On GARP

A. Townsend Peterson, 2011

The GARP program and a rudimentary manual can be downloaded from
http://www.nhm.ku.edu/desktopgarp/Download.html.

Preparing Environmental Data

The environmental data layers for input into GARP must be in raster format, and must have data (i.e., not have the value specified for 'no data') across the entire study region. What is more, all of the layers must be cast on exactly the same grid system, as will be illustrated below. Once the layers are exported as ASCII raster files, they are then processed with GARP Dataset Manager, a small module that downloads with the program. These steps are shown in greater detail in this section.

Environmental data for niche modeling can be drawn from many sources, reflecting the many suites of environmental factors that may affect distributions of species (see **Que Datos Debo Incluir?**). Some common sources of environmental data are the following:

The original version of desktop GARP is prepared for use with ArcView, and many of these modules are still the best and most functional means of preparing data for niche modeling in GARP. The following is a brief summary of how to export data using these modules from ArcView. For users with ArcGIS, suggestions are given below.

Upon downloading and installing desktop GARP, an extension "GARPDatasets" is automatically installed in the extensions folder of ArcView. This extension should be available by choosing File < Extensions in ArcView, and clicking it on.
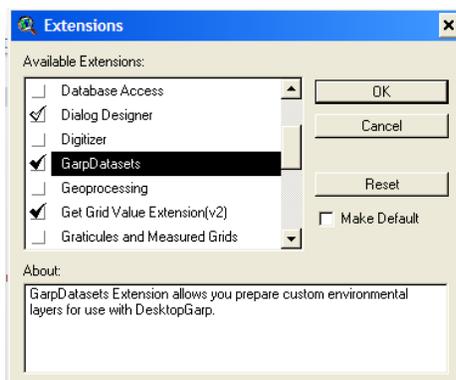

Once the extension is turned on, it shows up as "GARP" in the main menu of ArcView.

With the raster datasets of interest loaded as grids in ArcView, one initiates the extension, and then must provide a path where the datasets are to be deposited. Note that the path *must* end with a "\".


Then, one must choose a clipping method by which to reduce the environmental data grids to the study area, which (as is justified in the **conceptual section**). The simplest are user-input coordinates, or a rectangle drawn on the map using the graphics draw capabilities in ArcView (the rectangle must be selected). A slightly more complicated option is that of providing a polygon shapefile that has a single element in its attributes table. Regardless, one chooses among these options on the next page:
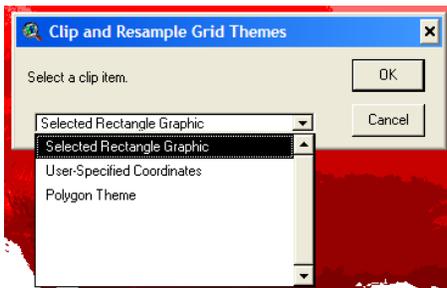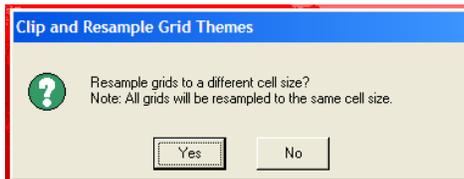

Then, one selects the grids to be clipped:
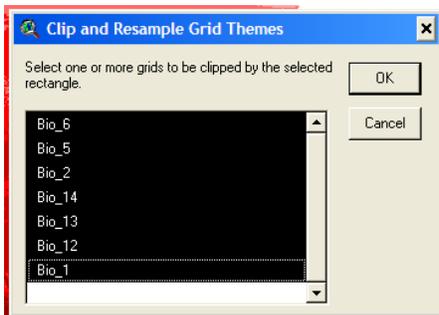
The program then provides a summary of the spatial resolutions associated with each of the grids:
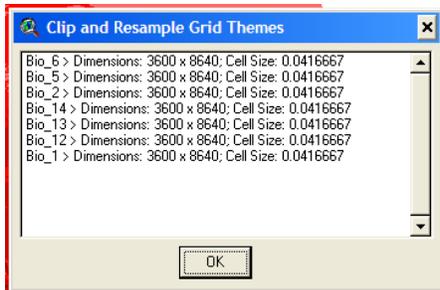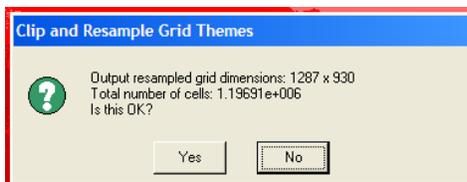
One is then asked if one wishes to resample grids to a different cell size (answer yes!):
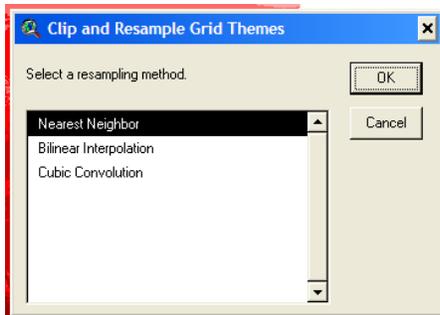

And then one specifies the resolution to which all of the grids should be resampled:
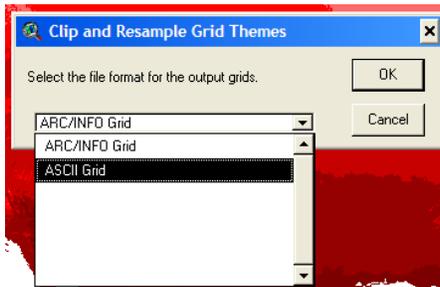

And the program provides some information about the grids that will result:

Then, one is asked about the manner of resampling. One normally would choose "Nearest Neighbor," but "Bilinear Interpolation" and "Cubic Convolution" are provided as additional options.

One then chooses the format of the grids to be output, as follows:

The key feature of the ASCII raster grids that are output is that they have exactly the same 'header' information, that is the specifications of the position and size of the grid. Opening one of the ASCII raster grids, this header information looks like this:


Note that GARP requires an additional layer, which must be named "mask.asc". This layer serves to restrict the analyses that GARP develops to a particular area, which (see conceptual section) should be equivalent to the "M" in the BAM diagram. Frequently, this layer is simply a copy of one of the other environmental layers.

The next step involves using the program module *GARP Dataset Manager*, which downloads and installs with the main GARP program. This program has two functions: (1) it rescales all values in each of the environmental grids to the range 0-254, and (2) it removes all line breaks, creating in effect a single vector of values across the entire region of analysis. The program is initiated from the main menu.

To start this program in processing the datasets for use in GARP, select Load Layers < from Ascii Raster Grids
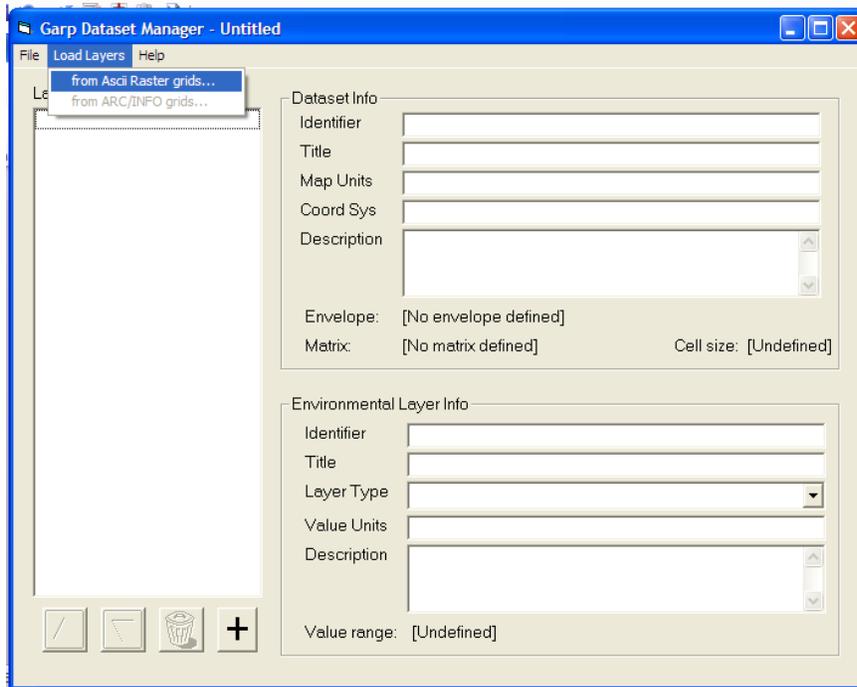
One then navigates to the folder in which the ASCII raster grids are stored, and one clicks on any of the grids. All ASCII raster grids in the folder will load automatically. Note that this process can be time-consuming, particularly if the grids that you are using are large.

One must then provide an identifier for the data set … this is accomplished by filling in the top two lines on the left side of the program interface with the same identifier:
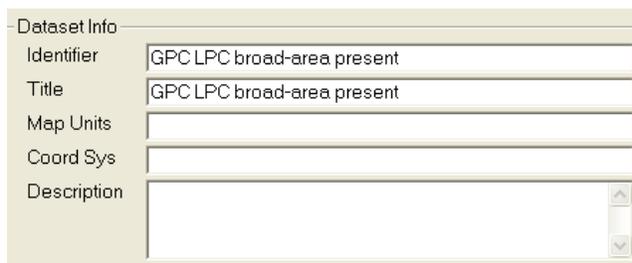
This identifier is what will appear in the main GARP program as the tag for this particular data set.

Finally, the user must recall that GARP's use of these datasets is "dumb"; that is, if you are planning to project among different environmental datasets, the only guide that GARP has for connecting different datasets is the order of the data layers in the GARP data file (i.e., what is produced by this program). As a consequence, you must put the different data layers in the same order, among different GARP datasets. The program has a panel of buttons for reordering data layers, as well as removing and adding them to the data set (recall, though, that the headers of the different environmental layer ASCII files must be *exactly* the same, or this program will bomb!):

Note also that Maxent's use of different data layers when projecting from one dataset to another is different, so you must pay careful attention to this difference between the two programs. After the data layers are in the proper order, one simply saves (File < Save) the GARP dataset, generally in the same folder as the data layers, and one is ready to roll on actual analyses using GARP.

Using Desktop GARP

The desktop GARP interface is divided into 6 panels. For each of these panels, a description of the parameters involved will be provided. The basic interface looks like this:
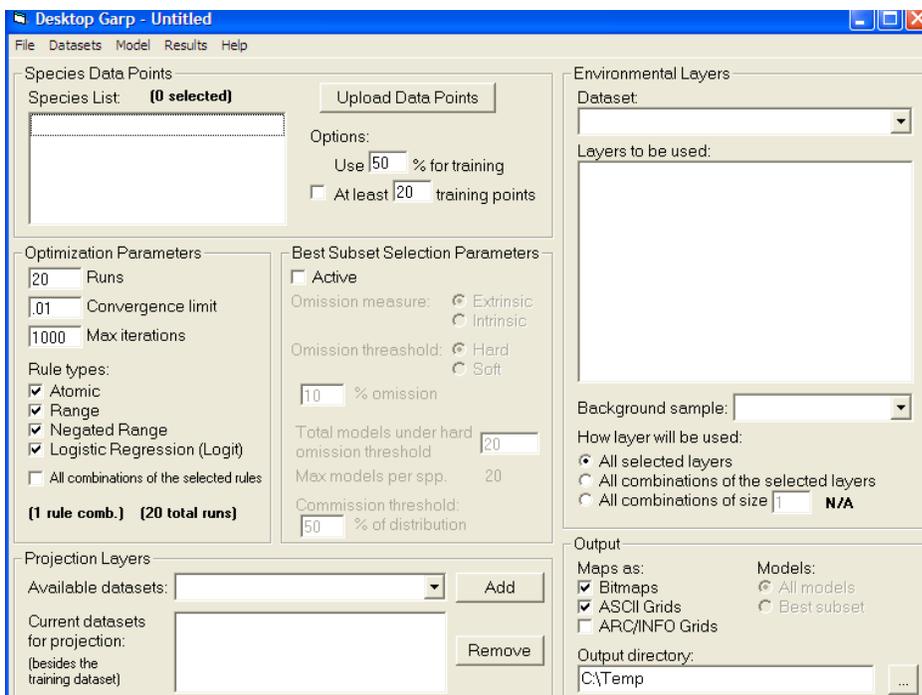
1. Species Data Points:

Several options are available for inputting occurrence data. First, one can use an ASCII format text file, but this option can be messy, in light of the relaxed formatting in such files. Also, one can use a point shapefile from Arc—this files must have one field that identifies the species, and the GARP program will automatically pick up on the latitude and longitude. Finally, and most simply, one can use an Excel 2003 file (suffix .xls) to provide the data to the program, in which three columns are present—the first is the species identifier, and the 2nd and 3rd are the longitude and latitude. There are two points that must be noted: (1) as the program reads from top to bottom in the file, any change in the species identifier will indicate another species for analysis; as a result, one must sort the file by the species identifier, such that all records of a given species are together; (2) again, the program is a bit "dumb," and so the 2nd column must be longitude, and the third must be latitude. An example of a formatted data file is this:

Hitting the button "Upload Data Points," the file loads automatically. If there are multiple species in the list, then multiple lines will appear in the "Species List" window.

2. Optimization Parameters

Here, one specifies the number of replicate models ("runs") that one wishes to develop. The convergence limit is rarely changed, but can be reduced to make the algorithm process longer and in greater detail—if this is done, then the Max iterations should be increased accordingly. Rule types can be turned on and off, but all are known to contribute significantly, so it is recommended that all be left on.

3. Environmental Layers

To add environmental datasets to an analysis, one chooses Datasets < Scan Directory,
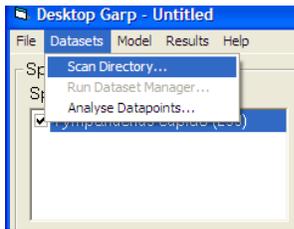
and then one identifies the directory in which all of the environmental datasets of interest are located. In this case, 5 environmental datasets are contained within the folder "Future." By typing any character in the File Name area, and hitting open, the program will detect all datasets in any folder or subfolder of this directory.
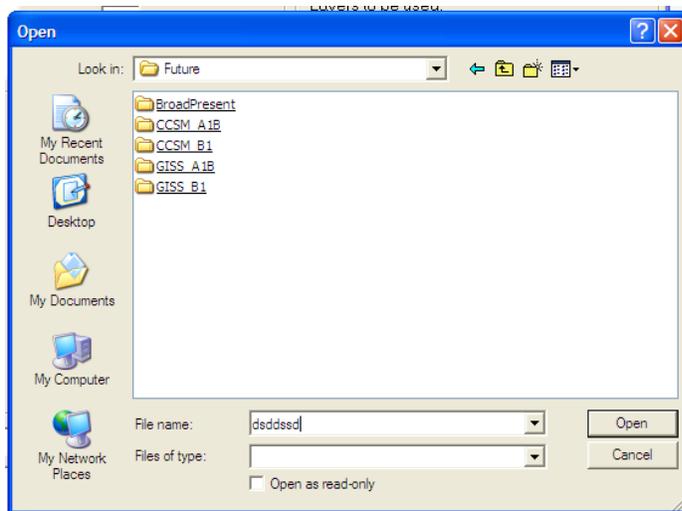
Once one has scanned the directory, one can use the picklist at "Dataset" to choose a dataset. The dataset in this window will be used for model calibration. One can also turn on and off specific layers via checking an unchecking the boxes next to each particular layer. Finally, one can do a variety of jackknife manipulations of inclusion or exclusion of layers in the analyses, with a greater flexibility than is offered by Maxent.
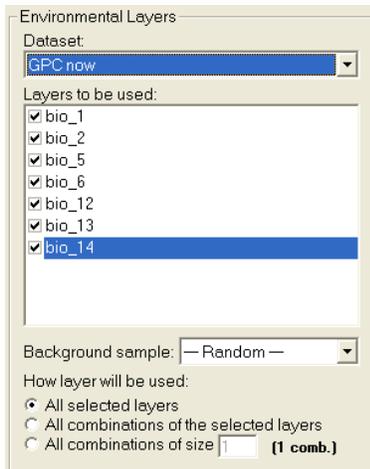
4. Projection Layers

In this box, one can specify up to 9 environmental datasets onto which the models trained in the analyses will be projected. They will appear in the results with suffixes "_1", "_2", etc., in order corresponding to their order in this list.
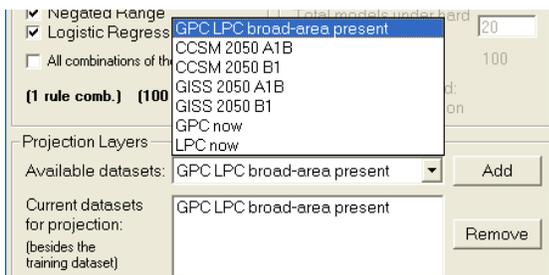
5. Output:

In this window, one specifies the path of the folder in which one wishes for the results to be placed. One can choose among three types of output, although ASCII Grids and ARC/INFO Grids are the only useful ones. Specifications for enabling ARC/INFO Grid output are provided in the manual to desktop GARP. The buttons under Models are not enabled and make no difference in the output.

6. Best Subset Selection Parameters

This panel includes the most complicated set of options and selections that must be made. The user is referred to a detailed publication (Anderson et al. 2003) for a close understanding of these options. However, a basic summary is as follows:

a. Omission measure: indicates whether the data set aside from modeling in panel 1 are used for calculations of omission rates, or whether the actual data used in model calibration are used. "Extrinsic" is almost always used in this case.

b. Omission threshold: indicates whether an absolute ("hard") or relative ("soft") criterion is used for omission criteria.

c. % distribution: this percentage is either the absolute value (i.e., omission rate <5%) or the relative value (i.e., the 20% of replicate models that have the lowest omission rates), depending on (b).

d. Commission threshold: this field indicates the proportion of the low-omission models identified in (c) to be retained in the best subsets section.

```
Best Subset Selection Parameters
☑ Active
Omission measure:     ⦿ Extrinsic
                       ○ Intrinsic
Omission threashold:  ○ Hard
                       ⦿ Soft
 20    % distribution

Total models under hard
omission threshold        20
Max models per spp.      100
Commission threshold:
 50    % of distribution
```

The results are organized as follows. Within the target directory (in the example, "G1_LPC"), there is a directory "BestSubsets" that includes a subdirectory for each species, and that for each species includes a subdirectory for each projection. The projection to the environmental layers on which the model was calibrated is the "_0" subdirectory. Subsequent projections are in the same order as on the configuration page.
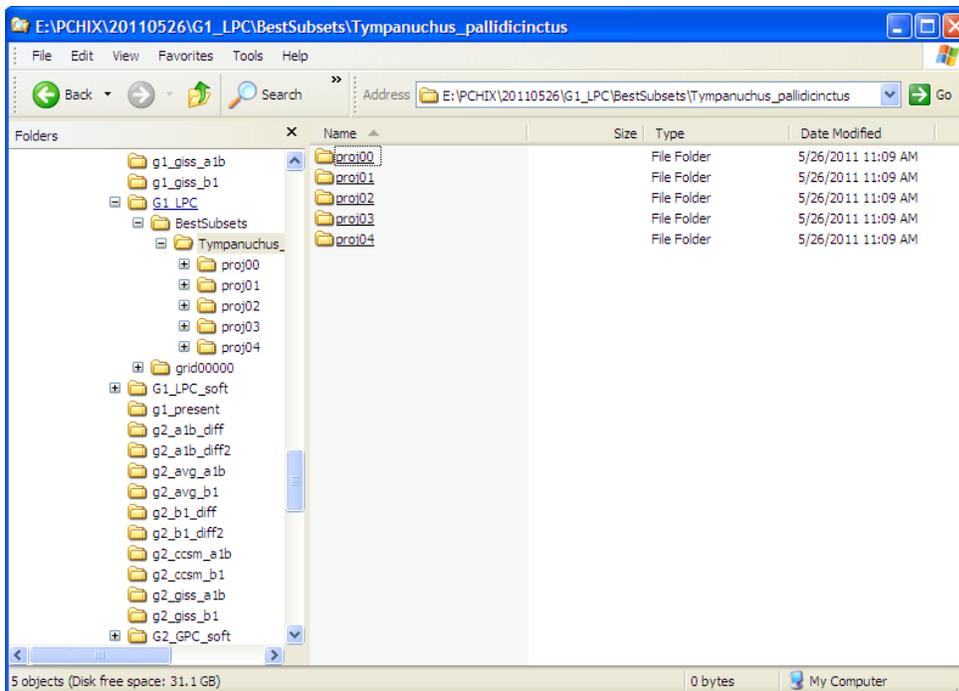
Finally, the program exports a "Results" file, both in ASCII and Excel formats. This file provides quite a lot of information about the model, including the following fields:

Field name	Description
Task	A running count of the models having been run as part of this analysis.
Run	A running count of the models run for this particular species.
Species	The species that is under analysis in this particular case.
Atomic Rules	Inclusion of atomic rule types in this analysis.
Range Rules	Inclusion of range rule types in this analysis.
Negated Rules	Inclusion of negated range rule types in this analysis.
Logit Rules	Inclusion of logit rule types in this analysis.
Iter.	The iteration at which the model evolution ceased.
Conv.	The change in the optimization parameter that caused the analysis to stop.
Train Acc	Overall correct prediction of training data.
Pr:Pr/Ac:Pr	Elements of the confusion matrix based on training data.
Pr:Ab/Ac:Pr	Elements of the confusion matrix based on training data.
Pr:Pr/Ac:Ab	Elements of the confusion matrix based on training data.
Pr:Ab/Ac:Ab	Elements of the confusion matrix based on training data.
Test Acc	Overall correct prediction of the intrinsic testing data.
Pr:Pr/Ac:Pr	Elements of the confusion matrix based on intrinsic testing data.
Pr:Ab/Ac:Pr	Elements of the confusion matrix based on intrinsic testing data.
Pr:Pr/Ac:Ab	Elements of the confusion matrix based on intrinsic testing data.
Pr:Ab/Ac:Ab	Elements of the confusion matrix based on intrinsic testing data.

Total Area	The total area (in pixels) that is involved in the analysis (i.e., calibration area).
Presence Area	The total area (in pixels) that is predicted as suitable.
Absence Area	The total area (in pixels) that is predicted as unsuitable.
Non-predicted Area	The total area (in pixels) that is not predicted as suitable or unsuitable, taken as predicted as unsuitable
Yes	The number of extrinsic testing points that was correctly predicted as suitable.
No	The number of extrinsic testing points that was incorrectly predicted as unsuitable.
ChiSq	A chi-square statistic based on the preceding 5 fields--not in use these days.
p	The p-value associated with the chi-square statistic in preceding field.
Commission	The proportion of the overall study area that was predicted as suitable.
Omission (int)	The proportion of training occurrence points that was predicted incorrectly as unsuitable.
Omission (ext)	The proportion of the extrinsic testing occurrence points that was predicted incorrectly as unsuitable.
Status	Whether the model was processed.
Message	Whether the model processing was successful.
Layers:	The remaining columns indicate the inclusion or exclusion of different environmental layers in the analysis.