

# ***BioMapper 3***

***User's manual***



# ***Introduction***

This document is actually a print-out of the help file. This is exactly the same information. Personally, I find help files easier to use, as you can search for key-words and click on links to jump directly to a related page. However, some have asked me for a paper version of this file and here it is.

I didn't edit it and thus it may look a little strange. A few remarks on the notations will perhaps help you to understand what is going on:

When a word is underlined, that means you can find its definition at the end of the manual.

When a word is double-underlined, that means there is a chapter related to this subject. However, the chapter might not have the same name.

The chapters are sorted in logical sequences, to a certain point.

This document is probably the last one I shall update in the future. If you need brand new information, consider looking first in the on-line FAQ (<http://www.unil.ch/biomapper/faq.html>), then in the help file or here. Don't forget the Biomapper-List discussion group on Yahoo.

I wish you success in the use of Biomapper.

Alexandre Hirzel

Lausanne, the 9<sup>th</sup> of January 2004

# Overview of BioMapper 3

*BioMapper* is a kit of GIS- and statistical tools designed to build habitat suitability (HS) models and maps. It is centred on the Ecological Niche Factor Analysis (ENFA), which allows to compute HS models without the need of absence data.

More precisely, it can achieve the following tasks:

- Preparing the ecogeographical maps in order to use them as input for the ENFA (e.g. computing frequency of occurrence map, standardisation, masking, etc.)
- Exploring and comparing them by mean of descriptive statistics (distribution analysis, etc.)
- Visualising them in 2D or 3D.
- Computing the Ecological Niche Factor Analysis and exploring its output
- Computing a Habitat suitability map
- Evaluating its prediction accuracy by means of cross-validation

*BioMapper* is designed to be autonomous but as it uses the same files format as the GIS software *Idrisi*, they can transparently work in conjunction.

Please be sure to read the Rules of use before to work with *BioMapper*.

## BioMapper at first glance

When you open it, *BioMapper* is composed of two parts: on the left you have a column with two compartments entitled *ecogeographical maps* and *Workmaps*. The first compartment will contain the environmental predictors used to build the model. The workmap compartment will mainly contain maps resulting from your *BioMapper*'s computations, but this is also here that you must place the map with the species observations. You may actually place here any map you want often to access.

On the right part you have the result window. In this window will be logged all your operations, along with their results. This window is part of the project and allows you to trace back what you have done. It is stored in file with the same name as your project, but with a ".log" extension. This file is actually an RTF file, which you can import into any word-processor. This window is fully editable to help you to keep track of your investigations. But this window is also "active". Whenever your mouse cursor hovers above a map filename, it will take the shape of pointing hand. By double-clicking it, you can open a visual display of the map. A right-click will pop-up a menu with several operations you can do on the map.

In the status bar (the bar at the foot of the *BioMapper* window) you will get various information, including short help message about the command or menu your mouse cursor is hovering over.

Finally, the menu bar is organised so as to present the operation in chronological order. As you progress with your analysis, you are going to use modules presented in menu from left to right. When a special window is active (a graph, a map, etc.), new menus may appear in this bar, proposing special operations.

Finally, by clicking the F1 key will display this help file, sometimes at the page corresponding to process you are about to launch.

## Modus operandi

Here are described step-by-step how to conduct a complete analysis, beginning with environmental data and ending with a validated Habitat Suitability map. The Frequently Asked Question (FAQ) page (<http://www.unil.ch/biomapper/faq.html>) on the *BioMapper* website gives also a lot of additional and up-to-date information. Another source of help is the *BioMapper* discussion group at <http://groups.yahoo.com/group/Biomapper-List/>.

Click on the subtitles below to get more detailed information on each procedure.

### A. Preliminaries

You begin by converting the data you have collected (field sampling, official databases, etc.) into *Idrisi*-formatted maps. These operations can be done with *Idrisi* or *BioMapper*, or any other tool you want (I frequently use *Corel Draw*, *MapInfo*, ASCII editors as well as home made programs).

These operations depend on the source format and it is impossible to give here an exhaustive description of the procedures; each database will probably need a different kind of treatment before to be ready to use.

Once you have got maps that are usable by *BioMapper* and *Idrisi*, you can work with them. There are two kinds of operations you want to do:

1. Deriving the available maps into biologically-relevant maps (by reference to the focal species), which we shall call EcoGeographical Variables (EGV).
2. Transforming the maps in order to make them suitable for the ENFA.

## **B. Preparing the maps**

Now that your EGV maps are ready, you must prepare them for the ENFA. Basically, it means to make them overlayable and, as far as possible, with an unimodal and symmetrical distribution. You must also verify that there are no discrepancies between them. You create a project that will store all important data, operations and results.

## **C. Ecological niche analysis**

You are now ready to perform the Ecological Niche Factor Analysis. This will give you several outputs (eigenvalues, scores matrix, etc.) that you should examine attentively in order to verify them and to get ecological understanding about the focal species. Now, you have got your ecological niche model.

## **D. Habitat Suitability computation**

Finally, you can use this model to compute a HS map.

## **E. Evaluation and post-production**

The predictive accuracy of this map can then be evaluated by a cross-validation procedure. And finally, you may want to reclassify or smooth the HS map.

### **Shortcut**

The steps C, D and E may be done sequentially in one shot by using the *Automatic analysis* dialog box (Menu Multivariate analyses/Automatic analysis...). Although it can speed up the operations and make them easier, you should first follow the operational processes "manually" a few times in order to better understand how things are interconnected. By the way, the first time you study a species, you should do it manually in order to keep an eye on each part of the whole process. Afterward, when you will be tuning the model, trying various approaches, the automatic analysis will prove a very practical functionality.

# **Preliminaries**

First you have to choose which data are relevant (and available (field sampling, official databases, etc.)) for your focal species. ENFA is not sensible to irrelevant data nor to their order. All useful information will be extracted and summarised into the ecological niche factors. Thus, don't fear to use too much data. Irrelevant data will increase the computation time and the memory needs, but will not significantly influence the accuracy of the result. Nevertheless, they could decrease the generalisation of your model.

Then you have to convert those data into *Idrisi*-formatted maps. These operations can be done with *Idrisi* or *BioMapper* (Maps/Convert), or any other tool you have (I frequently use *Corel Draw*, *MapInfo*, ASCII editors as well as home made programs). If you are working with ArcView or Arc/Info, there exist several tools enabling you to convert ESRI grids into *Idrisi* Raster. The easiest is probably the extension made by Holger Schäuble (Look for Grid Converter (av2idrisi.zip) on <http://arcscripts.esri.com/>). It works for ArcView 3.x. See the FAQ for more info.

Those operations can be very time-consuming and it is impossible to give here a full step-by-step description of the procedures; each database will probably need a different kind of treatment before to be ready to be used. Typically, 75 to 90% of the time devoted to a project will be devoted to this task.

When you have got maps usable by *BioMapper* and *Idrisi*, you can work with them. You first want to derive biologically-meaningful maps (by reference to the focal species) from them, and make them quantitative (The ENFA needs quantitative data).

Here are a few guidelines that will lead you on your way:

- Prepare the maps containing the ecogeographical variables
- Prepare a Boolean map (containing 0 and 1 only), the 1 indicating the cells where the species is present (Maps/Convert). Alternatively, you can also attribute an integer weight to the species observations. In this case, replace the 1 by this weight. But you must be sure that your data are the product of a homogeneous collecting effort.
- Mask all these maps to be sure they cover exactly the same area (Maps/Formatting/Mask)

## ***Tips to make ecogeographical maps quantitative***

An important step is to make the ecogeographical data quantitative. They may belong to three types, each needing a different preparation method:

### **Quantitative data**

Examples: Altitude, slope, mean July temperature, amount of yearly rain, etc..

These data are ready to be used by the ENFA. You may want to transform them in order to make them more relevant to the species (smoothing, buffering, averaging, etc.) but they are technically acceptable.

## Qualitative data

Examples: Soil type, land cover, vegetation type, etc..

These maps cannot be used in this format. There are two methods to extract quantitative data out of them. Let's imagine we are dealing with a map of vegetation types to illustrate them:

The first one is to determine which is the important feature of the represented categories, to reorder them using a semi-quantitative scale determined by this feature and to code it with a integer numerical value. For instance, vegetation could be ordered by height of the canopy as follows: 1.Bare ground, 2.Grassland 3.Bushes 4.Forest

The second method implies first to transform the map into several **Boolean maps**, each describing a relevant category. All available categories can be used, or only a few of them, or it is also possible to pool several categories into one Boolean map. These operations can be easily done by using Idrisi's *Image calculator*. These maps will then be used as described in the next paragraph. For instance, we could consider that only forests and bushes are relevant for our species and compute two Boolean maps representing presence/absence of these two entities.

## Boolean data (=binary data)

Examples: Presence/absence of a species, cultivated areas, towns, roads, lakes, forests, etc..

These maps cannot be used in this format. There are four main methods to extract quantitative data out of them:

The first method is preferentially used when the map represents a resource or a shelter for the focal species, when you guess that the species needs a minimal amount of it or cannot live when it is too important in the landscape. The method implies to choose a radius of influence and to compute the frequency of occurrences into a circular area around the focal cell. Generally, the radius is chosen in order to produce a circle area equal to the home range of the species. This analysis can be performed by the BioMapper's module *Circan*, option *Frequency* ([Maps/Contextual/Circular analysis](#)). For instance, the quantitative map could represent the frequency of forests into a 1km radius buffer around each cell.

The second method is typically biological. Many species live at the interface between two habitats (forest edges, lake shores, coasts, etc.). The principle is then to compute the length of the boundary parting them in a circular area defined as in the first method. This can be performed by the BioMapper's module *Circan*, option *Edge length* ([Maps/Contextual/Circular analysis](#)). A few other landscape-ecology indices are available in *Circan*.

The third method is ideal when the map represent the locations of disturbance sources (towns, roads, etc.). It consists to compute a distance map, attributing to every cell the distance to the nearest occupied cell. This can be done with BioMapper's *Distan* module ([Maps/Contextual/Distance](#)) or Idrisi's *Distance* module. The real distance can be used, or a cost distance or even a function of the real distance (to lower the influence of the longer distances).

The fourth method is more difficult to justify biologically. It consists in smoothing the Boolean map to get values between 0 and 1. This can be done with Idrisi's *Filter* module or BioMapper's module *Circan*, option *Gaussian* or *Mean* ([Maps/Contextual/Circular analysis](#)).

# Preparing the maps

Now that your EGV maps are at hand, you must prepare them for the ENFA. Basically, it means to make them overlayable and, as far as possible, unimodal and symmetrical. You will also verify that there are no discrepancies between them. You will create a project that will be used for all the subsequent operations.

Here is the step-by-step procedure:

- Create a list of the ecogeographical maps ([Files/Ecogeographic maps/Add maps...](#))
- You will perhaps have to change the background value for each map; this value will not be used in the ulterior analyses. Furthermore, it is better to assign a "Biomapper extension" to each map, indicating the kind of its data (Boolean, number, frequency, ...). Although it is not strictly required, it will allow you to find more quickly the relevant maps in the jungle of all of them.
- Normalise the ecogeographical maps ([Maps/Formatting/Transformation](#)) using the Box-Cox function. It may happen that the transformation generates a constant map (all cells have the same value) or nearly-boolean maps (almost all cells belong to a small range of values.); as these maps would create problems during the subsequent computations, it is better to revert to the original map (untransformed) or to discard them. If you have checked Replace, the new normalised maps will replace the original ones in the project's EGV list (they will now have a "-box" suffix).
- Verify the consistency and usability of these maps ([File/Ecogeographic maps/Verify maps](#)). This operation verifies that all maps have the same background and non-background cells and that there are neither constant or nearly-boolean maps. If any, it will list the problematic maps.
- Save the project for future use ([Files/Save project as...](#))
- Add the species map among the Work Maps ([Files/Work maps/Add map...](#)) and mark it as "Species map" (right click, [Mark as species map](#))

## ***Ecological niche analysis***

You are now ready to perform the Ecological Niche Factor Analysis. This will give you many outputs (scores matrix, etc.) that you should examine attentively in order to verify them and to get ecological understanding about the focal species. Now, you have got your Ecological Niche model.

- Compute the covariance matrix ([Multivariate analyses/Matrix/Covariance matrix](#))
- Save the project to include the covariance matrix (thus you will not have to compute it again the next time) ([Files/Save Project As...](#))
- Do an Ecological Niche Factor Analysis (ENFA) ([Multivariate analyses /Factors/ENFA](#)). The scores of this analysis (=the eigenvectors) will be saved automatically and can be got again later by the mean of the menu [Multivariate analyses /Factors/Open](#). Note that invoking this menu will automatically launch the covariance matrix computation if it was not yet done; you can therefore jump over the two first steps of this process.
- Check the eigenvalues (they must all be greater than or equal to zero)
- Study carefully the results of the analysis. See the [FAQ](#) for further information about these results.

## ***Habitat suitability computation***

Finally, you can use this ecological niche model to compute a HS map.

- Compute a Habitat Suitability map ([Habitat suitability/Habitat Suitability map](#))
- Here you can choose the HS algorithm you want to use (for now only the “medians” algorithm is available, as described by Hirzel et al. (*Ecology*, in press).
- Select the factor maps you want to include in the HS computation. When you click on the browse (...) button, a dialog box appears allowing you
  - To choose how many factors you want to include (comparing the eigenvalue distribution to Mc-Arthur’s broken-stick gives you a clue about this problem).
  - Then you can compute the chosen factor maps. They will be saved. As long as you do not compute again the ENFA, they will be available for HS computation, otherwise, you will have to compute them again (this is the longest part of the HS computation).
  - You can now modify the weight assigned to each factor. By default, the weights are computed from the eigenvalues and represent the amount of information explained by each factor. You should not modify them.
- Select the HS algorithm. Four of them are available, each covering a different case (my favourite are Medians and Geometric mean):
  - **Medians**: Quick, accurate in most situations, good generalisation power. The species distribution on each factor must be unimodal and more-or-less symmetrical.
  - **Distance geometric mean**. Slow, good generalisation power. No assumption is made on the species distribution.
  - **Distance harmonic mean**: Slow, medium generalisation power. No assumption is made on the species distribution. It gives a high weight to each single observation and therefore could give better results when the sample size is very small and each observation might bring relevant information to the model.
  - **Minimum distance**: Medium speed, low generalisation power, low predictive accuracy. No assumption is made on the species distribution. It gives a very high weight to each single observation and therefore might give better results when the sample size is very small and each observation might bring relevant information to the model. However it produces usually bad results.
- More information about these algorithms [here](#).
- Choose a name for the HS map. By double-clicking on the field, a name will be automatically filled.
- Now you can compute the HS map.
- You can visualise this map through the menu [View/Map](#) (or in Idrisi), or by double-clicking on its name in the output window.
- You can also visualise the ecological niche in two dimensions (two factors) with the menu [habitat Suitability/Niche 2D visualisation](#). That’s with this tool (and Idrisi’s CONTOUR operation) that I made the figures in the Environmental Management paper (2004).

## Evaluation of the HS map

You may now evaluate the predictive power of your HS map. This is done by a cross-validation process. It computes a confidence interval about the predictive accuracy of the HS model. The species locations are randomly partitioned into  $k$  mutually exclusive but identically-sized sets.  $k-1$  partitions will be used to compute a HS model and the left-out partition will be used to validate it on independent data. This process is repeated  $k$  times, each time by leaving out a different partition. This process results in  $k$  more-or-less different HS maps. By comparing these maps and how they fluctuate, one can assess their predictive power.

BioMapper does this following the method described by Boyce et al. (Boyce, M.S., P.R. Vernier, S.E. Nielsen, & F.K.A. Schmiegelow (2002) Evaluating resource selection functions. *Ecological Modelling* 157: 281-300). Each map is reclassified in  $b$  bins (by default,  $b = 4$ ). Each bin  $i$  covers some proportion of the map's total area ( $A_i$ ) and contains some proportion of the validation points ( $N_i$ ) (validation points are those observation left out during the cross-validation process). One computes then the area-adjusted frequency for each bin as  $F_i = N_i / A_i$ . If the HS map is completely random, one expects  $F_i = 1$  for all the bins. If the model is good, low HS should have a low  $F$  (below 1) and high HS a high  $F$  (above 1) with a monotonic increase in between. A way to measure the monotonicity of the curve is to compute a Spearman rank correlation on the  $F_i$ ; however, we didn't find this correlation much useful. We are currently trying to develop a better measure.

Practically, one often get a sigmoid curve, the  $F$  increasing exponentially and then stabilising and oscillating around a maximum value. Looking at the shape of this curve allows one to define where is the threshold between suitable and unsuitable habitat, from which point the model doesn't add significant information, etc.. Also, the variance of the  $F_i$  among the cross-validated curves reflects the prediction power of the model.

- Open the cross-validation (CV) dialog box ([Habitat suitability/Area-adjusted frequency cross-validation](#))
- This dialog box is similar to the [Habitat Suitability one](#) and should already be filled correctly if you have just computed your HS map. The only new field is entitled "k-fold cross-validation". Here you can define the CV parameters.
- The most important one is the number of partitions ( $k$ ). You can also choose to keep the temporary HS maps for further analysis or compute confidence limits maps.
- Click on the [Compute](#) button.

The cross-validation may take some time as  $k$  HS-maps must be computed. Once the process is completed, a new dialog box appears, entitled "area-adjusted frequency cross-validation". The upper graph shows the  $F_i$  curves along with the Spearman correlations. The panel just below the graphs allows the user to select various display options. The lower graph displays the bins. By default, there are four of them and they have an equal width. The histograms represent the average number of cross-validation points in each bin (in green) and the average area (or number of map cells) covered by each bin (in red). You can change the number of bins with the spin editor below the graph. You can also modify the sizes of the bins, either manually by drawing the histogram bar borders with the mouse, or by pressing on one of the buttons in the lower panel. You can choose between equal width (all bins cover the same HS range), equal counts (BioMapper tries to set bin limits so as they all have about the same number of cross-validation points) or equal area (BioMapper tries to set bin limits so as they all cover about the same area). The equal counts is somewhat similar to the operation described in the paper by Boyce et al. and seems to give the best results. There is still some work going on this part too.

Once you have chosen the relevant bin limits, you can reclassify the HS map to present only these new categories by clicking on the [Reclassify HS map](#) button.

You can also save the graphs (in WMF or EMF format, which you can use for instance in Word or Powerpoint) or the underlying data (in text format, which you can use in any statistic software for further analysis).

Finally, by leaving the cross-validation box, you will find a summary of the cross-validation statistics, along with the bin limits, in BioMapper's main result window.

## Post-production

Once you have got the perfect HS model and map, you can further process it in several ways:

- You may reclassify it to show only a few classes (e.g. "unsuitable", "marginal habitat", "core habitat"). Indeed, the 100 HS classes produced by the algorithms provide a misleading precision and a coarser scale would better represent what the model really mean (look at the [area-adjusted-frequency cross-validation](#) to decide which classes are significant). Moreover, displaying less classes may improve the interpretability of

- your map, as well as making it better suited for a black-and-white publication.
- Inversely, you may want to smooth the HS map to remove some stair steps that may occur in some cases. Smoothing will probably not improve the scientific value of your map, but may enhance its visual appeal.
- Finally, you might be asked to provide a Boolean map displaying good versus bad habitat. The menu [Post-production/Booleanise](#) HS map will provide you with a few tools to guide you in the process of choosing the right threshold (note that the [area-adjusted-frequency cross-validation](#) could be used too).

## **Idrisi**

Although *Biomapper* is entirely autonomous, it is designed to work in conjunction with the GIS-software *Idrisi*. As they both use the same file structure, there are no problems of conversion between them. *Idrisi* is developed at the Clark University (USA) by J. Ronald Eastman. I found *Idrisi* to be very efficient and powerful; moreover, as it is produced by a university, it is sold at a low cost.

*Biomapper* is fully compatible with the two last version of *Idrisi for Windows*, namely *Idrisi 2.0* and *Idrisi32*. You can switch between them by mean of the menu command [Ecogeographical maps/Options...](#), tab [files](#).

Further information can be found at <http://www.clarklabs.org> .

## **Projects**

Projects are the working units of *Biomapper*. They contain all the information needed to conduct an analysis from the input EGV maps to the HS map. A project is composed of the following elements:

- A set of [overlayable EGV maps](#)
- A [species presence map](#)
- The [covariance matrix](#) of these maps
- The [scores matrix](#) computed by the [ENFA](#)
- A [results listing](#)
- A set of [work maps](#) (not necessarily overlayable)

Projects can be saved, which is very useful as a complete analysis can be a long work. As the results listing is also saved, you can always remember how any given map was treated or computed.

## **Ecogeographical maps list**

The first elements to be part of a project are the EGV maps.

An ecogeographical variable (EGV) is a spatially defined variable that is associated to every location of a study area. It describes quantitatively features that can be ecological (e.g. frequency of forest, nitrates concentration), topographical (e.g. altitude, slope) or anthropic (e.g. distance to the nearest town, roads density).

An EGV map is a raster map whose cells contain the value of a given ecogeographical variable measured on the underlying area.

In a given project, all the EGV maps have to be overlayable, that is, they have to cover the same area, to have the same resolution and to use the same spatial unit.

## **Species map**

This is a [boolean](#) raster map whose cells contain a 1 when there is a proof of the focal species presence in the underlying area, and a 0 when such a proof doesn't exist. Note that a 0 does NOT mean that the underlying area is unsuitable for the species. Alternatively, you can assign an integer weight to the observation locations. This weight will be used for the ENFA computation and HS map modelling and evaluation.

To add a species map to a project, you must first add it to the [work maps list](#). Then, you mark it by right-clicking on it and selecting it as species map. There can be only one species map at a time in a given project and it must be [overlayable](#) on the [EGV maps](#). You can change the current species map by right-clicking on any map of the work map list (but remember: at the end, there can be only one).

## Work maps list

In this list you can store any map you want to have at hand during your analysis. In this list lies also the species map of your project. When an Biomapper's operation produces a new map, it will also automatically be added to this list.

Contrarily to the EGV maps, the works maps have not to be overlayable.

## Covariance matrix

This matrix stores the covariances of the EGV maps. The diagonal contains the variances.

This matrix is needed to perform the ENFA and as it is a highly time-consuming computation, this matrix is stored in the project file. When this matrix has already been computed, a flag "COV" appears on the status bar, at the bottom of the Biomapper's window.

You can compute the covariance matrix separately through the command menu Multivariate/Matrices/Covariance matrix. If this matrix was already computed, this command will only display it on the results listing.

To recompute the matrix, you have first to empty it by mean of the command menu Multivariate/Matrices/Reset covariances. This will turn off the "COV" flag in the status bar and you can now compute the matrix as mentioned above.

If the covariances matrix is not yet computed when you launch the ENFA, this will be done automatically.

## Ecological Niche Factor Analysis (ENFA)

This analysis is the central part of *Biomapper*. It is designed to compute the factors (like the Principal Components Analysis) that explain the major part of the ecological distribution of the species. Like in the PCA, the extracted factors are totally (by construction) uncorrelated but in this case they have biological signification: the first factor is the marginality factor, which describes how far the species optimum is from the mean habitat in the study area. The tolerance factors are sorted by decreasing amount of explained variance; they describe how specialised the species is by reference to the available range of habitat in the study area. Therefore, only a few of the first factors explain the major part of the whole information.

You launch the ENFA by mean of the command menu Multivariate/Factors/Ecological Niche Factor Analysis (ENFA). If the covariance matrix is not yet computed, this will be done automatically.

ENFA will display the following output;

- The global correlation matrix
- The global mean vector
- The species-covariance matrix
- The species mean vector
- The eigenvalues and their explained variance
- The scores matrix (eigenvectors matrix)

## Habitat Suitability algorithms

There are currently four HS algorithms available in BioMapper:

### Median algorithm

Before Biomapper 3.0, it was the only algorithm available. It gives good results in most situations and is the quickest of all. It was first introduced in Hirzel, A.H., J. Hausser, D. Chessel, & N. Perrin (2002) Ecological-niche factor analysis: How to compute habitat- suitability maps without absence data? Ecology 83: 2027-2036.

To compute the median algorithm, we divide the species range on each factor in 25 classes, in such a way that the median would exactly separate two classes. For every point in the environmental space, we count the number

of observations that are either in the same class or in any class farther apart from the median. To achieve normalisation, twice this number divides the total number of observations. Thus, a point belonging to one of the two classes directly adjacent to the median will get a value of one, and a point lying outside the observation distribution will get a value of zero. Lastly, the overall suitability index for this point is computed by the weighted average of its scores on each dimension, the weights being given by the amount of information explained by each dimension.

This algorithm makes the assumption that the best habitat is at the median of the species distribution on each factor, and that these distributions are symmetric. Although this is often true, in some case it's wrong: when the distribution is bimodal, it is even completely wrong. You can also get sub-optimal results when the study area is at the border of the species distribution.

### **Distance geometric mean algorithm**

This algorithm is described in a paper that is not yet published at the time I'm writing this help file: Hirzel, A.H. & R. Arlettaz (2004) Modelling habitat suitability for complex species distributions by the environmental-distance geometric mean. Environmental Management. In press.

The principle of this method is to draw in the factor space the influence field, or *suitability field*, of each species observation point in such a way that, when they are close together, their attraction powers reinforce each other. For any point  $\mathbf{P}$  in the factor space, one computes the geometric mean  $H_G$  of the distances to all observations  $\mathbf{O}_i$ . Thus, the denser the species points in the environmental space, the higher the habitat suitability. This actually comes down to calibrating a model in the environmental space to apply it to the geographic space.

$$H_G(\mathbf{P}) = \sqrt[N]{\prod_{i=1}^N \delta(\mathbf{P}, \mathbf{O}_i)}$$

This algorithm makes no assumption on the shape of the species distribution, but the density of observations must be representative of the species niche. The geometric mean produces a smooth set of envelopes around the observations points and provides a good generalisation of the niche.

### **Distance harmonic mean algorithm**

This algorithm is described in a paper that is not yet published at the time I'm writing this help file: Hirzel, A.H. & R. Arlettaz (in press). Environmental-envelope based Habitat-suitability models. In the proceedings of the 1<sup>st</sup> Conference on Resource Selection by Animals, Laramie, Wyoming, USA.

This algorithm is similar to the geometric mean one but uses the harmonic mean of the distances instead:

$$H_H(\mathbf{P}) = \frac{1}{\frac{1}{N} \sum_{\substack{i=1 \\ \mathbf{P} \neq \mathbf{O}_i}}^N \frac{1}{\delta(\mathbf{P}, \mathbf{O}_i)}}$$

The effect of this mean is to give a (too) high weight to all observations while keeping the information of observation density in the factor space. Therefore, it has a tendency to overfit the data, which might be good when you have a small sample size.

### **Minimum distance algorithm**

This algorithm is described in a paper that is not yet published at the time I'm writing this help file: Hirzel, A.H. & R. Arlettaz (in press). Environmental-envelope based Habitat-suitability models. In the proceedings of the 1<sup>st</sup> Conference on Resource Selection by Animals, Laramie, Wyoming, USA.

Again this algorithm is similar to the geometric mean one but uses the minimum of the distances instead:

$$H_M(\mathbf{P}) = \text{Min}_{i=1}^N \{\delta(\mathbf{P}, \mathbf{O}_i)\}$$

That means that the density of observations is not taken into account anymore. Each observation has the same weight, and the closest you are from one of them, in the factor space, the more suitable your habitat. There is no generalisation at all. Actually, on all examples I have tested this algorithm never gave good results. However, as it is described in the above paper, I decided to include it in this version of BioMapper. And perhaps someone will find a case where it proved useful.

## **Score matrix**

The score matrix contains the contribution of each EGV map to each factor. The coefficients give a weight to each EGV map and they will also be used to compute the HS map.

### **Marginality factor**

The marginality ( $M$ ) is a measure of the departure between the average of the species distribution ( $m_S$ ) and the average of the global distribution ( $m_G$ ). Mathematically, it is defined as:

$$M = \frac{|m_S - m_G|}{1.96s_G}$$

Where  $s_G$  is the standard deviation of the global distribution.

The coefficients of the scores matrix related to the marginality factor indicate the correlation between each EGV and the factor. The greater the absolute value of the coefficient, the higher this EGV contributes to the marginality. A positive value means that the focal species “prefers” the high values of this EGV (by reference to the whole study area); a negative value means that focal species “prefers” the low values.

## Results listing

On this page *Biomapper* displays all the operations you perform, including key-parameters and results. Those results are saved with the project and you can thus keep a track of what you have done and how. The result listing is saved in a separate file having the same name as your project but with a “.log” extension; it is in “RTF” (Rich Text File) format and can so be viewed with most word processors.

You can also copy parts of the results and paste them in any word processor or spreadsheet software.

**Input maps** are printed in green and **output maps** in red. The listing is active in the sense that if you double-click on any map name, a display window will be launched to show you that very map. A right-click will pop up a menu with common operations you can apply to the pointed at map.

You can also edit this page to remove any of its part or add some comments.

## Sorry, no help available

Sorry, no help was written on this subject. You can try to understand the pop-up hints that appear when you leave one second your mouse pointer on any field. I tried to make it as clear as possible... Hope you will catch my tortuous mind.

And there is perhaps useful information in the FAQ

Now, you are on your own...

## License and conditions of use

Please read this page attentively:

### Rules of use

By using *Biomapper* you subscribe to the following rules:

- **FOR NOW, BIOMAPPER IS INTENDED FOR SCIENTIFIC PURPOSE ONLY. ANY COMMERCIAL USE IS FORBIDDEN.**
- *Biomapper* is a postcard-ware. You can test it freely and, should you find it useful and make use of it for your research, you were to send a postcard to its author at the following address:  
Dr. Alexandre HIRZEL  
University of Bern  
Zoological Institute  
Division of Conservation Biology  
Baltzerstrasse 6  
CH-3012 BERN  
SWITZERLAND
- You cannot freely distribute *Biomapper*. But feel free to put a link on the download page on the *Biomapper*'s website at <http://www.unil.ch/biomapper>.
- If you use *Biomapper* to analyse your data, please quote it in your publications as follows: “Alexandre H. Hirzel, Jacques Hausser and Nicolas Perrin (2000). *Biomapper* 2.0. Laboratory for Conservation Biology, University of Lausanne.”
- You may also want to quote the paper describing the ENFA: Hirzel, A., J. Hausser, D. Chessel, and N. Perrin.

in press. Ecological-Niche Factor Analysis: How to compute habitat-suitability maps without absence data? Ecology.

- Alexandre H. Hirzel nor the Laboratory for Conservation Biology assumes any responsibility for any errors that may appear in this document, or in *Biomapper*.

## **Bugs and errors**

Biomapper is not yet fully completed and is in its test phase. Thus errors and bugs may occur. The computation procedures have been intensively tested on various data sets so there is little probability of remaining errors here; however, it is always a good idea to look carefully at the results exiting of a "black box". "Structural" bugs can nevertheless happen, creating potentially strange behaviour or crash of Biomapper. Please report me any bug or strange thing so I can correct it.

In the menu Help/Beta version history, you will find a historical review of the improvements and fixed bugs along the successive beta-versions.

Check the beta-version number to know where you are in the history. Always give me this number when you report me a bug. The two first numbers will remain 1.0 and only the two last are subjected to be incremented. The third number indicates the implementation of a new tool or a major change in the software. The fourth number will never be reset to zero and is incremented each time I recompile entirely the software; so it can indicate orthographic corrections, minor bug fixes or totally new foundation of the concept.

Alexandre H. Hirzel nor the Laboratory for Conservation Biology assume any responsibility for any errors that may appear in this document, or in *Biomapper*.

## **Credits and References**

### **References**

The following sources were used to realise this software :

- - (1985). Formulaires et tables de mathématique, physique, chimie. Genève, Tricorné.
- Eastman, J. R. (1997). Idrisi for Windows 2.0 user's guide. Worcester, Clark University.
- Hausser, J. (1993). ADIGE : Programme d'Analyse de la Distribution Géographique. Lausanne.
- Hausser, J. (1995). Analyse de l'habitat. Mammifères de Suisse. Birkhäuser. Bâles: 2-14, 475-481.
- Perrin, N. (1984). Contribution à l'écologie du genre *Cepaea* (Gastropoda): Approche descriptive et expérimentale de l'habitat et de la niche écologique. Institut de Zoologie et d'Ecologie Animale. Lausanne.
- Legendre, L. and P. Legendre (1984). Ecologie numérique. Québec, Masson.
- Press, W. H., B. P. Flannery, et al. (1989). Numerical recipes in Pascal : The art of scientific computing. Cambridge, Cambridge University Press.
- Sokal, R. R. and F. J. Rohlf (1981). Biometry : The principles and practice of statistics in biological research. New York, W.H. Freeman & co.

### **Thanks**

I want to thank all those people who helped me in the realisation of *Biomapper*.

- Many thanks to Prof. J.Hausser who gave me the FORTRAN code of his software "ADIGE 5C"
- To N. Perrin and Jérôme Goudet for all their useful advices.
- To L. MBayo for the Biomapper's logo
- To the testers: Alexandre Hirzel, Patrick Patthey, Myriam Saudan, François Metral, Véronique Helfer, Antoine Guisan, Fridolin Zimmermann, Christoph Plutzer, Elisabeth Zaniewski, Sébastien Sachot, Raoul Vega, Patrick Presi, Isabelle Dunand, Asier R. Larrinaga, Gwenaëlle Le Lay, Lluís Brotons, Thomas Sattler, Samuel Ehrenbold, Tany Leverett, Richard Zink, and many others...

### **Contacting the author:**

Dr. Alexandre Hirzel  
Conservation Biology

Baltzerstrasse 6  
University of Bern  
CH-3012 BERN  
Switzerland

E-mail address: <mailto:Alexandre.Hirzel@nat.unibe.ch>

## File maintenance

As each map is composed of two files (see [file structure](#)) it could be fastidious to perform some basic maintenance tasks like renaming and deleting maps. This is why this command is made for.

You can launch it by mean of the menu command [File/Files maintenance...](#)

Simply click on the [browse...](#) button, select the files you want to maintain. They will appear in the list box where you will be able to select or deselect them. The buttons on the right of the box execute the maintenance operation you want to do on them.

- Delete will delete the selected maps from the hard disk
- Change ext. will change their extension (a dialog box will appear)
- Rename will rename them (a dialog box will appear)

## Map properties

This box allows you to consult the basic map properties like maximum and minimum values and data type. You can also consult the lineage of the map and read or modify comments and title.

Here you can also optionally define the [background value](#).

Finally, here you can modify the [Biomapper's extension](#) of the map.

You can access this box through the menu command [File/Ecogeographic map/Properties...](#) or by right-clicking on the Ecogeographic map list and the popup menu entry [Properties...](#)

## Biomapper's extension

*Biomapper* introduces a new property for the *Idrisi* map files. It is named *Biomapper's extension*. It consist in adding to the end of the file name (before the dot window extension ) two letters preceded by a underscore character ("\_"), which describe the data type stored by the map. *Biomapper's* modules produce automatically such extension and it is very useful when you begin to have a lot of maps. Available extensions are:

- “\_FQ”: Frequency/probability data. It is used when the data are comprised in the range [0,1] or [0,100 %]. The map is stored in byte format and values are comprised between 0 and 100. The optional background value is 255.
- “\_BL”: Boolean data. It is used to store binary (0/1) (absence/presence) or logical (false/true) data. The map is stored in byte format and values are 0 or 1. The optional background value is 255.
- “\_NB”: Count data. It is used to store numbers. They cannot be less then zero and are discretely distributed. The map is stored in a byte, integer or real (rounded) format depending on the maximum value. The optional background value is -1.
- “\_QT”: Quantitative data. It is used to store any cardinal data. Many [EGV maps](#) are a good example of them (altitude, distance to towns, etc.). The map is generally stored in real format.
- “\_ST”: Statistical data. It is a subset of the quantitative data. This extension is used when the map was obtained by some statistical operation on a more basic map. The map is generally stored in real format.
- “\_CL”: Qualitative data. It is used to store any nominal data (land-cover, type of soil, phytosociology unit, etc.). The map is generally stored in byte or integer format.

# File structure

*Biomapper* uses the same file structure as the GIS-software *Idrisi*. Here I give you the structure of those files as it could be useful to convert from and to other software. Note however that *Biomapper* can only work with binary files (no ASCII files).

Each map is composed of two files:

- A metadata file (or documentation file) that contain important information about their corresponding map data files.
- A data file that stores the actual information layer.

*Biomapper* can read and write two kinds of files: the one used by *Idrisi 2* and the one used by *Idrisi32*. You can switch between them by mean of the menu command *Ecogeo.map/Options...*, *tab files*.

By default, *Biomapper* will use the *Idrisi32* format. If you want to use by default the *Idrisi 2* format, place *Biomapper* in the *Idrisi* directory.

## Metadata file structure

Image Documentation files contain important information about their corresponding Image data files (i.e., its metadata, or header information). Whenever a *Biomapper* module accesses a map file, it also accesses the accompanying Image Documentation file. Metadata files are specified by a ".doc" extension.

### File Contents

Following are the fields that are stored in Image Documentation files. They may be broken down into four major groups:

#### **Information about the image as a whole:**

- title: A descriptive name of the file.
- data type: The type of numbers stored in the file. Allowable entries are byte, integer and real. (See the description for Data files.)
- file type: The format in which the Image file is stored. Allowable entries are ASCII, Binary and Packed Binary (See Data files.) but *Biomapper* can only read binary file type.
- columns: The number of columns in the image. This is extremely important as it tells *Biomapper* modules how to construct the rectangular image from the stored values. (See Data files.)
- rows: The number of rows in the image.

#### **Information about the georeferencing system of the file:**

- ref. system: The name of the geographic referencing system used with the file. This may be Plane, Lat/Long, or a specific referencing system defined by a Reference System Parameter file.
- ref. units: The unit of measure used in the specified reference system. Allowable entries are m, ft, mi, km, deg and radians.
- unit dist.: The scaling factor between the given coordinates and actual measurements on the ground. This will almost always be 1. The unit distance answers the question, "If I move one unit in the reference system described here, how far have I moved on the ground, measuring in reference units?"
- min X: The minimum X coordinate (left edge) of the map.
- max X: The maximum X coordinate (right edge) of the map.
- min Y: The minimum Y coordinate (bottom edge) of the map.
- max Y: The maximum Y coordinate (top edge) of the map.
- pos'n error: A measure of the accuracy of the positions in the image. Not used by *Biomapper*.
- resolution: The inherent resolution of the image.

#### **Information about the values stored in the file:**

- min value: The minimum value in the map.
- max value: The maximum value in the map.
- value units: The unit of measure of the values in the map. It is suggested that the term classes be used for all qualitative data sets, and that whenever standard linear units are appropriate, that the same abbreviations that are used for reference units should also be used (m, ft, mi, km, deg, rad).
- value error: This field records the error in the data values that appear in image cells. It is not used by

*Biomapper*.

- flag value: Any value in the image that is not a data value, but rather has a special meaning. If there is no flag value, this entry should remain blank. *Biomapper* uses this flag to mark the background cells.
- flag def'n: Definition of the above flag value. When background cells are present, this field must contain "background".
- legend cats: The number of legend categories present. Not used by *Biomapper*.

#### **Other information about the file:**

The following four entries are optional, and any number of each may be entered at the end of the file, so long as each has the correct term in the 14 character descriptive field to the left. These are all text fields and are included to facilitate complete documentation of the Image file.

- comment: Any additional information about the data may be recorded here.
- lineage: Description of the history by which the values were recorded/derived. Most *Biomapper* modules write into this field.
- completeness: The degree to which the values describe the subject matter indicated.
- consistency: The logical consistency of the file.

## **Creation**

Documentation files may be created using any text editor. They are created automatically by any *Biomapper* module that produces a map file.

## **Structure**

Map Documentation files are stored in ASCII format. The first 14 characters of each line describe the contents of the line, while the remaining characters contain the actual information.

### **Example**

For example, the documentation file for a soils image might look like this:

```
file title      : Major Soils Groups
data type      : byte
file type      : binary
columns        : 512
rows           : 480
ref. system    : plane
ref. units     : m
unit dist.     : 1
min. X         : 0
max. X         : 15360
min. Y         : 0
max. Y         : 14400
pos'n error    : unknown
resolution     : 30
min. value     : 0
max. value     : 3
value units    : classes
value error    : 0.15
flag value     : 0
flag def'n    : background
legend cats    : 0
lineage        : Soil polygons derived from 1:5000 scale color air photography
lineage        : and ground truth, with the final compilation being adjusted to
lineage        : the map base by hand.
comment        : Value error determined by statistical accuracy assessment
comment        : based on a stratified random sample of 37 points.
```

## **Idrisi32**

The only crucial change between the file structures of *Idrisi 2* and *Idrisi32* is the adjunction of a new field at the beginning of the file:

```
file format      : IDRISI Raster A.1
```

Furthermore, they are now specified by a ".rdc" extension.

# Map file structure

Map files store the raster data layers used in analyses. They are specified by a “.img” file extension.

## File Contents

Map files contain the attribute value of each cell in the map.

## Creation

Map files are created by most *Biomapper* modules.

## Structure and Examples

While the logical structure of an Map file is a grid, the actual structure, as it is stored, is a single long column of numbers. For instance, a map consisting of 3 rows by 5 columns is stored as a single column of 15 numbers. It is the Metadata file that allows *Biomapper* modules to construct the grid from this list. An image that looks like this:

10	15	9	10	1
1	14	10	11	13
14	13	11	10	12

has a map file that looks like this:

10  
15  
9  
10  
1  
1  
14  
10  
11  
13  
14  
13  
11  
10  
12

The numbers in an map file may be integer, byte or real. This is termed the data type.

1. Integers are whole numbers within the range - 32768 to + 32767.
2. Byte values are positive integer numbers ranging from 0 to 255.
3. Real numbers have a fractional part, or are whole numbers outside the integer range.

The real data types can store values within a range of  $\pm 1 \times 10$  to the power of 38 with a precision of 7 significant figures.

*Biomapper* can only use binary map files.

## Idrisi32

The only change between the file structures of *Idrisi 2* and *Idrisi32* is the extension of the file: it is now “.rst”

# What to do if an eigenvalue is negative?

After the ENFA, the eigenvalues are the first thing to check: they must all be greater than 0. It may happen that one or more are negative (check the last eigenvalues as they are sorted by decreasing order). This means that either the global or the specie correlation matrix was nearly singular and that the inversion algorithm produced absurd results. A matrix is singular is a matrix where one of the columns (or rows) is a linear function of one or

several other, or where one of the columns is filled with zeros. Generally this is the latter happens in our case.

It means that two or more EGV maps are too highly correlated and they must be removed. To decide which map must be removed, examine the correlation tree ([View/Correlation tree](#)) to see which are the most correlated maps. Remove one of the most redundant pair ([Right click](#) on the map then [Remove](#)) and launch the ENFA again (as the covariance matrix is already computed, the process will be far shorter). Repeat these check-remove-compute operations until all the eigenvalues are null or positive. Don't fear a loss of information: the removed maps contain mostly redundant data.

### **Habitat suitability model (HS model)**

This is a mathematical function, an algorithm or a set of rules that assigns – for a focal species – a habitat quality to a given set of ecogeographical variables (EGV). The habitat suitability represents how close the predictor values are to the ecological needs of the species.

### **Habitat suitability map**

This is a raster map whose cells contain habitat suitability index computed by the mean of a HS model and several Ecogeographical maps.

### **Ecogeographical variable (EGV)**

This is a spatially defined variable that is associated to every location of a study area. It describes quantitatively features that can be ecological (e.g. frequency of forest, nitrates concentration), topographical (e.g. altitude, slope) or anthropic (e.g. distance to the nearest town, roads density).

### **Ecogeographical map**

This is a raster map whose cells contain the value of a given ecogeographical variable measured on the underlying area.

### **Species map**

This is a boolean raster map whose cells contain a 1 when there is a proof of the focal species presence in the underlying area, and a 0 when such a proof doesn't exist. Note that a 0 does NOT mean that the underlying area is unsuitable for the species.

### **GIS**

**Geographical Information Systems** are programs designed to input, maintain and analyse spatially referenced numerical data, in order to simulate natural phenomena and to support management and decision-making.

### **Marginality**

The marginality ( $M$ ) is a measure of the departure between the average of the species distribution ( $m_S$ ) and the average of the global distribution ( $m_G$ ). Mathematically, it is defined as:

$$M = \frac{|m_S - m_G|}{1.96s_G}$$

Where  $s_G$  is the standard deviation of the global distribution.

### **Specialisation**

The specialisation ( $S$ ) is a measure of the choosiness of the species about the available range of the EGVs. It

mathematically defined as the ratio between the standard deviations of the species distribution( $s_s$ ) and the global distribution ( $s_g$ ):

$$S = \frac{s_g}{s_s}$$

S varies between 1 and the infinity.

Note that we can for more convenience also define the *tolerance* ( $T$ ) as the inverse of the specialisation and which varies between 0 (very specialised) and 1 (ubiquitous):

$$T = \frac{s_s}{s_g}$$

### **Tolerance**

The Tolerance ( $T$ ) is a measure of the choosiness of the species about the available range of the EGVs. It mathematically defined as the ratio between the standard deviations of the global distribution ( $s_g$ ) and species distribution( $s_s$ ):

$T$  varies between 0 (very specialised) and 1 (ubiquitous).

$$T = \frac{s_s}{s_g}$$

Note that we can also define the *specialisation* ( $S$ ) as the inverse of the tolerance and which varies between 1 and the infinity:

$$S = \frac{s_g}{s_s}$$

### **Species Distribution**

This is the distribution of the EGVs for the localities where there is a species presence proof (see species map). By contrast, the *global distribution* is the distribution of the EGVs for the whole study area.

### **Global Distribution**

This is the distribution of the EGVs is the distribution of the EGVs for the whole study area. By contrast, the *species distribution* is the distribution of the EGVs for the localities where there is a species presence proof only.

### **Project**

Projects are the working units of Biomapper. They contain all the information needed to conduct an analysis from the input EGV maps to the HS map. A project is composed of the following elements:

- A set of overlayable EGV maps
- A species presence map
- The covariance matrix of these maps
- The scores matrix computed by the ENFA
- A results list
- A set of working maps (not necessarily overlayable)

Note that those elements are not necessarily present; they will be added one after the other during the analysis process.

### **Overlayable maps**

To be overlayable, two maps must comply with the following rules:

- The same area thou shalt cover
- The same resolution thou shalt have
- The same spatial unit thou shalt use

### ***Discrepancies between maps***

Two overlayable maps are discrepant if one can find at least one cell that is a background cell in one map and a value cell in the other.

### ***Background***

Every Biomapper's map can have two kind of cells:

- 1° **value cells** contain a meaningful values (binary, qualitative or quantitative)
- 2° **background cells** contain no usable value (they are identified by an arbitrary **background value**).

### ***Boolean map***

A boolean (after the mathematician and logician George Boole (1815-1864) ) map is a raster map whose value cells contain only 0 or 1. Generally, 1 indicates the presence of an item and a 0, its absence.

### ***FAQ or Frequently Asked Questions***

This is a document I update frequently to report all questions asked by Biomapper users, along with my answers. A lot of information can be found there, both practical and statistical.

You can access a local version of the FAQ through the [Help/FAQ](#) menu in Biomapper. There you will find the FAQ in its latest version when you downloaded Biomapper. To find a brand new version, check the Biomapper website at <http://www.unil.ch/biomapper/faq.html>.

### ***Factor map***

This is a map that summarise all ecogeographical variables according to the score matrix computed by the ENFA. Whilst you can look at them as any other map, they are mainly temporary files used by Biomapper to compute a habitat suitability map.