The Habitat and Biodiversity Modeler is intended to model species and habitat impacts, including biodiversity. Those familiar with Land Change Modeler will note that the tools in this application are those previously available in Land Change Modeler. Because of the extensiveness of these tools, we have now included them in a single vertical application. The tools include:

1. Habitat Assessment
2. Habitat Change and Gap Analysis
3. Species Modeling including Maxent
4. Biodiversity Analysis and the ability to import IUCN species range data
5. Landscape Pattern Analysis and Change Process
6. Corridor Planning
7. Reserve Planning using Marxan

**Species Tab**

The species tab models habitat, species gap analysis, refines existing species polygons, and habitat suitability and distribution.

**Habitat Assessment Panel**

The Habitat Assessment panel allows one to assess the status of habitat on an animal species-specific basis. Based on any existing or predicted land cover maps and an optional map of species-specific habitat suitability, the habitat assessment tool develops a map with five categories: primary habitat, secondary habitat, primary potential corridor, secondary potential corridor and unsuitable. Important parameters that control this process include home range sizes, buffers based on sensitivity to humans and the ability to cross gaps within home ranges and during dispersal. The resulting map can be used to estimate maximum populations and serves as a primary resource in the planning for corridors. The habitat suitability map can be created with the Habitat Suitability / Species Distribution panel.

Important terms and parameters that need to be specified include:
Habitat and Potential Corridor

The habitat assessment map produced by this analysis includes five categories of habitat status. Below, they are indicated with a possible interpretation. However, they can be interpreted in any way that seems appropriate to the study under consideration.

1. **Primary Habitat.** This is habitat that meets all the necessary life needs in terms of home range size, access to summer and winter forage, etc. Issues other than minimum area and required buffer size are specified by a minimum suitability on a habitat suitability map.

2. **Secondary Habitat.** This includes areas which have the designated habitat cover types, but which are missing one or more requirements (such as area or minimum suitability level) to serve as primary habitat. Secondary habitat areas provide areas of forage and safe haven for dispersing animals as they move to new areas of primary habitat.

3. **Primary Potential Corridor.** Areas of primary potential corridor are non-habitat areas that are reasonably safe to traverse, such as at night.

4. **Secondary Potential Corridor.** There are areas that are known to be traversed by the species in question, but which constitute much riskier cover types.

5. **Unsuitable.** These are areas that are not suited for habitat or corridors.

**Include as Potential Habitat**

Any land cover type from the input land cover maps can be included in the study and can be selected for those that are associated with habitat for the species in question.

**Gap Distance Within Range**

This parameter is concerned with gaps within the home range of the species of concern. Gap distances do not need to be specified by cover types included as potential habitat components.

**Gap Distance Outside Range**

This parameter is concerned with gaps that the animal is capable of crossing when dispersing. This parameter is important in determining which areas can serve as potential corridors. In addition, this parameter effectively establishes the maximum length of the corridor.

**Minimum Core Area**

This constitutes, in the case of primary habitat, the minimum home range area of the species involved, exclusive of any buffers (hence the use of the term core). For secondary habitat areas, the core area is more likely related to forage abundance.

**Minimum Edge Buffer**

This is the size of buffer needed as distance from human activity. For potential corridor areas, this therefore constitutes half the necessary corridor width.

**Minimum Habitat Suitability**

The inclusion of a habitat suitability model is optional but strongly recommended. For each of the main habitat/corridor categories, a minimum suitability can be specified for inclusion in that category. A general strategy for development of this layer is as follows:

1. Develop separate suitability maps for each of the primary and secondary habitat and potential corridor categories. The Habitat Suitability / Species Distribution panel provides a variety of tools for empirically developing this. However, the multi-criteria
evaluation (MCE) option will most often be the tool of choice since the suitability mapping will be based on published reports of species/landscape associations.

2. Rescale the range of the primary habitat suitability map to a range of 0.75-1.0 using the STRETCH module. Then rescale the secondary habitat map to a 0.5-0.75 range; the primary potential corridor map to a 0.25 - 0.5 range and the secondary potential corridor map to a 0 - 0.25 range. Combine these four maps using the cover option in OVERLAY. The result will be a single map layer that ranges in value from 0.0-1.0. The default thresholds in HBM are set for 0.75, 0.5 and 0.25 in the decision for allocating land to the basic categories (before consideration of minimum area, gap crossing and buffer considerations). All areas with a value of 0 are by definition unsuitable.

In practice, the user is free to establish whatever thresholds are meaningful and logical in the context of their study.

Habitat Change / Gap Analysis Panel

This panel is used for two kinds of analyses: an analysis of change in habitat status (created by means of two runs of the Habitat Assessment panel) and Gap Analysis by comparing the results of one run of the Habitat Assessment panel and a protection layer map. In the case of habitat change, a graph is produced of gains and losses that can be altered with one of net change.

With gap analysis, the protection map can be either a simple Boolean image showing areas that are protected or not, or a multi-level integer map showing various protection levels. The result is simply a crosstabulation of habitat categories and protection levels.

Species Range Polygon Refinement Panel

This panel allows for the refinement of range polygon maps of species distributions developed by experts who draw the ranges onto map bases. This information is exceptionally valuable, but subject to error as a result of imprecision in the base maps, projection and geodetic datum errors, and limited geographical extent of expertise (i.e., the expert delineates only in the areas where she or he has expertise).

The underlying principle of the refinement process is to uncover the common environmental logic of the areas delineated by the range polygon. It does this by creating clusters of environmental conditions according to a set of environmental variables that the user believes can characterize the niche of the species. It then compares these clusters with the range polygon to determine the proportional inclusion of clusters within the range polygon. Clusters that fall wholly or largely within the polygon are assumed to describe essential components of that niche. Those that fall mostly or wholly outside are assumed to be unlikely components. The polygon is thus refined by removing areas that fall below a designated confidence. In addition, another option is provided to simply output a confidence map that can be used in conjunction with the original range polygon by the Weighted Mahalanobis Typicality procedure in the Habitat Suitability / Species Distribution panel. This is the default option and the one we generally recommend.

Environmental Variables and Cluster Development

The critical component of this analysis is the production of environmental clusters. For this you will need to supply a set of environmental variables that can describe basic environmental conditions. Because of the clustering technique used, this is limited to a maximum of seven variables. To stay within this limit, we strongly recommend the use of Principal Components Analysis as a way of reducing a larger set of variables to a smaller set of highly informative components. That said, you should avoid the inclusion of components with very low explanatory power.
What variables should be used? This should be decided in the context of the species being modeled. Variables can be divided into direct, resource or indirect gradients. Direct gradients are those variables that affect directly the physiology of the species (e.g. temperature), while resource gradients are those that are consumed (e.g. water, nutrients), and indirect gradients are those that do not have a direct effect on the species, but affect the distribution of a direct or resource gradient (e.g. elevation, aspect). Common variables used are elevation, slope and aspect (because of their relationship to temperature and soil moisture), the standard deviation and mean annual NDVI (as a statement of realized long term and seasonal habitat conditions), the standard deviation and mean annual temperature (as a statement of variations in energy availability), the standard deviation and mean annual precipitation (as a statement of variations in water availability), and minimum and maximum temperature and precipitation (which represent extreme environmental conditions).

Output Options

Four output options are provided:

1. **Presence.** This is a refined range polygon where areas that are poorly associated with the core environmental characteristics of the original range polygon are removed.

2. **Presence/Pseudo-Absence.** The output is the same as the above except that areas that are extremely unlikely to be associated with the core environmental characteristics of the original range polygon are treated as absence while only those that have a close association are considered as presence. This option is provided to allow the use of modeling procedures that require absence data (such as logistic regression). However, bear in mind that the absences are really pseudo-absences. To account for sampling issues, the absence pixels are chosen as a random sample of those that meet the lower threshold criterion such that the number matches (given some variance associated with the random selection process) the number of presence pixels.

3. **Confidence.** This is the default option and the one we generally recommend. Each pixel within the original polygon is assigned a confidence value from 0-1 based on how well it fits the general nature of a coherent pattern of environmental conditions (as will be explained further below).

4. **Thresholded Confidence.** This option is the same as the above, except that areas that fall below a minimum specified confidence are forced to have a confidence of zero.

For all options except the Confidence output, an upper and/or lower threshold must be specified, to establish areas of presence or absence. The default thresholds will serve as a general guideline of the values that would be used. In general, for presence, you are looking for a value that separates a clear group of clusters that strongly overlap the range polygon, while for absence you want to isolate clusters that have very little or no presence in the polygon. In many instances, this is very hard to do, which is why we recommend the use of the Confidence option coupled with the Weighted Mahalanobis Typicality procedure in the Habitat Suitability / Species Distribution panel. Using this option, no decision needs to be made.

There is also a background mask option which is quite important to the use of this procedure. If you are modeling a land species and are working in an area with significant ocean areas, you should provide a mask image to remove these from the calculations of proportional areas. For marine species, clearly the opposite applies.
Habitat Suitability / Species Distribution Panel

This panel provides a set of tools for developing habitat suitability and species distribution maps. The specific options available depend upon the nature of the training data, if any, that will be used: presence only, presence/absence, abundance or none. In all cases, you will need to specify a set of environmental variables that define the species habitat or niche.

Environmental Variables: Habitat Suitability Mapping

For habitat suitability mapping, the variables used likely relate to habitat land cover types, proximity to summer and winter foraging areas, proximity to human disturbance and so on. All variables specified must be continuous variables unless the multi-criteria evaluation (MCE) option is used. For all but the MCE option, categorical variables should be converted to a series of Boolean layers (also known as dummy variables). For the instance where MCE is used, an assignment procedure is provided that assigns suitabilities to categorical variable classes. Also with the MCE option, Boolean constraints can be added.

Environmental Variables: Species Distribution Modeling

The variables that should be used for species distribution modeling should be decided in the context of the species being modeled. Generally you would include variables that relate to the seasonal and interannual availability of energy and water. Commonly used factors include elevation and slope (because of their relationship to temperature and soil moisture), the first and second principal components of mean monthly Normalized Difference Vegetation Index (NDVI) imagery (as a statement of realized long term and seasonal growing conditions), the long term coefficient of variability in NDVI (as a statement of interannual variability), and the first two components of mean monthly precipitation and temperature.

No Training Data - MCE

The Multi-Criteria Evaluation option is designed for cases where training data are not available but where studies are available to guide the development of a suitability or distribution map by means of a multi-criteria evaluation. This option utilizes the special techniques available in the TerrSet modules FUZZY and MCE.

When using the MCE option, the first and very important stage in the analysis is to convert each of the environmental variables to factors. The difference between the two is that a variable is unscaled with respect to the model while a factor is scaled to a specific numeric range using a scaling procedure that is directly related to the expression of suitability. For example, if one were modeling a species that is sensitive to humans, a distance from human settlement layer might be used. Suitability would clearly be worst within and immediately next to areas of human occupation. As you move farther away, the land is becoming increasingly better, up to a limit. It might be that, once one reaches a distance of 2 kilometers, being further is now irrelevant - it is far enough away. In this case, maximum suitability (on the basis of this variable alone) will have been reached. Thus, we should rescale the variable such that suitability is 0 at the edge of human occupation and increases in value until it reaches its maximum at 2 km, and remains at that value for all greater distances. In the transition of multi-criteria evaluation, this process is known as standardization, but in reality one is recasting the data into an expression of membership in the fuzzy set of suitable lands.

Two options are provided for standardization: a call to the FUZZY module in TerrSet or a call to the ASSIGN module. The former is designed for the standardization of continuous variables such as in the example above while the latter is intended for the standardization of categorical variables. Note that in contrast to the standardization used in TerrSet's multi-objective decision making procedure, standardization here uses a 0.0-1.0 scaling range.
As with the MCE module itself, there are several aggregation options that dictate how the factors will be combined to create a single suitability map. The default is weighted linear combination (WLC), which is appropriate when you wish the factors to trade off (i.e., to allow poor qualities to be compensated by good qualities). The Minimum operator allows no trade-off and characterizes each location by its worst quality. This is clearly the most conservative operator. The Maximum operator also allows no trade-off, but characterizes locations by their best quality.

**Presence Data**

Presence data is probably the most common form of training data for species modeling - it records where the species has been observed, but not where it has been observed to be absent. Three procedures are available for dealing with these data: MaxEnt, Mahalanobis Typicality, and a Weighted Mahalanobis Typicality procedure.

- **MaxEnt**

  Habitat and Biodiversity Modeler provides an interface to the MaxEnt program for modeling presence-only species data. MaxEnt is a widely used species distribution model that employs a maximum entropy approach to estimate the probability distribution of a particular species. Maximum entropy is a machine learning algorithm that estimates the probability distribution of a species by finding the maximum entropy distribution subject to the constraint that the expected mean equals the empirical mean of the distribution. This method falls into the category of a use-availability model, as it uses a sample of the environmental conditions present in the study region (called background) to define the species probability distribution. MaxEnt has been widely used as a presence-only species distribution model, as it has been described to have higher predictive accuracy than other presence-only methods, although careful attention should be placed when sample bias exists. One advantage of MaxEnt is that it is able to represent complex species-gradients relationships, as it uses transformation of the variables, called features (e.g., if product features is selected, the product of all possible pair-wise combinations of variables is used, allowing for the fitting of simple interactions).

  The MaxEnt software, developed by Steven Phillips, Miroslav Dudik and Robert Shapire, can be downloaded for free from: www.cs.princeton.edu/~schapire/maxent/, and must be installed before using this option in HBM. All MaxEnt functionality has been implemented in this version.

- **Mahalanobis Typicality**

  The Mahalanobis Typicality option assumes that the underlying species distribution is normal with respect to environmental gradients. However, our tests have shown that it performs reasonably even with mildly skewed data. The output is in the form of typicality probabilities - an expression of how typical the pixel is of examples it was trained on. Thus a value of 1.0 would indicate a location that is identical to the mean of environmental conditions that were evident in the training data. However, be careful about the interpretation of low typicalities. Since typicalities express the full range of variability, a low typicality may be unusual, but still legitimately a location that is part of the species’ range. If you are looking for a threshold for when to consider an area as being unlikely to be part of its range, it is likely to be a very low value (e.g., 0.001). As an illustration of this concept, consider the case of a blue lobster. Blue lobsters are very rare, but they are still lobsters! See the on-line Help System for MAHALCLASS for further information.

- **Weighted Mahalanobis Typicality**

  This option requires both a training site file and a confidence (weight) file. It was intended that this option would be used with the confidence output of the Species Range Polygon Refinement panel. A confidence/weight image contains values from 0.0-1.0 that express the degree of confidence that the pixel is truly a member of the species’ range. Habitat and Biodiversity Modeler uses this file along with a corresponding training file and submits them to the FUZSIG module for developing the signature statistics that are needed by MAHALCLASS. FUZSIG creates a weighted multivariate mean and variance/covariance matrix based on the confidence weights.
Presence / Absence Data

Two modeling approaches are available for presence/absence data: multi-layer perceptron neural network and logistic regression.

- **Multi-Layer Perceptron**
  This option uses a subset of the options available in the MLP module utilizing the loaded environmental variables. Please refer to the on-line Help System for MLP regarding this option.

- **Logistic Regression**
  This option uses a subset of the options available in the LOGISTICREG module utilizing the loaded environmental variables. Please refer to the on-line Help System for LOGISTICREG regarding this option.

Abundance Data

Using abundance data, the sole modeling approach is multiple regression. This option calls the MULTIREG module in TerrSet using the loaded environmental variables. Please refer to the on-line Help System for MULTIREG regarding this option.

Species Data Formatting

The input to these procedures can be vector, raster, XY-Text or XY-CSV. XY-Text is a text file format suitable for presence point data, where each location is referenced by an X and Y coordinate separated by one or more spaces or tabs. XY-CSV (comma separated values) is similar except that the X and Y pair are separated by a comma. For all other tabular formats, we recommend that you load the data into Database Workshop and output the data as a vector file. Database Workshop can accept a wide range of formats (including DBF, ACCDB, XLS and CSV) and allows you to sort and subset before outputting to a vector or raster layer.

Biodiversity Tab

The Biodiversity tab allows for the import of IUCN species range polygons and the assessment of species biodiversity.

Subset IUCN Species Ranges Panel

This tool allows you to import IUCN species data and to subset the data according to the IUCN Red List status: least concerned, near threatened, vulnerable, endangered, critically endangered, extinct in the wild, extinct, and data deficient. Resulting data can be used to create biodiversity maps using the Biodiversity Analysis Panel.

Biodiversity Analysis Panel

The Biodiversity Analysis panel provides the ability to produce a spatially explicit mapping of:

1. **Alpha Diversity**: the total number of considered species at each location. Alpha Diversity is computed simply as the richness of species at each location - i.e., it is the total number of species found at each location.
2. Gamma Diversity: the total number of considered species over a large region. Gamma Diversity is calculated as the richness of species over a region. Thus the value recorded at any pixel represents the richness within the region to which it belongs and not the richness at that particular spot.

3. Beta Diversity: the ratio of Gamma to average Alpha Diversity over a large region, and thus a measure of the turnover of species. There are many measures of beta diversity that have been proposed. The measure used here is the original Whittaker’s beta diversity.

4. Sorensen Dissimilarity: a measure of species compositional dissimilarity. Sorensen’s Dissimilarity is measured as 1 minus Sorensen’s Index, where Sorensen’s Index is computed as the number of species that are common between the pixel and the region to which it belongs divided by the average alpha within the region.

5. Range Restriction: a continuous measure of vulnerability that can also be interpreted as a measure of endemism. The Range Restriction Index is based on a comparison of the area over which the species is found relative to the entire study region. It is intended for continental or global scale analyses and should include a mask file to mask out water areas for land species or vice versa for marine species. The index ranges from 0-1 with high values indicating that the majority of species present at that location have restricted ranges. Note that the index is continuous and does not rely on a threshold area to define range restriction. The formula for the Range Restriction Index is as follows:

\[
RRI = \frac{\sum_{i=1}^{n} \left(1 - \frac{\text{range}}{\text{total area}}\right)^2}{\text{Alpha Diversity}}
\]

where alpha diversity is expressed as richness (the number of species) and the total area is the total area of the image minus any masked areas.

In all cases, the input for this analysis is in the form of species range polygons. Three input formats are supported. The first is a vector composite polygon where all species polygons are contained within the same vector file. The second is a vector group file that lists the names of a set of vector files that contain the range polygons for a single species. The third is a raster group file that lists a set of raster files that contain the rasterized range polygons of a single species.

With the exception of Alpha Diversity, the data must ultimately be converted to a raster form for analysis. Thus, if a vector group file is supplied, each file is rasterized (using the spatial characteristics of the reference file) and a raster group file is created with the same name as the vector group file. If a vector composite file is used, it is first broken out into a set of separate vector files, along with a vector group file of the same name as the vector composite file. These vector files are then in turn rasterized.

- **Regional Definition**

All measures except Alpha Diversity and the Range Restriction Index require the definition of a region over which the index is calculated. Three options are provided. The vector and raster region polygon options will yield a mapping where all pixels within a region (such as an ecoregion) will have the same index value. The focal zone option, however, is quite different and can produce a different value at each pixel location.

The focal zone option calculates values by comparing the species composition in each pixel to those in a circular zone surrounding it. To use the focal zone option, you must set the focal zone diameter (e.g., 50 km). This focal zone is moved successively over every pixel in the image. As a consequence, the analysis does take considerable time to complete. Continental scale analyses at a moderate resolution (e.g., 1 km) are probably best set up at the end of the day so that they can run overnight.
Landscape Analysis Tab

The Landscape Analysis tab performs landscape pattern and change process analysis.

Landscape Pattern Analysis Panel

This panel permits analyses of landscape pattern of any landcover maps, or landcover prediction. Options include:

Normalized Entropy

This measure is Shannon's Entropy measure normalized by the maximum entropy for the number of landcover classes involved. Another common term for this measure is Diversity. It is calculated over the local neighborhood of each pixel, defined as a 3x3, 5x5 or 7x7 neighborhood. The formula is as follows:

\[
E = -\sum \left( \frac{p \times \ln(p)}{\ln(n)} \right)
\]

where \( p \) is the proportion of each class within the neighborhood, \( \ln \) is the natural logarithm and \( n \) is the number of classes. The result is an index that ranges from 0-1 where 0 indicates a case where the landcover is uniform within the neighborhood and 1 indicates maximum diversity possible of landcovers within the neighborhood.

Relative Richness

This is another measure of diversity of cover classes, measured as:

\[
R = \frac{n}{n_{max}} \times 100
\]

where \( n \) is the number of different classes present in the neighborhood and \( n_{max} \) is maximum number of classes possible.

Edge Density

Edge Density is a simple measure of fragmentation. Edge density is tabulated as the number of adjacent pairs of pixels within the neighborhood that are different from each other relative to the maximum number of different pairs possible.

Patch Area

Patch Area groups adjacent pixels of similar landcover category into patches, calculates their areas, and outputs an image where each pixel expresses the area of the patch to which it belongs.

Patch Compactness

Patch Compactness groups adjacent pixels of similar landcover category into patches, calculates their compactness, and outputs an image where each pixel expresses the compactness of the patch to which it belongs. Compactness is calculated as:
\[ C = \text{SQRT}(\frac{A_p}{A_c}) \]

where \( \text{SQRT} \) is the square root function, \( A_p \) is the area of the patch being calculated, and \( A_c \) is the area of a circle having the same perimeter as that of the patch being calculated.

**Landscape Change Process Analysis Panel**

The Landscape Change Process Analysis panel compares two landcover maps and measures the nature of the change underway within each landcover class. It does this by using a decision tree procedure that compares the number of landcover patches present within each class between the two time periods to changes in their areas and perimeters. The output is in the form of a map where each landcover class is assigned the category of change that it is experiencing. The interpretation of the categories is as follows:

- **Deformation**: the shape is changing.
- **Shift**: the position is changing.
- **Perforation**: the number of patches is constant but the area is decreasing.
- **Shrinkage**: the area and perimeter are decreasing but the number of patches is constant.
- **Enlargement**: the number of patches is constant but the area is increasing.
- **Attrition**: the number of patches and the area are decreasing.
- **Aggregation**: the number of patches is decreasing but area is constant or increasing.
- **Creation**: the number of patches and area are increasing.
- **Dissection**: the number of patches is increasing and the area is decreasing.
- **Fragmentation**: the number of patches is increasing and area is strongly decreasing.

Note, however, that while the output is in the form of a map, it is not spatially explicit - i.e., the process attributed to a landcover category is uniform over the entire study area.

**The Planning Tab**

The Planning tab offers a corridor planning tool to develop biological corridors as well as an interface to the Marxan software for reserve planning.

**Corridor Planning Panel**

This panel is used to build biological corridors. The primary inputs are Boolean maps of the two terminal regions and a habitat suitability map. Optional inputs include a development suitability map, a conservation value map and a protected lands map. There are options to specify the ideal corridor width and the number of branches. The first branch is by definition the best route. Successive branches are of lower quality.
HBM builds corridors using a cost distance procedure. The first step involves an aggregation of the various suitability/value maps. In general, the effect is such that conservation value increases suitability for the corridor while development value decreases it (although only in unprotected lands). Once an aggregate suitability map has been created, the suitabilities are converted to frictions and a cost distance is calculated from one of the terminal regions. A least-cost path is then run from the other terminal region back to the first. After this, a second cost distance is run from the least-cost path, after which it determines the mean relationship between cost distance and spatial distance to determine a cost threshold to use in constructing the corridor. If additional branches need to be built, the suitability of already selected corridor areas is reduced to zero and the process is repeated.

Marxan Panels

Two Marxan panels within the Habitat and Biodiversity Modeler Planning tab are meant to interface with the Marxan software. Marxan is freeware developed at The University of Queensland intended for conservation planning. Marxan provides techniques for reserve system design and performance as well as tools for developing multi-use zoning plans for natural resource management. Marxan can be used to identify areas that meet biodiversity targets, taking into account minimum costs.

Marxan is not provided with IDIRISI but can be downloaded for free from The University of Queensland website at: http://www.uq.edu.au/marxan.