidisciplinary Conference on Sinkholes and the Engineering and Environmental Impacts of Karst,* 469–480. https://doi.org/PNR61

RTeam. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Shaban, A., & Darwich, T. (2011). THE ROLE OF SINKHOLES IN GROUNDWATER RECHARGE IN THE HIGH MOUNTAINS OF LEBANON. *Journal of Environmental*

Hydrology, *19*(19). Retrieved from http://www.hydroweb.com

Shamet, R. M., Perez, A., & Nam, B. H. (2017). Sinkhole Risk Evaluation: Detection of Raveled

Soils in Central Florida's Karst Geology Using CPT. In *Geo-Risk 2017* (pp. 257–266).

Reston, VA: American Society of Civil Engineers.

https://doi.org/10.1061/9780784480717.024

Sinkhole Boundaries. (n.d.). Retrieved March 21, 2018, from http://gisdata-

cosmo.opendata.arcgis.com/datasets/ccb2ab8ba2904d6091a15d8c32c94d2e_27

Sørensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness

index: evaluation of different methods based on field observations. *Hydrology and Earth*

System Sciences, *10*, 101–112. Retrieved from www.copernicus.org/EGU/hess/hess/10/101/

Tihansky, A. B. (1999). SINKHOLES, WEST-CENTRAL FLORIDA. *U.S. Geological Survey,*

Tampa, Florida, 121–140.

Tu, T. (2016). Sinkhole Monitoring Using Groundwater Table Data. *Electronic Theses and*

Dissertations. Retrieved from http://stars.library.ucf.edu/etd/5236

US Department of Commerce, N. O. and A. A. (n.d.). What is LIDAR. Retrieved from

https://oceanservice.noaa.gov/facts/lidar.html

Waltham, T., Bell, F. G. (Frederic G., & Culshaw, M. G. (2005). *Sinkholes and subsidence :*

karst and cavernous rocks in engineering and construction. Springer.

Wu, Q., Deng, C., & Chen, Z. (2016a). Automated delineation of karst sinkholes from LiDAR-

derived digital elevation models. *Geomorphology*, *266*, 1–10.

https://doi.org/10.1016/J.GEOMORPH.2016.05.006

Wu, Q., Deng, C., & Chen, Z. (2016b). Automated delineation of karst sinkholes from LiDAR-derived digital elevation models. *Geomorphology*, *266*(May), 1–10. https://doi.org/10.1016/j.geomorph.2016.05.006

Xavier Robin, Natacha Turck, Alexandre Hainard, N. T., & Frédérique Lisacek, J.-C. S. and M. M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77. https://doi.org/10.1186/1471-2105-12-77

Xiao, H., Yong Je Kim, B., Boo Hyun Nam, B., & Dingbao Wang, B. (n.d.). Investigation of the impacts of local-scale hydrogeologic conditions on sinkhole occurrence in East-Central Florida, USA. https://doi.org/10.1007/s12665-016-6086-3

Zhu, J., Taylor, T. P., Currens, J. C., & Crawford, M. M. (n.d.). IMPROVED KARST SINKHOLE MAPPING IN KENTUCKY USING LIDAR TECHNIQUES: A PILOT STUDY IN FLOYDS FORK WATERSHED. https://doi.org/10.4311/2013ES0135