10-13-2005

Creation, Analysis, and Evaluation of Remote Sensing Sinkhole Databases for Pinellas County, Florida

Larry D. Seale Jr.
University of South Florida

Follow this and additional works at: https://scholarcommons.usf.edu/etd
Part of the American Studies Commons

Scholar Commons Citation

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Creation, Analysis, and Evaluation of Remote Sensing Sinkhole Databases
for Pinellas County, Florida

by

Larry D. Seale, Jr.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science Department of Geology College of Arts and Sciences University of South Florida

Major Professor: H. Leonard Vacher, Ph.D.
Robert Brinkmann, Ph.D.
Mark T. Stewart, Ph.D.

Date of Approval:
October 13th, 2005

Keywords: GIS, karst, airborne, laser, mapping, ALSM, light, detection, ranging, LIDAR

© Copyright 2005, Larry Don Seale, Jr.
TABLE OF CONTENTS

LIST OF TABLES iii

LIST OF FIGURES iv

ABSTRACT vi

INTRODUCTION 1

Setting 1
Motivation 3
Background 4
Purpose and organization 5

PART ONE: COUNTY-WIDE SURVEY OF SINKHOLES USING 2000 ALSM DATA 7

Methods 7

Data management 7
ArcGis 8
Delineating sinkholes 8

Results 12

Distribution 12
Size and shape 14

PART TWO: ANALYSIS OF THE ALSM TECHNIQUE FOR MAPPING SINKHOLES IN PINELLAS COUNTY 17

Analysis of results 17

Problems with ALSM analysis 18

ALSM and contemporaneous air photos 18
Robustness of sinkhole delineations 19
ALSM in urban areas 21
ALSM in vegetated areas 22
Multi-path reflections  
Data volume  
Problems with sinkhole database comparisons  
Air photo interpretation  
Intersection of ALSM sinkholes with the 1995 dataset  
Intersection of ALSM sinkholes with the 1926 dataset  
Comparison of size and circularity among datasets  
Difficulties that arise from changes in time and different operators  

PART THREE: NORTHEASTERN PINELLAS COUNTY FOCUS AREA  
Comparisons of databases  
Sinkholes in northeastern Pinellas County  
Intersection of composite-analysis sinkholes with ALSM and air-photo databases  
Statistical reliability of ALSM and air-photo sinkhole databases  
Morphometrics of sinkholes in northeast Pinellas County identified with composite analysis  

SUMMARY AND CONCLUSIONS  
REFERENCES
LIST OF TABLES

Table 1  Location and density of sinkholes in relation to developed and undeveloped land in Pinellas County  12
Table 2  Number and density of sinkholes identified using ALSM in Pinellas County alphabetically by municipality  13
Table 3  Average area and average circularity of sinkholes identified in the ALSM, 1995, and 1926 sinkhole databases  29
Table 4  Number of intersecting features among databases  34
Table 5  Number of intersecting sinkholes among databases  37
Table 6  Spreadsheet calculating accuracy of ALSM sinkhole database versus the composite analysis  41
Table 7  Spreadsheet calculating accuracy of 1995 sinkhole database versus the composite analysis  42
Table 8  Spreadsheet calculating accuracy of 1926 sinkhole database versus the composite analysis  44
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pinellas County location map</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Circularity index ($I_c$) for six shapes</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>The municipalities of Pinellas County, Florida</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>Distribution of sinkholes in Pinellas County</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>Grid and contours resulting from ALSM processing</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>Discrepancy between ALSM and air photo</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>Sketch illustrating the potential underestimation of idealized sinkhole radius</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>Changes in vegetation indicate a possible sinkhole</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>Background signal reduction in urban areas</td>
<td>22</td>
</tr>
<tr>
<td>10</td>
<td>Swimming pool detected by ALSM</td>
<td>23</td>
</tr>
<tr>
<td>11</td>
<td>Individual data points in an undeveloped, vegetated region</td>
<td>24</td>
</tr>
<tr>
<td>12</td>
<td>ALSM fails to detect all apparent sinkholes</td>
<td>24</td>
</tr>
<tr>
<td>13</td>
<td>Wetland area contoured with ALSM data</td>
<td>25</td>
</tr>
<tr>
<td>14</td>
<td>Venn diagram depicting the intersecting area (km$^2$) of sinkholes in the three sinkhole databases for the focus area</td>
<td>33</td>
</tr>
<tr>
<td>15</td>
<td>Idealized representation of the possible classifications resulting from the intersection of the ALSM or air-photo databases with the composite analysis</td>
<td>38</td>
</tr>
<tr>
<td>16</td>
<td>Histogram of the sinkhole area frequency for features in the focus area located by composite analysis</td>
<td>47</td>
</tr>
</tbody>
</table>
Figure 17  Histogram of equivalent diameter frequency for features in the focus are located by composite analysis  48

Figure 18  Histogram of the circularity of sinkholes in the focus area identified by composite analysis  49

Figure 19  Histogram of nearest neighbor frequency for composite-analysis sinkholes  50
Creation, Analysis, and Evaluation of Remote Sensing Sinkhole Databases for Pinellas County, Florida

Larry Don Seale, Jr.

ABSTRACT

A database of likely sinkholes in Pinellas County, Florida, created using airborne laser swath mapping (ALSM a.k.a LIDAR for light detection and ranging), correlates poorly with other databases of likely sinkholes created from modern and historic aerial photographs. Urbanization appears to be the cause of the poor correlation. Buildings obscure much of the ground surface in urban areas, and many man-made depressions can be confused with natural sinkholes. Additionally, the lack of air photos contemporaneous with the ALSM data hinders ALSM analysis in rapidly developing areas.

Selecting a lightly-developed portion of the county for further study reduced the effects of urbanization. Air photos of this focus area, taken two years after the ALSM data were collected, image essentially the same surface as the ALSM data; therefore, ALSM and the air photos can be considered concurrent. While correlations among the two databases in the focus area were better than in the county-wide comparisons, the incongruencies were still numerous and the validity of the databases was unsubstantiated.

An additional database of likely sinkholes in the focus area, created using all available information, represents the most exhaustive search for sinkholes in Pinellas County to date. By assuming it is correct (i.e. it identifies “true sinkholes”), this composite analysis is used to assess the validity of the ALSM database and the air photo
databases. Measuring the ALSM and air photo databases against the composite analysis reveals that, while ALSM outperforms the air photo methods, the ALSM and air photo analyses each fail to recognize true sinkholes more than 50% of the time. However, it also demonstrates that, while flawed, using the databases allows for a better-than-random chance of selecting a site free of sinkholes.
INTRODUCTION

Setting

The Florida Peninsula is part of one of the largest carbonate systems in the Earth’s geologic history (Hine, 1997). With the development of the Gulf of Mexico during the middle Jurassic Period, thick sequences of carbonates, evaporite, and siliciclastic sediments began to accumulate on a crystalline allochthonous remnant from the collision of the North African and North American Plates (Scott, 1991). This relatively stable passive-margin tectonic setting and the higher-than-present sea level from the Cretaceous to the Paleogene produced a broad, shallow-water marine platform now present in the Yucatan Peninsula, the Bahamas, and as far north as North Carolina – including Florida. There was little tilting or disturbance, and flat-lying sedimentary sequences accumulated to thicknesses of as much as 7 km (Randazzo, 1997). Approximately 25 Ma., near the beginning of the Miocene Epoch, siliciclastic sediments began to be transported onto the platform in sufficient quantities to suppress carbonate deposition, and, by the mid-Pliocene Epoch, siliciclastic sediments covered virtually the entire Florida Platform (Scott, 1997). Today, active areas of carbonate sedimentation are restricted to the south-southwestern parts of the Florida Platform and the Bahamas Bank (Hine, 1997).

Pinellas County, Florida, is the most densely populated county in the state. Covering approximately 750 sq. km., the county is located mostly on a small peninsula in west-central Florida bounded by the Gulf of Mexico on the west, Pasco County to the
north, and Old Tampa Bay and Hillsborough County to the east (Figure 1). The carbonate bedrock beneath Pinellas County is overlain by siliciclastic, clay-rich sediments of the Miocene-age Peace River and Arcadia Formations and undifferentiated Pliocene and Pleistocene-age quartz sand (Fig 1). The result is a covered-karst terrane. Maximum elevations are just above 30-m and local relief is generally small. Sinkholes are scattered across the landscape. One of the predominant landforms in Florida, sinkholes pose a significant hazard to property and the environment (Tihansky, 1987).

**Figure 1** Pinellas County location map. Pinellas County (red) is located on the west-central coast of the Florida Peninsula and is bounded by the Gulf of Mexico to the west, Pasco County to the north, and Old Tampa Bay and Hillsborough County to the east. It has a total land area of approximately 750 sq. km. Simplified stratigraphic column for Pinellas County is shown (left).

Most sinkholes in Pinellas County are cover-subsidence sinkholes that form when the unconsolidated surficial sediments are piped downward into an underlying void dissolved in the limestone bedrock. This process produces shallow, circular depressions at the surface. Cover-subsidence sinkholes form slowly; in contrast, cover-collapse sinkholes form dramatically. Cover-collapse sinkholes form when the overlying
sediment collapses suddenly into large voids in the limestone bedrock. This can occur when cohesive sediments form a structural arch above the void dissolved in the bedrock. The structural arch can collapse into the void when environmental factors, such as soil moisture or groundwater levels change, or when the void enlarges to the point that the cohesiveness of the surficial sediments can no longer support the arch. The sediment collapses into the underlying void rapidly and often with little warning (Tihansky 1987).

**Motivation**

A succession of recent, high-profile sinkhole collapses, most notably the collapse of a section of the elevated cross-town expressway in neighboring Hillsborough County, has highlighted the importance of locating sinkholes. While photo-worthy, these singular events are small when compared to the cumulative economic impacts that sinkholes have throughout Florida. Florida law requires insurance companies to pay for home damages caused by sinkhole activity. As a result, many insurance companies are refusing to offer policies in sinkhole-prone areas and many of the companies that do offer policies have increased their premiums to near prohibitive levels, forcing homeowners to rely on state-subsidized insurance policies (Harrington, 2004). In 2003, Citizens Property Insurance, a state-run, high-risk insurance company, paid $6-million in sinkhole claims. They projected that number to rise to $50-million by 2005. To offset the cost, Citizens has raised its premiums by as much as 35% in sinkhole-prone areas (Harrington, 2004). Such policies have led to efforts to create databases of sinkholes. This study is the first to use airborne laser swath mapping (ALSM) to create a complete, county-wide database of sinkholes.
Background

In September, 2003, the Board of County Commissioners, Pinellas County, Florida, and the Department of Environmental Science and Policy (ESP), University of South Florida, Tampa, entered into an agreement to conduct a geographic, remote-sensing assessment of karst features in Pinellas County. One goal of the project was to create a comprehensive database of sinkholes in Pinellas County.

Phase one of this project was completed by ESP graduate student Kelly Wilson. Her work (Wilson, 2004) analyzed historic (1926) and modern (1995) air photographs for visible sinkholes in order to characterize the effects of urbanization on karst landforms. Wilson (2004) found 2,703 sinkholes and possible sinkholes on the 1926 aerial photographs, and she found 900 sinkhole and possible sinkholes on the 1995 aerial photographs. Wilson (2004) concluded that urbanization has dramatically obscured the sinkholes in Pinellas County.

Phase two, the subject of this study, was to independently create a database of likely sinkholes identified with ALSM. ALSM is a relative new remote-sensing technique using an aircraft-mounted laser range finder to measure elevations of the ground surface quickly and accurately (Carter et al., 2001). Depending on the equipment, more than 25,000 measurements per second can be collected along a swath beneath the aircraft. These elevation data are combined with data about the aircraft’s flight path to determine the coordinates of each elevation measurement (Carter et al., 2001).

In the case of Pinellas County, the Geosensing, Engineering, and Mapping (GEM) Research Center, Department of Coastal and Civil Engineering, University of Florida
collected approximately 400-million \(x,y,z\) points. They conducted two separate data gathering flights; the first took place in May 1999 and the second in September 2000.

Post-flight, filtering algorithms developed by the GEM Research Center processed the data to remove elevation measurements returned from buildings and vegetation (Shrestha and Carter, 2000). In the simplest sense, the algorithms work by detecting multiple elevation modes in the data. The lowest elevation mode is assumed to be the ground surface, and higher modes are assumed to be vegetation or buildings. Data within these higher modes are removed from the dataset. The filtered dataset, processed by the GEM Research Center, contains approximately 150-million \(x,y,z\) data points covering Pinellas County. GEM Research Center provided Pinellas County with both the unfiltered and filtered datasets.

The filtered data was provided as 157 space-delimited text (.txt) files. Each .txt file contains data from a square parcel of land, 10,200-ft on a side. The file names reference the 10,000-ft grid of the Florida State Plane coordinate system, west zone, with the name corresponding to the southwestern-most point of the square. For example, the southwest corner of file 380_1270.txt was 380,000-ft east and 1,270,000-ft north in the state plane grid.

**Purpose and organization**

The purpose of this thesis is to compile and analyze a sinkhole database from the ALSM survey of Pinellas County. The following questions will be considered: Does the ALSM technology allow us to assess the effect of urbanization since the time of the air photo surveys analyzed by Wilson (2004)? What does the ALSM-sinkhole database tell us about the distribution, density, and geometry of sinkholes in Pinellas County?
This thesis is presented in three parts. The first part is a report of findings from the county-wide study of sinkholes found from the ALSM survey. It addresses the differences from the findings of Wilson (2004), and discuss the possible implications with respect to the question of the effects of urbanization.

In the second part of the thesis, an argument is made that a detailed look at the data in this study indicates that the database of sinkholes created using ALSM is suspect, and therefore, comparisons such as those in the first part of the thesis are problematic.

In the third part, Sinkholes are characterized in the undeveloped northeastern corner of Pinellas County where a case can be made that – given the low level of urbanization there – the ALSM data can be utilized as one component of a composite approach to create a reliable database of sinkholes.
PART ONE: COUNTY-WIDE SURVEY OF SINKHOLES USING 2000 ALSM DATA

Methods

Data management

A desktop workstation served as a dedicated GIS computer for the project. The computer is a Dell Precision 650 dual-processor Intel Xeon 3.2-MHz computer with 4-GB SDRAM, a 70-GB primary hard-drive, a 110-GB secondary hard-drive, and a 20-in. Sony Trinitron monitor interfaced with a Nvidia Quadro FX 3000 video card with 256-MB video ram. Processing the 150-million data points required using the fastest processor reasonably available. Unfortunately ESRI’s suite of ArcGIS 8.0 software cannot utilize dual processors; however, this left us with a free processor on which to run other applications while processing ALSM data.

Microsoft’s Access software was used to convert the provided .txt files into database (.dbf) files that can be read by ESRI’s ArcGIS. This conversion is a multi-step process that allows one to format the data into correctly sized columns and combine multiple .txt files into a single .dbf file. Database files were constructed so that each contained data for a strip of land which is 10,200-ft wide (including 200-ft overlap) and extended north and south to the county boundaries. The result was nine tables sequentially named 380 through 460. Table 380, for example, is the seamless assembly of the \(x,y,z\) coordinates contained in the files that begin with “380.” If a specific area is
to be analyzed, one can assemble any combination of files into a table to create a seamless data set.

ArcGIS

After converting the .txt files into .dbf format, ESRI’s ArcMap was used to open and convert them into shape (.shp) files. These .shp files containing x,y,z point data are the source data for the grids, contour maps, and ultimately the sinkhole database.

The inverse distance weighting (IDW) method, which assigns an elevation to each grid cell based on the elevation value of nearby points, was used to create nine elevation grids corresponding to the nine point-data tables. As the name implies, the farther the distance a point lies from the center of the grid, the less weight it is given when calculating the elevation of the grid cell (Isaaks and Srivastava, 1989). Based on user-defined parameters, each cell is a 7-ft square and is assigned an elevation based on the twelve nearest points.

The ArcGIS extension Spatial Analyst was used to contour the elevation grid with a 1-ft contour interval (CI). This small contour interval provided a higher level of resolution than the USGS topographic maps of the county which have, at best, 5-ft contour intervals. As the contour interval increases, fewer sinkholes are portrayed on topographic maps (Applegate, 2003). A detailed workflow, from initial data management through GIS analysis is given in a technical report submitted to Pinellas County (Seale et al., 2005).

Delineating sinkholes

A systematic search of the ALSM-based contour map was conducted to locate depressions in order to build a database of sinkholes in Pinellas County. It is recognized
that features referred to as *sinkholes* in this study are, in fact, depressions that are likely sinkholes; however, many landforms such as dunal depressions, oxidized wetlands, and remnant costal features can be mistaken for karst sinkholes when using remote sensing techniques.

The contour lines generated by Spatial Analyst do not have hachured internal lines indicating closed depressions. To overcome this, the contour lines were viewed in conjunction with the corresponding elevation grid. The grid is color coded with color changes indicating positive or negative changes in elevation.

The depression features selected for the database meet a set of criteria established in consultation with my graduate advisors, H. Len Vacher, Robert Brinkmann, and Mark T. Stewart. Foremost, the depression must be discernable using the ALSM data. Features that appear as depressions in aerial photography, but are not recognizable with ALSM, are not included in the database. For example, a change in vegetation may be recognizable in the aerial photograph with no ALSM-based contour lines indicating a depression. Such visible vegetation changes that do not correspond with ALSM-recognized depressions occur most frequently in low-elevation areas with very little relief.

Second, the ALSM-recognized depression must appear to be of natural origin. In relatively undeveloped areas, 2002 aerial photographs were used in conjunction with the ALSM data in 1999 and 2000 to evaluate the depression in terms of circularity, depth, presence of water, and changes in vegetation. In developed areas, the same 2002 aerial photographs were used to consider shape, apparent topographic modifications, and the arrangement of nearby structures to distinguish natural from man-made depressions.
In addition to creating a simple database; the size, shape, and distribution of ALSM-identified sinkholes throughout Pinellas County were calculated using GIS. The XTOOLS extension was used to add area, perimeter, and centroid coordinate attributes to each sinkhole identified in the various databases. To characterize shape, the circularity index was calculated based on the formula

$$\frac{A_m}{\pi \left(2 \frac{A_m}{P_m}\right)^2}$$

where $A_m$ and $P_m$ are measured area and measured perimeter, respectively. Circularity has been used by many authors (Bahtijarevic, 1996 and Denizman, 2003) to represent how a shape departs from a circle. Denizman (2003) uses the formula

$$\frac{A_m}{\pi \left(\frac{P_m}{2\pi}\right)^2}$$

the inverse of the formula above, to calculate the circularity index. Both authors (Bahtijarevic, 1996 and Denizman, 2003) use circularity index for “the ratio between the measured depression area and the area of a circle with the same perimeter.” Both authors also say that “elongate features have values smaller than one whereas convoluted shapes present values greater than one”. This cannot be true. Both formulae calculate, as a ratio, the efficiency for fitting the most area into the smallest perimeter. The most efficient area-to-perimeter arrangement, a circle, has a circularity index of one. While these formulae are a valid characterization of the amount of departure from a circle, neither can return values of both more than unity and less than unity for any range of possible
geometric features. This is why water pipes are circular rather than some other shape; there is no more efficient area-to-perimeter ratio than a circle (Figure 2).

**Figure 2** Circularity index ($I_c$) for six shapes. **A** is a perfect circle with the circularity index equal to one. All other shapes represent less efficient area to perimeter ratios, and their circularity indexes are greater than one. **E** and **F** represent the same depression identified in the ALSM database (**E**) and the 1995 air photo database (**F**).
The size, density, distribution, and circularity of ALSM-identified sinkholes were compared to information compiled by Wilson in her 2004 study of sinkholes located in 1926 and 1995 aerial photographs.

Results

Distribution

The survey of the 2002 ALSM data produced 1,561 sinkholes that appear to be of natural origin in the 749 km$^2$ of dry land in Pinellas County. The depressions have an aggregate area of 5.77 km$^2$, representing about 0.77 percent of the total land area. The sinkhole density is 2.09 depressions per km$^2$. The depressions are distributed nonuniformly across the county.

The areas identified as undeveloped in the study of 1995 air photos (Wilson 2004) have a combined aggregate area of 97 km$^2$. Of the 1,561 ALSM-identified sinkholes, 617 (40%) are located in areas classified as undeveloped by Wilson (2004). Depression density in the undeveloped areas is 6.38 depressions per km$^2$. The remaining 944 sinkholes are found in the remaining 652 km$^2$ consisting of developed land. The depression density in developed areas of the county is 1.45 depressions per km$^2$ (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>Developed Land</th>
<th>Undeveloped Land</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land Area</td>
<td>652 km$^2$</td>
<td>97 km$^2$</td>
<td>749 km$^2$</td>
</tr>
<tr>
<td>Number of ALSM-Identified Sinkholes</td>
<td>944</td>
<td>617</td>
<td>1561</td>
</tr>
<tr>
<td>Depression Density</td>
<td>1.45 per km$^2$</td>
<td>6.38 per km$^2$</td>
<td>5.77 per km$^2$</td>
</tr>
</tbody>
</table>

*Table 1* Location and density of sinkholes in relation to developed and undeveloped land in Pinellas County.
Sinkholes found in undeveloped areas are slightly larger (4,662 m² per sinkhole), on average, than depressions found in developed areas (3,180 m² per sinkhole). By visual inspection, it is clear that the largest number and highest density of ALSM-identified sinkholes are in the northern one-quarter of the county. Within that area, the northeastern corner of the county has the highest density.

The northeastern corner of Pinellas County is not only classified as undeveloped, but it is also the largest contiguous area of unincorporated land in Pinellas County (Figure 3). All told, the unincorporated land in Pinellas County aggregates to 214 km² (Table 2). There are 941 ALSM-located sinkholes in this area. The depression density for unincorporated Pinellas County is 4.4 sinkholes per km². There are 622 ALSM-identified sinkholes within the city limits of various municipalities within Pinellas County (Table 2).

Of the municipalities, Tarpon Springs has the highest sinkhole density with 5.13 sinkholes per km². Oldsmar, with 4.09 sinkholes per km², has the second highest sinkhole density within a city. Both of these communities are located in the northern one-quarter of the county, and approximately two-thirds of Oldsmar was classified by Wilson (2004) as undeveloped. Tarpon Springs and Oldsmar aside, incorporated areas of Pinellas County have a much lower sinkhole density than the surrounding unincorporated land.

Distribution of ALSM-identified sinkholes across the county is also evident on a map showing sinkhole distribution (Figure 4). The distribution of ALSM-identified sinkholes closely resembles that of sinkholes found in the 1995 air photographs. Both
techniques found the most sinkholes in the undeveloped northeast corner of the county (Fig 4).

Size and shape

The ALSM-identified sinkholes have an average area of 3,736 m$^2$ per sinkhole. Their average equivalent diameter is 58.3 m. The average circularity index is 2.25.

<table>
<thead>
<tr>
<th>Municipality Name</th>
<th>Land Area km$^2$</th>
<th>2000 ALSM</th>
<th>Depressions per km$^2$ ALSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clearwater</td>
<td>73.14</td>
<td>79</td>
<td>1.08</td>
</tr>
<tr>
<td>Dunedin</td>
<td>28.75</td>
<td>30</td>
<td>1.04</td>
</tr>
<tr>
<td>Gulfport</td>
<td>6.23</td>
<td>1</td>
<td>0.16</td>
</tr>
<tr>
<td>Kenneth City</td>
<td>2.29</td>
<td>1</td>
<td>0.44</td>
</tr>
<tr>
<td>Largo</td>
<td>39.53</td>
<td>29</td>
<td>0.73</td>
</tr>
<tr>
<td>Oldsmar</td>
<td>21.29</td>
<td>87</td>
<td>4.09</td>
</tr>
<tr>
<td>Pinellas Park</td>
<td>34.87</td>
<td>42</td>
<td>1.20</td>
</tr>
<tr>
<td>Safety Harbor</td>
<td>25.23</td>
<td>36</td>
<td>1.43</td>
</tr>
<tr>
<td>Seminole</td>
<td>5.12</td>
<td>4</td>
<td>0.78</td>
</tr>
<tr>
<td>South Pasadena</td>
<td>3.72</td>
<td>2</td>
<td>0.54</td>
</tr>
<tr>
<td>St. Petersburg</td>
<td>244.04</td>
<td>183</td>
<td>0.75</td>
</tr>
<tr>
<td>Tarpon Springs</td>
<td>24.93</td>
<td>128</td>
<td>5.13</td>
</tr>
<tr>
<td>Other municipalities</td>
<td>9.80</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Sub Total</strong></td>
<td><strong>518.94</strong></td>
<td><strong>622</strong></td>
<td></td>
</tr>
<tr>
<td>Unincorporated Land</td>
<td>214.20</td>
<td>941</td>
<td>4.39</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>733.14</strong></td>
<td><strong>1563</strong></td>
<td></td>
</tr>
</tbody>
</table>

*Table 2* Number and density of sinkholes identified using ALSM in Pinellas County alphabetically by municipality. Incorporated municipalities with no ALSM-identified sinkholes are not listed.
Figure 3 The municipalities of Pinellas County, Florida. ALSM-identified sinkholes in Pinellas County are shown as red polygons.
Figure 4 Distribution of sinkholes in Pinellas County. Sinkholes located with ALSM are shown to the left and sinkholes located in the 1995 aerial photographs by Wilson (2004) are shown to the right.
PART TWO: ANALYSIS OF THE ALSM TECHNIQUE FOR MAPPING SINKHOLES IN PINELLAS COUNTY

Analysis of results

Wilson (2004) and this report use a variety of remote-sensing techniques to identify sinkholes in Pinellas County. Wilson (2004) studied the impacts of urbanization on sinkholes by first locating all of the sinkholes visible on pre-urbanization 1926 aerial photographs and then comparing these findings with sinkholes visible on post-urbanization 1995 aerial photographs. Wilson located 2,703 sinkholes and possible sinkholes on the 1926 aerial photographs, and 900 sinkholes and possible sinkholes on the 1995 aerial photographs.

The hope underlying this study was that the increased resolution of ALSM would allow detection not only of the surviving sinkholes detected in previous studies, but also the more-subtle features that were undetectable on aerial photographs. The result of the search through the ALSM data is 1,561 ALSM-identified sinkholes in Pinellas County.

At first glance, the numbers seem to affirm the hypothesis that ALSM is a superior remote-sensing technique for recognizing and delimitating sinkholes. The numbers seem to imply that ALSM should be the method of choice to find the subtle surface expressions of sinkholes. Indeed, one can construct a good story from the gross statistics and patterns of the datasets. The 1926 dataset contains a large number (2,703) of sinkholes distributed rather uniformly throughout Pinellas County. As urbanization spreads, many of these sinkholes were obscured. Thus, only 900 sinkholes were detected
in 1995, mainly in the few remaining undeveloped areas of the county. Then, ALSM, with its superior ability to resolve more subtle features, detected 1,561 sinkholes in Pinellas County.

However, upon detailed examination of the data behind the summary statistics, this study finds that these summary statistics do not provide an accurate assessment of the ALSM method.

**Problems with ALSM analysis**

There are a few publications detailing the difficulties in using ALSM, but none have dealt with the problems associated with locating discrete features over a broad area. Davenport et al. (2004) studied the temporal consistency of laser altimetry data. Toyra et al. (2003) characterized the errors observed in topographic measurements in deltaic wetland environments.

Close examination of the sinkhole data of this study has revealed a number of problems with remote sensing of sinkholes and making comparisons between remotely sensed sinkhole databases. Although ALSM (a.k.a. LIDAR) has been used in previous studies to locate small numbers of sinkholes in well-defined study areas (Montané, 2002; Carter et al., 2001), a literature search did not find any published accounts of comprehensive searches for previously unrecorded sinkholes in urban environments with ALSM. It is instructive, therefore, to elaborate in some detail on the difficulties encountered when using ALSM to locate sinkholes.

**ALSM and contemporaneous air photos**

ALSM is of little value without corresponding aerial photographs. The air photos allow the operator to interpret the contour lines, but also introduce the problem of photo
interpretation (Figure 5). It is now common practice to take color air photos concurrently with ALSM (Shrestha, pers comm.), but for this study, no such photos existed. Photos taken from 27 to 43 months after the ALSM flights were used during the assessment. In rapidly expanding areas such as Pinellas County, much of the land surface can be modified in two to four years (Figure 6).

Robustness of sinkhole delineations

ALSM under-represents the area of depressions. Let’s assume that ALSM correctly detects the bounds of a depression. With a one-foot contour interval, the uppermost closed-contour line can fall anywhere from the upper rim of the sinkhole to one foot below the rim of the sinkhole. Assume a circular, conical sinkhole with an area of 4000 m$^2$ (radius, 35.68 m), and a depth of 1 m. If the outermost, circular closed-contour line lies in the worst-case position of one foot below the rim of the sinkhole, the radius would be 24.80 meters (Figure 7) and the area would be 1932 m$^2$, or approximately half of the full area. The problem is accentuated in low-relief depressions because the slope of the sinkhole is reduced, allowing the one-foot contour line to lie farther inward of the rim, further reducing the measured area.

Heavy vegetation within sinkholes also tends to cause ALSM to under-represent the area of an interpreted depression. High point density in sparsely vegetated regions surrounding a depression effectively limits the maximum outward extents of the closed-contour lines, while low point density within the more heavily vegetated depressions requires a higher degree of interpolation. The result is that crenulated contour lines
**Figure 5** Grid and contours resulting from ALSM processing. Depressions are shown in lighter colors and contour lines (CI = 1ft) show elevations, but without accompanying aerial photos, these depressions cannot be classified.

**Figure 6** Discrepancy between ALSM and air photo. Contour lines (CI = 1ft) indicate a depression were none is evident in the air photographs taken at a later date. Note what appears to be a bulldozer parked in the center of the closed-contour lines.
delineating depressions do not pass beyond the outward boundary of the depression, but they do recurve inward, cutting deeply towards the center of the depression (Figure 8). The more crenulated a closed-contour is, the larger the under-representation of the sinkhole’s full area.

**Figure 7** Sketch illustrating the potential underestimation of idealized sinkhole radius.

**Figure 8** Changes in vegetation indicate a possible sinkhole. ALSM located the depression, but does not accurately represent the area. Crenulations do not pass beyond the apparent boundary, but they do cut deeply into the depression causing an under representation of the area and a high circularity index.

**ALSM in urban areas**

This study finds that ALSM does not work well in urban areas. Once the high-elevation modes (usually laser returns from rooftops) are filtered from the dataset, too few data points remain to resolve depressions. Making matters worse, the vast majority
(likely > 95% in urban areas) of the laser returns composing the low-elevation-mode are from paved roads and parking lots (Figure 9).

ALSM cannot distinguish natural from man-made depressions. This obvious fact prevents automated searches for depressions in the ALSM data and makes photo interpretation necessary. While swimming pools and hot tubs can be identified and excluded from the database (Figure 10), there are many man-made features that can be mistakenly classified as naturally occurring.

**Figure 9** Background signal reduction in urban areas. In this urban area, the post-filtering data points (red dots) are found primarily on roadways and a baseball field. Laser returns from rooftops and trees have been removed during the filtering process. Virtually no post-filtering laser returns are recorded from the actual bare earth.

**ALSM in vegetated areas**

The ability of ALSM to detect small, low-relief features is tied directly to the density of $x,y,z$ data points. The data points are filtered to remove laser returns from vegetation. In heavily vegetated areas, point density is low and, therefore, the ground surface is poorly resolved (Figure 11). Sinkholes and heavy vegetation tend to occur
together (cypress domes, for instance); therefore, many sinkholes are not located with ALSM.

Figure 10 Swimming pool detected by ALSM. The depression could not be classified as a man-made object without the aerial photograph.

Along the same lines, ALSM’s ability to correctly delimit the boundaries of depressions is affected by point density. There are many cases where topographic lines generated from ALSM data do not correspond well to what is evident in the corresponding aerial photograph, especially for small discrete features such as sinkholes. In many areas with low point density, one can see ALSM indicating a depression, but its true dimensions clearly are not represented; therefore, morphometric calculations of sinkholes delineated by ALSM are incorrect (Figure 12).
Figure 11 Individual data points in an undeveloped, vegetated region. White areas are heavily vegetated and have low point density. Dark areas are sparsely vegetated and have a high point density.

Figure 12 ALSM fails to detect all apparent sinkholes. Changes in vegetation indicate numerous sinkholes, but ALSM identified only a few. Arrows point to changes in vegetation that indicate possible sinkholes, but no depressions are detected with ALSM. Note that ALSM has under-represented the size of the apparent sinkholes that it did detect.

Also in heavily vegetated areas, ALSM identifies depressions where none are evident in air photos. On one hand, these may be the sinkholes with subtle surface expressions that only ALSM can detect, but in dense vegetation there are laser returns from all elevations making identification of the low-elevation mode difficult to detect.
during the filtering process. These depressions that appear in ALSM but are not recognized in air photos may be true depressions, or they may be artifacts of the ALSM technique.

**Multi-path reflections**

Multi-path reflections occur when a laser light-beam from the ALSM transmitter strikes multiple reflective surfaces and thus takes a tortuous path back to the ALSM receiver. This multi-path reflection has a longer travel time than a direct reflection. The result is a data point with an erroneously low elevation. Multi-path reflections are common in urban areas, where reflective materials are common, but standing water can also create multi-path reflections. Some sinkholes hold standing water which can lead to over-estimations of depth. Wetlands holding standing water also tend to return a large number of multi-path reflections, giving the appearance of sinkholes where none exist (Figure 13).

![Figure 13](image_url)  
**Figure 13** Wetland area contoured with ALSM data. Wetlands are among the most difficult to image with ALSM. The ground tends to be uneven and standing water can lead to multi-path reflections. This results in a high density of erroneous contour lines. Sinkholes are not imaged due to background scatter.
Data volume

ALSM will constitute data overload for most large-scale projects. Going from 150-million $x,y,z$ points to a usable contour map was, by far, the most time-consuming step of this project, requiring approximately three months of data processing. Once the contour map was generated, the depressions in the county were quickly located; however, mistakes and/or omissions are typically found whenever the project is reviewed. This indicates that, especially with a very large project, operator error can affect the final results.

Problems with sinkhole database comparisons

It is well established that it is difficult to make comparisons between databases of landscape morphology (Wills and McCrink, 2002). Kastning and Kastning (2003) specifically address the issue of delineating sinkholes:

There is a widespread discrepancy in interpreting sizes, extents, and densities of sinkholes, based merely on geometry: sometimes sinkholes are defined as having an arbitrary radius from their lowest point; often their sizes are defined by the uppermost, closed contours based on a given contour interval on a topographic map; sinkholes may be delimited by contours representing the lowest elevations along their rim, or sinkholes may be interpreted to include all surficial area that drains internally through them; that is, a contributing drainage basin.

Many of these issues exist when comparing the sinkhole databases created by different operators, at different times, and using different methodology.
Air photo interpretation

Photo quality is a major concern. While analyzing the 1926 photos was an ambitious undertaking, the results were clearly limited by photo quality. This limitation is understandable considering that the 1926 photos were taken only 38 years after photographic film was introduced and 23 years after the Wright brothers began making powered flights. The features identified as sinkholes in the 1926 database tend to be very large, poorly defined areas that compare poorly with the discrete features identified in 1995 and by ALSM.

A second problem is the need for interpretation. The operator is called upon to classify features based on best judgment. This problem is exacerbated in urban areas, and especially in the case of recognizing sinkholes, because landscapers often strive to create natural-looking landscapes, in this case, artificial ponds.

Intersection of ALSM sinkholes with the 1995 dataset

Analysis of the 1,561 ALSM identified sinkholes reveals that only 357 intersect with the 900 sinkholes and possible sinkholes located on the 1995 aerial photographs. Of the 900 features that Wilson (2004) located in the 1995 aerial photographs, only 261 are categorized as definite sinkholes. The remaining 639 features are “probable sinkholes.” Comparing the 261 “definite sinkholes” to the Southwest Florida Water Management District (SWFWMD) land-used map reveals that 135 (51%) intersect with areas that SWFWMD classified as reservoirs. Of the 639 “possible sinkholes”, 288 (45%) intersect with SWFWMD-classified reservoirs. Assuming that SWFWMD has correctly classified man-made features (mainly retention ponds) as reservoirs, then Wilson (2004) was more successful (albeit only slightly) identifying natural features in the “possible” category
than in the “definite” category. This highlights the difficulty in distinguishing natural from man-made depressions in air photos.

ALSM sinkholes are approximately half the size of the 1995 features and their circularity is almost twice that of the 1995 features (Table 3). Circularity of features from the air photos is biased towards a value of unity because of the manual digitization. Wilson (2004) drew circles, more or less, when delineating depressions; on the other hand, ALSM contour interpolation was a completely automated process and had no bias for any particular shape (Figure 2 E and F). However, Wilson (2004) did tend to capture the full extent of the depressions as seen in the air photograph, while the ALSM database relied on the uppermost closed contour, which fails to capture the full extent of the depression, as already noted.

Intersection of ALSM sinkholes with the 1926 dataset

Analysis of the 1,561 ALSM-identified sinkholes reveals that only 582 intersect with the 2,702 sinkholes and possible sinkholes located on the 1926 aerial photographs. It is unclear if this difference is due to obfuscation of sinkholes by the spread of urbanization, or if it reflects the difference in technologies.

Comparison of size and circularity among databases

Sinkholes located with ALSM are almost half the area of those in the 1995 sinkhole database and less than a quarter the size of the sinkholes in the 1926 database (Table 3). The circularity index of ALSM sinkholes is larger than the circularity index of sinkholes in either of the air photo databases (Table 3).
<table>
<thead>
<tr>
<th></th>
<th>Average Area m²</th>
<th>Average Circularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALSM</td>
<td>3,736</td>
<td>2.25</td>
</tr>
<tr>
<td>1995</td>
<td>6,482</td>
<td>1.26</td>
</tr>
<tr>
<td>1926</td>
<td>16,450</td>
<td>1.29</td>
</tr>
</tbody>
</table>

Table 3 Average area and average circularity of sinkholes identified in the ALSM, 1995, and 1926 sinkhole databases.

Undoubtedly, the large size of the sinkholes in the 1926 database reflects the technology of the time. Due to the poor quality of the 1926 air photographs, Wilson (2004) could identify only large depressions in many areas. In many cases, these large depressions are coalesced, composite sinkholes (uvalas). These uvalas are characterized by dark soil tones and distinct hydric vegetation (cypress domes). The sinkholes in the 1995 database are significantly smaller than those in the 1926 database. In some cases, the higher-quality photographs allowed Wilson to identify sinkholes as individuals instead of uvalas, but even by 1926, many uvalas were connected by ditches to promote drainage. This, coupled with a lowering of the water table, erased the soil tonal patterns, and the vegetation within the depressions shifted from hydric vegetation to more xeric vegetation.

For reasons discussed in earlier sections of this thesis, the reduction in size between the 1995 and ALSM sinkholes does not indicate further resolution of features as in the 1926-1995 comparison but rather a failure of ALSM to capture the full extent of a sinkhole.
The sinkholes delineated in the 1926 and the 1995 databases have a similar circularity index because the manual delineation tended to create similar, roughly circular shapes. The ALSM-identified sinkholes have a higher circularity index because ALSM creates complex, crenulated shapes which have high circularity values (Figure 2 E and F).

**Difficulties that arise from changes in time and different operators**

Many of the problems that have been discussed have arisen because of either (1) differences over time or (2) methodology problems. Differences over time are consistent with the spread of urbanization and changes in technology, both of which prevent us from making valid comparisons among databases. Methodology problems result in incorrect classification of depressions, inability to recognize some depressions, and mistaking artifacts (e.g., multi-path reflections) in the ALSM data for sinkholes. Some of these problems were self-imposed in the project because of the attempt to create a sinkhole database of Pinellas County solely from ALSM data and independently from Wilson’s (2004) analysis. That is not to say that ALSM is not useful when used under favorable conditions and in conjunction with, instead of independently from, all other available information. Thus, in a positive vein, we are in a position to establish protocols for searching for sinkholes in covered-karst regions with ALSM.

Urbanization is the primary hindrance to identifying sinkholes with remote sensing. It obscures sinkholes, reduces background information necessary to recognize patterns of sinkhole development, changes frequently and rapidly, and creates depressions of uncertain origin.

Failure to capture photographs concurrently with ALSM was a source of confusion. ALSM without accompanying photographs does have applications, but no
broad-ranging, systematic classification of discrete features, such as sinkholes can take place without concurrent photographs.

ALSM can not be relied on to correctly delineate the boundary of depressions, especially where point data density is reduced. Morphometric analysis of sinkholes can not be based on the upper-most closed contour based on ALSM data, and to do so would be a misapplication of ALSM.
PART THREE: NORTHEASTERN PINELLAS COUNTY FOCUS AREA

Comparisons of databases

The northeastern corner of Pinellas County (Figure 14), located east of Lake Tarpon and north of the municipality of Safety Harbor, covers an area of 65.75 km\(^2\). This area is only lightly developed. Working in this focus area will reduce the number of problems associated with both differences in time and the methodology as detailed in Part Two of this thesis. In an undeveloped area, landscape modifications are at a minimum, so photos taken two to four years prior to ALSM are effectively concurrent.

First, comparisons will be made among the ALSM, 1995-photo, and 1926-photo datasets in the focus area. This will allow for comparisons of the databases in an undeveloped region of the county. Second, the focus area will again be analyzed for sinkholes, using all available information to locate sinkholes in a focus area, thus creating a fourth database of sinkholes in the focus area. Sinkholes within this area will be located using best judgment.

Within the focus area, there are 426 sinkholes and possible sinkholes in the 1926-photo sinkhole database; 241 sinkholes and possible sinkholes in the 1995-photo sinkhole database; and 575 sinkholes in the ALSM-identified sinkhole database. By land area, these sinkholes occupy 14.09 km\(^2\), 1.73 km\(^2\), and 2.38 km\(^2\) respectively. The GIS intersections of these features indicate the following with respect to area: 12.81 km\(^2\) (91%) of the 1926 sinkholes occur only in the 1926 sinkhole database; 0.72 km\(^2\) (42%) of
the 1995 sinkholes occur only in the 1995 sinkhole database; and 1.02 km$^2$ (43%) of the
ALSM sinkholes occur only in the ALSM sinkhole database (Figure 15). These numbers
show that identification of sinkholes depends on the technique.

If (1) sinkholes located in 1926 are still present and undisturbed in the
undeveloped focus area, and (2) ALSM is superior to standard aerial photos for locating
sinkholes, then one would expect ALSM to locate a larger percentage of the features
identified in the 1926 photographs. Instead, only 170 (40%) of the 425 focus-area

Figure 14 Venn diagram depicting the intersecting area (km$^2$) of sinkholes in the three sinkhole databases
for the focus area. Circles and intersections are proportional to their respective areas. Numbers inside
intersections are areas in km$^2$.
sinkholes in the 1926 database intersect with sinkholes identified in the ALSM database. There are comparably low intersections among the other databases for the focus area (Table 4). With this dismal finding for intersecting sinkholes in the datasets for the most ideal, undeveloped setting, one has to question the validity of all of the county-wide databases.

<table>
<thead>
<tr>
<th></th>
<th>ALSM</th>
<th>1995</th>
</tr>
</thead>
<tbody>
<tr>
<td>1926</td>
<td>335</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>170</td>
<td>129</td>
</tr>
<tr>
<td>1995</td>
<td>183</td>
<td>168</td>
</tr>
</tbody>
</table>

Table 4 Number of intersecting features among databases. Databases titles are shown in bold. Total number of sinkholes in each database is shown outside of the database titles. Because of the uvala effect, many sinkholes in one database can intersect with a single sinkhole in a second database, therefore, both intersection values are given. The top number in a cell is the number of sinkholes of the column intersecting with sinkholes of the row, and the bottom number in a cell is the number of sinkholes of the row intersecting with sinkholes in the column. For instance, 335 of 575 ALSM-identified sinkholes intersect with 425 sinkholes identified in the 1926 database, and 170 of 425 1926 air-photo sinkholes intersect with the 575 sinkholes identified in the ALSM database.

With respect to the reliability of datasets, there are three possibilities:

1. The aerial photo analysis and the ALSM each detect a different population of true sinkholes with some overlap. In this case, all techniques are valuable in detecting sinkholes.
2. One technique is vastly superior to the others, i.e., one dataset is correct while the others are partially correct. In this case, one method is valid while the others can be disregarded.

3. Each technique is severely flawed. Each detects some features correctly, and incorrectly identifies many other features. In this case, no one method can be relied on to provide reliable data.

In order to determine which of these three possibilities is true for this study, a detailed study of sinkholes in the focus area was completed using all available information in a composite analysis.

**Sinkholes in northeastern Pinellas County**

*Intersection of composite-analysis sinkholes with ALSM and air-photo databases*

The focus area was analyzed using all available data, essentially creating a fourth database of sinkholes for this corner of Pinellas County. Also, there is a chance to check repeatability of the ALSM-identified sinkhole database, and it allows for a determination of the reliability of the ALSM and air photo datasets.

The search of sinkholes in the focus area using the composite analysis located 479 sinkholes. One would expect to find most of the ALSM-identified sinkholes in this composite analysis, since the same areas is searched by the same operator and using the same technology. Furthermore, one would expect to find additional sinkholes not identified by ALSM for the reasons discussed in this report.

Surprisingly, only 358 of 575 (62%) sinkholes from the ALSM-identified database intersect features identified with composite analysis (Table 5). From visual inspection, it appears that most of the 33% of the ALSM-identified sinkholes that were
not located are found in heavily vegetated areas where ALSM indicated a depression but no indications of a sinkhole appear in the air photos. It is difficult to assess if these apparent depressions are actually present or if they are artifacts produced from a reduced background signal and resulted in a low degree of repeatability.

Of the 241 sinkholes in the 1995 sinkhole database, 172 (71%) intersect features identified in the composite analysis (Table 5). While this is a high intersection percentage, especially considering the different operators employing different methods, only half as many (241 vs. 479) sinkholes were located in the 1995 air photos as compared to the composite analysis.

The intersection of sinkholes identified on the 1926 air photos with composite analysis is difficult to assess. As discussed, the poor-quality air-photos made identification of discrete depressions difficult; therefore, the results of that study are not well suited for direct comparisons among datasets derived from more recent studies. It is noteworthy, however, that only 185 (44%) of the 425 sinkholes located on the 1926 air photos intersect with sinkholes located with composite analysis (Table 5). If most of the sinkholes located on the 1926 air photos are uvalas, one would expect a larger intersection percentage, especially considering that an intersection would be counted if any one of the constituent sinkholes is identified in the composite analysis. It appears that many of the sinkholes located on the 1926 air photos are shallow wetlands that are not composed of coalescing sinkholes.

Comparing the databases shows that using either aerial photo analysis or ALSM analysis alone to locate sinkholes produces flawed results; each method detects some features correctly and incorrectly identifies many other features. ALSM detects
depressions where the composite analysis dictates none exist, and standard air photo analysis fails to identify all of the sinkholes in the area. No one method can be relied on to provide reliable data.

<table>
<thead>
<tr>
<th></th>
<th>ALSM 1995</th>
<th>ALSM 1926</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite analysis</td>
<td>358</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>356</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td>329</td>
<td>185</td>
</tr>
</tbody>
</table>

**Table 5** Number of intersecting sinkholes among databases. Data format is identical to that used in Table 4.

*Statistical reliability of the ALSM and air-photo sinkhole databases*

If the composite database is correct, one can calculate the accuracy of the ALSM and air-photo databases by comparing them to the composite analysis. By intersecting the ALSM analysis or one of the air photo analyses with the composite analysis, every point in the focus area can be classified as: (1) true positive (classified as a sinkhole by composite analysis, correctly classified by single-method analysis), (2) false positive (classified as non-sinkhole by composite analysis, incorrectly classified by single-method analysis), (3) true negative (classified as non-sinkhole by composite analysis, correctly classified by single-method analysis), and (4) false negative (classified as sinkhole by composite analysis, incorrectly classified by single-method analysis). The four possibilities are illustrated in Figure 16.
Figure 15 Idealized representation of the possible classifications resulting from the intersection of the ALSM or air-photo databases with the composite analysis. The focus area is shown within the bounding box. The area lying outside the circles is classified as non-sinkhole in both analyses. The sinkhole and non-sinkhole areas in the composite analysis are considered to be correct for the assignment of classifications. Any point within the focus area will be classified as either true positive, true negative, false positive, or false negative.

The areas for each subset in Figure 16 are given in Rows 20:23 and as percent area in the $2 \times 2$ matrix of numbers shown in Block B8:C9 of the spreadsheets in Tables 6, 7, and 8. With this information, one can calculate sensitivity, specificity, positive predictive value, negative predictive value, and efficiency for each of the ALSM and air-photo databases (for background on terminology, see Vacher 2003 and Grimes and Schulz 2002). Sensitivity is the probability that a point is classified as sinkhole in the single-method analysis, given that it intersects a point classified as sinkholes in the composite analysis. Sensitivity is calculated in Tables 6, 7, and 8 by dividing the percent area correctly classified as sinkhole in the single-method analysis by the percent area classified as sinkhole in the composite analysis ($\frac{B_8}{D_8}$). Specificity is the probability that a point is classified as non-sinkhole in the single-method analysis, given that it
intersects a point classified as *non-sinkhole* in the composite-analysis. Specificity is calculated in Tables 6, 7, and 8 by dividing the percent area correctly classified as *non-sinkhole* in the single-method analysis by the percent area classified as *non-sinkhole* in the composite analysis \( \frac{C_9}{D_9} \). Positive predictive value is the probability that a point classified as *sinkhole* in the single-method analysis intersects a point classified as *sinkhole* in the composite analysis; it is the probability that a point classified as *sinkhole* by the single-method analysis is correct (i.e. “true positive”). It is calculated in Tables 6, 7, and 8 by dividing the percent area correctly classified as *sinkhole* in the single-method analysis by the total area classified as *sinkhole* in the single-method analysis \( \frac{B_8}{B_{10}} \).

Negative predictive value is the probability that a point classified as *non-sinkhole* in the single-method analysis intersects a point classified as *non-sinkholes* in the composite analysis; it is the probability that a point classified as *non-sinkhole* by the single-method analysis is correct (i.e. “true negative”). It is calculated in Tables 6, 7, and 8 by dividing the percent area correctly classified as *non-sinkhole* in the single-method analysis by the total area classified as *non-sinkhole* in the single-method analysis \( \frac{C_9}{C_{10}} \). Efficiency is the probability that any point is classified correctly during the analysis; it is the sum of the true positives and true negatives \( \frac{B_8 + C_9}{D_{10}} \).

ALSM analysis has a sensitivity of 43% (Table 6) due to its poor performance in heavily vegetated areas. Often in these areas, no points were classified as *sinkhole* in the ALSM analysis, but sinkholes were recognized during the composite analysis by coincident changes in vegetation type or color.
The positive predictive value indicates that just over half (55%) of the points classified as *sinkhole* in the ALSM analysis lie within sinkholes identified in the composite analysis. This result provides an interesting measure of the repeatability for the original ALSM analysis. The composite analysis reveals that 45% of the points identified as *sinkhole* only months earlier were false positives.

ALSM analysis has a specificity of 98%. This is due, in small part, to the fact that the boundaries of ALSM-identified sinkholes seldom extend beyond the boundaries of the corresponding sinkholes in the composite database (recall that ALSM under-represents the area of sinkholes). More to the point, however, specificity is high because the area of sinkholes in the ALSM database is small relative to the focus area (2.383 km$^2$ and 65.748 km$^2$, respectively).

The high negative predictive value is due to the fact that the area of composite-analysis non-sinkholes is large relative to the focus area. There is a 97% chance that a point classified as *non-sinkhole* in the ALSM analysis intersects a point classified as *non-sinkhole* in the composite analysis; most points are classified as *non-sinkhole*, and there just are not that many sinkholes to begin with.

Efficiency is high (95.7%) because the largest area category, non-sinkhole, is frequently classified correctly (specificity). Picking a point at random, there is a 93.7% chance that one would pick a point correctly classified as *non-sinkhole* (Table 6, D22), and there is a 2.0% chance that one would pick a point correctly classified as *sinkhole* (Table 6, D20).
### Table 6

<table>
<thead>
<tr>
<th></th>
<th>Classified sinkhole</th>
<th>Classified non-sinkhole</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Area of focus area km²</td>
<td>65.748</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Area of composite-analysis sinkholes km²</td>
<td>3.073</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Sinkhole prevalence %</td>
<td>4.674</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Area of ALSM sinkholes km²</td>
<td>2.383</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Intersecting area km²</td>
<td>1.322</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Sinkholes area %</td>
<td>2.011</td>
<td>2.663</td>
</tr>
<tr>
<td>7</td>
<td>Non-sinkholes area %</td>
<td>1.613</td>
<td>93.713</td>
</tr>
<tr>
<td>8</td>
<td>Totals</td>
<td>3.624</td>
<td>96.376</td>
</tr>
<tr>
<td>9</td>
<td>Sensitivity</td>
<td>43.025</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Specificity</td>
<td>98.308</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Predictive value positive</td>
<td>55.496</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Predictive value negative</td>
<td>97.237</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Efficiency</td>
<td>95.724</td>
<td></td>
</tr>
</tbody>
</table>

The 1995-sinkhole analysis has a sensitivity of 32% (Table 7). This is largely because few points were classified as *sinkhole* in the 1995 analysis when compared to the composite analysis. The positive predictive value indicates that just over half (57%) of the points identified as *sinkhole* in the 1995 analysis intersect points classified as *sinkhole* in the composite analysis. The 1995 air-photo analysis has a specificity of 99%. As in the ALSM analysis, specificity is high because the area of sinkholes identified in the 1995 analysis is small relative to the focus area (1.730 km² and 65.748 km², respectively).
The low positive and high negative predictive values in the 1995 analysis, like their counterparts in the ALSM analysis, are due to the relatively few points classified as *sinkhole*, and the relatively many points classified as *non-sinkhole*, respectively. As in the ALSM analysis, efficiency is skewed by the large number *non-sinkhole* points, thus, it remains high despite the low positive predictive value.

<table>
<thead>
<tr>
<th>1995</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Area of focus area km²</td>
<td></td>
<td></td>
<td>65.748</td>
</tr>
<tr>
<td>2</td>
<td>Area of composite analysis sinkholes km²</td>
<td></td>
<td>3.073</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Sinkhole prevalence %</td>
<td></td>
<td>4.674</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Area of 1995 sinkholes km²</td>
<td></td>
<td>1.730</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Intersecting area km²</td>
<td></td>
<td>0.989</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Classified sinkhole</td>
<td></td>
<td>Classified non-sinkhole</td>
<td>Totals</td>
</tr>
<tr>
<td>8</td>
<td>Sinkholes area %</td>
<td>1.505</td>
<td>3.170</td>
<td>4.674</td>
</tr>
<tr>
<td>9</td>
<td>Non-sinkholes area %</td>
<td>1.126</td>
<td>94.200</td>
<td>95.326</td>
</tr>
<tr>
<td>10</td>
<td>Totals</td>
<td>2.631</td>
<td>97.369</td>
<td>100.000</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Sensitivity</td>
<td></td>
<td>32.191</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Specificity</td>
<td></td>
<td>98.819</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Predictive value positive</td>
<td></td>
<td>57.198</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Predictive value negative</td>
<td></td>
<td>96.745</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Efficiency</td>
<td></td>
<td>95.704</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Classified sinkholes that are sinkholes (true positive)</td>
<td></td>
<td>0.989</td>
<td>1.505</td>
</tr>
<tr>
<td>20</td>
<td>Classified sinkholes that are non-sinkholes (false positives)</td>
<td></td>
<td>0.740</td>
<td>1.126</td>
</tr>
<tr>
<td>21</td>
<td>Classified non-sinkholes that are non-sinkholes (true negative)</td>
<td></td>
<td>61.934</td>
<td>94.200</td>
</tr>
<tr>
<td>22</td>
<td>Classified non-sinkholes that are sinkholes (false negatives)</td>
<td></td>
<td>2.084</td>
<td>3.170</td>
</tr>
</tbody>
</table>

| Table 7 | Spreadsheet calculating accuracy of 1995 sinkhole database versus the composite analysis. User inputs are highlighted. |

The 1926 analysis has a sensitivity of 41% (Table 8). This is larger than the 1995 analysis because of the large number of points identified as *sinkhole*; i.e., the many points classified as *sinkhole* in this single-method analysis increase the odds of intersecting
points classified as *sinkhole* in the composite analysis. While the odds of intersection are increased, there are a significant number of incorrectly classified points in the 1926 database. As a result, the positive predictive value is 8%. The many points classified as *sinkhole* also have an effect on the specificity of the 1926 air-photo analysis. While 80% specificity is still high, it is 18% less than the specificity of the other databases. The points identified as *sinkhole* in the 1926 analysis could be better classified as sinkhole-prone areas or areas of high sinkhole density. As in the other analyses, the negative predictive value of 1926 analysis gives a relatively high degree of certainty that points designated as *non-sinkhole* intersect *non-sinkhole* points in the composite analysis.

The low sensitivity and low positive predictive value of the three analyses indicate that no single remote-sensing method can be relied on to locate all *sinkhole* points in the focus area; however, the databases do provide a useful result.

To appreciate the significance of these findings, consider a homebuyer who wishes to purchase a piece of property without a sinkhole (a single-home lot is sufficiently small to be considered a point). Can either the ALSM or one of the air-photo analyses be relied to find a home site with no sinkholes? [In this case, the most important number is the negative predictive value.] Picking a point at random, one would have a 5% chance of selecting a home site intersecting a sinkhole in northeastern Pinellas County (sinkhole prevalence; Cell D3 in Tables 6, 7, and 8). The negative predictive values, 97% for all databases (Cell D16), means that if the point selected by the homebuyer is classified as *non-sinkhole* in one of the databases, there would a 3% chance that it was a false negative. By using one of the databases, the homebuyer would reduce the degree of uncertainty by 40%.
<table>
<thead>
<tr>
<th>Area</th>
<th>1926</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area of focus area km²</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65.748</td>
</tr>
<tr>
<td>Area of composite-analysis sinkholes km²</td>
<td></td>
<td></td>
<td></td>
<td>3.073</td>
<td></td>
</tr>
<tr>
<td>Sinkhole prevalence %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.674</td>
</tr>
<tr>
<td>Area of 1926 sinkholes km²</td>
<td></td>
<td></td>
<td></td>
<td>14.087</td>
<td></td>
</tr>
<tr>
<td>Intersecting area km²</td>
<td></td>
<td></td>
<td></td>
<td>1.267</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classifications</th>
<th></th>
<th>Classified sinkhole</th>
<th>Classified non-sinkhole</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinkholes area %</td>
<td>8</td>
<td>1.927</td>
<td>2.747</td>
<td>4.674</td>
</tr>
<tr>
<td>Non-sinkholes area %</td>
<td>9</td>
<td>19.499</td>
<td>75.826</td>
<td>95.326</td>
</tr>
<tr>
<td>Totals</td>
<td>10</td>
<td>21.426</td>
<td>78.574</td>
<td>100.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Calculations</th>
<th></th>
<th>% Area</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Classified sinkholes that are sinkholes (true positive)</td>
<td>20</td>
<td>1.267</td>
<td>1.927</td>
<td></td>
</tr>
<tr>
<td>Classified sinkholes that are non-sinkholes (false positives)</td>
<td>21</td>
<td>12.820</td>
<td>19.499</td>
<td></td>
</tr>
<tr>
<td>Classified non-sinkholes that are non-sinkholes (true negative)</td>
<td>22</td>
<td>49.854</td>
<td>75.826</td>
<td></td>
</tr>
<tr>
<td>Classified non-sinkholes that are sinkholes (false negatives)</td>
<td>23</td>
<td>1.806</td>
<td>2.747</td>
<td></td>
</tr>
</tbody>
</table>

**Table 8** Spreadsheet calculating accuracy of 1926 sinkhole database versus the composite analysis. User inputs are highlighted.

On the other hand, consider someone selling a home site in northeast Pinellas County. The three remote sensing databases have positive predictive values of 55%, 57%, and 9% respectively. This means that in the best case, more than 40% of the points classified as *sinkhole* in the databases are false positives. While many of false positives could be correctly reclassified with ground-truthing, the stigma of a positive test would remain. The property owner will likely face a significant reduction in property value, and expensive geophysical investigations would be required to challenge the results of the remote sensing analysis.
The caveat to this reasoning is that home sites are not chosen at random. Buildings are preferentially located on uplands. Indeed, zoning regulations require this in many cases. This study does not analyze risk associated with development in upland areas where sinkholes commonly develop and home sites are more frequently located.

*Morphometrics of sinkholes in northeast Pinellas County identified with composite analysis*

The reliability of the sinkhole databases is calculated by assuming that the composite analysis accurately locates sinkholes. That is, it is the standard of “truth” for comparison. Future studies will assess the validity of that standard. The following morphometric statistics for sinkholes identified with the composite analysis are made available so that future researchers can compare their finding to the focus-area study.

The average area of the composite-analysis sinkholes (6,416 m$^2$) is remarkably close to the average area of the sinkholes in Wilson’s (2004) 1995 sinkhole database (6,482 m$^2$). This indicates that high-quality air photos are best when assessing the true extent of sinkholes and affirms that ALSM under-represents the areal extent and poor-quality photos overestimate the areal extent. The histogram of the area of composite-analysis sinkholes shows there is a single mode of sinkhole area and a definite lower limit to the size of sinkholes located with ALSM (Figure. 17).

The average equivalent diameter of ALSM-identified sinkholes in the focus area is 84.86 m (Figure 18).

The histogram of circularity (Figure 19) illustrates that sinkholes delimited by manual digitization have a small range of values. While manual digitization fails to resolve the small-scale perturbations of a sinkhole’s planimetric shape, it is sufficient to
distinguish radial symmetry from bilateral symmetry (Figures 2 A and B). The mean circularity of 1.11 and the small range of values indicate that there is no preferential orientation of development for individual sinkholes in the focus area.

The nearest-neighbor distance was calculated for each sinkhole located in the focus area (Figure 20). Nearest-neighbor calculations for composite analysis sinkholes in the focus area were performed using a nearest-neighbor calculator in ArcGIS.

The nearest-neighbor index is the ratio of the average actual distance ($L_a$) and the expected distance between randomly distributed points ($L_e$), where

$$L_e = \frac{1}{2 \times \sqrt{D}}$$

and

$$D = \frac{n}{A}$$

(also known as the Clark and Evans’ (1954) test). The result can vary from zero, indicating maximum clustering, to 2.149, indicating evenly distributed as widely spaced as possible. A value of 1.00 indicates complete randomness. The nearest neighbor index of the focus area is 1.016, indicating near complete random distribution of sinkholes. This indicates that there are no areas of preferential sinkhole development in the focus area.
### Figure 16

Histogram of sinkhole area frequency for features in the focus area located by composite analysis.

![Histogram of Focus Area Sinkholes](image)

<table>
<thead>
<tr>
<th>Area (sq. m.)</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2000</td>
<td>10</td>
</tr>
<tr>
<td>2001-4000</td>
<td>20</td>
</tr>
<tr>
<td>4001-6000</td>
<td>30</td>
</tr>
<tr>
<td>6001-8000</td>
<td>20</td>
</tr>
<tr>
<td>8001-10000</td>
<td>10</td>
</tr>
<tr>
<td>10001-12000</td>
<td>5</td>
</tr>
<tr>
<td>12001-14000</td>
<td>5</td>
</tr>
<tr>
<td>14001-16000</td>
<td>5</td>
</tr>
<tr>
<td>16001-18000</td>
<td>5</td>
</tr>
<tr>
<td>18001-20000</td>
<td>5</td>
</tr>
<tr>
<td>20001-22000</td>
<td>5</td>
</tr>
<tr>
<td>22001-24000</td>
<td>5</td>
</tr>
<tr>
<td>24001-26000</td>
<td>5</td>
</tr>
<tr>
<td>26001-28000</td>
<td>5</td>
</tr>
<tr>
<td>28001-30000</td>
<td>5</td>
</tr>
</tbody>
</table>

- Mean: 6416.08
- Standard Error: 214.13
- Median: 5268.76
- Mode: #N/A
- Standard Deviation: 4686.37
- Sample Variance: 21962087.99
- Kurtosis: 3.15
- Skewness: 1.56
- Range: 28951.09
- Minimum: 46.57
- Maximum: 28997.67
- Sum: 3073303.61
- Count: 479.00
Figure 17 Histogram of equivalent diameter frequency for features in the focus area located by composite analysis.
**Figure 18** Histogram of the circularity of sinkholes in the focus area identified by composite analysis.
Figure 19 Histogram of nearest neighbor frequency for composite-analysis sinkholes.
SUMMARY AND CONCLUSIONS

At the beginning of this project the hope was to identify the sinkholes that existed in 1926 prior to widespread urbanization. Surviving sinkholes would be located in the 1995 aerial photographs along with new sinkholes that had formed in the intervening 69 years to characterize the effects of urbanization of karst landscapes. ALSM, a supposedly superior technique, would then be used to detect known sinkholes, the subtle depressions associated with reactivation of sinkholes covered by urbanization, and previously undetected sinkholes.

Urbanization presents a significant challenge to locating sinkholes with remote sensing. After filtering the data points in urban areas, too few ALSM data points remain to locate subtle, redeveloping sinkholes. Known sinkholes in urban areas often show signs of modification and are difficult to distinguish from man-made retention ponds designed to resemble natural features. There are many urban areas where ALSM data do not correspond to what is evident in air photos taken two to four years after the ALSM flights. Comparisons among the sinkhole databases revealed that the county-wide databases are severely flawed.

To lessen the effect of urbanization, comparisons among the databases were made in a lightly developed focus area. A fourth database of sinkholes in the focus area was created using a composite analysis. Comparisons between the composite analysis and other databases reveal that the ALSM and air photo databases are incomplete and
inaccurate to varying degrees. The poor quality of the 1926 air photos made interpretation difficult. Instead of locating discrete sinkholes, large areas of coalescing sinkholes were identified. The 1995 air photos were much more suited to locating discrete features as opposed to the uvalas identified in the 1926 air photos; however, many sinkholes in the focus area were not detected in the 1995 air photos, and many sinkholes included in the 1995 sinkhole database are suspect. The lack of concurrent air photos to accompany the ALSM data has also been a source of error. In heavily vegetated areas many apparent sinkholes are not detected by ALSM. It is also demonstrated that in heavily vegetated areas, those sinkholes that are detected are poorly delineated, resulting in an underestimation of area.

This work indicates that ALSM is not the superior technique for locating sinkholes, but it is an important component in a composite analysis. ALSM did locate some sinkholes that are not apparent in air photos alone.

While it is shown that a single remote sensing technique has limited success locating sinkholes, the composite analysis has resulted in the most comprehensive database of sinkholes in the focus area created thus far. We have been able to calculate the density of sinkholes in the focus area and the accuracy of the ALSM and air-photo analyses. Sensitivity calculations indicate that any single remote-sensing technique will correctly classify less than half of the points that lie within sinkholes. However, the negative predictive values show that by using a single remote-sensing technique, one can select non-sinkhole points with a better-than-random level of certainty.
Future considerations for sinkhole research should include comprehensive
ground-truthing and geophysical exploration of the sinkholes classified in the focus area.
Only then can the validity of the composite analysis be tested.
REFERENCES


