Predictive modeling of cave entrance locations: relationships between surface and subsurface morphology

William Blitch 1, Adia R. Sovie 2, and Benjamin W. Tobin 1*

1Kentucky Geological Survey, University of Kentucky, 310 Columbia Ave., Lexington, KY 40506, USA
2Department of Fisheries and Wildlife, Michigan State University East Lansing, MI, USA

Abstract: Cave entrances directly connect the surface and subsurface geomorphology in karst landscapes. Understanding the spatial distribution of these features can help identify areas on the landscape that are critical to flow in the karst groundwater system. Sinkholes and springs are major locations of inflow and outflow from the groundwater system, respectively, however not all sinkholes and springs are equally connected to the main conduit system. Predicting where on the landscape zones of high connectivity exist is a challenge because cave entrances are difficult to detect and imperfectly documented. Wildlife research has a similar issue of understanding the complexities of where a given species is likely to exist on a landscape given incomplete information and presence-only data. Species distribution models can address some of these issues to create accurate predictions of species or event occurrence across the landscape. Here we apply a species distribution model, MaxEnt, to predict cave entrance locations in three geomorphic regions of Kentucky. We built the models with cave locations from the Kentucky Speleological Survey database and landscape predictor variables, including distance from sinkholes, distance from springs, distance from faults, elevation, lithology, slope, and aspect. All three regional models predict cave locations well with the most important variables for predicting cave entrance locations consistent between models. Throughout all three models, sinkholes and springs had the largest influence on the likelihood of cave entrance presence. This unique use of species distribution modeling techniques shows that they are potentially valuable tools to understand spatial patterns of other landscape features that are either ephemeral or difficult to identify using standard techniques.

Keywords: MaxEnt, karst, caves


INTRODUCTION

Globally, over one billion people live or rely on karst landscapes for water (Goldscheider et al., 2020). However, these landscapes contain geological hazards including sinkhole collapse, flooding, and are highly susceptible to water contamination (De Waele et al., 2011, Goldscheider, 2019). In epigenic karst landscapes, the surface topography and groundwater are bound together in a dynamic relationship; changes in one affect the other. Caves, current or remnant groundwater flow paths, are among the most direct and accessible connection points between the surface and subsurface environments (Kambesis, 2007). Researchers can enter these environments and, when active water is present, directly observe karst forming processes (e.g., Palmer, 1990; Kelly et al., 2009; Schwartz et al., 2013; Heimel & Tobin, 2022). In hydrologically active cave systems, the cave entrance provides an opportunity to easily observe how contaminants, sediment, and solutes move between surface water and groundwater (White, 2007). Conversely, hydrologically inactive caves may give insight into past hydrogeologic conditions and the geomorphic evolution of karst regions (Plan et al., 2009). Understanding the regional distribution of cave entrances can put local information from a single cave into a regional context and provide insight into the complex relationship between surface topography and groundwater that shapes karst landscapes (Kambesis, 2007).

While caves are some of the most well-known features of karst landscapes, other features, such as sinking streams, sinkholes, and springs, are often the easiest to identify (De Waele et al., 2011). However, these surface features are not always directly tied to larger cave systems that act as main conduits through
the subsurface. Cave entrances are often challenging to locate since they can be remote, ephemeral, and frequently obscure (Weishampel et al., 2010). This means that only a small fraction of all cave entrances can be documented (Kambesis, 2007). To expand our ability to identify cave systems, knowing where cave entrances are likely to exist is a critical next step. To address this next step, non-traditional tools have potential to provide greater insight into cave entrance distribution.

One such set of tools, Species Distribution Models (SDM), are an increasingly common tool used to understand spatial patterns of natural resources, predict phenomena, and understand landscape-level ecological processes (Valavi et al., 2022). Even though these models were popularized for modeling wildlife distributions, SDMs can be applied to a variety of environmental features. Mattaquin et al. (2019) used them to predict the location of paleolithic archaeological sites associated with caves. Other researchers have utilized these models for understanding transferability of soil movement models (Hjort et al., 2014), landslide risk (Kerekes et al., 2018), assessing the fractal nature of topography (Tate, 1998), and hydrometeorological hazard risks (Dwivedi et al., 2022). In each of these cases, the models were able to use the distribution of related surface features to predict the locations of features of interest. The ability of SDMs to create predictions with presence only data and associated surface features makes them an especially relevant tool for modeling cave distribution in epigenic karst.

Among the various approaches to species distribution modeling, the Maximum Entropy model (MaxEnt) is especially powerful. MaxEnt is a regression-based machine-learning method that has high predictive capabilities and minimal overfitting when compared to other SDMs (Valavi et al., 2022). MaxEnt estimates the relationships between occurrences from a spatial distribution (i.e., response-in our case cave entrance locations), and a set of landscape features (i.e., predictors) distributed across the same space. MaxEnt is a probabilistic method used to estimate the distribution of a target variable by maximizing its entropy subject to a set of constraints (Phillips et al., 2004). The constraints are defined by the empirical averages of the variables or features that are being studied. This enables us to model complex, non-linear relationships between variables and estimate their joint probabilities.

MaxEnt’s performance means that, unlike some other SDM approaches, it can produce high quality spatial predictions with a limited number of landscape variables. Often, to maximize predictive capability, an SDM must use many potentially related predictor variables (Elith & Leathwick, 2009). This makes it difficult to examine causal relationships between environmental predictors and responses. By choosing specific and relevant environmental predictors with MaxEnt it is possible to create a high performing predictive surface that does not obscure the specific relationships between environmental factors and responses (Elith & Leathwick, 2009).

Here, we demonstrate the use of MaxEnt models to investigate the factors predicting cave entrances in three distinct karst regions in Kentucky. Karst in Kentucky spans a range of lithology and geomorphology that allow us to assess the viability of an SDM approach to understanding cave entrance locations on the landscape. Across all karst regions of Kentucky, sinkholes (including sinking streams) are inlets into the active cave systems and springs are the modern outlets. Therefore, we hypothesized that proximity to surface karst features, specifically sinkholes and springs, will be the primary driver of cave entrance probability. Due to the variety of karst in Kentucky, we utilized three unique karst regions as a proof-of-concept regarding the use of MaxEnt to predict cave locations.

METHODS

Study site description

We focused on three unique geomorphic settings in Kentucky: the Western Pennyroyal (WPR), the Red River Gorge (RRG), and the Inner Bluegrass (IBG) (Fig. 1). While the IBG and WPR study areas have similar topography, the IBG has Ordovician-aged bedrock compared to Mississippian-age rock in the WPR. The IBG study area has the highest number of documented springs, while the WKY has more sinkholes. The RRG study area is unique among the three areas because of its steep elevation gradient and relatively few documented karst features.

Fig. 1. Location of the three study sites: the western Pennyroyal (WPR), Inner Bluegrass (IBG), and Red River Gorge (RRG) areas. Gray-scale on map represents amount of karstification across the state with white representing highest level of karstification and black representing no karst present.
Western Pennyroyal (WPR)
The WRP study area (Breckinridge, Meade, and Hardin Counties) is part of the sinkhole plain landscape of central Kentucky (White et al., 1970). It is dominated by massive Mississippian limestones to the east and capped by younger Mississippian Sandstones towards the western end of the study area. There are two large normal fault systems that trend northeast to southwest and intersect with the large Rough Creek fault system outside of the study area. When compared to the other study areas, the WPR had the least variability in slope and elevation and the highest sinkhole density (Fig. 2, Table 1). More than half of the study area has a high potential for karst development (Paylor & Currens, 2001). Cave development in this region includes 367 documented entrances and multiple cave systems over 15 km in surveyed length (Kentucky Speleological Survey, 2022).

Red River Gorge (RRG)
The RRG study area (Lee, Powell, and Estill Counties) lies along the Pottsville escarpment, the western boundary of the Cumberland Plateau (Martin, 2013). This region is known for the steep elevational gradient along the escarpment, with local relief of over 300 m. While the eastern portion of the study area is dominated by non-soluble Pennsylvanian lithologies, the western portion of the study area is underlain by massive Mississippian limestone units up to 33 m thick. Along the escarpment the Pennsylvanian strata cap the uplands while the Mississippian layers form the lower slopes and valleys, giving rise to steep slopes and cliff faces. The area also contains the Irvine-Paint Creek Fault system which strikes roughly northeast to east through the study area. RRG has the highest elevation and slope among study sites and the greatest variability between these factors (Fig. 2). This area is considered to have minimal karst potential due to the region having a small number of sinkholes and springs. It does, however, have the largest number of cave entrances, with 381 documented entrances, and mapped caves with over 4 km of passage.

Inner Bluegrass
The IBG region (Fayette, Woodford, Jessamine, and Franklin Counties) is dominated by thin, interbedded Ordovician limestones and shales and lies along the axis of the Cincinnati Arch (Simpson & Florea, 2009). The Lexington and Kentucky River fault systems run through the southeast corner of the study area, trending northeast-southwest (Fig. 2). Due to high sinkhole density, the area has high karst potential. The IBG has 228 entrances and mapped caves up to 5 km in length.

Fig. 2. Slope and elevation and location of features (springs, sinkholes, and faults) for the three focus areas of this project: Western Pennyroyal, Red River Gorge, and Inner Bluegrass.
MaxEnt modeling

To predict cave entrance location, we utilize multiple data sets of karst, hydrologic, geologic and topographic data as predictors. In epigenic karst, sinkholes and springs act as input and output locations for subsurface systems (Shofner et al., 2001). There are a variety of other regionally specific factors that combine to control where surface and subsurface morphologies intersect on the landscape (Florea, 2005, Panno et al., 2013). To capture these specific conditions across Kentucky’s karst regions we also chose surface flow paths, surface flow direction, faults, geology, slope, and elevation as relevant predictors.

We took cave entrance locations from the Kentucky Speleological Society database (KSS). KSS maintains a database of known cave features reported by citizen scientists, researchers, and the public. Except for surface flow paths, which we sourced from the USGS National Hydrography dataset, we gathered data from the Kentucky Geological Survey online database (Kentucky Geological Survey, 2022). We rasterized and resampled data to 2.8 m² resolution to match the elevation data. We treated lithology as a categorical variable, based on geologic unit names. Using the Euclidean distance tool in ArcPro we calculated distance from the cave entrance to faults, sinkholes, flow paths, and spring features. We created the flow direction datasets using the Flow Direction tool in ArcPro. All geospatial data manipulation was conducted in ArcPro (ESRI, 2020).

We fit the MaxEnt model using established code in the package Dismo (Hijmans et al., 2017) in program R (R Core Team, 2022). We used the area under the receiver operating characteristic curve (AUC) to test model fit. AUC is a summary statistic that represents the probability that the model will rank a randomly chosen positive instance higher than a randomly chosen negative instance. AUC values range from 0.5 to 1.0, where 1.0 represents a perfect differentiation between presences and absences (Philips et al., 2004, 2017). An AUC value of 0.5 would correspond to completely random prediction (Philips et al., 2017). We used permutation importance and response curves to understand the relationship between each environmental variable and cave entrance locations. To calculate permutation importance the values of each predictor are randomly permuted (i.e., shuffled) in the test data while keeping the target variable unchanged. The MaxEnt model is then retrained on this data, and the decrease in the model’s performance (in our case AUC) is measured compared to the original data. The larger the decrease in AUC, the more important the variable. Response curves show the relationship between each predictor value and the probability value of the response. Finally, we used the full model to create a predictive surface of the probability of cave locations across our study areas.

RESULTS

Western Pennyroyal

Sinkholes and slope were the most important contributing predictors to predicting cave entrances in the WPR (Figs 3, 4). Faults, springs, and lithology were all similarly important to the model. Cave entrance occurrence was positively related to increasing slope and negatively related to increasing distance from springs and sinkholes (Figs 3, 4). Probability of cave entrance occurrence decreased with increasing distance from faults, however faults had relatively low permutation importance. The model accurately predicted cave occurrence in the WPR with an AUC of 0.87.

Inner Bluegrass

Sinkholes and springs were the most important predictors in the IBG model, with sinkholes accounting for over half of the overall permutation importance (Fig. 5). Cave entrance occurrence was positively related to increasing slope and negatively related to increasing distance from springs and sinkholes (Figs 3, 4). Probability of cave entrance occurrence decreased with increasing distance from faults, however faults had relatively low permutation importance. The model accurately predicted cave occurrence in the WPR with an AUC of 0.87.

Red River Gorge

Sinkholes and springs were the most important predictors in the RRG model, with a density of 0.2.

Table 1. Summary of basic information of each study area.

<table>
<thead>
<tr>
<th>Study area</th>
<th>Elevation (m asl)</th>
<th>Slope (degrees)</th>
<th>Caves</th>
<th>Springs</th>
<th>Sinkholes</th>
<th>Density (% area)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western Pennyroyal</td>
<td>202.5</td>
<td>34.4</td>
<td>19.2</td>
<td>18.6</td>
<td>366</td>
<td>127</td>
</tr>
<tr>
<td>Inner Bluegrass</td>
<td>262.2</td>
<td>35.6</td>
<td>18.1</td>
<td>20.5</td>
<td>226</td>
<td>469</td>
</tr>
<tr>
<td>Red River Gorge</td>
<td>282.2</td>
<td>59.7</td>
<td>43</td>
<td>32.6</td>
<td>381</td>
<td>16</td>
</tr>
</tbody>
</table>

Fig. 3. Predictive surface of cave entrance locations generated from the MaxEnt model of the Western Pennyroyal study area, Kentucky (USA), with variable permutation importance.
DISCUSSION AND CONCLUSIONS

Across our study areas, proximity to sinkholes and springs are the most important landscape features for predicting cave entrance locations (Figs 3, 5, and 6). This relationship was consistent between study areas despite regional differences in lithology, topography, and structure (Fig. 4). This relationship fits with the current understanding of epigenic karst and conduit network formation (An et al., 2019). The

Red River Gorge

Sinkholes were the most important predictor in the RRG model with springs, faults and elevation also important predictors (Fig. 6). Cave entrance occurrence was negatively related to increasing distance from faults, sinkholes and springs (Fig. 4). The relationship between slope and cave entrance occurrence was non-linear with a peak in probability at roughly 70° (Fig. 4). The model accurately predicted cave occurrence with an AUC of 0.91.

Fig. 5. Predictive surface of cave entrance locations generated from the MaxEnt model of the Inner Bluegrass study area, Kentucky (USA), with variable permutation importance.

Fig. 6. Predictive surface of cave entrance locations generated from the MaxEnt model of the Red River Gorge study area, Kentucky (USA), with variable permutation importance.
other important predictors were proximity to faults, elevation, and slope. The relationship between these predictors and cave entrance location were not as consistent as proximity to karst features, likely due to regional differences in topography and structure. Both the WPR and RRG models had faults among their most important predictor variables. Doctor et al. (2008) demonstrated a connection between geologic structure and the formation of karst networks, likely explaining the importance of faults in our models. Unlike the WPR and RRG study areas, the majority of faults in the IBG study area are concentrated along the eastern and southeastern margin: this may explain why faulting was not an important predictor.

The regional abundance of sinkholes and springs affected the specific permutation importance in each model. In general, if a predictor is widely distributed and highly abundant, it will be less useful for the model to use in discriminating between where presences and absences are likely to occur. This is important to consider when evaluating the results for a specific model in isolation. However, the relationships shown by the response curves for proximity to sinkholes, springs, and faults as well as the effect of topography remain consistent between models despite differences in abundance. Additionally, not all sinkholes and springs were shown to be highly probably cave entrances, suggesting that only some may be closely connected to larger conduits. This is best highlighted in the IBG region where only some sinkholes have high probability of a cave entrance (comparing Figures 3 and 5).

Our model suggests that in karst environments, cave entrances are most likely to occur where water is directly input into the subsurface or where groundwater is output at the surface. This enhances previous attempts to identify cave entrances with Lidar data (Weishampel et al., 2010) and gives potential insight into the relationship between surface and subsurface environments in karst. This relationship between surface and subsurface environments in karst terrains is dynamic and controlled by the interdependent behavior of topography and groundwater systems (Palmer, 2003). The consistent response curves (Fig. 3) for the most important predictors in our model suggest that the relationships between surface and subsurface in epigenic karst are largely similar across regions.

Differences in regional geomorphology did not seem to have impacts on the overall predictive performance of each regional model. All models performed well at identifying cave entrance locations in each training set with AUC values greater than 0.80. Our results suggest that MaxEnt can be a helpful tool to identify new areas with high cave entrance potential and help prioritize survey effort and conservation. Nevertheless, there are some limitations to the application of our study to understanding karst resources. In particular, MaxEnt assumes there are relationships between current features on the landscape surface and the feature of interest (Phillips et al., 2017), in this case, cave entrances. As a result, if the surface morphology is not related to the development of the underlying conduits, then the model may not predict caves well. For example, caves of hypogene origin that do not have any direct connection to surface morphology (Klimchouk, 2015) may not be predictable with this methodology. Additionally, MaxEnt models are highly correlative and as a result may not provide causal relationships between variables unless an underlying relationship is known a priori (Phillips et al., 2017).

Despite these challenges, our work highlights the high potential to predict cave entrances across landscapes when karst processes link surface and subsurface features. The similarities between regional models suggest that the most important predictors identified here are likely useful factors in predicting karst features in other regions. We suggest that subsequent modeling efforts focus on the relationship between karst features and higher resolution topographic data in predicting cave locations. These models show that the complex interaction between groundwater systems, conduit systems, and surface topography is essential to understanding the development of karst landscapes through time. It is also clear that the application of SDM methods can be useful in examining geomorphologic relationships and predicting discrete geologic features. These relationships are critical for understanding contaminant movement between surface and subsurface, Jeminez-Sanchez et al. (2008) suggested that proximity of livestock farms to caves plays a role in the amount of contamination present in a cave. Through predicting cave entrance locations, land managers may be better able to buffer against contamination in these systems, as is done with surface riparian buffers (Cole et al., 2020). Additionally, these models have potential to assist in identifying likely cave locations in any landscape where surface morphology data is available but humans may not be able to easily access, including other planetary bodies.

ACKNOWLEDGEMENTS

Cave location data was provided via a data request from the Kentucky Speleological Survey.

Authorship statement: WB conducted the geospatial analysis and wrote the manuscript. ARS conducted the statistical analyses, wrote, and reviewed the manuscript. BWT provided project supervision, wrote, and reviewed the manuscript.

REFERENCES


Predicting cave entrance location


