GIS-Based Sinkhole Survey
Summary of Methodology

Abstract :

With the ongoing development of high-precision remote sensing technologies and the increasingly widespread implementation of such technologies as means of data collection, new possibilities are opening up in the field of GIS. A prime example of this is LiDAR — a relatively new remote sensing technology which (among other things) allows the ground surface to be identified beneath a structure, forest canopy or other intervening surface. As LiDAR has come to be developed and implemented more widely in recent years, the availability of high-resolution bare-earth elevation data has come to present a unique opportunity for sinkhole detection in karst landscapes.

Overview :

The basis of this sinkhole survey was a 5-foot resolution bare-earth DEM (Digital Elevation Model) derived from LiDAR data collected in 2010/2011 by Ayres Associates & Quantum Spatial. The data was collected for the Wisconsin Region Orthography Consortium (WROC) using federal CDBG funding. This data was prepared by the WI DNR and is distributed freely through WisconsinView, which was the original source for the DEM used in this work. The DEM was processed and analyzed with free and open-source software — with special reliance on GRASS GIS and Quantum GIS for data processing tools. Also used were Python, GDAL and R.

This document outlines the methodologies used to produce the sink points dataset which will be provided to the CSP by Legion GIS, LLC, and which will be vetted with the aim of producing a first-of-its-kind comprehensive, countywide dataset of likely and potential sinkholes in Crawford County, WI.

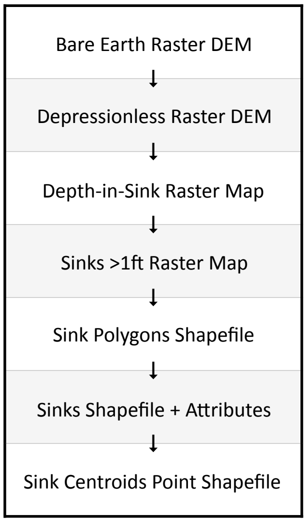
Methodology :

› IDENTIFYING SINKS IN GRASS GIS :

The raster DEM was loaded into GRASS GIS, and the computational region was set to match the raster extents and cell size [1]. Hydrological analysis was then performed using `r.fill.dir` to isolate closed depressions and fill them to their point of overflow [2]. This process generated a depressionless or 'filled' DEM raster product.

The GRASS raster calculator (`r.mapcalc`), was then used to subtract the original DEM values from the filled DEM values to produce a "depth-in-sink" raster map [3].

Upon examination of the data, it was deemed necessary to apply a depth threshold to limit the number of resulting points due to the fact that the analysis identified some $\sim 3.72 \times 10^{6}$ discrete sinks. To this end, the raster calculator was used again to reclassify the depth-in-sink map to produce a raster map of all cells having a sink depth of at least (≥) 1 foot [4]. The resulting sink map was converted into vector polygon features using the *r.to.vect* module [5]. During the process, each feature was assigned a unique



numeric identifier (‘`cat`’). The resulting dataset contained 10,639 features, a reduction of about 97.2% from the original count.

The polygons corresponding to sink features were isolated with `v.extract` [6]. This produced a vector layer containing polygon outlines of each sink feature. Sink depth values were attached to each feature by calculating raster zonal statistics for the depth-in-sink raster map within the bounds of the feature’s geometry [7]. Maximum values, reflecting the greatest sink depth within a given polygon, were attached to the attribute table.

Elevation values were attached to the vector features in the same way. Zonal statistics were calculated for the original DEM [8], and maximum values were attached to each feature as an estimated elevation value.

At this point, the vector layer was exported to an ESRI Shapefile using `v.out.ogr`, and Quantum GIS was used to do further processing.

› GENERATING CENTROID POINTS / MANAGING ATTRIBUTES IN QUANTUM GIS :

The sink polygons shapefile was opened using QGIS, and centroids were calculated for each polygon using the OGR geoprocessing toolbox; this process produced a point for each sink polygon and attached it to the existing attributes.

A this point, some additional information was attached to the attribute table. For instance, latitude (‘`lat`’) and longitude (‘`long`’) fields were created and populated with WGS84 coordinates for each point, to facilitate location of the sinks in the field with GPS.

GRASS GIS Commands Used to Generate Sink Points :

[1]:
```
g.region -p -a raster=CR_DEM@mapset align=CR_DEM@mapset
```

[2]:
```
r.fill.dir input=CR_DEM@mapset output=CR_fill direction=CR_flowdir
```

[3]:
```
r.mapcalc “CR_sinkdepth = CR_fill@mapset - CR_DEM@mapset”
```

[4]:
```
r.mapcalc “CR_sinks = (CR_sinkdepth@mapset >= 1)”
```

[5]:
```
r.to.vect -s input=CR_sinks@mapset output=CR_sinks_vect type=area
```

[6]:
```
v.extract input=CR_sinks_vect@PERMANENT type=area where=value is 1
output=CR_sinks_poly
```

[7]:
```
v.rast.stats map=CR_sinks_poly@PERMANENT raster=CR_sinkdepth@PERMANENT
column_prefix=dp method=maximum percentile=100
```

[8]:
```
v.rast.stats map=CR_sinks_poly@mapset raster=CR_DEM@mapset column_prefix=ev
method=maximum percentile=100
```

› FILTERING THE DATA :

Because a substantial number of points were observed to fall within ditches along roadsides, it was determined that the number of points could be significantly reduced by discarding those which fell within the right-of-way of a public road. A TIGER/Line shapefile containing road geometries was buffered to 30 feet using the geoprocessing toolbox in QGIS. A spatial query was used to identify all points lying within the buffered zone.* These points were marked by assigning them a value of '1' in the newly-created '`row_flag`' field.

In spite of the relatively poor quality of the TIGER/Line geometries, preliminary assessment suggests that it was an effective method, with no sinkholes found to occur among the discarded points (though an exhaustive and methodical assessment was not carried out). All in all, 1,488 features (about 14% of all sink points) were found to fall within the ROW and were de-prioritized for further examination.

** It is worth noting that this method will only capture areas within the right-of-way of a town road or other roadway which is not more than 60 feet wide. In order to capture the full rights-of-way of highways or other roadways wider than 60 feet, those features would have to be first extracted from the dataset and then buffered to the appropriate distance.*

A similar method was used to flag all points lying within flood zones. It has been noted by geologists that sinkholes are less likely to be found in valley bottoms for a variety of reasons, including the regional bedrock topology and water table depth. Furthermore, visual examination of the data showed that many of the erroneous sink points had been placed in wetlands, marshes, flood zones and riparian areas, apparently due to hummocky surface topology and fluvial landforms. Based on consultation with Kelvin Rodolfo (Prof. Emeritus with the University of Illinois–Chicago), we determined it would be prudent for the sake of expedience and resource limitations to ignore these points.

A NFHL shapefile containing 100-year flood zone geometries was acquired through FEMA. A spatial query was used to find all points lying within flood zones (zones A & AE) and these were assigned a value of 1 in the '`nfhl_flag`' field. 1,796 features (nearly 17% of all points) were identified as lying within the floodplain and were de-prioritized for further examination.

By eliminating almost a third of all features (3,284 or ~31%) from our initial examination, we were able to focus our efforts while, in theory, discarding only points which would not be expected to mark genuine sinkholes. Ideally, these points would be examined after all others have been checked, to account for the possibility that sinkholes might in fact occur along roadsides or in floodplains. In practice, it is not expected that this task will be given priority or that resources will be present to carry it out. Nonetheless, the number of sinkholes discarded along with these points is expected to be very small.

› LIMITATIONS :

There are a number of limitations imposed by the nature of the methodology used here. First and foremost, this effort is bound by the limitations of the elevation data upon which it was based. This includes limitations imposed by the margin of error involved, as well as the resolution of the data itself.

In the official LiDAR ground control survey report, the vertical margin of error was estimated to fall within ±0.74284 feet of true ground level at a 95% confidence level. This indicates that some small number of cells in the original DEM may contain elevation values nearly 9 inches deviant from ground truth. In some situations, this margin of error would certainly be significant enough to affect the processing outcomes of tasks such as hydrological analysis.

The LiDAR DEM used for this sinkhole survey was of 5-foot resolution, meaning that topological features smaller than 5 feet cannot be reliably resolved. Despite the unprecedented high resolution of this data, it cannot be assumed to be a perfect representation of the ground surface. Inaccuracies can be expected to result from the fact that hydrological analysis was performed over a gridded model of surface topology, and not over the surface itself; thus, we can only suppose the resulting data to represent an approximation (albeit a very good approximation) of real-life landforms and hydrology.

The ground control report, as well as other metadata which accompanied the original dataset, will be delivered to the CSP.

point assessments were performed preceding from the assumption that the sink points were placed effectively

for one, it must be noted that each cell contains an average of the elevation within that cell; and since furthermore, there are limits imposed by the resolution of the data… a 5ft DEM, despite being the best data available, can not be assumed to be a perfect representation of the ground surface.

For these reasons, as well as due to limitations of the algorithm itself, some sinks may not have been detected by the process

Depth values should be understood to be a GIS-based estimation, and may be inaccurate for a number of reasons;

› TEST OF ACCURACY :

A qualified test of accuracy of the final data would be in order
in order to be an acceptable test of accuracy, the data should be compared to a reference dataset of higher accuracy, namely the ground surface through field surveys and so-called ‘ground-truthing’.

NOTES : For those interested in replicating this process, it should be understood that these are very memory-intensive calculations (the sink-filling process in particular). The raster used in this study contains nearly 1×10^9 cells. While these analyses were performed on a machine with 8GB of RAM*, this is often seen as insufficient for intensive GIS processing, and based on the author’s experience 16GB of RAM would have been highly desirable.

Some issues were encountered in getting GRASS GIS to successfully execute the sink-filling process; these were determined to be most likely memory-related, though the ultimate cause has not been identified.

* Windows 10, 4-Core x64, 3.5GHz, 8GB RAM (6.95GB usable)… r.fill.dir completed in approximately 10.6 hours… found 475054 unresolved areas
*
*

… larger, basin-like sinkholes did not show up so well
… quite a few smaller sinkholes were missed; the majority by virtue of the 1-ft cutoff, but at least a few were missed entirely by fault of the algorithm or of the data (which is as of yet unclear)