Desktop Garp

Users Manual

Table of Contents

Introduction

DesktopGarp is a software package for biodiversity and ecologic research that allows the user to predict and analyze wild species distributions.

DesktopGarp is a desktop version of the GARP algorithm. The acronym stands for Genetic Algorithm for Rule-set Production. GARP was originally developed by **David Stockwell** at **ERIN Unit of Environment Australia** and enhanced at the **San Diego Supercomputer Center**. An interactive version of the algorithm, implemented via web, can be found at the **Biodiversity Species Workshop** website.

GARP is a genetic algorithm that creates ecological niche models for species. The models describe environmental conditions under which the species should be able to maintain populations. For input, GARP uses a set of point localities where the species is known to occur and a set of geographic layers representing the environmental parameters that might limit the species' capabilities to survive.

GARP searches iteratively for non-random correlations between species presence and absence and environmental parameter values using several different types of rules. Each rule type implements a different method for building species prediction models. Currently there are four types of rules implemented: atomic, logistic regression, bioclimatic envelope, and negated bioclimatic envelope rules. For a comprehensive description of GARP algorithm, read David Stockwell's **GARP Technical Manual and Users Guide**.

System Requirements

The minimum system configuration to run **DesktopGarp** is:

- Intel/Pentium based computer;
- Windows 98 SE (Second Edition), Edition (ME), NT 4, 2000 (all versions) and XP (all versions);
- 64 MB of RAM;

- 5 GB on the hard disk.

Additionally, the recommended system configuration, based on average size experiments, is:

- 256 MB or more RAM - The more the better: the optimal amount of RAM depends on how large your dataset is (number of cells), the number of layers you have in it and the number of tasks within an experiment. The amount of memory can be estimated as: (#cells per layer) X (#layers in datasets) X (10KB X # of tasks)
- 40GB of hard disk, preferably a fast SCSI disk. The results can be quite numerous and large, depending on the setup of your experiment;
- Microsoft Excel 97, 2000 or XP - to process input data and output results;
- ESRI ArcView with Spatial Analyst Extension - to create custom datasets and analyse resulting distribution maps;
- Multi-processor machines **WILL NOT** improve performance of a single experiment, as DG is currently completely sequential (not parallel). The gain will only be felt if two or more experiments are running at the same time (each experiment will take one processor).

Installation

Run the installation executable downloaded from our website. Locate **DesktopGarpSetup_1_1_3.exe** using Windows Explorer and double click it. If you choose to open the file directly from the website, the installation process will start immediately after the download is complete.

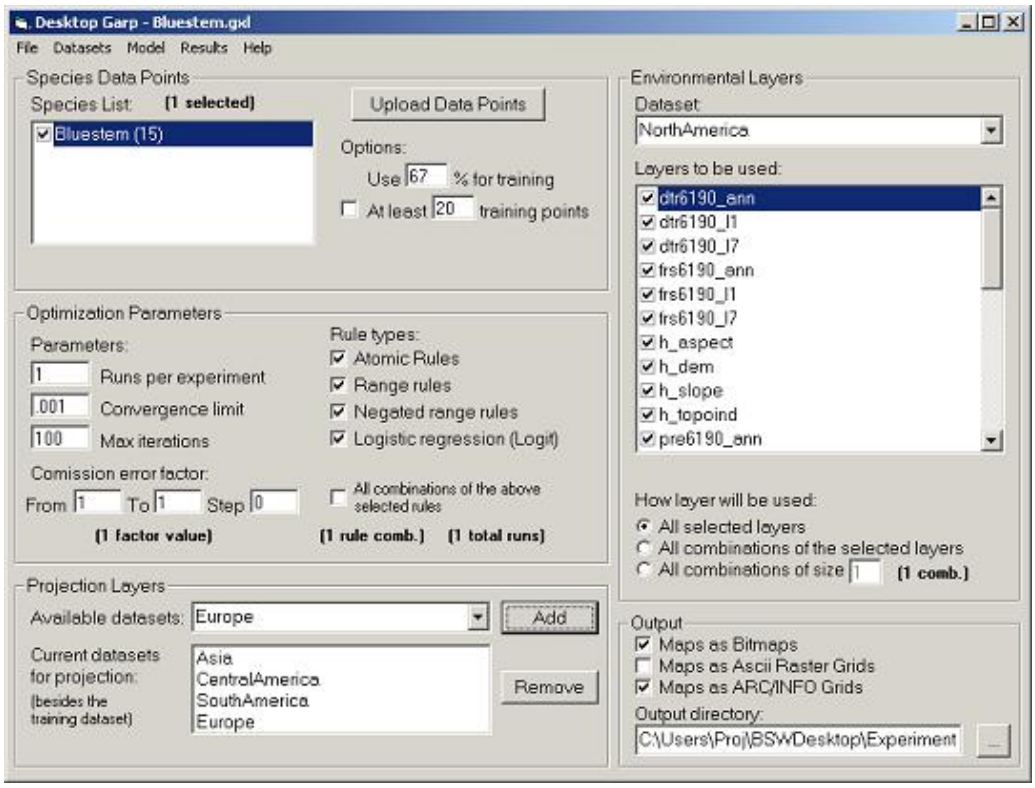
When the installation process starts, follow the instructions displayed on the screen.

During installation, you will be prompted to install an additional software package called Microsoft XML Parser. Read and accept the license agreement if you agree with it, and go through the steps as described on the screen. Note that DesktopGarp will not work if you do not have this Microsoft package installed on your system.

Once both packages are installed, you are ready to use DesktopGarp.

Using DesktopGarp

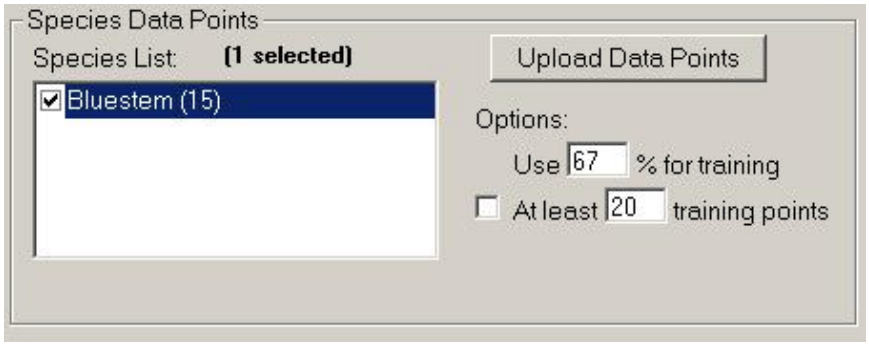
The DesktopGarp user interface is relatively simple. It contains just one window where the user specifies all the parameters and data to be used in the experiment. Below is a sample of the interface.



Below is a detailed functional description of each user interface panel.

Input of Data Points

The Species Data Points panel handles the species occurrence (point) data. A sample of this area is shown below.



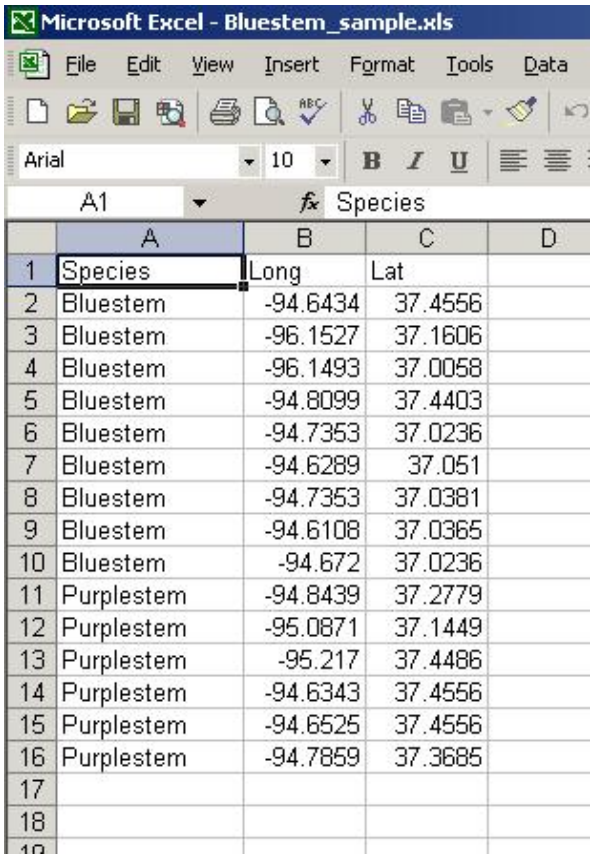
New species occurrence information can be entered by clicking the Upload Data Points button. It will open a dialog box to specify the location of the occurence data file. Currently three formats are supported: Comma delimited, MS Excel Spreadsheets and ArcView Shapefiles.

Comma delimited and Excel files should contain three columns: the first one for species name, the second for longitude, and the third for latitude. The first line is ignored, so it can be used for labels.

In this version, DesktopGarp accepts only files in this format, so make sure the columns are ordered: species name, longitude and latitude. Notice that longitude comes before latitude.

Each line of the file represents a single data point entry for the species. Data points for the same species must come together. Different species names define different species for the software.

Below is a sample Excel worksheet and the corresponding comma delimited file containing species information for two species. The actual files can be downloaded at **MS Excel sample** and **comma delimited sample**.



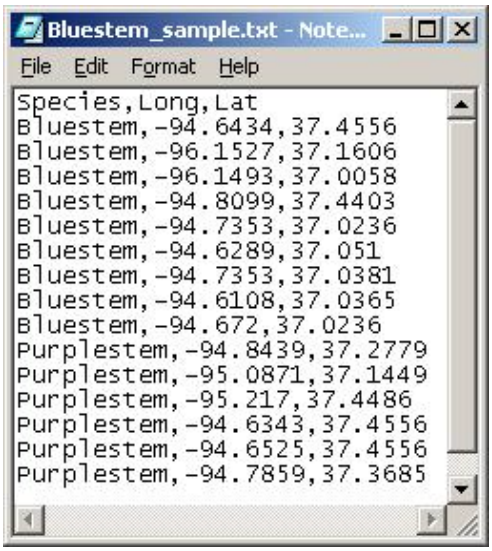
Microsoft Excel - Bluestem_sample.xls

File Edit View Insert Format Tools Data

Arial 10 B I U

A1 fx Species

	A	B	C	D
1	Species	Long	Lat	
2	Bluestem	-94.6434	37.4556	
3	Bluestem	-96.1527	37.1606	
4	Bluestem	-96.1493	37.0058	
5	Bluestem	-94.8099	37.4403	
6	Bluestem	-94.7353	37.0236	
7	Bluestem	-94.6289	37.051	
8	Bluestem	-94.7353	37.0381	
9	Bluestem	-94.6108	37.0365	
10	Bluestem	-94.672	37.0236	
11	Purplestem	-94.8439	37.2779	
12	Purplestem	-95.0871	37.1449	
13	Purplestem	-95.217	37.4486	
14	Purplestem	-94.6343	37.4556	
15	Purplestem	-94.6525	37.4556	
16	Purplestem	-94.7859	37.3685	
17				
18				



Bluestem_sample.txt - Note...

File Edit Format Help

```
Species,Long,Lat
Bluestem,-94.6434,37.4556
Bluestem,-96.1527,37.1606
Bluestem,-96.1493,37.0058
Bluestem,-94.8099,37.4403
Bluestem,-94.7353,37.0236
Bluestem,-94.6289,37.051
Bluestem,-94.7353,37.0381
Bluestem,-94.6108,37.0365
Bluestem,-94.672,37.0236
Purplestem,-94.8439,37.2779
Purplestem,-95.0871,37.1449
Purplestem,-95.217,37.4486
Purplestem,-94.6343,37.4556
Purplestem,-94.6525,37.4556
Purplestem,-94.7859,37.3685
```

Choosing ESRI Shapefile brings up a prompt for the species name column. In this case, the feature class should be of type Point.

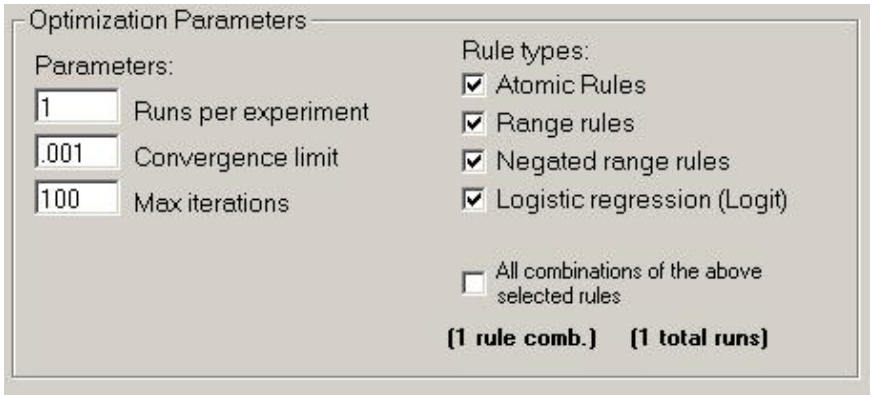
The list box will present all species loaded and the number of data points (in parenthesis) for each one. The check box to the left of each species in the list allows the user to control which species from the list will be used in the experiment.

In this panel, the user can also specify two parameters that define how the data will be sampled and used. The first option allows the user to specify what percentage of points will be used for training, i.e., model construction. The remaining points will be used for testing. If 100% is specified for training, no significance test will be performed on the models.

The second option allows the user to specify a minimum number of points to be used for training. To enable it, check the box to the left of this option. When enabled, it will override the percentage value and use at least the specified number of points for training. This option is useful for species with few data points because it forces the program to use a minimum number of points for analysis. The algorithm typically does not perform well with fewer than 20 data points for training.

Optimization Parameters

On the Optimization Parameters panel, the user can specify some parameters that control the overall behavior of the genetic algorithm. A sample of this panel is shown below.



Optimization Parameters

Parameters:

1 Runs per experiment

.001 Convergence limit

100 Max iterations

Rule types:

- [x] Atomic Rules
- [x] Range rules
- [x] Negated range rules
- [x] Logistic regression (Logit)

- [ ] All combinations of the above selected rules

(1 rule comb.) **(1 total runs)**

The number of runs per experiment defines how many times each distinct task will be performed within the experiment. For example, for two species and 10 runs per experiment, 20 runs in that experiment will be executed: 10 for the first species and

10 for the second one.

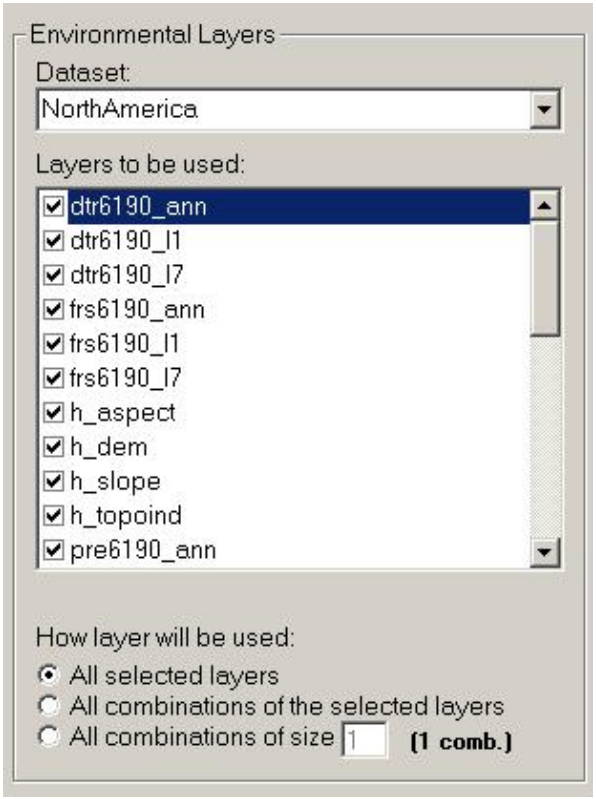
The convergence limit establishes a stop condition for iterations within the genetic algorithm. Its behavior varies depending on how difficult or easy the problem is. Usual values are between 0.01 and 0.10. If this parameter is set to 0, the algorithm will stop only when the maximum number of iterations is reached.

Max iterations value establishes another stop condition for the genetic algorithm. It forces the optimization to stop at the specified iteration, even if the convergence limit has not been reached yet. More iterations tend to yield more stable results. Usual values are between 100 and 1000.

The rule type checkboxes allow the user to specify which algorithm is used to produce rules in the species model. For a complete discussion on each rule type, please refer to David Stockwell's **GARP Technical Manual and Users Guide**.

The all combinations checkbox generates one task for each combination of the checked rules. For example, if range, logit and atomic rules are checked, DesktopGarp will create tasks where only each of those rules are used, then one for range and logit rules, one for range and atomic rules, one for logit and atomic rules, and one for all three rules combined. This is useful for analyzing the impact of each particular rule on the results. The labels below the checkbox show how many combinations will be created and also the total tasks or runs that will be executed (combinations times runs).

Native Range Datasets



The Native Range Datasets or Environmental Layers panel allow the user to define the environmental coverages that will be used as input for the prediction. The algorithm will try to correlate the input data points to the values on those layers to get the final prediction.

The dataset combo box displays the choices for the dataset that will be used on the experiment. The datasets listed on this combo box are the ones scanned using the menu option Datasets->Scan directory....

Once the dataset has been chosen, DesktopGarp will automatically list all layers present on that dataset on the layers to be used list box. There, the user can control which layers will be used by clicking on the checkbox that appears to the left of each layer name.

Note: There is a sample dataset that ships with DG. It is located on directory **SampleDataset**, under the installation directory (by default it is C:\Program Files\DesktopGarp\SampleDataset).

Below the layers list, there are three radio buttons that define how the selected layers will be used. The first one, all selected layers, will force DesktopGarp to use all selected layers in the optimization.
All combinations of selected layers will cause the experiment to have one task for each possible combination of the selected layers.
The all combinations of selected size N radio button has similar effect, but will limit the experiment to the combinations that contains exactly N layers.
The last two alternatives using combinations of layers are useful for determining which layers are important to a species. A method for analyzing that would be using linear multiple regression to predict the error values (omission and commission), using the information on whether a particular layer was used on a task as an independent variable.

Note: The use of combinations of layers may cause the number of tasks within the experiment to be too large. There is a label next to the bottom of the panel that shows how many combinations that setup will yield. Tests have shown that DesktopGarp can handle well up to 10,000 tasks in the same experiment.

Output Parameters

The Output panel specifies the output prediction map format and the output directory for maps and other generated documents.

The prediction maps can be generated in three formats:

- **bitmaps**: MS Windows bitmaps, with extension "*.bmp*";
- **ASCII raster grids**: ASCII text format, with extension "*.asc*";
- **ESRI Arc/Info grids**: ESRI proprietary format for grid spatial data storage and management. A separate directory is created for each grid.

Note: In some systems, the last checkbox (map output as Arc/Info grids) will be disabled (grayed out) due to the lack of an important ESRI library, the **Grid I/O** library, on the system. To have this option enabled, follow these instructions:

- Install **ESRI ArcView GIS**;
- Install **ESRI Spatial Analyst Extension**, choosing **Custom Setup** and including **GRID I/O** on the installation;
- Add the GRID I/O library to the system path, as follows:
  - On Windows 2000 and XP operating systems, right-click the **My Computer** icon on the desktop and select the menu item **Properties**. Click the **Advanced** tab and click the **Environmental Variables...** button. On the **System variables** list (on the bottom), select variable **PATH**, by double-clicking it. Add the directory to ESRI ArcView **bin32** directory to what is already defined for that variable, separated by a semi-colon ";". For default installation, you should add: **;C:\ESRI\AV_GIS30\ARCVIEW\BIN32** to what is already set for the PATH variable. Make sure you don't remove what was already set for the PATH variable, otherwise some software on your system might stop working properly.
    Ask for assistence from your systems administrator or technical support if you are not sure how to make these modifications to your system.
  - On Windows 98/ME, ask your system administrator to do it for you.
- It is done. Next time you open DG, the **Arc/Info Grid** output option should be available.

Another important file that is stored on the output directory is the file result.xls which stores a summary of all tasks, error messages, result parameters, statistical tests, accuracy, and more. A description and a method for analysis of this data file is available at the **Results Table** section.

All *.bmp*, *.asc* and other result files are stored under the directory specified on the text field Output directory. This must be a valid folder (local or remote) accessible through the computer being used. ESRI Arc/Info grids are stored in subdirectories of the Output directory and called sequentially grid00000, grid00100, grid00200 and so on. The directory grid00100 for example, stores all grids resulting from tasks 100 through 199. This is because of an ESRI limitation on the number of grids allowed in a directory.

For more details on the files generated as the results of an experiment, please see the section on **Results**.

Projection Datasets

In the Projection Datasets panel, the user specifies which datasets will be used in the projection phase of the experiment. A sample picture of this panel is shown below.

At the end of each task, DesktopGarp will project the rule set obtained during optimization onto every dataset specified on the Current Datasets list. DesktopGarp will also project the rule set onto the native range dataset, defined on the Native Range Dataset panel.

A list of available datasets is shown on the Available Datasets combo box. The Add button adds the dataset selected on the combo box. To remove a dataset from the Current Datasets list, highlight it and then click the Remove button.

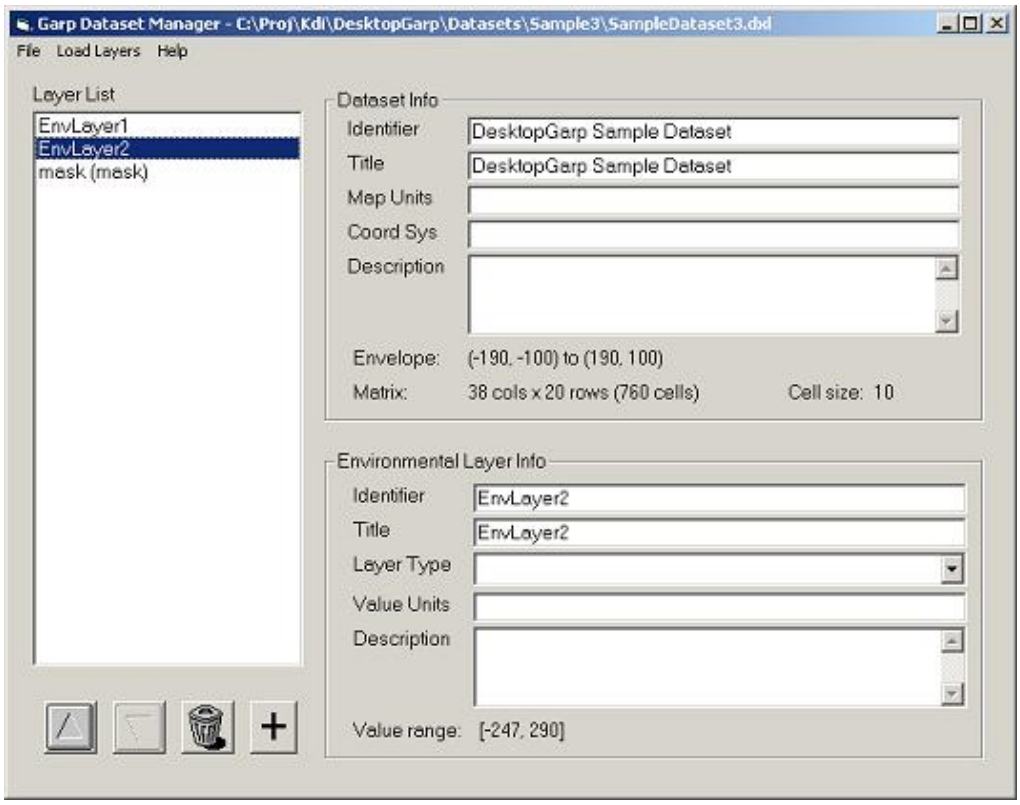
Using different datasets on the experiment is useful when researching invasive species, climate changes, and time analysis. For a more detailed description of methods for those studies, please read the **Invasive Species, Climate Changes and Time Analysis** section of this manual.

Create custom datasets with Dataset Manager

Although DG comes with a sample dataset for demonstration purposes, it might not be appropriate for all experiments. In that case, it is necessary to create a custom dataset that is compatible with DG format.

There is a tool called **Dataset Manager** that is installed with DG that can be used for creating new datasets. It is available from the **Start**->**Programs**->**Desktop Garp**->**Dataset Manager** menu.

Dataset Manager interface with an open dataset is shown below:



The **Dataset Info** panel allows the user to set generic metadata about the dataset. Currently, the most useful fields are **Identifier** and **Title**. Those values will be displayed in DG once the datasets are scanned from the hard disk. On the bottom of this panel, there are some information on the dataset, such as its envelope and grid cell count and size.

The **Environmental Layer Info** panel allows the user to set generic metadata about each individual layer on the current dataset.

The **Layer List** shows all available layers on the open dataset. In that case, there are two regular layers (EnvLayer1 and EnvLayer2), and a mask layer. The buttons right below this list can be used to move layers up and down, or to remove or add layers manually.

To produce DG datasets, this tool expects the layers to be in **ASCII Raster Grid** format, and placed on the same directory. They must have the same geographical boundaries and the same cell size, i.e., if you stack the layers together, all cells must match perfectly on top of each other.

Note: There is an ESRI ArcView extension that ships with DG install package that was designed to help ArcView users to produce ASCII Raster grids in the format described above, from GRID themes on a view. That extension is placed on DG installation directory (which by default is C:\Program Files\DesktopGarp), and is named **GarpDatasets.avx**. To install and use this extension, copy the file **GarpDatasets.avx** to ESRI ArcView **ext32** directory (on standard installations, that directory is: C:\ESRI\AV_GIS30\ARCVIEW\EXT32\). Then, click your project window, and click the File->Extensions menu. Check the box next to

GarpDatasets extension, and click the **Ok** button. That will create a new sub-menu on the **View** menu, called **Garp**. The only menu item on that sub-menu allows the user to create ASCII raster grids for Dataset Manager from Grids in a theme.

Once the environmental layer (.asc) files are ready to be converted to DG format, click the Load Layers->From Ascii Raster Grids menu item, and application will prompt the user for the directory where the asc files are stored. Choose one of the asc files and the conversion will begin. As each layer is processed, they will show up on the **Layer List**. The application might hang up for some time, from seconds to minutes, depending on the size of each layer, but that is normal. Just wait until it finishes.

Note: In order to restrict (or mask) the area of study to a particular region within the datasets, create an Ascii raster grid called **mask.asc**, with NO_DATA values on the cells that should be excluded from the Garp experiment. If you have NO_DATA values on the masked region of your regular layers, you can just copy one of the .asc files to the same directory and rename it to **mask.asc**. Dataset Manager application will automatically convert that layer and use it as the mask for that dataset, during the conversion process.

After all layers are converted, the users needs to click the File->Save menu item, choose a file name for the dataset. This file must be saved on the same directory as the other asc files were originally stored). The dataset is ready to be used in DG.

Results

During execution, DesktopGarp creates many output objects on the file system, as files and directories, depending on the choice of output parameters. Those objects are the results for the experiment. The main resulting objects are: prediction maps and the results table.

Prediction Maps

These are created in three formats, depending on the options selected on the **output parameter panel**. The formats are

- **bitmaps**: MS Windows bitmaps, with extension "*.bmp*";
- **ASCII raster grids**: ASCII text format, with extension "*.asc*";
- **ESRI Arc/Info grids**: ESRI proprietary format for grid spatial data storage and management. A separate directory is created for each grid.

The prediction maps, regardless of the format in which they are created, are graphical representations of the predicted geographic distribution of the species on a particular task within the experiment.

Note: The user is usually required to do some further processing and selection of the resulting maps to get more accurate and reliable results. Although a single task can generate a good prediction map, there are situations where the genetic algorithm within GARP gets stuck on local optima, which causes the result to be poor in quality. To avoid such pitfalls, please follow one of the recommended **methods** described later in this manual.

The bitmap format is intended for use on low end systems that have no geographic information system available for further analysis and visualization of the results. It allows the user to see the results of each task individually, using standard operating system image visualization tools.

Below is a sample bitmap image of a prediction in Brazil.

The other two formats, **ESRI** proprietary format for grid datasets on its GIS and ASCII raster, allow the user to do further analysis and data processing to get more complete and robust results.

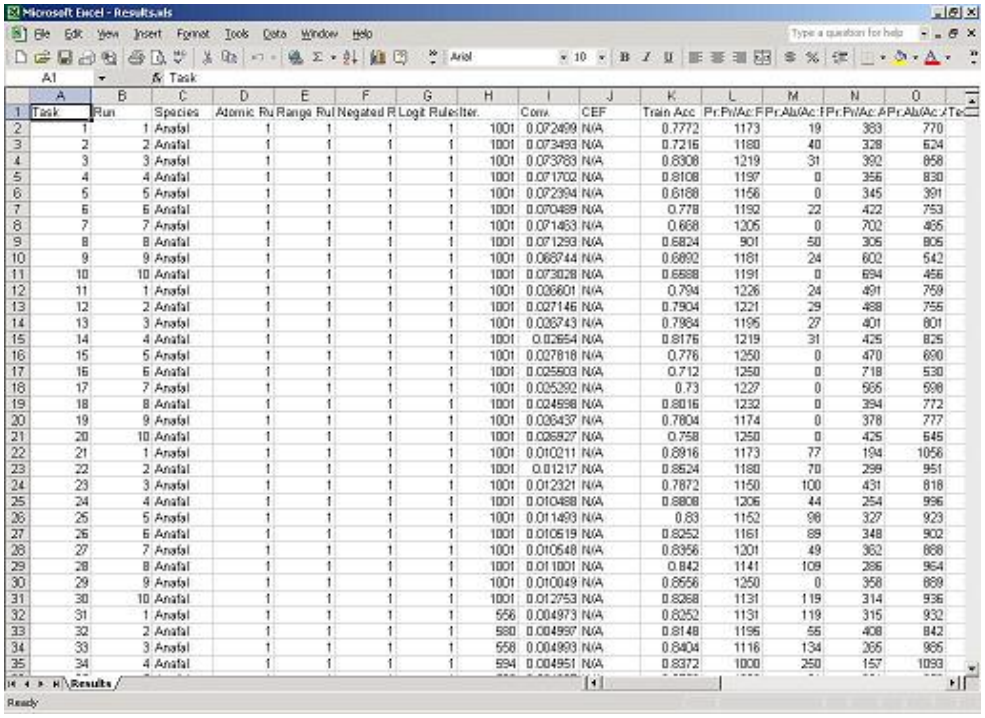
The ASCII format must be imported into a GIS software to prepare maps for display, whereas ESRI ARC/INFO Grid format allows the user to immediately display and handle maps using ESRI or other GIS software.

All *.bmp*, *.asc* and other result files are stored under the directory specified on the text field Output directory. This must be a valid folder (local or remote) accessible through the computer being used. ESRI Arc/Info grids are stored in subdirectories of the Output directory and called sequentially grid00000, grid00100, grid00200 and so on. The directory grid00100 for example, stores all grids resulting from tasks 100 through 199. This because of an ESRI limitation on the number of grids allowed in a directory.

Results Table

All information on the execution of the experiment and each GARP task within it is stored on an MS-Excel file named result.xls in the output directory. This file shows a summary of all tasks, error messages, result parameters, preliminary statistical tests, accuracy, and more.

Below is a snapshot of a sample result file.



Task	Run	Species	Atomic Ru	Range Rul	Negated R	Logit Rule	Iter.	Conv	CEF	Train Acc	Pr.Pr/Ac.F	Pr.Ab/Ac.F	Pr.Pr/Ac.A	Pr.Ab/Ac.A	Te
1	1	Anafal	1	1	1	1	1001	0.072499	N/A	0.7772	1173	19	383	770	
2	2	Anafal	1	1	1	1	1001	0.073493	N/A	0.7216	1180	40	328	624	
3	3	Anafal	1	1	1	1	1001	0.073783	N/A	0.8308	1219	31	392	858	
4	4	Anafal	1	1	1	1	1001	0.071702	N/A	0.8108	1197	0	356	830	
5	5	Anafal	1	1	1	1	1001	0.072394	N/A	0.6188	1156	0	345	391	
6	6	Anafal	1	1	1	1	1001	0.070489	N/A	0.778	1192	22	422	753	
7	7	Anafal	1	1	1	1	1001	0.071483	N/A	0.668	1205	0	702	466	
8	8	Anafal	1	1	1	1	1001	0.071293	N/A	0.6824	901	50	306	806	
9	9	Anafal	1	1	1	1	1001	0.068744	N/A	0.6892	1181	24	602	542	
10	10	Anafal	1	1	1	1	1001	0.073028	N/A	0.6888	1191	0	694	456	
11	1	Anafal	1	1	1	1	1001	0.026601	N/A	0.794	1226	24	491	759	
12	2	Anafal	1	1	1	1	1001	0.027146	N/A	0.7904	1221	29	488	755	
13	3	Anafal	1	1	1	1	1001	0.026743	N/A	0.7984	1195	27	401	801	
14	4	Anafal	1	1	1	1	1001	0.02654	N/A	0.8176	1219	31	425	825	
15	5	Anafal	1	1	1	1	1001	0.027818	N/A	0.776	1250	0	470	690	
16	6	Anafal	1	1	1	1	1001	0.025603	N/A	0.712	1250	0	718	530	
17	7	Anafal	1	1	1	1	1001	0.025292	N/A	0.73	1227	0	565	598	
18	8	Anafal	1	1	1	1	1001	0.024598	N/A	0.8016	1232	0	394	772	
19	9	Anafal	1	1	1	1	1001	0.026437	N/A	0.7804	1174	0	378	777	
20	10	Anafal	1	1	1	1	1001	0.026827	N/A	0.758	1250	0	425	645	
21	1	Anafal	1	1	1	1	1001	0.010211	N/A	0.8916	1173	77	194	1056	
22	2	Anafal	1	1	1	1	1001	0.01217	N/A	0.8524	1180	70	299	951	
23	3	Anafal	1	1	1	1	1001	0.012321	N/A	0.7872	1150	100	431	818	
24	4	Anafal	1	1	1	1	1001	0.010488	N/A	0.8808	1206	44	254	996	
25	5	Anafal	1	1	1	1	1001	0.011493	N/A	0.83	1152	98	327	923	
26	6	Anafal	1	1	1	1	1001	0.010619	N/A	0.8252	1161	89	348	902	
27	7	Anafal	1	1	1	1	1001	0.010648	N/A	0.8356	1201	49	362	888	
28	8	Anafal	1	1	1	1	1001	0.011001	N/A	0.842	1141	109	286	964	
29	9	Anafal	1	1	1	1	1001	0.010049	N/A	0.8556	1250	0	358	889	
30	10	Anafal	1	1	1	1	1001	0.012753	N/A	0.8268	1131	119	314	936	
31	1	Anafal	1	1	1	1	556	0.004973	N/A	0.8252	1131	119	315	932	
32	2	Anafal	1	1	1	1	980	0.004997	N/A	0.8148	1196	55	408	842	
33	3	Anafal	1	1	1	1	558	0.004993	N/A	0.8404	1116	134	265	985	
34	4	Anafal	1	1	1	1	894	0.004951	N/A	0.8372	1000	250	157	1093	

The first line of the file holds a label for each column. Each subsequent line shows information on a single task within the experiment. Below are column descriptions.

Task	Sequential number or identifier assigned to each task. The task identifier is attached to each prediction map file name or directory as a link back to the task information.
Run	Sequential number assigned to a group of tasks that are repetition of the same input parameters. For example, if 5 runs are selected on the optimization parameters panel, each unique task will be repeated 5 times, and they will be represented with different numbers in this column, from 1 to 5 in this example.
Species	Species name as provided in the input file and displayed on the list box on the species data points panel.
Atomic Rules	Indicates whether the atomic rules were used on the experiment. Possible values for that column are 0 (false) or 1 (true).
Range Rules	Indicates whether the range rules were used on the experiment. Possible values for that column are 0 (false) or 1 (true).
Negated Rules	Indicates whether the negated rules were used on the experiment. Possible values for that column are 0 (false) or 1 (true).
Logit Rules	Indicates whether the logistic regression rules were used on the experiment. Possible values for that column are 0 (false) or 1 (true).
Iter.	Number of iterations executed by the GARP algorithm for that task. The value is limited by the maximum number of iterations. The value may be lower if the convergence limit was reached before the maximum number of iterations. Because the final iteration sums up optimization results, this column often brings the maximum number of iterations plus one.
	Value of the convergence control variable when the task finished. Can be slightly below the convergence limit specified on the optimization parameter panel in cases where the task ended because the limit was reached. Can

	also be greater than that limit in cases where the maximum number of iterations was reached before the convergence limit.
CEF	Not used.
Train Acc	Accuracy calculated using the training data points. Accuracy is calculated using the expression: (a+b)/(a+b+c+d), where the variables a, b, c and d are explained below.
Pr:Pr/Ac:Pr	Number of points where the model predicted presence and the input point (actual point) was a presence record. In this case, the model has predicted the point successfully. This is variable a on the accuracy expression. The mneumonics on the column label represents: Pr(edicted):Pr(esent)/Ac(tual record):Pr(esence).
Pr:Ab/Ac:Pr	Number of points where the model predicted absence and the input point (actual point) was a presence record. In this case, the model has predicted the point incorrectly. This value usually represents a real error, because of either low model quality or a misidentification of the specimen record. This is true when using presence only data, which is the case of the current release of DesktopGarp. This is variable c on the accuracy expression. The mneumonics on the column label represents: Pr(edicted):Ab(sent)/Ac(tual record):Pr(esence).
Pr:Pr/Ac:Ab	Number of points where the model predicted presence and the input point (actual point) was an absence record. At a first glance, the model has predicted the point incorrectly. However, when using presence points only, this error can be due to insufficient sampling of the interest area in cases where the point might be suitable for the species but no information is available about the occurrence of the species at this location. This is variable d on the accuracy expression. The mneumonics on the column label represents: Pr(edicted):Pr(esent)/Ac(tual record):Ab(sence).
Pr:Ab/Ac:Ab	Number of points where the model predicted absence and the input point (actual point) was an absence record. In this case, the model has predicted the point successfully. This is variable b on the accuracy expression. The mneumonics on the column label represents: Pr(edicted):Ab(sent)/Ac(tual record):Ab(sence).
Test Acc	These columns show the accuracy calculated using the data points set aside for accuracy testing, and that have not been used for training the model. The value is calculate by the same expression described for training points and the columns Pr:Pr/Ac:Pr, Pr:Ab/Ac:Pr, Pr:Pr/Ac:Ab and Pr:Ab/Ac:Ab represent the same concepts described above, applied to the test dataset.
Total Area	Total number of non-masked cells on the interest area.
Presence Area	Total number of cells where the species is predicted to be present.
Absence Area	Total number of cells where the species is predicted to be absent.
Non-predicted Area	Total number of cells where the algorithm could not define whether the species is present or absent; that is, no rule applied to that area.
Yes	Number of test points that fall inside the presence area.
No	Number of test points that fall outside the presence area.
ChiSq	Intermediate value for calculating the chi-square test value.
p	Chi-square test. Probablility of a random predictions being similar, that is, having the same number of correct predicted points as the one generated by GARP on that task.
Commission	Percentage of the prediction area that exceeds the recorded occurrence.
Omission (int)	Intrinsic omission. Percentage of the training points that are omitted from the prediction; that is, those that are predicted absent but are presence records.
Omission (ext)	Extrinsic omission. Percentage of the test points that are omitted from the prediction; that is, those that are predicted absent but are presence records.
Status	Status of the task. Can be "Waiting to be processed," "Successfully processed" or "Failed."
Message	More detailed message to explain the status message.
Layers	Indicates which layers have been used on the prediction. Represented by 0 (false) or 1 (true) below the layer name.

Methods

This section describes methods for getting results out of DesktopGarp for some ecological and biodiversity studies. For a more complete and formal discussion on methods using DesktopGarp for biodiversity research, please refer to the **bibliography**.

Native Range Distribution

The most straightforward use of DG is to get predictions of species' distribution for their native range. This can be accomplished in several different ways.

First, the user can set DesktopGarp parameters to run a single task. The result will consist of a single prediction map in the output directory. This is similar to the output from David Stockwell's **GARP web interface** at the San Diego Supercomputer Center.

Although that result can be useful on its own, it may be of limmited quality because of GARP algorithm limitations. Because it is an optimization algorithm, GARP can get stuck on local optima during execution, yielding poor results.

To overcome this problem, the user can specify a high number of runs on the parameters panel and let DesktopGarp produce a number of results using the same input parameters. Then, using a GIS, the user can stack the results together, summing the resulting grids into a result dataset. Although GARP can be unstable sometimes, when the same task is run several times, GARP often predicts values for some areas in a very stable manner. Other areas are predicted with less frequency. Adding up many results decreases the chance of skewed results because of optimization problems.

There are other more advanced methods for predicting species distributions for their native range, for example best subset selection developed by A.T. Peterson and other methods developed by E. Wiley and McNyset using a statistical method called ROC Analysis. Please see references in the **bibliography** for further details on those methods.

Invasive Species, Climate Changes and Time Analysis

Sorry. This section is under construction.

Environmental Layers Selection

Another use of DesktopGarp is to point out what environmental factors are more significant or important than others for a given species. This operation is called environmental layer jackknifing.

This can be accomplished by using the controls on the **environmental layers** panel. Three radio buttons on the bottom of that panel control how the layers are used within the experiment.

By running multiple tasks with combinations of layers, either all combinations or a a subset of all combinations with a fixed number of layers, the user can analyze the results of the experiment using multiple linear regression to check which layers have a significant impact on the resulting errors.

DesktopGarp writes the result file in a format that is relatively easy to input into a multiple linear regression analysis. See the description for the **layers columns** on the result file.

When the experiment is finished, DesktopGarp stores values on the layers columns, which can be 0 or 1, indicating whether the specific layer was used on the task. Those values can be used as independent variables on a multiple linear regression analysis. These values can be used to predict some of the accuracy and error values.

It is generally better to use commission and omission errors as dependent variables instead of using accuracy values. Accuracy mixes together both commission and omission, which reflect different errors.

Then the user can analyze each layer to determine the existence of a significant positive or negative correlation with errors.

The next step is to prepare and run another experiment using just those layers that are known to be significant on the model generation, minimize omission, and keep commission down to a reasonable level.

For a discussion on omission and commission errors and the concepts of predictive modeling of species, please refer to the **bibliography** or the resources listed on the **links** page.

Note: With a large number of layers, the use of combinations of layers can lead to an exponential or factorial explosion on the number of resulting tasks to be run on the experiment. Bear in mind that DesktopGarp can currently run about 2,000 to 10,000 tasks within an experiment in a reasonable time using a standard workstation. There are projects in development that aim to create a parallel version of DG. When these are complete, it will be more feasible to run larger experiments using combinations of layers.

Rule Type Selection

Just as the user can select the significant environmental layers for prediction, the user can also try to identify the most relevant rule types for a particular species.

This approach may be more useful for algorithm development than for use in particular predictions, however, it is available on the software user interface.

During development of DesktopGarp, many experiments were executed using this approach to determine the significance of each particular rule type in predicting distributions for many different species. Although those experiments showed that atomic rules are less significant than range, negated and logit rules in most situations, on most occasions atomic rules contribute to the results as well.

Recent Development History, Known Problems and Issues

Known problems and issues on the current version:

- In some systems, the installation package does not perform properly and is unable to set up the software. This is usually related to the lack of some dynamic link libraries on the Windows 98 (first edition). A package for correcting this situation is in preparation and will be released in a future version.
- A memory leak has been detected during the execution of experiments, which causes the program to waste some of the system memory, especially if it is left running a large number of iterations for long time. This problem might be felt after 2,000 to 3,000 iterations.

  To correct the problem, interrupt the experiment when DesktopGarp has taken too much memory from the system and close the application. Reopen the application, reopen the interrupted experiment, and click **Run unfinished experiment** from the **Model** menu.

  This bug should be fixed in the next release.

Bibliography

GARP Algorithm

Stockwell, D. R. B. 1999. Genetic algorithms II. Pages 123-144 *in* A. H. Fielding, editor. Machine learning methods for ecological applications. Kluwer Academic Publishers, Boston.

Stockwell, D. R. B., and D. P. Peters. 1999. The GARP modelling system: Problems and solutions to automated spatial prediction. International Journal of Geographic Information Systems **13**:143-158.

Stockwell, D. R. B., and I. R. Noble. 1992. Induction of sets of rules from animal distribution data: A robust and informative method of analysis. Mathematics and Computers in Simulation **33**:385-390.

Application

Anderson, R. P., M. Laverde, and A. T. Peterson. 2002a. Geographical distributions of spiny pocket mice in South America: Insights from predictive models. Global Ecology and Biogeography 11:131-141. **(*)**

Anderson, R. P., M. Laverde, and A. T. Peterson. 2002b. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos 93:3-16. **(*)**

Anderson, R. P., D. Lew, and A. T. Peterson. In press. Evaluating predictive models of species' distributions: Criteria for selecting optimal models. Ecological Modelling, v. 162, p. 211 232. **(*)**

Chen, G., and A. T. Peterson. 2000. A new technique for predicting distributions of terrestrial vertebrates using inferential modeling. Zoological Research 21:231-237.

Chen, G., and A. T. Peterson. In press. Prioritization of areas in China for biodiversity conservation based on the distribution of endangered bird species. Bird Conservation International.

Feria, T. P., and A. T. Peterson. 2002. Using point occurrence data and inferential algorithms to predict local communities of birds. Diversity and Distributions 8:49-56.

Godown, M. E., and A. T. Peterson. 2000. Preliminary distributional analysis of U.S. endangered bird species. Biodiversity and Conservation 9:1313-1322.

Papes, M., and A. T. Peterson. In press. Predictions of the geographic course of the invasion of Ageratina adenophora in China. Journal of Wuhan Botanical Research. (*)

Peterson, A. T. 2001. Predicting species' geographic distributions based on ecological niche modeling. Condor 103:599-605.

Peterson, A. T., L. G. Ball, and K. C. Cohoon. 2002a. Predicting distributions of tropical birds. Ibis 144:e27-e32.

Peterson, A. T., and K. C. Cohoon. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecological Modelling 117:159-164.

Peterson, A. T., M. A. Ortega-Huerta, J. Bartley, V. Sanchez-Cordero, J. Soberon, R. H. Buddemeier, and D. R. B. Stockwell. 2002b. Future projections for Mexican faunas under global climate change scenarios. NATURE 416:626-629.

Peterson, A. T., V. Sanchez-Cordero, C. B. Beard, and J. M. Ramsey. 2002c. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. Emerging Infectious Diseases 8:662-667. (*)

Peterson, A. T., V. Sanchez-Cordero, J. Soberon, J. Bartley, R. H. Buddemeier, and A. G. Navarro-Siguenza. 2001. Effects of global climate change on geographic distributions of Mexican Cracidae. Ecological Modelling 144:21-30.

Peterson, A. T., J. Soberon, and V. Sanchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. Science 285:1265-1267.

Peterson, A. T., D. R. B. Stockwell, and D. A. Kluza. 2002d. Distributional prediction based on ecological niche modeling of primary occurrence data. Pages 617-623 in J. M. Scott, P. J. Heglund, and M. L. Morrison, editors. Predicting Species Occurrences: Issues of Scale and Accuracy. Island Press, Washington, D.C. (*)

Peterson, A. T., and D. A. Vieglais. 2001. Predicting species invasions using ecological niche modeling. BioScience 51:363-371.

Stockwell, D. R. B., and A. T. Peterson. 2002a. Controlling bias in biodiversity data. Pages 537-546 in J. M. Scott, P. J. Heglund, and M. L. Morrison, editors. Predicting Species Occurrences: Issues of Scale and Accuracy. Island Press, Washington, D.C.

Stockwell, D. R. B., and A. T. Peterson. 2002b. Effects of sample size on accuracy of species distribution models. Ecological Modelling 148:1-13.

(*) Articles using DesktopGarp.