Estimating Cave Entrance Density Using GIS Analysis: Quintana Roo, MX

Slide 1 (Intro): This project began as an exercise for a class in GIS analytical methods, but became an insightful study in how to manage uncertainty in data sets. The central question is 'how does cave entrance (or sinkhole) density vary across a karst landscape?' with engineering applications in mind for sinkhole risk assessment. The question is complicated by the fact that survey data of known entrances is dependent on our ability to access them, meaning that there is strong exploration bias inherent this type of data. I've used several geostatistical methods in GIS to quantify exploration bias with the assumption that remaining density differences reflect actual spatial differences due to varying geological properties of the bedrock.

Slide 2 (Site Location): The site location is a stretch of coastline along the eastern side of the Yucatan Peninsula (between Cancun and Tulum) in the state of Quintana Roo, MX. What we commonly think of as the Yucatan peninsula is actually a part of a much larger carbonate platform that extends into the Gulf of Mexico. What is currently above modern sea-level was deposited in shallow seas between 23 and 2 million years ago, forming a carbonate sequence over 1 km thick.

Slide 3 (Geology): The oldest rocks on the platform are Miocene in age, with younger formations occurring in bands along the coastline representing different depositional environments. This region is widely known for its extensive cave networks, including 4 of the 5 longest underwater caves in the world, which continue to be explored.

Slide 4 (Underwater caves): Since exploration by cave diving began in the 1980's, over 1,200km of underwater passage has been mapped. Most of these caves are relatively shallow, between 10-20m below the surface. Passages form complex mazes and often contain speleothem growth from past sea-level lows.

Slide 5 (Dry caves): Exploration of dry caves has largely been overlooked until recently. Currently 150km of dry passage has been mapped, and passage development appears to converge inland. These caves are extremely shallow, often within only 1-2 meters of the surface. One of the striking features of caves in Quintana Roo is frequent ceiling collapse forming entrances.

Slide 6 (Cenotes): When sinkholes open into the water table they are called cenotes. Many cenotes are developed commercially for tourism and contribute significantly to the local economy. In a rapidly developing area like Quintana Roo, it is especially important to understand where and how sinkholes and cenotes form because these offer direct pathways into the aquifer, on which the population depends for their water supply. Very thin cave roofs mean there is high risk for collapse when constructing large buildings, roads, or railroads across the peninsula.

Slide 7 (Survey Data): This map shows nearly 2,000 surveyed cave entrance locations from the Association for Mexican Cave Studies and the Quintana Roo Speleological Survey. The study area is a 140km by 70km rectangle parallel to the coastline that has been clipped to the political boundary of Quintana Roo. A geological map is overlain showing Miocene age limestone towards the interior and Pleistocene limestone along the coastline.

Slide 8 (Cave Entrance Distribution): The most striking thing we notice about this data set is that it is clustered very much along the coastline. This might reflect differences in geology, or it could be a result of exploration bias along roads that also follow the coastline. If there were no data bias and the platform were entirely geologically uniform, then entrance distribution would be random. One way to quantify data clustering is to compare an observed data set to a similar but random set. On the right I've plotted the same number of entrances in a random simulation.

Slide 9 (Ripley's K-function): We can use a geospatial statistic in GIS that analyzes differences in point density at varying scales. This is sometimes called Ripley's K fuction, and it compares expected density assuming random distribution to observed density within intervals around an arbitrarily chosen point. If a data set is not significantly different from the random set, then it will plot as a line with a slope of 1. If it plots above, than point density would be higher than expected at projected distances, and the data is said to be clustered. If the opposite is the case, then the data is dispersed. The results from nine iterations around different arbitrary points show very clearly that the survey data is significantly clustered, but we need to consider context to determine why this is the case.

Slide 10 (Accessibility Bias): GIS analysis provides a method called cost-mapping to quantify the difficulty of moving across terrain. This method uses a raster grid with cells assigned weighted values according to different factors that increase the 'cost' of moving across them. In Quintana Roo, some of these factors include land ownership, topography, vegetation, and the presence of roads. For this analysis landownership was considered negligible because large stretches are owned in common and most landowners are agreeable to exploration since cenotes may be a source of income.

Slide 11 (Weighted Overlay): We divided slope into three classes by intervals of 10% grade. Vegetation, for this purpose, was classified as either not present or agriculture or jungle. Distance from roads was considered in km intervals for up to 3 km around roads. All factors were assigned equal weight. The resulting map aligns overwhelmingly with roads, indicating that proximity to roads is a dominant influence in accessibility. We can look at differences in cave density along roads in more detail using multi-ring buffers:

Slide 12 (1km buffer): Within a 1km distance of roads, cave entrance density was assessed to be about 0.81 ent/ sq km. This area contained about 70% of surveyed entrances.

Slide 13 (2km buffer): Between 1 and 2 km from roads, known cave entrance density dropped to about half of that at 0.4 ent/sq km. The area contained about 20% of the survey data set.

Slide 14 (3km buffer): Between 2 and 3km from roads, entrance density dropped by about half again to 0.18 ent/km. This area contained about 6% of the surveyed entrances.

Slide 15 (Density Graph): When we plot these results, and results from smaller intervals, the trend is a negative power function with a strong R value. In geography this is known as a distance decay curve, and is conventionally used to assess the distances that people are willing to travel to retail stores. These results strongly suggest that data clustering is driven by access to roads.

Slide 16 (Geological Units): While it is not always practical or possible to eliminate data bias, we may still use it for purposes of comparison if we understand how it is skewed. Reconsidering the original question asking how entrance density might vary across karst, we can compare similarly explored zones in both older and younger units of rock. When entrance density is calculated only within 1 km of roads, we see that within the younger, Pleistocene limestone along the coast entrance density is almost 4 times that of what is estimated inland. This suggests that true differences exist due to geological variation, but also indicates that more exploration is needed inland to ensure that data sets are as similar as possible.

Slide 17 (Remote Sensing): A much faster way to assess entrance density may utilize remote sensing. Even in GoogleEarth it is possible to identify an manually count potential entrances. While this is simple to do, it is also time consuming and depends a great deal on the quality of imagery in an area. High-resolution LIDAR is another option, and this has the advantage of being able to rapidly process a DEM in GIS to identify sinks (although too much objectivity introduces problems of noise and scale). Ultimately, even remote sensing data must be field verified to evaluate its effectiveness.

Slide 18 (Results): In summary, the analysis of the existing cave entrance survey data gives us clearer knowledge of how accessibility bias affects density estimates. We also have a way to quantify that bias, and a technique to use similarly biased data for meaningful comparisons. In this study are, accessibility is most strongly influenced by the distance from a cave to a road. Comparing areas of assumed similar exploration, it appears that geologic differences do in fact influence entrance density.

Slide 19 (Conclusions): Returning to the big picture, we may be able to effectively assess sinkhole risk in unexplored areas if we thoroughly inventory small areas and are certain that cave-formation processes remains similar within a geologic unit. Testing limitations of survey data or remote sensing allows us to identify areas that require more detailed description. While a carbonate platform may seem to have relatively uniform geologic properties, observed clustering suggests that significant variations exist across geologic units, and possibly within them as well. Detailed geologic mapping would greatly help to evaluate sinkhole risk below the regional scale. The techniques discussed here provide a ‘first-pass’ look at data, but must be tailored to fit the features of a particular landscape and considered alongside environmental influences that may also contribute to the non-random spatial distributions of geological features.

Hidden slide 20 (Histogram): I recently added road data traced in Google Earth and compiled a Near Table in GIS that calculates the distance from each cave entrance in the survey set to the nearest road. The results show an increase in the number of caves within a 1 km buffer, from about 70 to nearly 80% of the survey data set.