

Modeling species habitat with environmental predictor variables

In this example, we demonstrate how to predict the occurrence of an animal by correlating observations of it with environmental conditions, building a statistical model that expresses that correlation, and creating maps of predicted occurrence by applying the model to maps of the environmental parameters. This type of analysis is sometimes called habitat modeling, species distribution modeling, species occurrence modeling, or range mapping.

Watch the original webinar (now somewhat obsolete)

This demonstration was originally given as a one hour webinar to the [EBM Tools Network](#) on 15 October 2008 using a beta version of MGET 0.7. The webinar consisted of a short PowerPoint presentation followed by a longer demonstration showing how to perform the analysis in ArcGIS using geoprocessing tools in MGET. You can download the PowerPoint slides in either [Microsoft PowerPoint format](#) or [Adobe PDF format](#). You can also watch a [recording of the webinar](#) (77 MB, MP4 format). We highly recommend this because the PowerPoint does not explain any details of the analysis. All of the details are captured in the ArcGIS demonstration.

Since that time we have made many improvements to MGET that simplify the processing required to complete the original example. The updated demonstration downloadable below is designed for a more recent version of MGET and only roughly matches the webinar. Please keep that in mind when watching the webinar.

At the end of the webinar, the EBM Tools Network encountered a technical problem with its conference calling system and had to terminate the audio right when I called for questions. Thus you will hear the audio cut out while the video keeps going. If you have any questions about the end of the presentation, please contact me (jason.roberts@duke.edu).

Download the original demonstration (MGET 0.7)

If you prefer to try the original demonstration presented in the webinar, please see [these instructions](#). You must use MGET 0.7 and ArcGIS 9.2, 9.3, or 9.3.1. Later versions of MGET and ArcGIS are not supported.

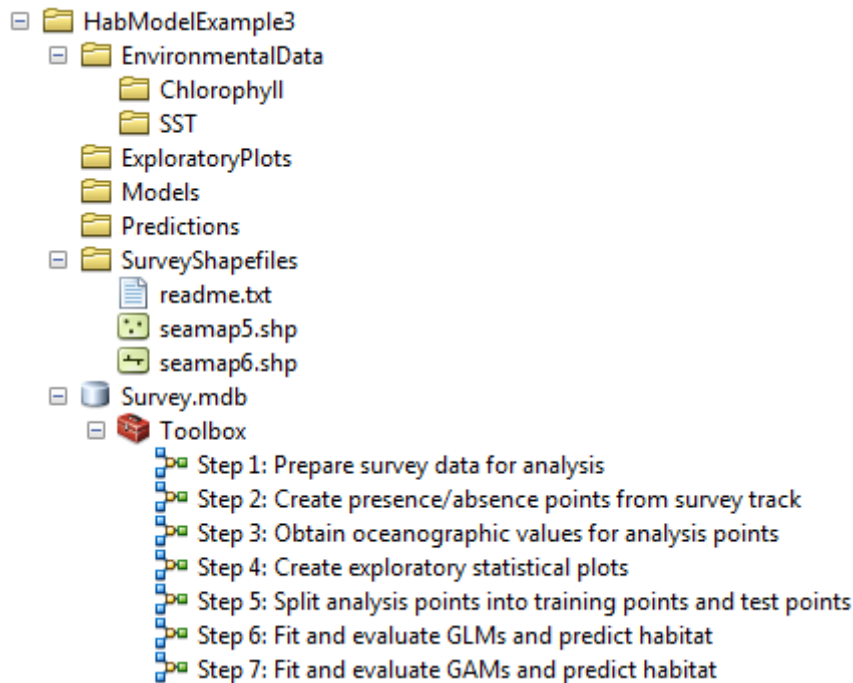
Download the updated demonstration (MGET 0.8a49 and later)

To run the updated demo, the following software must be installed. See the [MGET Installation Instructions](#) for step-by-step installation instructions. Please contact jason.roberts@duke.edu if you need help.

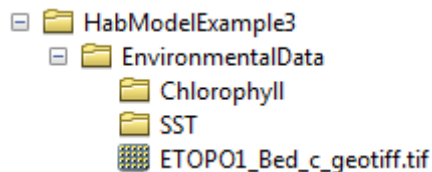
- ArcGIS 9.3, 9.3.1, 10, 10.1, or 10.2
- ArcGIS Spatial Analyst extension
- Python 2.5 (if ArcGIS 9.3 and 9.3.1), 2.6 (if ArcGIS 10), or 2.7 (if ArcGIS 10.1 or 10.2)
- Python pywin32 package (a.k.a. Python for Windows Extensions) installed for your version of Python
- [MGET 0.8a49](#) for your version of Python
 - [MGET 0.8a50](#) and [MGET 0.8a51](#) should work
 - MGET 0.8a52 and later will **NOT** work. MGET 0.8a52 introduced a number of improvements to the modeling tools that broke the workflows in this example. I need to fix them and release a new example. Please contact me (jason.roberts@duke.edu) for more information.
- R 2.12.0 or later
 - Earlier versions might work.

Running the demo

1. Download the file [HabModelExample3.zip](#) and save it to a folder of your choice.
2. Unzip the file. A single folder called **!HabModelExample3** should be created.
3. Start ArcCatalog and go to that **!HabModelExample3** folder. It should look like this:



4. The example uses the **NOAA ETOPO1 bathymetry**, which I did not include in the !HabModelExample3.zip due to the bathymetry file's large size. Download the **ETOP01 Bedrock cell-referenced georeferenced tiff** and store it in the !HabModelExample3\EnvironmentalData folder. Unzip the file into that folder. Now in ArcCatalog you should see a raster called **ETOP01_Bed_c_geotiff.tif**:



5. You are now ready to run the geoprocessing models in the toolbox in the **Survey.mdb** geodatabase. In ArcCatalog, right-click on the first model and select **Edit**. When the model diagram comes up, open the **Model** menu and select **Run entire model**. This first model import the shapefiles **seamap5.shp** and **seamap6.shp** (downloaded from OBIS-SEAMAP) into **Survey.mdb**.
6. After the first model completes, run each of the remaining models in sequence using the same method. Right-click, select **Edit**, open the **Model** menu and select **Run entire model**. Annotations within each model describe what it is doing.

Important changes from the original demonstration

- I changed how and when oceanographic rasters are created. In the original demo, Step 1 was to convert HDF files to rasters, so that they could be sampled in Step 3. MGET 0.8 now includes tools for directly obtaining values of popular oceanographic products at point features by communicating with servers over the Internet. These tools are used in the new Step 3 model. The Sample Rasters Listed In Fields tool no longer must be used, which greatly simplifies the sampling procedure.
- I changed how presence/absence points are created. In the original demo, I developed a complicated geoprocessing model for creating an arbitrary number of absence points randomly along the survey trackline that were not within a certain distance of a presence point. There is some support for this kind of technique in the modeling literature (e.g., Wisz and Guisan, 2009) but I did not like it because it made the map output from the final prediction step difficult to interpret. That map represents a probability from 0 to 1, but because the input points did not represent a consistent quantity of anything, the probability was uninterpretable, even if the final binary classification (habitat / not habitat) appeared reasonable. In the new demo, I used a new

tool to generate presence/absence points along the trackline using a consistent definition. Each point represents 15 minutes of ship survey effort. Therefore the final map represents the probability "that one or more Atlantic spotted dolphins will be sighted during 15 minutes of survey effort" (conducted using the survey protocol on the NOAA cruise the produced the input data, under the environmental conditions of that cruise).

- I added an additional covariate, eddy kinetic energy (EKE), just to see what would happen. It ended up being an unimportant predictor.
- I changed Step 4 to create more statistical plots. In particular, following Zuur et al. (2009), a Cleveland dot plot is created to explore the distribution of the predictor variables, select suitable transforms, and identify outliers. (I did not end up removing any outliers.)
- Before fitting GAMs, I fit GLMs. I also fit several GLMs to illustrate a manual model selection procedure. The MGET Fit GLM tool now computes the variance inflation factor (VIF) of all input variables. I used this to address the problem that chlorophyll and bathymetry are fairly correlated and really only one should be included in the model. Bathymetry worked slightly better and has the added advantage of being a static predictor that is unaffected by clouds, making it more useful for predictions. I also removed the eddy kinetic energy predictor after finding it was not significant. The final model used SST and bathymetry.
- Using the final model (SST and bathymetry), I fitted two GAMs using spline smoothers. The first one used the default smoother parameters. This overfitted both SST and bathymetry with unrealistic curves. The second model adjusted the k parameter of the smoothers to limit the complexity of the fits, producing a more realistic fit for a sacrifice in explained deviance.
- I removed DataExploration.mxd. You can make your own exploratory map or look at the one I made in the webinar.

Our long term plans with this example

Eventually we will produce a full explanation of the entire example as a series of web pages. At that time, I will remove the original webinar.