DIVA-GIS

Version 7.5

Manual

January 2012

Robert J. Hijmans, Luigi Guarino, and Prem Mathur

Conditions of use

The DIVA-GIS software can be used and distributed freely. It is provided "as is", without warranty of any kind, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose and noninfringement. In no event shall the authors or copyright holders be liable for any claim, damages or other liability, whether in an action of contract, tort or otherwise, arising from, out of or in connection with the software or the use or other dealings in the software.

Contributors

DIVA-GIS is maintained by Robert J. Hijmans. Edwin Rojas, Rachel O'Brien, and Mariana Cruz have written parts of the code for previous versions. DIVA-GIS was improved thanks to bug-reports and/or suggestions made by many individuals, particularly Luigi Guarino, Prem Mathur, Xavier Scheldeman, and others including Østein Berg, Israel Barrantes, Rachel O'Brien, Coen Bussink, Arthur Chapman, Dirk Enneking, Catherine Graham, Dave Hodson, Andrew Jarvis, Maarten van Zonneveld were also involved in the development of earlier versions.

Support for the early development has come from the Bioversity International, the UC Berkeley Museum of Vertebrate Zoology, the International Potato Center (CIP), SINGER/SGRP, FAO, and USDA.

We have benefited from the work of the following persons: M. Sawada (Rook's case), Gerald Everden and Frank Warmerdam (PROJ4); Andrew Williamson (Shapechk); the contributors to Delphi Zip Version 1.70
(http://www.geocities.com/SiliconValley/Network/2114/)

Warning

DIVA-GIS is a relatively new program that is under continuous development and not all parts have been tested completely. This means that you should never blindly believe the results of your analysis. Rather, you should always test if DIVA-GIS works well, for example by manually calculating the expected results for a small number of grid cells, or by first doing the calculations with a simple sample data set for which you know the results.

If you find a possible error, please be so kind as to report it!

Please send your comments to info@diva-gis.org.

Abstract

DIVA-GIS is a free computer program for mapping and analyzing spatial data. It is particularly useful for analyzing the distribution of organisms to elucidate geographic and ecological patterns. It is aimed at those who cannot afford generic commercial geographic information system (GIS) software, or do not have the time to learn how to use these often complicated programs, and for anyone else who wants a GIS tailor-made to analyze biological distributions. DIVA-GIS supports vector (point, line, polygon), image and grid data types. It can help improve data quality by finding the coordinates of localities using gazetteers, and by checking existing coordinates using overlays (spatial queries) of the collection sites with administrative boundary databases. Distribution maps can then be made. Analytical functions in DIVA-GIS include mapping of richness and diversity (including based on molecular marker (DNA) data; mapping the distribution of specific traits; identifying areas with complementary diversity; and calculating spatial autocorrelation. DIVA-GIS can also extract climate data for all locations on land. Ecological niche modeling can be carried out using the BIOCLIM and DOMAIN algorithms.

Table of Contents

1. INTRODUCTION

This manual explains how to use DIVA-GIS. Additional information, including an introductory tutorial, exercises and examples of the use of DIVA-GIS can be found on the DIVA-GIS website (http://www.diva-gis.org).

1.1 Conventions used

The following conventions are used in this manual:

Italics are used to refer to menusand sub-menus, and also to tabs and views. A forward slash (/) is used to relate a submenu to a menu.

"Quotes" are used to refer to buttons and boxes.

File types are referred to by their extension. For example, a dBase file, like data.dbf, is referred to as a DBF file

1.2 Installing DIVA-GIS

If you downloaded DIVA from the Internet, you will need to unzip the downloaded files first (use e.g. 7-Zip). Then you should click on **setup.exe** to install DIVA-GIS; also click on this file if you are installing from a CD-ROM. You will be asked in what directory (folder) you want to install the program. As you can install the program in any directory you like, in this manual we will refer to this directory as the **DIVA directory**, or `<DIVA dir>`. By default, DIVA-GIS will be installed in the **C:\program files\DIVA-GIS** directory.

1.3 The DIVA-GIS desktop

Once installation is complete, you can click on the DIVA-GIS icon, which will launch the DIVA-GIS **program**. The program consists of two overlapping parts, which we call *views*. The *Data view* is where you will do most of your work. The *Design view* is used to produce a graphical representation of the results of your work that can be saved as

a graphics file, printed or used in another application. You move between views by clicking on the appropriate tab at the bottom right hand corner of the window.

1.3.1 The Data view

The *Data view* is taken up by a map and, to its left, its associated table of contents (TOC) (Figure 1). The content of the map is managed and analyzed via functions that are spread over nine menus arranged across the top of the desktop: *Project, Data, Layer, Map, Analysis, Modeling, Grid, Stack,* and *Tools*. The menus are discussed in chapters 3 to 11 (one chapter per menu). There's also a *Help* menu (but there is no on-line help yet).

A map consists of geo-referenced databases called **layers** (sometimes also referred to as **themes**). For example, a map of the world may be made up of the following layers: national boundaries, altitude, main rivers and capital cities. Each layer on the map is listed in the **table of contents** (TOC). Layers can be made (in)visible on the map using the checkbox next to their name in the TOC. If you click on a layer's checkbox while holding the Ctrl key down, all layers on the map will be made (in)visible (depending on the state of the checkbox of the layer that you click on).

A single click on a layer in the TOC makes it the **active layer**, indicated by an uplifted layer bar (for example, the "countries" layer in Figure 1). The order in which different layers are drawn on the map is from the bottom upwards, so a layer higher up in the TOC will cover a lower one). You can change the order of a layer by dragging it up or down (i.e., by clicking on it and then moving your mouse while keeping the left button pressed).
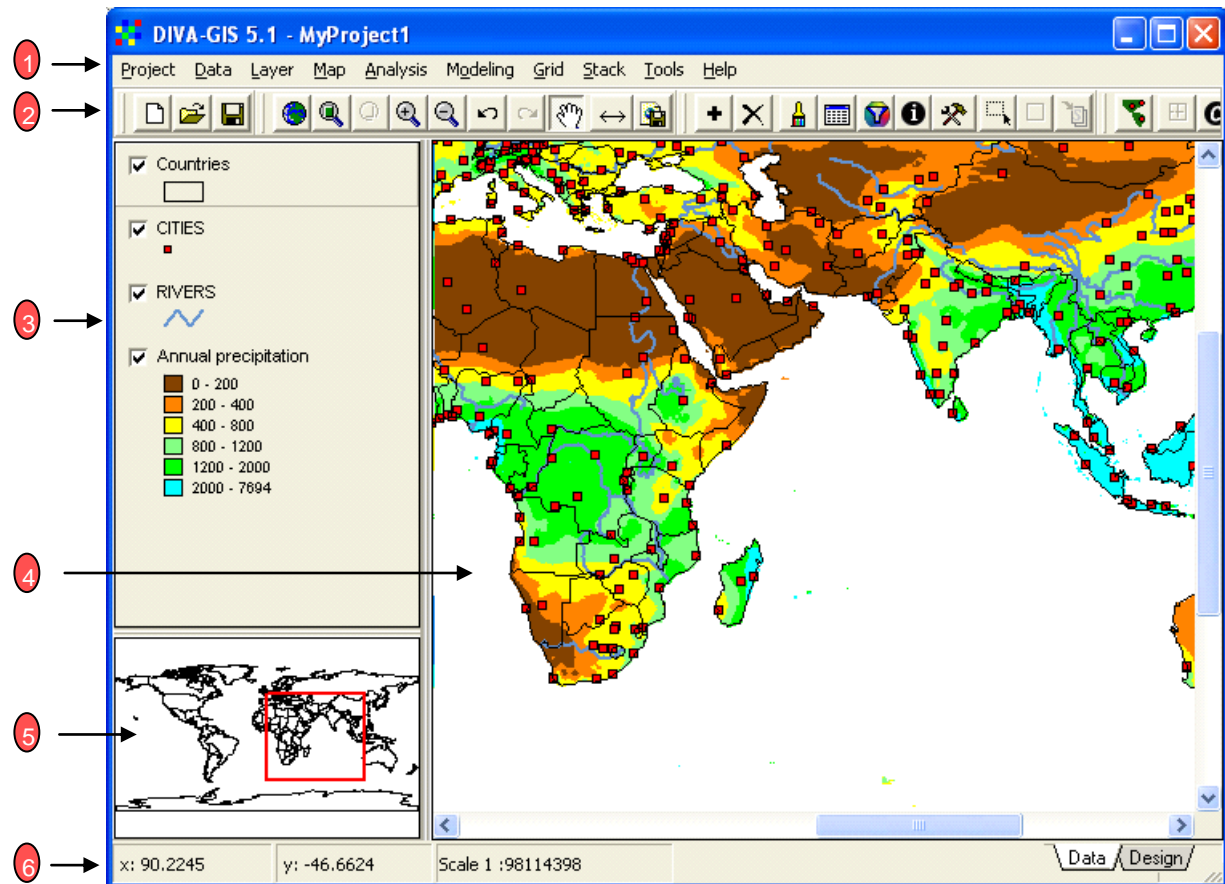
① **Menu bar** – Provides access to all DIVA-GIS functions.

② **Toolbar** – Quick access to some of the more commonly used functions.

③ **Table of contents** – Lists all the layers on the map.

④ **Map** – The map you're working on.

⑤ **Overview Map** – Shows the location of what is visible on the map in a wider region

⑥ **Status bar** –Indicates, from left to right, the location of the cursor (x and y coordinates), the scale of the map, the value and position (row, column) of a gridcell, and the distance between two points.

7 – Data/Design view tabs?

Figure 1. The DIVA-GIS main window in *Data view*

Double-clicking on a layer in the TOC (or clicking on *Layer/Properties*) brings up a small window that allows changes to be made to some of the layer's properties (such as its colors) on the map (Figures 2 and 3).

It is important that you explore these windows (click on the different tabs). See Section 5.2 for more details about them.
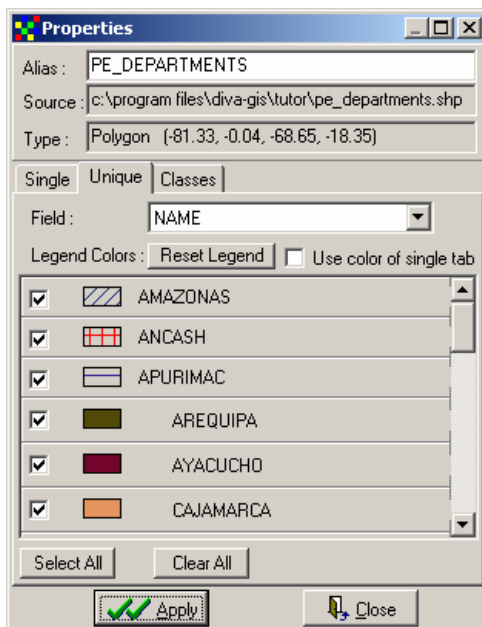
Figure 2. The *Layer/Properties* window for shapefiles.
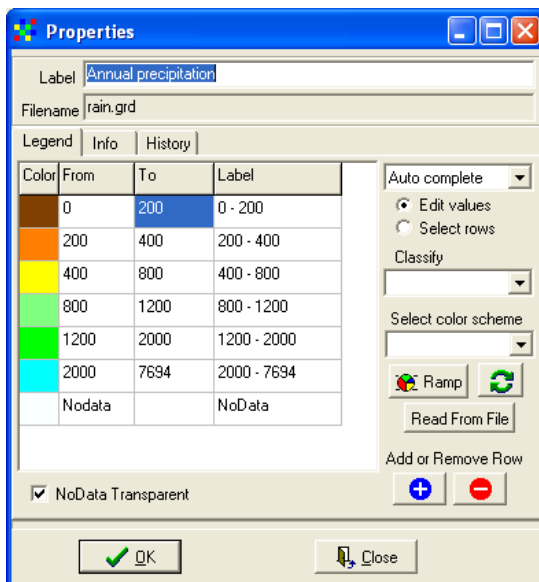
Figure 3. The *Layer/Properties* window for grids.

1.3.2 The Design view

The *Design view* of the main DIVA-GIS window allows you to put together a map that you can save as an image file. It is based on the layers that are visible in the *Data view*. The resulting image can be exported from DIVA-GIS, for example to be used as an illustration in a document or presentation.
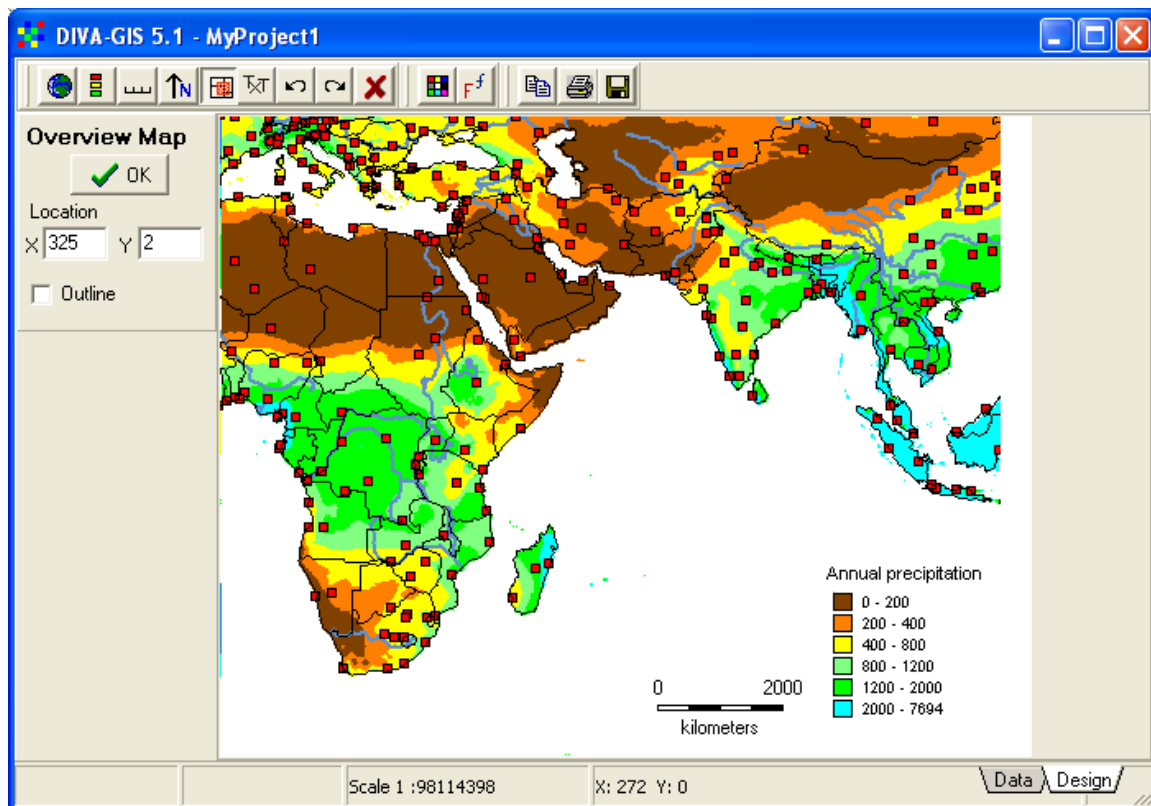
Figure 4. The DIVA-GIS main window in *Design view*.

Box 1. The *Design view* buttons.

Icon	Name	Short explanation
	Map	Adds the map from the *Data view* to the *Design view*
	Legend	Adds the legend from the *Data view* to the *Design view*
	Scale	Adds a graphic scale to the *Design view*
	North arrow	Adds a North arrow to the *Design view*
	Overview map	Adds the overview map to the *Design view*
	Add text	Adds text to the *Design view*
	Undo	Un-does last addition to the *Design view*
	Redo	Re-does last addition (after undo) to the *Design view*
	Clear all	Starts from scratch
	Set color	Sets the color for the text and lines to be added to the *Design view*
	Set font	Sets the font for text to be added to the *Design view*
	Copy	Copies the *Design view* to the clipboard
	Print	Prints the *Design view* to a specified printer
	Save	Saves the *Design view* to a graphics file (BMP or TIF)

Producing a nice map for presentation purposes is not very difficult but will take some trial and error. Start by adding the basic map from the *Data view*. Then, decide if there is too much white space around the image. If there is, go back to *Data view* and zoom and/or pan around to better fill up the area. If there is too much white space at the right of the image, you may need to remove a layer that causes this to happen (for

example, a layer that, although perhaps invisible, is mainly located to the east of the area of interest).

When you have the map the way you want it, add a legend. First, click on the "Add legend" button. A panel will appear on the left hand side of the window. It includes a number of options, and lists all the visible layers. Decide what layers to include in the legend, by unchecking the ones you do not want.

You can add the legend to the map by clicking on the image at the point where you want the legend to appear. One approach is to keep clicking until you find the spot you like; then press "Clear All", followed by "Add Map", then "Add legend" and finally "OK". The legend will be put where you last had it, because the coordinates were saved in the text boxes. You can also change these values in the text boxes before you press "OK". Another approach is to add a legend and use the "Undo" button if you do not like the place it gets, and try again.

The width of the TOC in the *Data view* determines its width in the *Design view*. Thus, if it is too wide or narrow, go back to the *Data view,* change it, and try again Design *view*.

Adding a scale, North arrow, and text to the image follows the same principles as outlined above for the legend.

You can set the color and font for any text that you add to the image. To set the font for the TOC, however, you must go to the *Tools*/*Options*/*Layer* menu in the *Data view*, change the default font, *Close* the project and then *Open* it again (in the *Project* menu).

To export your map from DIVA-GIS, you can print it, copy it to the clipboard (and paste it into another application), or save it to a graphics file, in bitmap (BMP) or TIF format.

1.4 File Types and Formats

DIVA uses files of various types and formats. The most important are

1. the shapefile, gridfile and image file formats for spatial databases and

2. the dBaseIV (DBF) format for reading and writing external (non-spatial) databases.

1.4.1 Shapefiles

Shapefiles are so-called vector databases, describing the location of points (e.g., collecting locations), polylines (e.g., roads) and areas (or polygons, e.g., soil types, countries). A shapefile actually consists of three separate files with the same name but with different extensions (SHP, SHX and DBF), but they are treated as one file. There are some shapefiles with additional files (extensions SBN and SBX), but these are not essential and are not used in DIVA-GIS. The shapefile format was developed by ESRI, a leading GIS software company. They were initially developed for use in ArcView, but now nearly all GIS programs can either use them directly, or import them.

1.4.2 Gridfiles

Grids are central to the analytical capabilities of DIVA-GIS. A grid divides (a part of) the world into equal-sized cells. The advantage of using grids as opposed to areas such as countries or other administrative units is that grid cells of the same size and shape allow more objective comparisons.

For grid databases, in which an area is divided into equally sized rectangles, DIVA-GIS gridfiles are used. A gridfile consists of two separate files *.GRI and *.GRD, but DIVA-GIS treats them as if they were one file. The *.GRI file has the actual data, and the *.GRD file has metadata and a number of parameters that are needed to read the *.GRI file properly.

From these two files, DIVA-GIS creates two more files, *.BMP and *.BPW. These files are derived from the *.GRI and *.GRD files and are used to visualize the data on the map, but are otherwise not essential. If the BMP and BPW are absent, DIVA-GIS creates these files automatically when opening a gridfile.

The BMP and BPW files can also be used to visualize gridfiles in ArcView and in ArcExplorer (as images). Unlike in DIVA-GIS, however, the underlying grid data will not

be accessible in these programs, and it will not be possible to change their legend categories. If you want to use the grid data in another program, you should export the gridfile to a suitable format (Chapter 3).

Gridfiles with the same extents and gridcell size can be manipulated together in a "stack." Stacks are discussed in Chapter 9.

1.4.3 Image files

Image files are special kinds of grids that can be displayed but not used for analysis, as the data associated with the different colours in the file are not accessible. A typical example of such a file would be an air photo or satellite image. DIVA-GIS supports three image formats: TIFF, JPEG and mrSID.

1.4.4 DBF files

DBF (version IV) is a commonly used database format. DIVA-GIS uses it to import and export tabular data. You can create a DBF file by exporting it from a database program such as Access. It used to be possible to easily do this from Excel but this is no longer the case, although it may still be possible with some third-party plugins. Fortunately you can use the (free and open source) OpenOffice spreadsheet Calc

If you use Calc, you must take care not to lose data, particularly not to lose precision (decimals) of coordinate data, or to create a DBF file with unsupported characteristics.

The field names (variable names) should be in the first row, and only there. Each column with data should have a field name. The names of fields may not be repeated, may not start with a number, and should only consist of letters and numbers. Field names should NOT contain characters such as * ^ % ? / - and >. Field names should not be longer than 11 characters. Always make a backup copy of the DBF file and save the new file in the native format <u>first</u>, so that you do not lose your data should they not be saved correctly as DBF.

1.5 Geographic coordinates

There are several ways to describe a location on the Earth. The most commonly used are degrees of longitude and latitude. A location can be between 180° West and 180° East, and between 90° North and 90° South. Degrees are often subdivided using a sexagesimal system (a calculus system with 60 as the basic number) of "minutes" and "seconds" (exactly as done with subdivision of hours). For example, a latitude can be described as e.g., 12°34´15´´S (12 degrees, 34 minutes, 15 seconds, Southern hemisphere).

This system worked fine in the days of paper maps, but it is not very suitable for the digital age. A decimal system is universally used in geographic computing. In the decimal system, latitude and longitude are described by a single number each, and no letters, with the sign indicating the hemisphere (+ = N or E, – = S or W) (e.g., – 12.57083 for 12°34´15´´S). To convert longitude and latitude in degrees, minutes and seconds to decimal degrees, the following formula is used:

$$DC = h \cdot \left(d + \frac{m}{60} + \frac{s}{3600} \right)$$

Where DC is the decimal coordinate; d is the degrees (°), m the minutes (´), and s the seconds (´´) of the sexagesimal system, h = 1 for the northern and eastern hemispheres and -1 for the southern and western hemispheres. For example, 30°30´0´´S = -30.500 and 30º15´55 ´´N = 30.265. You can do these calculations in a spreadsheet program or in DIVA-GIS using *Tools/Geo-calculator* (Chapter 10).

Decimal degrees should normally be recorded with 4 or 5 decimals. At the equator, one unit of the fourth decimal (0.0001 degrees) equals about 10 meters (less at other latitudes; not affected by longitude). That should be precise enough for most applications. If you are using high-precision GPS (with differential correction), 5 decimal places would be warranted. See Wieczoreck et al. (2004) for a thorough discussion of coordinate precision and uncertainty.

2. THE *PROJECT* MENU

The *Project* menu contains functions for the management of DIVA-GIS project files, and some related tasks (Box 2).

Box 2. The *Project* menu.

Icon	Sect.	Name	Short explanation
	2.1	New	Starts a new project (map and associated image)
	2.1	Open	Opens an existing project (file with .DIV extension)
	2.1	Close	Closes the current project
	2.1	Save	Saves the current project
	2.1	Save As	Saves the current project with a new name
	2.2	Export Project	Exports a project (including all data) to a DIVA-GIS export file (file with extension "DIX")
	2.2	Import Project	Imports a DIVA-GIS export file (DIX)
	2.1	-	Her you can find a list of the 10 last used DIVA-GIS projects
		Exit	Closes the project and shuts down DIVA-GIS

2.1 Projects

A DIVA-GIS project is a description of a DIVA-GIS map: it includes a collection of layers and their display properties, as well as some general parameters describing the map's scale and center. A project file can be closed, saved with a new name, and opened again using commands from the *Project* menu. To create a new project, select *New*. This creates an empty map that can be filled by adding features using *Layer/Add*. You can save a project by clicking on *Project/Save*. Project files have the extension DIV. The names of the last 10 projects you saved are listed at the bottom of the *Project* menu to allow quick access to these files. You can open a recently used project by selecting it from this list.

It is important that you clearly understand the difference between a DIVA-GIS project file (DIV) and the layers (files) that make up the map. The project file does not

contain the actual data, it only points to the files containing the data pertaining to the different layers included in the project, and stores map properties (such as scale). Relative paths (e.g., \diva\myshp.shp) are stored for data files that are below the project file in the folder structure. This allows sharing projects over a network, or saving them on a CD-ROM (different drive letters can be used). For all other files, the absolute path is stored (e.g. c:\mydata\diva\myshp.shp). It is also possible to use network paths (e.g., \\network\share\shape.shp).

This means that if you delete a project file, all your data will be still be available. However, if you delete, or rename, a data file, the project file will not be able to find it anymore, and DIVA-GIS will show a message indicating this.

2.2 *Import project* and *Export project*

Another way to share a DIVA-GIS project is to place all its contents into a DIVA-GIS export file (extension DIX). In contrast to a DIV file, the DIVA-GIS export file contains the project file **and** all the layers (data files). You can send such an export file to another DIVA-GIS user, who can then import it into DIVA, or you can use it to simply store all files pertaining to a project in one location. These files are compressed, and they do not take up much disk space; they can often be sent to another user via email.

To import a project file, you must indicate where the data should be expanded, and under which project name. Typically, you would make a new directory for this, so that it is clear which files belong to a specific project that you imported.

3. THE *DATA* MENU

The *Data* menu will help you to manage your data. It includeds functions to import and export data to and from DIVA-GIS and to check or prepare coordinate data for use.

Box 3. The *Data* menu

Icon	Sect.	Name	Short explanation
	3.1	Import Points to Shapefile	Creates a shapefile of points from a text file, DBF file or Access database.
	3.2	Import Text to Line/Polygon	Creates a shapefile of lines or polygons from a text file.
	3.3	Draw Shape	Creates a shapefile of points, lines or polygons by moving the cursor across the map
	3.4	Polygon to Grid	Assigns values from a polygon shapefile to a gridfile
	3.5	Points to Convex Polygon	Creates a convex polygon around all points in a shapefile
	3.6	Selection to new shapefile	Saves a selected part of a shapefile, or indeed a complete gridfile, to a new shapefile
	3.7	Extract values by points	Extracts values from a grid or a stack for all points in a shapefile and produces different types of text files
	3.8	Climate	1. Queries a location on the map for its climate 2. Creates grid files of climate variables 3. Creates CLM files
		Georeference	There used to be a module ("Assign Coordinates") for this. Now this will take you to the BioGeomancer website
	3.9	Check Coordinates	Compares locality descriptions in a point shapefile file with those in a polygon shapefile file
	3.10	Export Gridfile	Exports a gridfile to a number of other grid data formats
	3.11	Import to Gridfile	Imports a gridfile from a number of other grid data formats
	3.12	Write VRT file	Write a 'Virtual Raster' header file to open gridfiles in e.g. ArcGIS or Quantum-GIS.
	3.13	Export Shapefile	Export Shapefile to Google Earth KMZ file
	3.14	File Manager	Renames or deletes gridfiles and shapefiles.
	3.15	Download	Goes to the DIVA-GIS website to download data (needs an Internet connection).

3.1 Import Points to Shapefile

With this menu you can create a shapefile of points from a text file, a DBF file or an Access database.

The TXT file must have a header row containing the variable names. It does not matter whether the columns are separated by spaces or a symbol (such as a comma or tab): the importation wizard will read your data anyway when you tick the box which specifies the separator you are using (Figure 5). However, "tab-separated" is probably best. Using commas as separators causes problems when you have a field with locality descriptions, which may well include commas (this problem can be circumvented by using "quotes" around text.

DIVA will figure out for itself what type of data is present in each column of the database: text, integer (whole) or real (decimal) numbers. But you may change this automatically generated setting if you wish. The same goes for the maximum number of spaces that a value of the variable will need. If you indicate fewer spaces than are actually used, the data will be truncated (cut off at the position that you indicated), not rounded.

With *Import Points to Shapefile/From dBase IV file (DBF)* you can make a shapefile of points from a DBF file if that DBF contains fields with latitude and longitude (both in decimal degrees). First, you must indicate the filename of your DBF file. And you must provide an output filename that is different from this input filename.

The program then reads the input file and allows you to select the fields that have the X (longitude) and Y (latitude) coordinate data. By default, only numerical fields are listed for you to choose from. However, you can check the "Include Text Fields" box to see text fields as well. If you use a text field for the X and Y coordinates, DIVA-GIS will attempt to transform the text values to numbers. Where this is not possible, or where there is no entry at all, an "empty" record is created. That is, the record is copied to the DBF table of the shapefile, but no associated point is created.
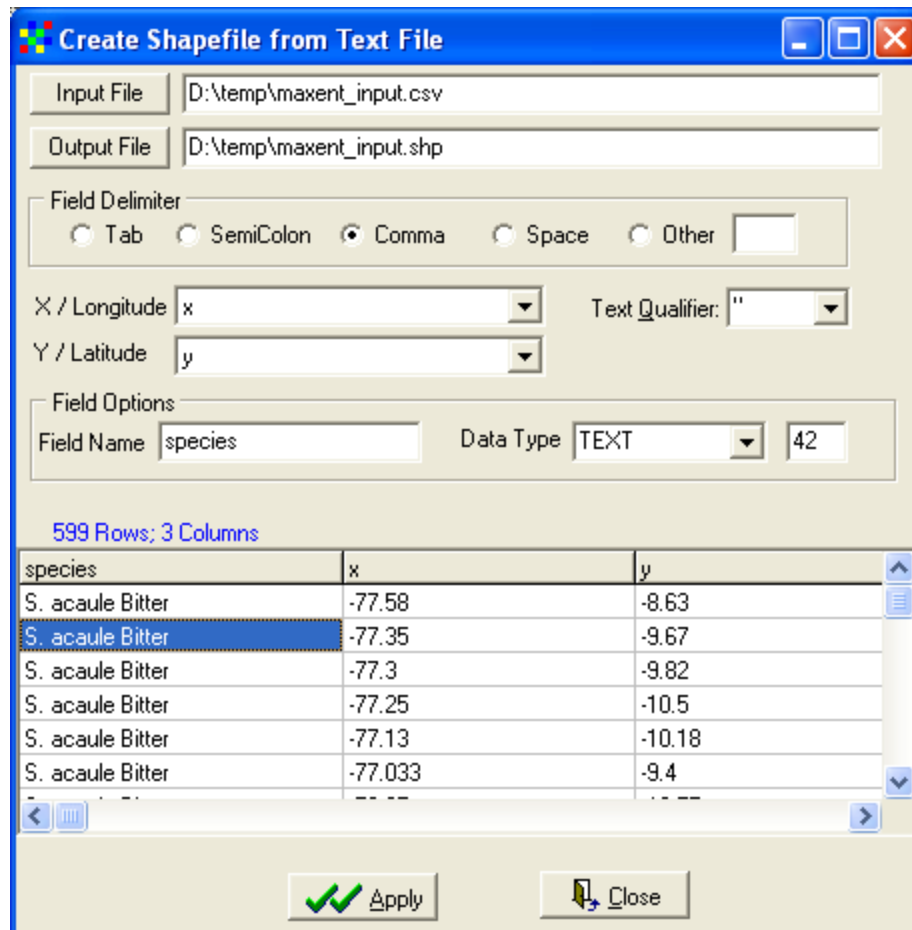
Figure 5. *Create Shapefile from Text File* window

3.2 Import Text to Line/Polygon

Here you can create a shapefile of lines or polygons from a text file (for example, with coordinates obtained with a GPS, or read from a map). For each object (line or polygon) in the text file, the first line should have an ID (number) followed by a pair of coordinates (X and Y, typically longitude and latitude), all separated by spaces.

On the following rows are coordinate pairs for the nodes of the object. After the last node of an object, put END, and start again for the following object. At the end of the file there should be an additional END (Box 4).

Box 4. The general format and an example of a text (.txt) file that can be imported to a shapefile of lines or polygons. "(…)" stands for any number of additional pairs of x and y. The map shows the result for the example data

Format

```
ID   x   y

x   y

 (…)

x   y

END

 ID   x   y

x   y

 (…)

x   y

END

END
```

Example

```
1 72 21
72 22
73 23
74 22
75 20
74 19
72 21
END
2 75 15
76 16
77 15
78 14
75 15
END
END
```
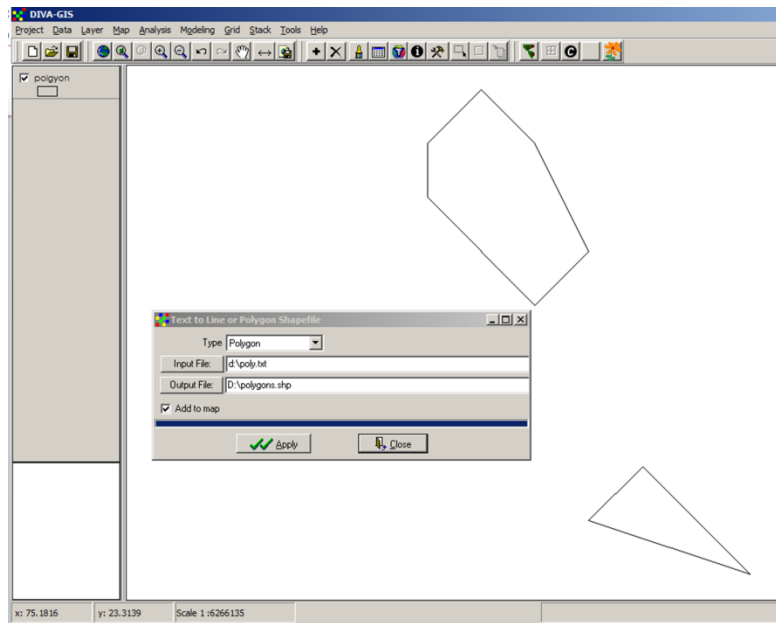
3.3 Draw Shape

You can use this function to create points, lines or polygons shapefiles by drawing (clicking) directly on the map (click on the "Draw" button). For each point, line or polygon that you have drawn, you can add two attributes that identify or describe the object. You can use a numeric attribute ("value") and/or a text attribute ("Label") by typing something in the box on the form. When you are done, press the Save button to save the results to a shapefile.
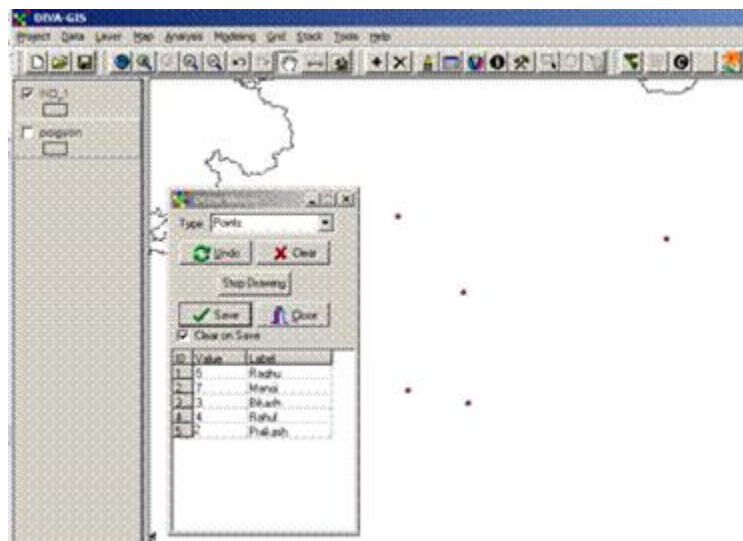
Figure 6b. *Create Shapefile by drawing on screen*

3.4 Polygon to Grid

This function creates a grid (output file) based on the features of a polygon shapefile (input file). If you do not select a particular field from the input shapefile, the grid will have the value 1 for the area covered by the polygon, and 0 otherwise. If you select a (numeric) field, each grid cell in the output file will be assigned the value of the polygon that covers the center of the cell.

Creating a grid from polygons can be useful to, for example, create a "mask" that can be used to remove cells (set to NODATA) from an existing grid that are outside of the

area of interest (also see the Grid / Reclass and Grid / Overlay). In this case you would convert the shapefile to a raster, using the existing raster as template, do not select a field such that you get a new grid with NODATA and 1, and then multiply (overlay) the two grids to only show values for the required region (e.g., a country).

3.5 Points to Convex Polygon

This creates a "bounding" convex polygon around a set of points. For example, you can use this to make a "range map" from a set of points representing point localities for a species. If you want to use all the points in the shapefile, use the *Single* tab. If you want to create more than one range map, use the *Multiple* tab and specify which field and which values in that field you want to use to distinguish the points that will be bounded. Once you have the polygon defining the range of an organism, you can use *Polygon to Grid* as described above to create a gridfile version of the range map.

3.6 Selection to New Shapefile

This saves a selected part (records) of the active (elevated in legend) shapefile to a new shapefile. Parts of shapefiles can be selected graphically, right on the map, or by querying the database (see the options available in the Layer menu; as explained in Chapter 5).

3.7 Extract Values by Points

The *Extract* tool assigns values to the locations specified in the active points shapefile. You can extract values from a polygon shapefile, a gridfile, a stack (of gridfiles) or a CLM file (see below). In all cases, the result is a text (.txt) file.

If you extract values from a stack, you can select a field in the point layer that will be used to "match the grids by class". That means that it will only extract values from a grid when the attribute of a point (for the selected field) matches the name of the grid. For example, you can use this when you have a shapefile of points with a field with species names, and a number of grids with filenames that match these species names.

3.8 Climate

DIVA-GIS comes with a climate data set for the whole world, excluding major water bodies (oceans) and Antarctica. These data are available at different spatial resolutions, and can be downloaded from the DIVA-GIS website. The climate data are stored in a special format (CLM files) to allow quick access and reduce storage space requirements. You can also use your own climate data in DIVA-GIS.

With *Climate/Point* you can click on the map (on land!) and DIVA-GIS will show climate data for that location. The window shows the altitude (in meters above sea level) and monthly average minimum and maximum temperature (°C) and monthly precipitation (mm), both as tables and graphs. It also shows the values of 19 bioclimatic variables derived from these monthly data.

Climate/Map provides a tool to map a specified climate variable over a specified area. When you press the "Read from layer" button in the *Climate/Map* window, the dimensions of the active layer will be copied to the minimum and maximum coordinates. By clicking "Adjust", the dimensions of the output gridfile are adjusted to the grid cells of the climate database. You can also draw a rectangle on the map to define the area (using the "Draw rectangle" button). Choose the variable that you need and whether you want a gridfile for the current or for the projected future climate. The result will be displayed automatically if you check "Add to map".

In the *Climate/Make CLM files* window you can construct CLM files from your own climate grids. CLM files are used to store grids of monthly climate data in DIVA-GIS.

All input files (gridfiles) must have the same number of rows and columns and the same origin. You must always provide a gridfile with altitude data. If you do not have such a file, you can replace it with any other file (as long as it has data and nodata in the right places, and of course, the output altitude file should be discarded). It is important to note that cells with "nodata" in the altitude file will not be stored in the CLM files.

The output of *Make CLM files* should always include the files index.clm and a *.cli file, without which the climate data files cannot be read properly. Store all files in a single

folder. Set the default path for the climate data to that folder (in *Tools/General Options*).

3.9   Check Coordinates

The *Check Coordinates* facility helps you to verify whether coordinates are correct.

The first time you make a shapefile of your database you are most likely to immediately spot some gross errors. For example, if you have a file of terrestrial bird specimen from the Solomon Islands, it is likely that some dots will fall in the ocean, and there might even be one in Siberia. These impossible or unlikely locations are easy to spot, and often also easy to correct. They may well be just typing errors. However, it is likely that there will also be other errors that cannot be spotted so easily.

The *Data/Check Coordinates* tool helps you identify such errors, using a method described by Hijmans et al. (1999). By simultaneously querying the specimen database and an administrative boundaries database, a new (temporary) database is created. For each record, this new database contains the location information according to both the specimen database *and* the administrative boundaries database. This information should be the same, and any mismatches probably reflect errors (or name or boundary changes!). This is illustrated in a simple example for an imaginary island that has three provinces, called A, B and C. Six germplasm accessions have been collected. The coordinates of the collection locations, according to the genebank's database, have been plotted on a map of the provinces (Figure 7). By querying the two databases, Table 3 is generated, pointing at accessions 2, 4 and 6 as possible errors.

In addition, it is possible to use altitude to spot likely errors in the coordinate data. In many cases, the passport data will include altitude. This figure can be compared with an estimate of the altitude of the locality, using *Data/Extract Values by Points*. This form of checking may not be particularly effective in highly-dissected mountainous terrain, because the estimated altitude will not be very precise. More precise estimates can be obtained using the SRTM elevation database (see the DIVA-GIS website). If coordinates are changed, it will be necessary to keep a record of the original coordinates, and of the reason and way the data were altered.
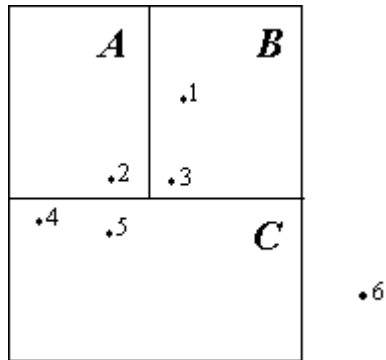
Figure 7. Imaginary island with three provinces (A, B and C) and the location of six collecting sites.

Table 1. The location of six imaginary collecting sites according to the genebank database and according to the administrative boundaries database

Genebank database		Administrative Boundaries
Accession	Province	Province
1	B	B
2	B	A
3	B	B
4	E	C
5	C	C
6	C	

3.10 Export Gridfile

With the *Export Gridfile* function you can export DIVA gridfiles to a number of different formats, listed in Table 4. You can either export a single file, or a group of files all at once.

Table 2. Export formats

Format	File extension
Text	TXT
Generic binary	BIL and HDR
Shapefile (points or polygons)	SHP, SHX, DBF
Google Earth	KMZ

IDRISI v2	IMG and DOC
GRASS (ASCII)	GRS
ESRI (Arc/Info, ArcView) shapefiles	SHP, SHX, and DBF
ESRI ASCII grids	ASC
ESRI binary grids	FLT and HDR

- A *Text File* has lines with the number and value of each grid cell. Cells are counted from the upper-left to the lower-right corner. "Nodata" values are exported as "-9999".

- When exporting gridfiles to the generic binary format BIL ("band interleaved by line"), a data file with the extension BIL is produced, as well as a header file with the extension HDR. These files can be imported into a number of GIS programs, including IDRISI, Arc/Info (with the "imagegrid" command) and ArcView (where they can be opened as an "image"). If you need a file in the similar formats BIP or BSQ, you can rename the extension of the output file, because these are the same when only one grid (or "band") is stored in a file.

- When exporting grids to IDRISI (version 2 and earlier), the result is a data file with extension IMG and documentation file with extension DOC.

- Files in the GRASS (ASCII) format can be imported to GRASS using the command "r.in.ascii".

Grids can also be exported to shapefiles (with rectangular polygons). This can be particularly useful when you want to use the data in ArcView but do not have the Spatial Analyst extension that allows you to visualize and manipulate grids (or the grid module in ArcInfo). In other cases, it would be more appropriate to export the gridfile to an "ASCII grid" or a "floating point" grid (extension FLT and HDR).

3.11 Import to Gridfile

With the *Import to Gridfile* module, you can import one or many grid data files into DIVA-GIS from the IDRISI (IMG or RST), generic binary (BIL/BIP/BSQ), and ESRI binary export formats.

3.12 Write VRT file

This option lets you create ".vrt" header files for existing gridfiles, such that they can treated like "VRT" ("Virtual") grid (raster) format files. These files can be read by ArcGIS, Quantum-GIS and other software that used the GDAL library. See http://www.gdal.org/gdal_vrttut.html for more info.

3.13 Export Shapefile

Save a shapefile to a Google Earth KMZ file (a compressed KML file). See above under Export Gridfile to export gridfiles rather than shapefiles.

3.14 File Manager

With the *File Manager* you can delete, copy and rename shapefiles and gridfiles. As both shapefiles and gridfiles in fact consist of more than one file, this can be a very helpful utility.

3.15 Download

This launches your Internet browser and the webpage http://www.diva-gis.org/data, from which you can download geo-referenced databases.

4. THE *LAYER* MENU

The *Layer* menu (Box 5) allows you to add and delete a layer (also known as "theme") to a project, and change a layer's properties. A layer can be a shapefile, a DIVA-GIS gridfile, or a georeferenced image (TIF, JPG or SID), but most functions in the *Layer* menu refer to shapefiles. Gridfiles and shapefiles for all countries of the world are available from the DIVA-GIS website.

Box 5. The *Layer* menu one symbol missing

Icon	Sect.	Name	Short explanation
✚	4.1	Add Layer	Adds a layer to the map
✕	4.1	Remove Layer	Removes the active layer from the map
🖌	4.2	Properties	Changes the style, color and size of lines of the active layer
abc		Add labels	Adds labels to a layer on the map, using one the fields in the shapefile database (undo by resetting in the *Properties* window)
❶	4.3	Identify Feature	Shows the attribute data of a geographic feature of the active layer on clicking on it
▦	4.4	Table	Shows the attribute data of the active layer (shapefiles only)
🔽	4.5	Filter	Displays a defined sub-set of the records from the active layer (shapefiles only)
🛠	4.5	Select Records	Selects records from the active layer (shapefiles only) that comply with a specific condition (query).
�textBox	4.5	Select Features	Selects point features by clicking or drawing a rectangle on the map
◻	4.5	Clear Selection	Unselects the currently selected point features of the active layer
📋	4.6	Copy	Copies the active layer to the clipboard
📋	4.6	Paste	Pastes a layer from the clipboard to the map
≠		Hide/Show Legend	Hides or shows information on the active layer in the legend

4.1 *Add Layer* and *Remove Layer*

Use *Layer*/*Add Layer* from the menu bar to add a layer to the map window. The new layer is added on top of the list of layers already in the legend. To remove a layer from the project, click on it once in the legend, to make it the active layer, and then click on *Layer*/*Remove Layer*. This will not delete the data; it will only remove the link to it from the current project. Multiple layers can be selected, and then removed together, by clicking on them (in the legend) while holding the down the Shift key.

4.2 Properties

Properties of the spatial objects of a shapefile, such as the size and shape of points and the colour of polygons, can be modified using *Layer*/*Properties*. You should first make it the active layer by clicking on it in the legend. Double-clicking on a layer in the legend also activates the *Properties* window. There are three ways to change the symbols:

1. you can change them all at once (use the *Single* tab)

2. you can give every unique entity a different symbol according to one of its attributes (*Unique*) or

3. you can classify the numeric attributes and give a different symbol to each class.

Use the reset button every time you choose a different attribute or number of classes. Here's a useful tip: if you want all polygons transparent except for one, you should first make the polygons transparent in the *Single* tab and then go to *Unique* to change the one polygon you want with a solid fill.

4.3 Identify Feature

When you click on an item (geographic feature) on the map's active layer, the *Identify Feature* window shows the variables in the database and their values for that item. If you click on more than one item, e.g., a number of points with (nearly) the same

location, the data for all these items are shown. The number of the visible record and the total number of records are indicated on top of the list of variables (e.g., "Rec 1 of 5"). The "↑" and "↓" buttons on the right of the list can be used to toggle between the selected records. If the active layer is a gridfile, the column, row, number and value of the grid cell that was clicked on are shown.

4.4 Table

Table allows you to view the database of the active layer (shapefile). If you click on a record in the table, the location of the corresponding geographic object will be highlighted. You can also move the center of the map to that object.

4.5 Select Records

Records can be selected for various purposes. For example, to save a subset of a layer to a new file, or to find geographic features that meet particular conditions. The selected items will be displayed in a different color (yellow is the default). You can make a selection by either making a query in the *Select Records* window, or by drawing on the map after clicking on the *Select Features* button. Only layers that are made from shapefiles can be selected from.

In the *Select Records* window you can either select by values, or by query. The first option is useful for variables with a limited number of values, which are typically non-numerical. Select the variable and all values will be listed. Then select the values to include in the selection. When using a query, you must select a variable, a criterion and the value. Use the "Add" button to put the query in the dialog box, and continue adding additional conditions using "AND", "OR" or condition parentheses. When you are ready, click "Apply".

Selections can also be made using *Select Features*, by clicking on a map item or drawing an area (click, keep your finger down, and move the mouse). The shape of the selection window (rectangle, circle, etc.) can be set in *Tools*/*General Options*.

Selections can be cleared using *Clear Selection*. The selection can also be converted to (saved as) a new shapefile with *Data/Selection to New Shapefile*.

The *Filter* facility also allows you to "select" features of a shapefile, but only one field at the time, and you can't save the result to a new shapefile.

4.6 *Copy* and *Paste*

The *Copy* and *Paste* functions can be used to copy and paste a layer in the legend. This is useful if you want to use one layer more than once on the same map. For example, a layer of a country can be used as the lowest layer, to give a background color, and – this time with transparent polygons – as the highest layer to place departmental boundaries on top of all other layers.

5. THE *MAP* MENU

The *Map* menu includes functions for changing the properties of the map extent, center and scale. These are explained in Box 6.

Box 6. The *Map* menu

Icon	Sect.	Name	Short explanation
		Properties	Sets map properties (background color, projection)
↔		Measure Distance	Measures a distance, by clicking twice or more on the map (double click to stop drawing)
	5.1	Map to Image	Exports the map or its TOC to the clipboard or to a graphics file
		Zoom to Full Extent	Zooms to the extent of all layers included in the TOC
		Zoom to Active Layer	Zooms to the extent of the currently active layer
		Zoom to Selected	Zooms to the extent of the current selection (if any)
		Zoom In	Zooms in to a point by a mouse click or to an area by drawing a rectangle (while holding the left mouse button down)
		Zoom Out	Zooms out from a point by a mouse click
		Previous	Goes back to the previous setting (zoom/pan)
		Next	Goes to the next setting (zoom/pan) (after Previous)
		Pan	Moves the visible zone of the map by holding down the left mouse button and moving the mouse
		Overview Map	Show or hide the overview map
		Design View	Switch to "design view" (from "map view") to create a map (e.g., with legend, text, scalebar)

5.1 Map to image

With *Map to Image* you can export either the current map or its table of contents to the clipboard or to a graphics file, in BMP (bitmap) or EMF (enhanced metafile) format. These images can subsequently be used as an illustration in other applications, such as a word processor or presentation software. BMP is a raster image format, while EMF is a vector format. You may want to convert these to JPG format using a graphics package to save space.

6. THE *ANALYSIS* MENU

This chapter describes the methods for analyzing biological distribution data available in DIVA-GIS. These analyses are all based on the location (latitude and longitude) and additional attributes of point data. The points represent locations where a specimen (e.g. herbarium, museum, genebank) was collected, or where any observation of the presence of a specific biological unit (e.g., species, landrace, genotype, allele) was made. These points should be in the active (shapefile) layer in the DIVA-GIS project (to remind you: a shapefile is made active by adding it to the map and clicking on its legend entry once, which highlights it). The output of the analysis routines can be a gridfile, a shapefile, or a database file (DBF).

Box 7. The *Analysis* menu

Icon	Sect.	Name	Short explanation
⊞	6.1-3	Point to Grid	Creates a grid with different indices (diversity, distance) or statistics from a points shapefile
◈	6.4	Point to Polygon	Creates a new shapefile (either from an existing one, or a rectangular or pentagonal grid) with diversity indices from a points shapefile
∷	6.5	Point to Point	Calculates diversity indices for a neighborhood around each observation in a points shapefile
Σ∷	6.6	Summarize Points	Calculates diversity indices for all the data in the database of a points shapefile
↘	6.7	Distance	Calculates distribution statistics from point data
◇	6.8	Autocorrelation	Assesses the presence of spatial autocorrelation from point or grid data
		Centroid	Calculates the centroid for each polygon in a shapefile.
▥	6.9	Histogram	Makes a histogram of the frequency distribution of the data in a grid
◹	6.10	Regression	Calculates a regression of the values in a grid against those in another grid
⌐	6.11	Multiple Regression	Calculates a regression of the data in a grid against those in two or more other grids

6.1 Point to Grid

Some of the main analytical functionality in DIVA-GIS is found in the *Analysis/Point to Grid* menu. As the name implies, this menu allows you to take point data in a shapefile and create gridfiles where each grid cell takes a value calculated based on the points that are found within the cell. The *Point to Grid* window is shown in Figure 8.
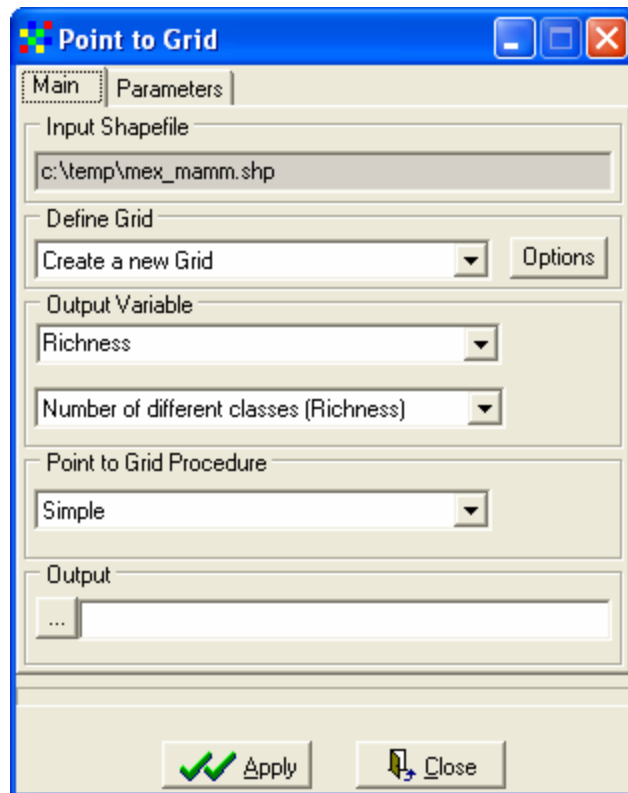
Figure 8. The *Point to Grid* window

6.1.1 Defining grids

The first step is to define the extent and resolution of the grid that will be used to make the calculations. You can either specify this from scratch or copy the parameters from an existing gridfile.

A grid has an "extent," which refers to the geographic area it covers. As a grid is rectangular, it can be defined by two opposite corners, e.g., the upper left and the lower right corner. The "resolution" is the area of each grid cell. Resolution is always expressed in the units of the geographic reference system. For example, in the case of the latitude/longitude system, a resolution of 1 means that each cell is 1 degree by 1 degree. The vertical and horizontal resolutions of a grid are usually the same, but could be different. If the grid extent and resolution are known, the number of rows and columns in the grid can be calculated. Alternatively, the resolution can be calculated when the other variables are known.

To define a grid, its geographic reference system must be specified and then its extent and resolution. Because maps are flat, while the Earth is not, a geographic reference system, or map projection, has to be used that renders the three-dimensional reality in two dimensions (a flat map) with minimal distortion. You can use any projection in DIVA-GIS (see Chapter 10 for a discussion of projections).

DIVA provides default parameters for the output grid that are calculated on the basis of the extent of the active points shapefile layer, and by setting the resolution to 1 (in the units of the reference system). However, when this would lead to less than 10 or more than 100 rows or columns, the default resolution is adapted accordingly.

You can change the default parameters by selecting any statistic from the *Analysis/Point to Grid* menu and clicking on the "Options" button under "Define Grid". On the window that appears when you do so (Figure 9), you can change the default grid extent (as defined by the coordinates of the four corners), resolution and number of columns/rows. As these parameters are all interrelated, you need to specify which other parameter should also change. For example, if you want to change the resolution, you must specify whether you want the number of columns to adapt, or the grid extent. When you want to adapt the grid extent, you can fix one corner of the grid, using the "Fix Corner" box. You can also define the grid extent by drawing a rectangle on the map after pressing the "Draw Rectangle" button.
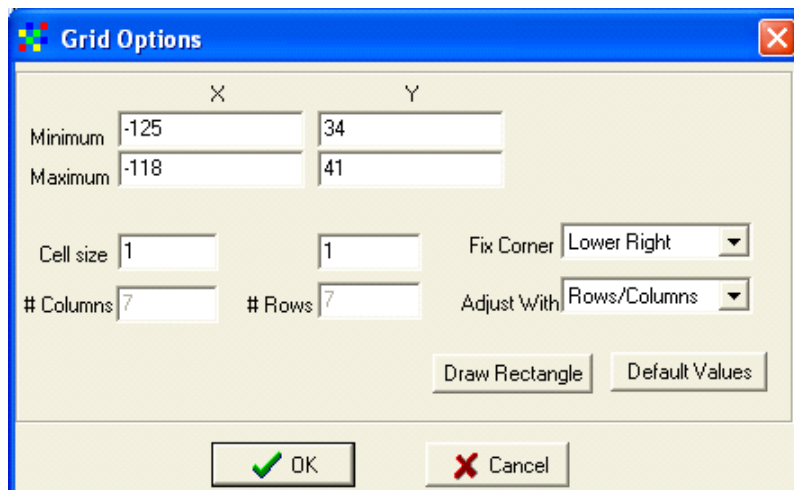
Figure 9. The *Grid Options* window

6.1.2 Using the parameters of an existing grid

The alternative to defining manually the parameters of a grid is to use the parameters of an existing grid. You can use the parameters from a gridfile produced in a previous analysis by scrolling down in the *Define Grid* bar of the *Point to Grid* window and then pressing "Options" to specify an existing grid. An existing gridfile can also be used as a "mask" to indicate what part of the grid should be ignored, or should get special treatment in the case of the Inverse Distance Weighted method (see section 4.3). The part of the grid that should be ignored can be indicated using numerical values or ranges.

6.2 Choosing output variables

There are a number of different output variables that can be calculated for a grid. The output variable you want can be selected under the *Point to Grid/Main/Output Variable* menu. Output variables are grouped, and associated with a number of options that can be set on the *Parameters* tab to the right of the *Main* Tab. The different output variables are discussed in the following sections.

6.2.1 Richness

In the *Richness* group there are four distinct output variables: Number of different classes, Number of observations, Presence/Absence and Rarefaction.

Number of different classes counts the different classes of a variable (e.g., the different species names in a dataset covering a genepool) present in each grid cell. The *Parameters* tab must be used to indicate which variable in the shapefile should be considered, and possibly to exclude irrelevant values.

The *Number of observation*s option calculates the number of points present in each grid cell. As there may be points in the shapefile that are not relevant, you may want to exclude these. In that case, you must select a variable from the database on the Parameters tab (Figure 10), and then you can exclude one or more of the values of that variable.

Presence/Absence simply gives cells in which a class is found the value 1 and cells in which it is not found the value 0.

The *Rarefaction* technique estimates the number of classes (species, for example) that would have been observed given a number of observations that is specified by the user (Sanders, 1968; Hurlbert, 1971; Magurran, 1988). Equation 1 gives the formula which is used to make this calculation.

A disadvantage of this method is that the estimate can only be calculated for those cells in which the actual number of observations is higher than that for which the estimate is calculated.

$$E(S) = \sum \left[1 - \left\{ \binom{N - N_i}{n} \Big/ \binom{N}{n} \right\} \right]$$
(Equation 1)

E(S) = Expected number of classes in the rarefied sample;
N = Total number of observations per cell, in the sample to be rarefied;
N_i = Number of individuals in the *i*-th class, in the sample to be rarefied;
n = User specified standardized sample size.
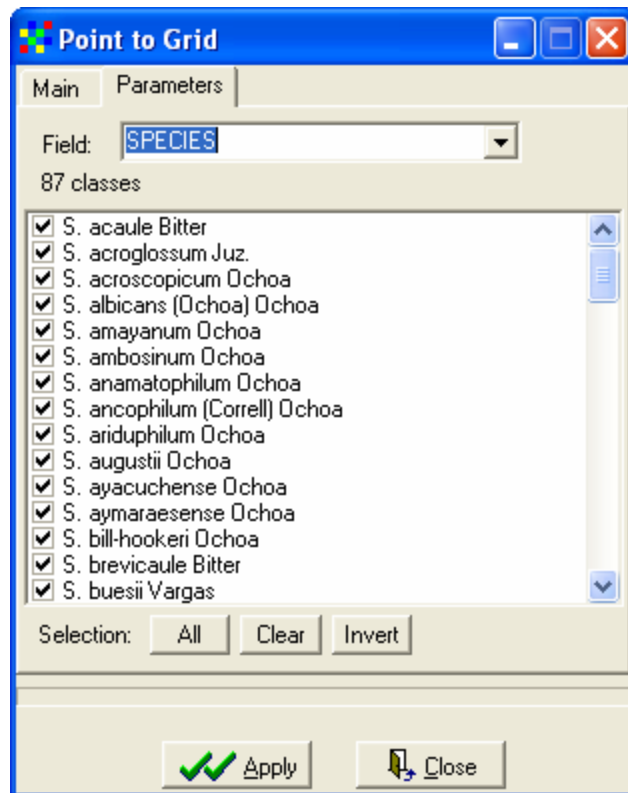
Figure 10. The *Parameters* tab in the *Point to Grid* menu.

6.2.2 Estimators of Richness

The number of species (or whatever other units) observed in an area depends to some extent on the effort invested in recording there. Because a complete census is rarely feasible, in most cases only a sample of an area is surveyed. An important problem that then arises is to estimate the total number of species (or whatever other units), S_{max}, for the area. This estimate can give both a measure of the completeness of the inventory and also allow for better (i.e., less biased by the number of observations) comparison with the species richness of other localities. An estimate of the maximum number of species is also useful when assessing if the further information to be gained from continued sampling justifies the cost. Various different approaches have been proposed to estimate S_{max}. Some of these have been implemented in DIVA-GIS based on the review by Colwell and Coddington (1994), and later authors.

Chao 1

Chao (1984) derived a simple estimator (S1 or "Chao-1") of the true number of species in an assemblage based on the number of rare species in the sample:

$$S_1 = S_{OBS} + \left(a^2 / 2b\right)$$

> S_{obs} = the observed number of species in a sample
> a = the number of observed species that are represented by only a single individual in that sample (i.e., the number of "singletons")
> b = the number of observed species represented by exactly two individuals in that sample (the number of "doubletons")

Chao 1 Corrected
This corrected version replaces the original Chao estimator (which is still included to allow for comparison with studies that have used this estimator). The corrected version is less biased.

$$S_{Chao1} = S_{obs} + \frac{F_1^2}{2(F_2+1)} - \frac{F_1 F_2}{2(F_2+1)^2}$$

> S_{obs} = the total number of species observed
> F_i = the number of species that have exactly i individuals (F_1 is the frequency of singletons, F_2 the frequency of doubletons)

Chao 2
Chao 2 is an incidence-based estimator of species richness (Chao 1987). Chao 2 and the Jacknife estimators are based on the number of samples for an area. To create samples, DIVA-GIS divides each grid-cell into 4 or 9 sub-areas.

$$S_{Chao2} = S_{obs} + \frac{Q_1^2}{2Q_2}$$

> S_{obs} = the total number of species observed in all samples pooled
> Q_j = the number of species that occur in exactly j samples (Q_1 is the frequency of uniques, Q_2 the frequency of duplicates)

Jacknife
"Jacknife 1" is the first-order jacknife estimator of species richness (incidence-based) (Burnham and Overton 1978,1979; Heltshe and Forrester 1983)

$$S_{jack1} = S_{obs} + Q_1 \left(\frac{m-1}{m} \right).$$

"Jacknife 2" is the second-order jacknife estimator of species richness (incidence-based) (Smith and van Belle 1984)

$$S_{jack2} = S_{obs} + \left[\frac{Q_1(2m-3)}{m} - \frac{Q_2(m-2)^2}{m(m-1)} \right].$$

S_{obs} = the total number of species observed in all samples pooled

Q_j = the number of species that occur in exactly j samples (Q_1 is the frequency of uniques, Q_2 the frequency of duplicates)

m = the total number of samples

Michaelis-Menten

This estimator is calculated by repetitive random sampling and fitting an asymptotic model, following the method of Raaijmakers (1987). For each sample size (from 2 to the number of observations –1), the average number of species in the sample is calculated over the random samples (the default is 100 samples for each sample size, but a higher number may be better for some data). The number of species is estimated from this generated species accumulation curve.

This asymptotic model assumes that the probability that the next individual captured will be a new species declines linearly with species number, and thus the species accumulation curve is the negative exponential function:

$$S(n) = S_{max} \left(1 - e^{-kn} \right) \qquad \text{(Equation 2)}$$

k is a fitted constant

n = the number of samples

The asymptotic behavior of the accumulation curve can also be modeled as the hyperbola:

$$S(n) = \frac{S_{max} n}{B + n}$$
(Equation 3)

S_{max} and B are fitted constants

This is the Michaelis-Menten equation used in enzyme kinetics and thus there is an extensive literature discussing the estimation of its parameters, which unfortunately presents considerable statistical difficulties (Colwell and Coddington, 1994). The method implemented in DIVA-GIS, favored by Raaijmakers (1987), is to calculate S_{max} and B using their maximum likelihood estimators as follows:

$$\text{If } X_i = \frac{S(n)}{n} \text{ and } Y_i = S(n) \text{ then}$$

$$\hat{B} = \frac{\overline{X} S_{yy} - \overline{Y} S_{xy}}{\overline{Y} S_{xx} - \overline{X} S_{xy}} \text{ and}$$

$$\hat{S}_{max} = \dot{Y} + \hat{B}\overline{X},$$
(Equation 4)

S_{yy}, S_{xx} and S_{xy} = the sums of squares and cross products of the deviations Y Y_i and X X_i

S-obs

This is simply the actual number of species observed per grid cell.

6.2.3 Turnover

Turnover (or beta-diversity) is a measure of the rate at which species assemblages change in space. It indicates how different a number of nearby areas are. Imagine two large areas with similar numbers of species overall, but one with different species in all its grid cells and another with the same species in all its grid cells. The first area would have a high turnover, the second area a low turnover.

At this point, only Whittaker's (1960) measure of beta-diversity is implemented in DIVA-GIS (Equation 9). It can be calculated for each grid cell considering its 8 neighbors (2 horizontal, 2 vertical, and 4 diagonal; "Queen's case") or considering its 4 closest neighbors (2 vertical and 2 horizontal; "Rook's case").

$$\beta_w = (S/\alpha) - 1 \qquad\qquad\qquad \text{(Equation 5)}$$

S = the total number of species over the grid cells considered
α = the average number of species in the grid cells considered

6.2.4 Diversity indices

DIVA-GIS can calculate a number of different diversity indeces for each grid cell. You must select a variable (field) from the input database for which you want to calculate an index, e.g species name. The formulas for all indices were taken from Magurran (1988), who provides a detailed description of their properties. See Table 5 for the mathematical description of the different diversity indices.

Table 3. Diversity indices

Index	Formula
Margalef	$D_{Mg} = (S - 1) / ln(N)$
Menhinick	$D_{Mn} = S / \sqrt{N}$
Shannon	$H' = -\sum p_i \ln p_i$
Simpson	$D = \sum(n_i (n_i - 1) / N/(N-1))$
Brillouin	$HB = (\ln N! - \sum \ln n_i!) / N$

S = number of unique classes (species) per cell

N = number of observations per cell

n_i = number of individuals in the i-th class

p_i = proportional abundance of the i th class = n_i / N

The Simpson index, D, decreases with increasing diversity, and hence it is usually expressed as 1-D or 1/D. In DIVA-GIS it is expressed as 1-D. You can use *Grid/Scalar* to calculate *D or* 1/D from this (see Chapter 9).

6.2.5 Molecular marker data

Molecular markers data associated with point localities can also be analyzed in DIVA-GIS. At this moment, only absence/presence data (e.g., of bands on an electrophoretic gel) can be analyzed. Molecular distance, molecular diversity and number of haplotypes can be calculated.

The fields that have the data must be selected using the *Parameters* window. The values in these fields can be checked in the "Values" tab. A band can either be "present" (value = 1 in the database), "absent" (value = 0 in the database), or "missing" (any other value).

Average or maximum molecular distance between the observations in a grid cell can be calculated using either Jaccard's (1908) distance (J), Nei and Li's (1978) distance (NL), and Sokal and Michener's (1958) distance (SM) (Equations 5-7).

$$J_{xy} = 1 - \frac{n_{11}}{n_{11} + n_{10} + n_{01}}$$
Equation 6

$$NL_{xy} = 1 - \frac{2 \times n_{11}}{2 \times n_{11} + n_{10} + n_{01}}$$
Equation 7

$$SM_{xy} = 1 - \frac{n_{11} + n_{00}}{n_{11} + n_{10} + n_{01} + n_{00}}$$
Equation 8

Where:
n_{11} = the number of bands shared by the genotpyes x and y (positive matching)
n_{10} = the number of bands present in x and absent in y
n_{01} = the number of bands present in y and absent in x
n_{00} = the number of bands absent both in x and y (negative matching)

Alternatively, molecular diversity can be calculated using Nei's diversity index (Equation 8).

$$NDI = \sum_{ij} x_i \cdot x_j \cdot NDI_{ij}$$
Equation 9

Where:
x_i = the frequency of the *i*-th allele in the population and
NDI_{ij} = the number of allele differences per locus between the the i-th and j-th loci

6.2.6 Choosing complementary areas

The *Reserve Selection* procedure aims to identify sets of grid cells that are complementary to each other, i.e. that capture a maximum amount of diversity in as few cells as possible. Instead of using simple richness, an adjustment can be made in which rare observations get a higher weight.

The procedure is based on the algorithm described by Rebelo (1994) (see also Rebelo and Sigfried (1992)). It has, for example, been used to determine priority areas for *in situ* conservation of species in a family of flowering plants in South Africa. The following discussion refers to species, but any multi-state variable could be used. The procedure is less straightforward than it might seem. Whereas the selection of the first cell is easy – it is the cell with highest species richness (or a random choice between ties if there are any) – the choice of the next cell depends on the previously selected cell. This is because the species in the cell with the second highest number of species may also be present in the first cell. In other words, the cell with the second highest number of species may not contribute very much to the overall number of species selected. To maximize the total number of species selected in as few cells as possible is a non-linear optimization problem. Rebelo (1992) developed a procedure that calculates an approximate optimal solution, and this is what has been implemented in DIVA-GIS.

An iterative procedure is used. In each iteration, the "value" of each grid cell is calculated, based on the observations in that cell, and in relation to the observations in the cells already selected. If there are two or more cells with the same "value", one is selected at random. Hence, this procedure can lead to slightly different results every time it is run.

Several options need to be selected under the *Parameters* tab. The first step is to select a variable, and possibly exclude some of the species (e.g., missing values) from the analysis. The second step is to choose a scoring approach. There are two options: "Equal weight" and "Rarity". With the "Equal weight" option, each category (e.g. species) has the same weight. With the "Rarity" option, the value of a cell is calculated using the "rarity value" for each observation. The rarity value equals the number of observations of a category divided by the total number of observations.

There are two ways in which you can limit the number of iterations. This is important because the procedure can take a long time with large databases (or a grid with many cells). You can indicate the "Minimum number of records per cell". This value should be 1 or higher. It allows you to exclude cells with what is, for your purposes, too low a number of observations. You can also indicate the maximum number of iterations that you want the procedure to cycle through.

When the process finishes, a "Process Report" appears, that tells you how many iterations were used (and hence the number of cells selected) and how many distinct observations are present in those cells.

Three new grids are created and added to the project, labeled "Sequence", "Classes" and "Additional Classes".

  − The Sequence grid indicates in which order cells were selected. The cell with most species is selected first, and has a value of 1. The next cell has a value of 2, etc.

  − The Classes grid indicates how many different classes (typically species) are in each selected cells.

  − The Additional Classes grid indicates how many *new* classes (species) are in that cell. These species were not present in any of the previously selected cells.

6.2.7 Statistics

If a numerical variable is selected, some basic statistics can be calculated for that variable, for each grid cell. The statistics included are listed in Table 6.

Table 4. Statistics

Coefficient of variation	Coefficient of variation of values
Maximum	Maximum value
Mean	Mean value
Median	Median value
Minimum	Minimum value
Mode	Mode (most frequent) value
Range	Difference between Max and Min

Range/Mean	Range divided by the mean
Standard deviation	Standard deviation of values
Sum	Sum of all values
Variance	Vriance of values

6.3 Methods of converting point data to grid data

At the core of DIVA-GIS is the conversion of the point data represented by the accession localities to the raster data represented by the diversity and other values of the different grid cells. Several GIS programs allow the conversion of point data to raster data. However, the way that this is generally done (as in the "Simple" procedure in DIVA-GIS) has the following important shortcomings:

- A point that is on a border between grid cells is assigned to one of those cells arbitrarily..

- The value of a point that falls within a grid cell is assigned to that cell only, irrespective of the proximity of the point to other cells.

- The origin of the grid (i.e. its bottom left-hand corner) is arbitrary. Given a coordinate system and resolution of the grid, slightly different origins could lead to quite different results of the point-to-grid conversion.

- Uncertainty about the location of points is not taken into account.

These shortcomings can largely be overcome with the "Circular Neighborhood" as implemented in DIVA-GIS for the *Point to Grid* analysis and described below.

6.3.1 Circular neighborhood

When the *Circular Neighborhood* option is chosen in *Analysis*/*Point to grid*/*Main*, calculations are made based not on the observations within each grid cells but rather on the observation found within a circle of specified radius with its center in the middle of each grid cell. The advantage of this method is that it produces a smoother surface. To avoid some observations not being used at all, the radius should not be smaller than the distance between the center of a square grid cell and its corners. A disadvantage may be that a number of observations (or even all, if the radius is

sufficiently large compared to the size of the grid cells) will used more than once in the calculations. See Bonham-Carter (1994) and Cressie (1991) for more information on circular neighborhoods.

When this method is used, you must specify in the *Options* tab the size of the neighborhood (i.e., the diameter of the circle), and indicate if you want to scale the result. Scaling means you correct the results for the difference between the size of the grid cell and that of the neighborhood.

6.4 Point to Polygon

Polygons, or closed lines that enclose an area, can also be used to summarize point distribution data. For example, you may want to make calculations for different countries.

Make a shapefile of points the active theme and click on *Analysis/Point to Polygon/Diversity.* Then specify which polygons are going to be used to make the calculations. You can either select an existent shapefile (e.g. with administrative boundaries or ecoregions) or define a new shapefile of squares (i.e., a grid) or hexagons. Note that if you define a grid of squares you will get the same output as with *Point to Grid,* except for minor differences due to the way points that fall on the borders between two cells are treated.

The output consists of a shapefile with the following fields:

- − DIVAID and ORDER: these can be ignored
- − OBS, the number of observations
- − SPP, the number of different classes, typically species
- − four diversity indices: Margalef, Menhinick, Shannon, Simpson (see section 7.3.3; Table 5)

The value "-1" in the Simpson column means "missing value" (i.e., the index could not be calculated).

As the size of polygons may vary, unlike that of grid cells, the results may need careful interpretation. Nevertheless, this can be an interesting option that allows summarizing the data by pre-defined, and often well-known, areas.

6.5 Point to Point

An alternative to the use of grids is the kind of point–based approach that was used by the Spatial Intra-specific Diversity (SID) software described by Nelson et al. (1997). The same procedure is also implemented in DIVA-GIS. Calculations are done for all observations lying within a user-defined circle around each observation point. The result is assigned to the location of the observation for which the calculation was done. For such a point-based analysis, you have to specify the variable in the points shapefile to be used. Unlike when creating a grid, you cannot choose the output variable. By default, the following variables are included in the output file:

- longitude
- latitude
- number of observations (NOBS)
- number of distinct values (NDIST)
- the diversity indices listed in Table 5

If you want, you can add to these any of the fields that are in the input shapefile. The results are stored in a shapfile.

6.6 Summarize Points

This calculates diversity indices, diversity estimators and basic statistics for the whole dataset in an input shapefile.

6.7 Distance

With the *Analysis/Distance/Statistic by Class* procedure, you can calculate distance statistics for each unique class of a multi-state variable; for example, for each species in a database. Currently, the following statistics area available:

- number of observations
- minimum distance between two points of each class (MinD)
- maximum distance between two points of each class (MaxD)
- average distance between two points of each class (AverageD)

Distances are expressed in meters. For correct calculations, you must specify whether the shapefile is in decimal degrees or projected (with meters as unit). The output is written to a DBF file.

The same calculation can be done for all points (i.e. without distinguishing among classes), and a complete distance matrix will be the produced (*Analysis/Distance/Matrix*).

6.8 Autocorrelation

Spatial autocorrelation exists when the values of nearby geographical objects are more similar than those of objects that are further away. For example, an altitude grid has a high spatial autocorrelation. The *Autocorrelation* option in DIVA-GIS calculates this relationship by using two common statistics, the Geary and Moran indices, the values of which may be interpreted as in Table 7. You can calculate the autocorrelation on a grid or a points shapefile. The spatial autocorrelation functionality implemented in DIVA-GIS is based on the "Rookcase" software by Sawada (1999).

Table 5. Autocorrelation statistics.

Geary	Moran	Interpretation
$0 < c < 1$	$i > 0$	Autocorrelation exists, data are clustered
$c = 1$	$i = 0$	Independent, at random
$c > 1$	$i < 0$	Negative autocorrelation

6.8.1 For points

When calculating autocorrelation for points, and you need to define which neighboring points to consider. This is done by specifying the lag (or neighborhood) distance. Only points within the lag distance are considered for calculation.

To determine a reasonable lag distance, a number of statistics are shown in the *Options* tab of the *Analysis/Autocorrelation/Points* window. These statistics include the number of observations (points) and statistics describing the distances among all pairs of points (min, max, mean, median and first and third quartile).

You must select a quantitative field from the shapefile's database to calculate autocorrelation. After selecting a field you can deselect individual values that you want to exclude. If you do that you need to recalculate the statistics, which will have changed, because you do no longer consider all points. The results will appear on clicking "Apply"
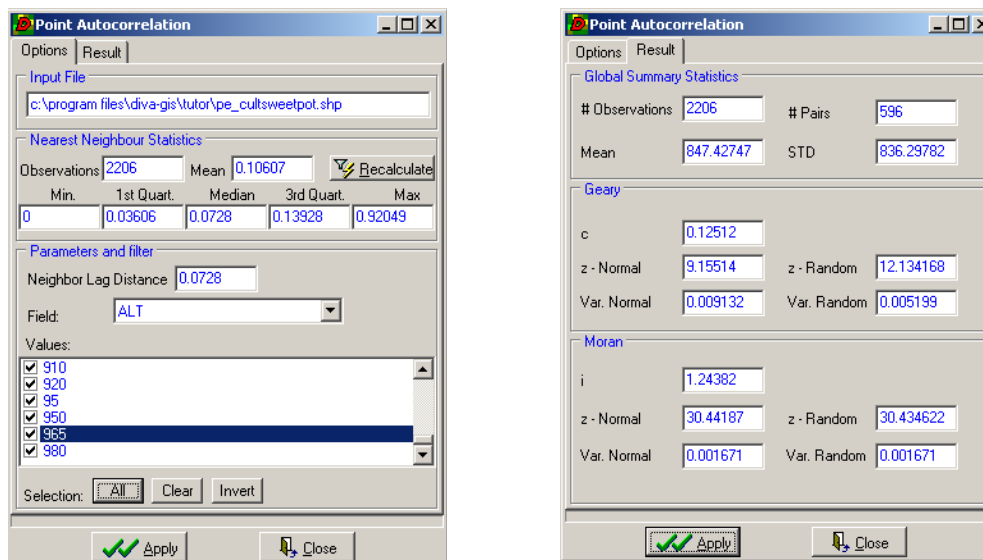
Figure 11. *Autocorrelation/Point* window (*Options*, left and *Results*, right)

6.8.2 For gridcells

No lag distance needs to be specified to calculate autocorrelation of grids. Only adjacent cells are considered, either 4 horizontally and vertically adjacent cells ("Rook's case"), or also including the 4 diagonally adjacent cells ("King's case").

6.9 *Histogram*

With the *Histogram* option your grid data are arranged in classes. The program offers you some common tools to improve the histogram (titles, show grid, etc.) and the result can be copied to the clipboard.

6.10 *Regression*

With the *Regression* tool, you can regress one grid against another (cell by cell) and plot the results. Choose one gridfile for X and another for Y, making sure that the grid extents and resolutions coincide. You can use linear, logarithmic, or exponential regression. For multiple regression and more advance statistical methods, it is convenient to use the *R* program (with raster package).

6.11 *Multiple Regression*

You can regress one grid against several others using the *Multiple regression* tool. If you find a satisfactory (statistically significant) model, you can use it to **predict** values for the same area, or for another area (by changing the independent variables).

7. THE *MODELING* MENU

The *Modeling* menu provides various options for niche modeling, or the prediction of species distribution patterns (Box 8).

Box 8. Functions in the *Modeling* menu

Icon	Sect.	Name	Short explanation
	7.1	Bioclim / Domain	Runs the BIOCLIM and DOMAIN models for predicting the presence of species
	7.2	External Models	Uses other models to predict the presence of species
	7.3	Evaluation	
	7.5	EcoCrop	Predicts the adaptation of a crop for a selected area
	7.6	Terrain Modeling	Creates grids of aspect or slope from an elevation grid

7.1 Bioclim / Domain

With the *Bioclim / Domain* tool you can extract data from the climate database for point locations (in the active points shapefile), make graphs of these data and use the extracted data for predictive niche modeling using the BIOCLIM (Busby, 1991; Nix, 1986) and the DOMAIN (Carpenter et al., 1993) approach.

You can only open this window when a points shapefile is the active layer on your Table of Contents. There are six tabs in this window:

7.1.1 *Input*

Here you can choose to run a model with the standard climate data (DIVA climate data) or with a stack of layers of your choosing.

You can also choose to consider all points as one class or distinguish different classes (e.g., species) by selecting an appropriate field. Duplicate points (with exactly the same coordinates and species) are removed from the analysis. You can choose whether

you want to also remove points that fall within the same grid cell (this is recommended).

7.1.2 *Frequency*

You can make cumulative frequency distribution plots for various climatic variables, for up to two classes (e.g., species), which are shown in different colors on the graph. This can be very useful in spotting outliers, which may be due to errors in the data. Click on a point on the graph (for the second class only) to see where the corresponding locality is on the map, and display attribute and/or climate data for that point. You can show percentiles and the 1.5 inter-quartile range (IQR; points within this range are not considered outliers) on the graphs, and change the size of the dots. You can save the graph to the clipboard with the "Copy" button at the top left hand corner (and then paste it into a document or presentation) and counts of the number of observations per class to a TXT file ("Save Stats" button in the bottom right hand corner).

7.1.3 *Outliers*

Here you can look for outliers in environmental space. You can include 1 to n variables in the evaluation (n being the number of variables in your stack; if you use DIVA-GIS climate files n =19). Suspect records are colored red (as opposed to black). You can click on a line on the graph to see where the corresponding locality is on the map, and display attribute and/or climate data for that point.

7.1.4 *Histogram*

Another way to explore the climate data for a points shapefile is to make histograms. Select a class, a variable and either an appropriate bar width or number of bars. You can restrict the minimum and maximum values to be shown. As usual, you can save the histogram to the clipboard.

7.1.5 *Envelope*

In this menu, you must choose a class (e.g., species) and two climatic variables. On pressing the "Apply" button, a graph will be drawn in which each point on the graph

represents the climate of a point on the map, each axis representing one of the climate variable that you chose.

You can set a percentile value to find out which are the extreme observations in terms of climate. The percentile is used to exclude the extreme values of all climate variables. A box is drawn on the graph to indicate, for the two climate variables selected, which points are within, and which points are outside the climatic "envelope" defined by the specified percentiles. These points will be different for each pair of variables.

The points that fall within the envelopes for *all* the climate variables are colored green on the graph. The points that fall outside one or more envelopes are shown in red. All the points inside the two-dimensional envelope shown on the graph are "selected" (highlighted in yellow) on the map. You can of course save the graph to the clipboard in the usual way. The "Extract" button saves a TXT file with the average values of each bioclimatic variable for each class.

7.1.6 *Predict*

This is the business end of this menu. The extracted climate data can be used to predict where a species is likely to occur, or would survive if it was brought there (this could be on a different continent, for example).

To select an area for mapping, you can enter the coordinates manually, or use "Draw rectangle", or you can press the "Read from layer" button, and the dimensions of the active layer will be copied. These may need to be adjusted a little to coincide with those of the climate database. This is done automatically, but you can also do it by pressing the "Adjust" button.

You can select the climate variables you want to use in the analysis. The fewer variables you select, the larger the predicted areas of suitability will be. You can also choose either the current climate or the future climate.

If you selected "Many classes" on the *Input* tab, you can use the "Batch" option under *Predict* to run the model for all different classes (typically species). Or you can select a single class and run the model for that only.

You must select an output grid. You have a choice of six different output variables for this grid:

Bioclim - Four types of areas are mapped. Areas completely outside the 0-100 percentile envelope for one or more climate variables get a code "0". The cells within the 5-95 percentile get a code "3", those outside this range but within the 2.5-97.5 percentile get a code "2", and the ones outside this but within the 0 -100 percentile for all climate variables get a code "1".

Bioclim True/False – All areas that are within the envelope described by the data points, cut off beyond a certain user defined percentile, are mapped as "true" (1); all other areas are mapped as "false" (0).

Bioclim Most Limiting Factor – For the cells that fall within the 0-100 percentile, the variable for which the percentile score was lowest (or highest) is mapped. The value in each grid square is a number that corresponds to the variables on the *Predict* tab.

Domain - The Domain procedure calculates the Gower distance statistic between each cell on the map and each point, using the values of the 19 climate variables. The distance between point A and grid cell B for a single climate variable is calculated as the absolute difference in the values of that variable divided by the range of the variable across all points. The Gower distance is then the mean over all climate variables.

$$d_{AB} = \frac{1}{p} \sum_{k=1}^{p} \frac{|A_k - B_k|}{range(k)}$$

The Domain similarity statistic is calculated as

$$D = 1 - d_{AB}$$

The maximum similarity between a grid cell and all points is mapped. In DIVA, this value is then multiplied by 100. A good match is thus a high number (e.g., above 95).

Domain (Mean Distance) - This is the same as above, but the mean distance to all points is used instead of the minimum distance.

Domain True/False - All areas that are within the specified threshold are mapped as "true" (1); all other areas are mapped as "false" (0).

Domain Most Limiting Factor - The variable with the lowest score (largest distance) is mapped. The value in each grid square is a number that corresponds to the variables on the *Predict* tab.

7.2 *External Models*

With this module you can export points from a shapefile to prepare input files for "external" distribution models, i.e. ones not supported directly by DIVA-GIS. In some cases, such as for MAXENT and GARP, this is in fact not very difficult, because they just use a list of species with coordinates. However, for programs that need a site by species matrix (such as GRASP), the preparation of input files is far from trivial. For GRASP, 3 files are made (YYY, XXX, and XXXpred).

Some models will produce grids in a format that DIVA can import via the *Data / Import to Gridfile* menu. GRASP, however, produces LookUp Tables, or LUT files, that can be imported in this module. You must select the LUTs you want, together with the stack that was used to make the input files (or another, but compatible, stack, if you study climate change).

7.3 Evaluation

Using this menu you can evaluate the performance of a niche model. It is a three step process, as described below, a sub-menu for each step.

7.3.1 Prepare points

This tool allows you to split point distribution data into a number of subsets that you can use to test a distribution model. You can specify what percentage of the points you want to use to "train" (i.e., run) the model (as a percentage of the total number of points: "Sample size training data (%)"), and how many separate times you want to do the sampling ("# Repetitions"). You can also sample the background environment to create "pseudo-absence" points ("Include Random Absence Data"). The output is

two shapefiles of points, the name of one ending with "_train" (the model training data, and the other with "_test" (the model testing data). Use the training data to run a model (see previous sections in this chapter), and the test data to evaluate the models (see below), using the "class" (e.g., species) field and the "presence/absence" field.

7.3.2 *Create Evaluation File*

Use a "_test" shapefile from the process described above to create a ROC file. You must provide a shapefile that has a "presence / absence" field. The field name does not matter, but the contents must be words that start with a "p" or with an "a" e.g., "pres" and "abs". A ROC file is a text file with values extracted for each species (grid) for the absence and the presence points. You can calculate kappa and ROC statistics with these (see below).

7.3.1 *Show ROC / Kappa*

You can visualize the results of model testing here. The input file is the ROC file generated above. You can asses the quality of a prediction by inspecting the ROC curve and you can find the optimal threshold for presence/absence by inspecting the kappa/threshold plot. The area under the ROC (Receiver Operating Characteristic) curve is commonly used to evaluate models in species distributional modeling. A ROC curve is created by plotting the true-positive fraction against the false-positive fraction for all test points across all possible probability thresholds (Fielding and Bell 1997). The curve goes from (0,0) to (1,1) and a model that produces a curve with a high true-positive fraction at low values of the false-positive fraction is considered good. This is commonly quantified by calculating the area under the curve (AUC). AUC ranges from 0 to 1 where a value of 0.5 indicates that a model is no better than random and a value of 1 indicates that the model can discriminate perfectly between presence and absence records.

7.4 *EcoCrop*

This module uses FAO's EcoCrop database of the environmental requirements of a long list of plant species, whichcan be used to identify possible crops to grow in a particular environment. In DIVA-GIS, we implement EcoCrop to predict the adaptation of a crop over geographic areas. Currently, only temperature and precipitation data are used to make this prediction.

In EcoCrop, the growing period is defined in days between Gmin and Gmax (start of growth and end of growth, respectively). The length of the growing season is defined as the average of Gmin and Gmax.

In DIVA-GIS, 12 possible growing seasons are considered, starting on the first of each month. To determine the suitability of a growing season for a certain crop, the following temperature parameters are used:

> KTMP: absolute temperature that will kill the plant
> TMIN: minimum average temperature at which the plant will grow
> TOPMN: minimum average temperature at which the plant will grow
> > optimally
> TOPMX: maximum average temperature at which the plant will grow
> > optimally
> TMAX: maximum average temperature at which the plant will cease to
> > grow

The suitability of a location (grid cell) for a crop is evaluated for each of the 12 potential growing seasons. If the average minimum temperature in one of these months is 4 °C or less above KTMP, it is assumed that, on average, KTMP will be reached on one day of the month, and the crop will die. The temperature suitability of that month is thus 0%. If this is not the case, the temperature suitability is evaluated for that month using the other temperature parameters. See the example for maize in Figure 12 for how this is done. The overall temperature suitability of a grid cell for a crop, for any growing season, is the lowest suitability score for any of the consecutive number of months needed to complete the growing season.
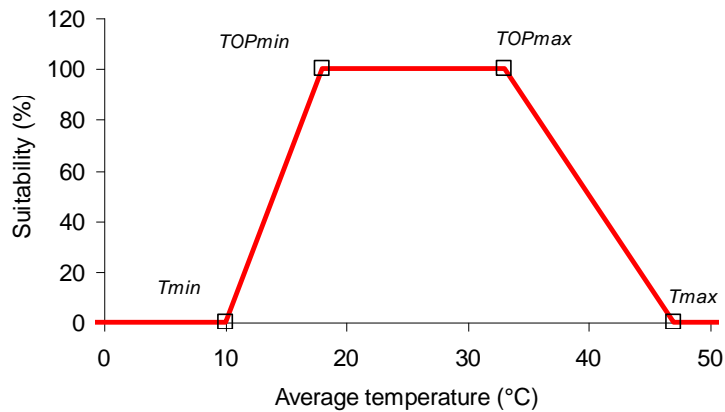
Fig ... ure data and calculated ... are for maize.

To estimate the suitability of the rainfall in an area for a particular crop, the following rainfall parameters are used:

Rmin: minimum rainfall (mm) during the growing season
Ropmin: optimal minimum rainfall (mm) during the growing season
Ropmax: optimal maximum rainfall (mm) during the growing season
Rmax: maximum rainfall (mm) during the growing season

The evaluation for rainfall is similar as for temperature, except that there is no "killing" rainfall and there is one evaluation for the total growing period (the number of months defined by Gmin and Gmax) and not for each month. See Figure 13 for an example. The output is the highest suitability score (percentage) for a growing season starting in any month of the year.

You can run the Ecocrop module for rainfall, or for temperature, or for both (the boxes to tick are in the *Predict* tab). In the last case, the minimum value of the two for each growing season is used to compute the suitability.

To run the module you must first select a species in the *Select* tab. You can search for species using the options in *Filter by*. Then select the crop you want by clicking on it (the selected record is highlighted in blue). On the next tab, *Parameters*, you can inspect the temperature and rainfall parameters from the EcoCrop database, and change them if you want. The changed parameters will be used, but not saved. If you want to make permanent changes you must edit the file faoecocrop.dbf in the <DIVA dir>\bin directory.

You can select the area for which you want output by filling the upper text boxes on the *Predict* tab. You can also use the extent of an existing shapefile or gridfile by making it active in the legend and then pressing the "Read from layer" button. The coordinates need to be adjusted to those of the climate grid cells. You can do this by pressing the "Adjust" button. If you do not do this, the program will do it when you press "Apply". You can run the EcoCrop module for current and for future climate (see
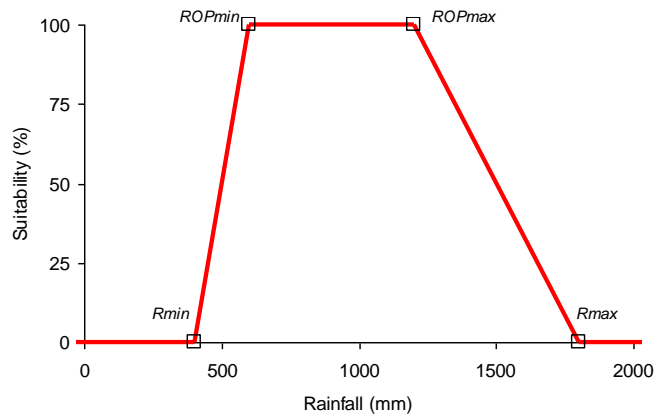
section 8.1). You must name an output gridfile.

Figure 13. Relation between rainfall during the growing season and calculated rainfall suitability values for that month. The parameters are for maize.

8. THE *GRID* MENU

Various functions are available in the *Grid* menu which allow you to create new grids from one or more existing ones. They are listed in Box 10.

Box 9. The *Grid* menu

Icon	Sect.	Name	Short explanation
⊞	8.1	*Describe*	Describes the contents and structure of a grid
	8.2	*Overlay*	Carries out arithmetic on the values of corresponding grid cells in two grids
	8.3	*Scalar*	Carries out arithmetic on a single grid
	8.4	*Reclass*	Classifies the values of a grid in a new way
	8.5	*Neighborhood*	Carries out neighborhood operations such as filter and pattern analysis
	8.6	*Aggregate*	Aggregates (combines) cells of a grid
	8.7	*Disaggregate*	Disaggregates (divides) cells of a grid
	8.8	*Cut*	Cuts a specified section out of a grid
	8.9	*Concatenate*	Combines two grids into one
	8.10	*New*	Makes a new grid with the same parameters as an existing one but different cell values
	8.11	*Transect*	Makes a graph of the values in a given column or row of a grid
	8.12	*Area*	Calculates the area of classes in a grid

8.1 *Describe*

You can use this function to get the following information on the contents and structure of a grid:

Structure

- type of data (integer, real)
- number of rows
- number of columns

Contents (across all cells)

- number of cells with data
- number of grid cells without data
- minimum value
- maximum value
- sum of grid cell values
- mean of grid cell values
- median of grid cell values
- Mode of grid cell values
- standard deviation of grid cell values
- variance of grid cell values

Describe operates on the active layer when the window is open, or else on any layer selected using the "Grid" button. Some of the information (e.g., the number of rows and columns) can also be obtained by double-clicking on a grid layer in the legend (and choosing the "Info" tab).

8.2 *Overlay*

Given two grids that have the same dimensions (number of columns and rows, resolution), and location (min and max X and Y coordinates), *Overlay* allows you to place them on top of each other (as it were), carry out some arithmetic on corresponding grid cells (i.e., cells directly above each other), and place the results in a new, third grid. The arithmetical operations included are addition, subtraction, multiplication, division and calculation of minimum and maximum. In the operation "Cover", the values of the cells in the first grid are used, unless the value of a cell in the first grid is "Nodata", in which case the value of the cell in the second grid is used.

8.3 *Scalar*

Scalar lets you create a new grid by applying an arithmetical operation to the values in each cell of an existing grid. For the following operations you need to also specify a number: add, substract, multiply, divide, raise to a power, and take the the logarithm with a specified base. Other functions include taking the absolute value, rounding, truncating and trigonometric operations.

8.4 *Reclass*

The *Reclass* option re-classifies data in a grid according to class limits which you specify. For example, you may want to give all grid cells with values between 0 and 1000 the value 1, and all cells with the values 1001-2000 the value 2. Class limits can be adjusted manually. Extra classes can be inserted with the button "+", and deleted with the "-" button. DIVA-GIS shows automatically the data type and minimum and maximum values of the input file to facilitate appropriate re-classification. You can save the new classification limits (using the "Save RCL" button) and then use them to apply the same re-classification to another grid ("Read RCL").

8.5 *Neighbourhood*

This function re-calculates the value of each grid cell according to the values of the cells in a 3x3, 5x5, 7x7 or 9x9 neighbourhood around that cell. Mean, median, mode, minimum, maximum, range, CV or the number of distinct classes (richness) within the neighborhood may be calculated. When the mean, median or mode are calculated, this is often referred to as a "filter", because it smooths the values on a grid. Filtering is sometimes used after disaggregating a grid, or to remove noise in a satellite image (mode filter).

The number of distinct classes is an indication of the pattern of the grid in that area, for example to what extent vegetation is heterogeneous.

Figure 14. How a 3x3 mean filter works: Original cell values (left) and the new

2	1	3
3	5	4
5	6	8

2.8	3.0	3.3
3.7	4.1	4.5
5.8	5.2	5.8

cell values (right).

8.6 *Calculate*

In the *Calculate* window, you can do mathematical calculations with one or more grids. The things you can do with *Calculate* can also be done with combinations of *Scalar* and *Overlay*. The advantage of *Calculate* is that you can do several steps at once. The disadvantage is that the calculations are much slower, so if you have very large grids you're better off using *Scalar* and/or *Overlay*.

Use the "Add Grid" button to choose the files you want to use, insert the operator(s) you require between grid names, and provide a name for the output grid. As with *Overlay*, to be able to use *Calculate* with multiple grids, these must all have the same dimensions (number of columns and rows, and resolution) and location (min and max X and Y coordinates).

8.7 *Aggregate*

The *Aggregate* procedure allows you to lump together or merge cells of a grid and create another grid with a lower resolution (larger grid cell size). You must specify the aggregation factor, which determines how many cells will be merged into one (and thus how big the new cells will be). For example, with factor 2 the new cells will have two times the length and two times the height of the original cell. In other words, four cells will be merged into one. The values in the cells of the aggregated grid will depend on the procedure chosen (mean, minimum, maximum, range, sum, mode, median, standard deviation).
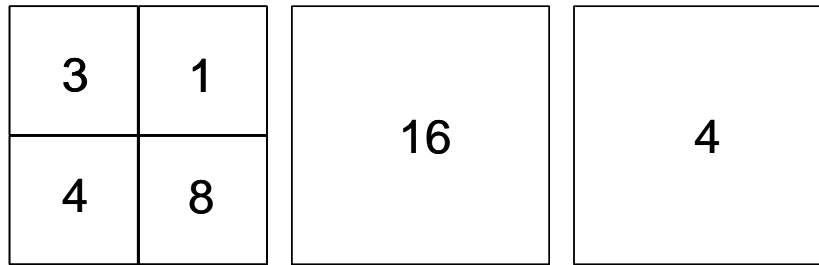
Figure 15. Grid values before (left), and after aggregation with procedure "sum" (middle) and "mean" (right).

If you choose the option "Ignore Nodata", aggregation is also done on groups of cells one or more of which has "Nodata" as value. If you do not choose this option, the value of the aggregated cells will become "Nodata" even if only one of its component cells has that value.

In some cases, the aggregation factor is not a multiple of the number of rows or columns. Say you have a grid of 3 by 3 cells that you want to aggregate with a factor two. In this case, the cells in the last column and row cannot be aggregated properly. DIVA-GIS gives you two options to deal with this: either ignore these cells and truncate the original grid (in the example, the aggregated grid would have one cell), or expand the grid by adding another row and column with Nodata values (in the example, the aggregated grid would have four cells).

8.8 *Disaggregate*

Disaggregate divides the cells of a grid into smaller cells. The value of the original cell is assigned to all of the smaller cells. You need to specify the disaggregation factor, which should be an integer greater than one.

8.9 *Cut*

Cut allows you to make a new grid consisting of a selected part of the area of an existing grid. You can define the area to be cut and placed in the new grid by the column and row numbers, or by coordinates. These parameters can also be selected by drawing a rectangle on the map. They can also be copied from an existing grid.

8.10 *Concatenate*

Concatenate combines the data of two grids with the same resolution and origin, but with different extents. For areas where both grids have values, the value of the first grids is maintained (as with *Cover*).

8.11 *New*

You can generate a new grid with this option, using the parameters from an existing grid. The grid cell values can be a single number, consecutive numbers starting with 0, or random numbers from a uniform or a Normal distribution.

8.12 *Transect*

Choose a row or column in a grid (either by specifying its number or by clicking on the grid on the map), and with *Transect* you get a graph showing on the vertical axis the changes in grid cell values with latitude or longitude. You can choose to fix the vertical axis so that different transects will be comparable. A moving average will be displayed if you tick the appropriate box.

The results may be saved in a number of ways. You can save both the graph and the data on which it is based in the clipboard. You can also save the data to a file by going to the *Data* tab and specifying an output file.

8.13 *Area*

Area lets you create a new grid in which the value in each cell represents the total area of the class in that cell in an existing grid. For example, you have a grid with values 1 to 10 (the classes). A cell with input value 1 will have as output value the combined area of all cells with that value.

9. THE *STACK* MENU

A **stack** is a set of grids with the same dimensions (number of columns and rows, and resolution) and location (minimum and maximum X and Y coordinates) which are handled together as a group. Grouping grids in a stack makes it easier to process many files in the same way, and allows a number of additional analytical procedures. You can carry out the following grid operations (i.e. from the *Grid* menu, see chapter 8) on stacks as well as on individual grids: *Scalar*, *Reclass*, *Neighbourhood*, *Aggregate*, *Disaggregate*, *Cut*, *Distance* and *Area*.

A stack is stored in a file with the extension GRS. A GRS file is nothing but a list of the grids that the stack contains. Hence, rather like a project or DIV file (remember them?), it does not contain any of the actual data in the grids. Rather, it points to the existing files with the data. Therefore, if you delete, rename, or move one its constituent grids to a different directory, the stack will become invalid.

Box 10. The *Stack* menu

Sect.	Name	Short explanation
9.1	Make Stack	Defines a stack by listing the grids it contains
9.2	Plot	Makes a histogram of cell values
9.3	Calculate	Carries out arithmetic on grids in a stack
9.4	Regression	Calculates regression on a cell by cell basis
9.5	Cluster	Classifies grid cells
9.6	Export to Text File	Places all grid cells values in a single text file

9.1 *Make Stack*

You make a stack by adding grids to a list and then naming the output GRS file. You can remove grids from the list individually or all at once if you make a mistake or change your mind. DIVA tells you about the dimensions and location of each grid you add to the stack. These must be identical for all grids in the stack. If you forget which grids are in a stack, or wonder whether it is still valid, or why it isn't, use the *Check Stack* tab to obtain a list of the grids included in a specified stack. You can edit stack files with text editors such as Notepad.

9.2 *Plot*

With this, you can make a histogram of the values of corresponding cells in all grids in a stack. Click on the map to choose which grid cell location to compare in this way.

9.3 *Calculate*

Using *Calculate*, you can produce a single grid from the multiple grids in a specified stack, the value in each cell of the output grid being the sum, mean, minimum, maximum, harmonic mean or sum of the reciprocals of the values in the grids in the stack.

If you select the "Calculate Area" option, a text file is produced with the area of each grid (expressed as number of cells) with values above a certain specified threshold value. If you select a grid to serve as a "mask", only the cells for which the mask has data (i.e., not "Nodata") will be considered.

9.4 *Regression*

You can use this function to perform regression on a cell basis, across a stack of n grids. That is, for each cell position, linear regression is done using 1…n as X values and the values in the corresponding grids as Y values. Make sure that the grids are in the right order (e.g., the grids could be the result of running models with increasing values of a parameter?). You can open the stack file with a text editor and check.

9.5 *Cluster*

You may want to classify grid cells based on the values they have in different grids, for example if the different grids represent the presence and absence of different species. You can do this in the *Cluster* menu in three steps:

1. In the *Distance Matrix* tab, calculate a matrix of distances (file extension DMT) between grid cells by specifying a distance measure.

2. In the *Cluster* tab, use the DMT file produced above to calculate a cluster file (file extension DTR) using either the weighted or unweighted UPGMA algorithm.

3. In the *Draw* tab, use the DTR file you just produced to draw a dendrogram of the distance relationships among grid cells, decide at what level of distance to prune the dendrogram, and finally produce a new grid in which the values of the grid cells will correspond to which cluster in the dendrogram the cells belongs to.

9.6 *Export to Textfile*

With the option *Export to Textfile,* all the gridfiles in the stack can be exported together to a single TXT file. Such a file can be used to make comparisons on a cell-by-cell basis, e.g. in a spreadsheet program. As an option, any row that has a null value for one or more of the grids may be omitted in the output file.

10. THE *TOOLS* MENU

The *Tools* menu (Box 11) has some additional miscellaneous functions.

Box 11. The *Tools* menu

Icon	Sect.	Name	Short explanation
	10.1	Projection	Projects a shapefile from geographic coordinates to a projection of choice
	10.2	Graticule	Adds a latitude/longitude graticule
	10.3	Shift Shape	Shifts the position of a shapefile
	10.4	Georeference image	Provides tools to help you georeference an image
	10.5	Geo-Calculator	− Transforms geographic coordinates in degrees/minutes/seconds to decimal degrees − Calculates the distance between two points
		Toolbars	Shows or hide toolbars
	10.6	Options	Sets general options (file locations, etc.)

10.1 *Projection*

With the *Projection* option you can transforming the data in a shapefile from geographic coordinates(i.e. lat/long) into any of a number of projections and back. There is no single best map projection. Some projections can be used for a map of the whole world; other projections are appropriate for small areas only. One of the most important characteristics of a map projection is whether it is "equal area" (a grid cell anywhere on the map will always cover the same amount of area) or "conformal" (the shapes of the geographic features are as they are seen on a globe). No map projection is both conformal and equal-area, and some are neither. For our kind of work, it is usually best to use equal area projections, particularly when large areas are being analyzed (see Table 8). This will assure that the grid cells are all of the same size, and therefore comparable to each other.

Projections are defined using the PROJ.4 notation. And example for longitude/latitude is "+proj=longlat +datum=WGS84 +ellps=WGS84 +towgs84=0,0,0".

Projection doesn't work on gridfiles yet. If you want to project a gridfile, you have to convert it to a shapefile first, using *Data/Export Gridfile*.

Table 6. Horizontal resolution and area of a 10-minute grid cell by latitude. The vertical resolution is 18.5 km at all latitudes.

Latitude [°]	Horizontal resolution [km]	Area per grid cell [km^2]
0	18.5	343
15	17.9	332
30	16.0	297
45	13.1	240
60	9.3	172
75	4.8	89
90	0	0

10.2 *Graticule*

With this function, you can make a shapefile of lines with fixed intervals (e.g. 10 degrees) in North/South and East/West directions. Adding such a graticule to a map makes it easier for the reader to determine where an area is. See Figure 4 for an example of a map with a graticule.

10.3 *Shift Shape*

Shift Shape helps you to move a shapefile in a vertical and/or horizontal direction to cope with displacements of the data. You can transform map units (for example from feet to meters) by multiplying.

10.4 *Georeference Image*

With *Georeference Image,* you can create a "world file" for an image. A word file is an attribute file associated with corresponding image file which allows the image to be imported into a GIS and located on the surface of the earth. TIF, mrSID and JPG are the supported image formats . TIF and SID are commonly used to distribute satellite images. The extension of the world file depends on the extension of the image. It is SDW for a SID file and TFW for a TIF file. A world file is a text file with 6 lines, with a number on each line denoting:

1. size of pixel in X direction
2. rotation term for row

3. rotation term for column
4. size of pixel in Y direction
5. X coordinate of centre of upper left pixel in map units
6. Y coordinate of centre of upper left pixel in map units

For example:

28.50
0
0
-28.50
174505.50
2769288.00

Some SID images do not need a world file. In many cases, these images already have world files (e.g. the SID images that can be downloaded from https://zulu.ssc.nasa.gov/mrsid/). If your image does not have a world file, you will need to create one, and you can use the *Georeference Image* tool for that. You will need to know how many rows and columns your image has. You can find this out with most graphics programs. You then fill this in, as well as the estimated coordinates of the extremes of the image. By overlaying a shapefile of the same area, you can see if the coordinates are correct. You will probably need to change them little by little a number of times before the image is correctly located.

10.5 *Geo-calculator*

With the *Geo-calculator* you can convert coordinates in the sexagesimal (degrees, minutes, seconds) system to a decimal system. This can be done for one location or for a file with many locations (use the *Batch* option). You can also calculate the distance between two points.

10.6 *Options*

The O*ptions* window lets you define a number of default settings, including the font used and how map features are selected.

On the second tab you can specify where DIVA should look for climate (CLM) files, and you can inspect and edit the metadata associated with these files.

REFERENCES

Bonham-Carter, G., 1994. Geographic information systems for geoscientists. Modelling with GIS. Computer Methods in the Geosciences 13. Pergamon/Elsevier, London, U.K.

Burnham, K.P., and Overton, W.S. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika 65, 623-633.

Busby, J.R. (1991). BIOCLIM - a bioclimatic analysis and prediction system. Pp. 64-68 in Margules, C.R. and Austin, M.P. (eds) Nature Conservation: Cost Effective Biological Surveys and data Analysis. Melbourne: CSIRO

Carpenter, G., Gillison, A.N. and Winter, J. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation 2: 667-680.

Chao, A., 1984. Non-parametric estimation of the number of classes in a population. Scandinavian Journal of Statistics 11, 265-270.

Chao, A., 1987. Estimating the population size for capture-recapture data with unequal catchability. Biometrics 43, 783-791.

Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society (Series B) 345:101-118.

Cressie, N. A. C., 1991. Statistics for spatial data. John Wiley & Sons, New York.

Fielding, A.H. and J.F. Bell, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24: 38-49.

Greene, S.L. and T. Hart, 1996. Plant Genetic Resource Collections: an Opportunity for the Evolution of Global Data Sets. Third International Conference/Workshop on Integrating GIS and Environmental Modeling Santa Fe, New Mexico, USA. (CD-ROM and web publication; http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/greene_stephanie/sgreene.html).

Heltshe, J. and N.E. Forrester, 1983 . Estimating species richness using the jackknife procedure. Biometrics 39, 1-11.

Hijmans, R.J., M. Schreuder, J. De la Cruz, and L. Guarino. 1999. Using GIS to check coordinates of genebank accessions. Genetic Resources and Crop Evolution 46: 291-296.

Hurlbert, S.H., 1978. The non-concept of species diversity: a critique and alternative parameters. Ecology 52: 577-586.

Intergovernmental Panel on Climate Change Data Distribution Center (1999). Providing climate change and related scenarios for impact assessment. CD-ROM. Version 1.0. Climate Research Unit, University of East Anglia, Norwich, UK.

Magurran, A.E. 1988. Ecological diversity and its measurement. Princeton University Press.

Nabhan, G.P. 1990. Wild *Phaseolus* ecogeography in the Sierra Madre Occidental, Mexico: aeorographic techniques for targeting and conserving species diversity. Systematic and ecogeographic studies on crop genepools 5. International Board for Plant Genetic Resources, Rome, Italy.

Nelson, A., G. LeClerc and M. Grum, 1997. The development of an integrated Tcl/Tk and C interface to determine, visualise and interogate infraspecifc bio-diversity. Internal CIAT document. GIS Laboratory, CIAT, Cali, Colombia.

Nix, H. A. 1986. A biogeographic analysis of Australian elapid snakes. Pp. 415 in R. Longmore (ed.). Atlas of Elapid Snakes of Australia. Australian Flora and Fauna Series No. 7. Australian Government Publishing Service, Canberra;

Raaijmakers, J. G. W. 1987. Statistical analysis of the Michaelis-Menten equation. Biometrics 43:793-803.

Rebelo, A.G. 1994. Iterative selection procedures: centres of endemism and optimal placement of reserves. Strelitzia 1: 231-257

Rebelo, A.G. and W.R. Sigfried. 1992. Where should nature reserves be located in the Cape Floristic Region, South Africa? Models for the spatial configuration of a reserve network aimed at maximizing the protection of diversity. Conservation Biology 6:243-252.

Sawada, M., 1999. Rookcase: an Excel 97/2000 Visual Basic (VB) add-in for exploring global and local spatial autocorrelation. Bulletin of the Ecological Society of America, October 1999: 231-243.

Sanders, H.L., 1968. Marine benthic diversity: a comparative study. American Naturalist 102:243-282.

Smith, E.P. and G. van Belle, 1984. Nonparametric estimation of species richness. Biometrics 40, 119-129.

Wieczorek, J., Q. Guo, and R.J. Hijmans, 2004. The point-radius method for georeferencing point localities and calculating associated uncertainty. International Journal of Geographic Information Science 18: 745-767.