arina Fisher-Phelps

GIST 5304: Final Report

My dissertation is about building better species distribution models for Southeast Asian bats. Species distribution models can be used to predict the range of a species based on environmental variables even if the data do not cover the entire range of the species. Before I can build distribution models I need to prepare maps that feature bat localities and landscape features that might impact the location of the bat localities. Thus, my CountryData geodatabase (GDB) is used to store spatial data that describes bat localities in Southeast Asia and the spatial relationships of those localities to different environmental features. The GDB was created for mapping bat localities so it is the principle FC that defines the theme of the GDB.

For this project, I focused on collating data for Peninsular Malaysia so all of the features classes are contained in a single feature data set (FDS) named PeninsularMalaysia. The projected coordinate system (PCS) for the FDS is Kertau Rectified Skewed Orthomorphic (RSO) Malaya Meters. This PCS has the underlying geographic coordinate system Kertau, which has an inverse flattening ratio of 300.8017. An RSO projection is an oblique cylindrical projection that preserves local shapes, local angles, and distances along the meridian. The Kertau RSO Malaya PCS was developed for Malaysia and is the national map projection for the country, hence its use for this GDB.

***Feature Classes***

The PeninsularMalaysia FDS contains nine feature classes that describe bat localities, major cities, city name annotation, karst locations, land cover type, protected area locations, roads, state boundaries, and changes in maximum temperature (Fig. 1). All data (from all sources) were originally in the geographic coordinate system World Geodetic System of 1984 so all FC’s were reprojected into the Kertau RSO Malaya PCS.

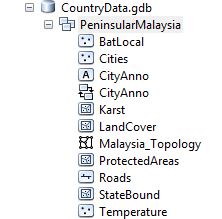


Figure 1. ArcGIS catalog diagram of CountryData GDB and PeninsularMalaysia FDS with nine FC’s and a topology.

The BatLocal FC contains point features that describe the locations where bats have been found. Bat locality data was pulled from the Global Biodiversity Information Network (GBIF), which is an open source database for species occurrence data. The locality data only existed as latitude and longitude coordinates so point features were created from the data and given attributes. The new FC was reprojected and imported into the FDS.

The Cities FC contains point features with the locations of major cities in Peninsular Malaysia. City data was pulled from a shapefile from the ESRI website. The original data had points for the entire globe so the data were clipped to Peninsular Malaysia and the attributes were simplified for use in this GDB. The updated FC was reprojected and imported into the FDS.

The CityAnno FC is an annotation features that has the names of the cities from the Cites FC. Name labels were created in a map document and then converted to GDB annotation.

The Karst FC contains polygon features that describe the possible location of caves. These data come from the World Map of Carbone Rock Outcrops v3.0 from the School Environment at the University of Auckland New Zealand. The East and South East Asia shapefile was downloaded from the website and clipped to Peninsular Malaysia. The clipped FC was reprojected and imported into the FDS.

The LandCover FS stores land cover types as polygons. These data were downloaded as raster data from GlobCover, which is an open source database managed by the European Space Agency. The raster data were converted into polygon data by land cover type. Multiple land cover polygons were merged into the simple types of cropland, mixed vegetation, forest, grassland, urban, water, and other. The new FC was reprojected and imported into the FDS.

The ProtectedAreas FC contains polygons features that store the locations of protected areas across Peninsular Malaysia. These data were downloaded as a global shapefile from ProtectedPlanet.net, the interface for the World Database on Protected Areas (WDPA), which is a joint project of the IUCN and UNEP. The shapefile was clipped to Peninsular Malaysia and the attributes were simplified for use in this GDB. The updated FC was reprojected and imported into the FDS.

The Roads FC stores road data as line features. The source of the data is OpenStreetMap, which is an open source map of global roads. The software BBBike was used to download the OpenStreetMap data as a shapefile (original road data are stored as a .pbf). A bounding box was used to select the approximate desired location from the BBBike website and then the shapefile was clipped to Peninsular Malaysia. The new FC was reprojected and imported into the FDS.

The StateBound FC stores the administrative state boundaries as polygon features. The data were downloaded as a shapefile from the DIVA-GIS website. The shapefile was for the entire extent of Malaysia so it was clipped to Peninsular Malaysia. The clipped FC was reprojected and imported into the FDS.

The Temperature FC stores weather station locations and changes in maximum temperature as point features. Maximum temperature data were downloaded from the NOAA climate portal and point features were created using latitude and longitude coordinates for the weather stations. Changes in maximum temperature were calculated using the R statistical software. Attributes were created and populated for the new FC then it was reprojected and imported into the FDS.

***Domains***

Seven domains were created for this GDB. All of the domains were created to facilitate faster data entry and decrease entry error. Having coded value domains creates a pull-down menu so that mistakes in data entry are nearly eliminated. Coded value domains also decrease data storage by decreasing the number of bytes needed for each attribute entry. The first is CityType, which describes the type of city in the Cities FC. CityType coded value domain was created to differentiate between province/state capitals, national capitals, and other cities. Cities should fall into one of these categories but there is also an unknown option if type is not recorded.

The FoodType coded value domain gives categories for the type of food eaten by a bat species. Almost all bats in Southeast Asia are insectivores, frugivores, or nectivores. Other is an option since some bats are more opportunistic and feed on multiple sources. Unknown is also a choice since the ecology of some bat species is not currently well known enough for placement into a category.

The GenusNames domain codes genus names for describing the species found at a locality. Genus names were given a domain so that there would be a dropdown menu for data entry. Misspellings are common in genus names and a domain eliminates such errors. Currently, there are only four genera in the GDB so there are only four genus categories; Kerivoula, Scotophilus, Pteropus, and Hipposideros.

The IUCNLand coded value domain categorizes the ProtectedArea FC. The International Union for Conservation of Nature (IUCN) has strict categories for protected areas that denote the amount of protection afforded to an area at an international scale. The seven categories with decreasing protection are; Strict Nature Reserve, Wilderness Area, National Park, Natural Monument, Habitat/Species Management Area, Protected Landscape/Seascape, and Protected Area with Sustainable Use. There is also a Not Reported option in the domain since some protected areas in Malaysia are not given an IUCN category.

The IUCNstatus coded value domain categorizes species by their vulnerability to extinction. There are eight categories with decreasing vulnerability; Extinct, Extinct in the Wild, Critically Endangered, Endangered, Vulnerable, Near Threatened, Least Concern, Data Deficient, and Not Evaluated.

The ProtectType coded value domain categorizes the ProtectedArea FC into different area types. Area type just describes what the area is being used for and not its level of protection like the IUCNLand domain. There are nine described types and also an other and unknown category. The nine described types are bird sanctuary, forest reserve, marine park, national park, heritage park, nature park, nature reserve, wildlife reserve, and wildlife sanctuary.

The RoostType coded value domain categorizes bats into where the different species roost. Many bats are highly dependent on their roost type and it is often an important indicator or distribution. There are three major roost types trees, caves, and buildings. The other category is for species that are opportunistic in roost selection. Unknown is also a choice since the ecology of some bat species is not currently well known enough for placement into a category.

***Subtypes***

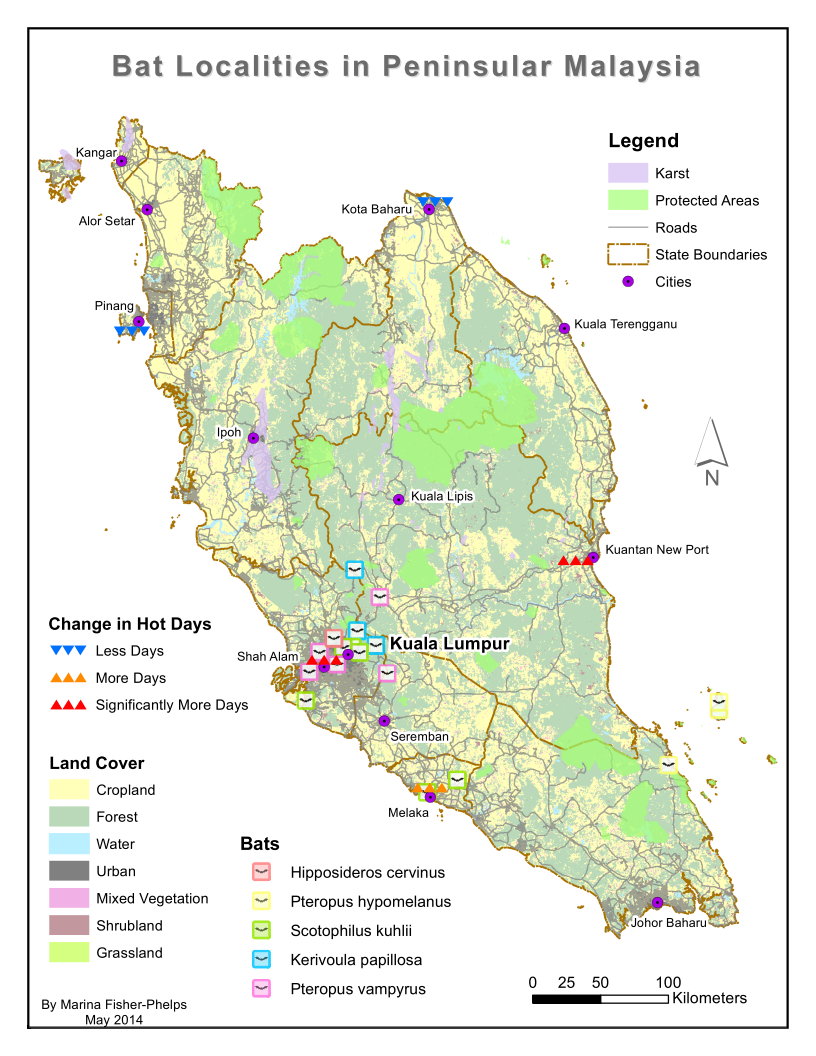
There are two subtypes in the PeninsularMalaysia FDS. The first is the LandType subtype that is part of the LandCover FC. This subtype categorizes land cover into eight different types; cropland, mixed vegetation, forest, shrubland, grassland, urban, water, and other. I created this subtypes because I wanted the different types to automatically have different symbology when I loaded it into a map document. In addition, thinking ahead to other possible attributes, subtypes would make it possible to assign different default values to the land cover types. The second domain is SpeciesNames, which applies to the BatLocal FC. This domain divides bat localities into five different species (papillosa, kuhlii, vampyrus, hypomelanus, and cervinus) and an unknown category. This subtype was created so that different species could have different default values for the domains that apply to the BatLocal FC. The GenusNames, FoodType, RoostType, and IUCNSpecies domains will not change for a given species thus having a subtype with default values causes data entry to be faster and occur with fewer errors.

***Topology***

The PeninsularMalaysia FDS has the Malaysia topology that has a single rule. The rule is that bat locality points must be properly inside the StateBound FC. This rule is important because the focus of the study is on Peninsular Malaysia so bat localities should be within the boundaries of the country. Bat localities outside these boundaries will not be covered by the environmental features within Peninsular Malaysia, which would cause errors in my species distribution models.

*Conclusion*

The schema of the CountryData GDB and the PeninsularMalaysia FDS organizes the data such that data storage space is minimized and data entry errors are minimized. Future FDS will have data for other Southeast Asian countries. Incorporating the data into a geodatabase also makes it possible to visualize the relationship between bat localities and environmental features through maps (Fig.2). Having the data organized in such a way will make it efficient to run species distribution models for bats of Southeast Asia.

Figure 2. Map for Peninsular Malaysia showing the bat localities and all environmental features.