



Lake Tanganyika Project: Zambia National Site Characterisation and Catchment Management Design Workshop

GIS training module – Lesson 5

Thomas Gumbricht, ICRAF



GLOBAL ENVIRONMENT FACILITY
INVESTING IN OUR PLANET

Lake Tanganyika Regional Integrated Management Project

A United Nations Development Programme (UNDP) / Global Environmental Facility (GEF) project covering the Lake Tanganyika riparian countries, Burundi, the Democratic Republic of Congo, Tanzania and Zambia.

Training module created by Thomas Gumbricht, ICRAF

Last updated: October, 2010

LESSON 5 – VEGETATION AND TIME SERIES DATA

Lesson overview

In this lesson you will learn about satellite derived vegetation data, and how to use it for assessing land degradation; how to visually explore changes in both space and time using GIS; how to analyze time series data using DIVA-GIS; and more advanced methods for grid calculations. The lesson also gives an introduction to different data formats for grid data.

Vegetation changes and land degradation

The lack of adequate ground data has led to most studies utilizing the growing archives of satellite data for mapping vegetation changes and land degradation. In this lesson, time series data of vegetation for the period 1981 to 2009 is calculated from 8 km resolution Normalized Difference Vegetation Index (NDVI) data obtained from the Advanced Very High Resolution Radiometer (AVHRR) sensor flown on a series of satellites operated by the National Oceanic Atmospheric Administration (NOAA) of the United States of America (USA). From 1981 until today AVHRR instruments have been carried on board, NOAA-7, NOAA-9, NOAA-11, NOAA-14, NOAA-16, and NOAA-17. The AVHRR sensor records electromagnetic radiation in five bands or channels: visible, near infrared, middle infrared and two thermal bands. The resolution at nadir (straight down) is 1.1 km. The spectral bands used for monitoring vegetation are the visible (580-680 nm) and the near infrared (725-1100 nm). Reflectances in these two bands are combined to calculate the Normalized Difference Vegetation Index (NDVI):

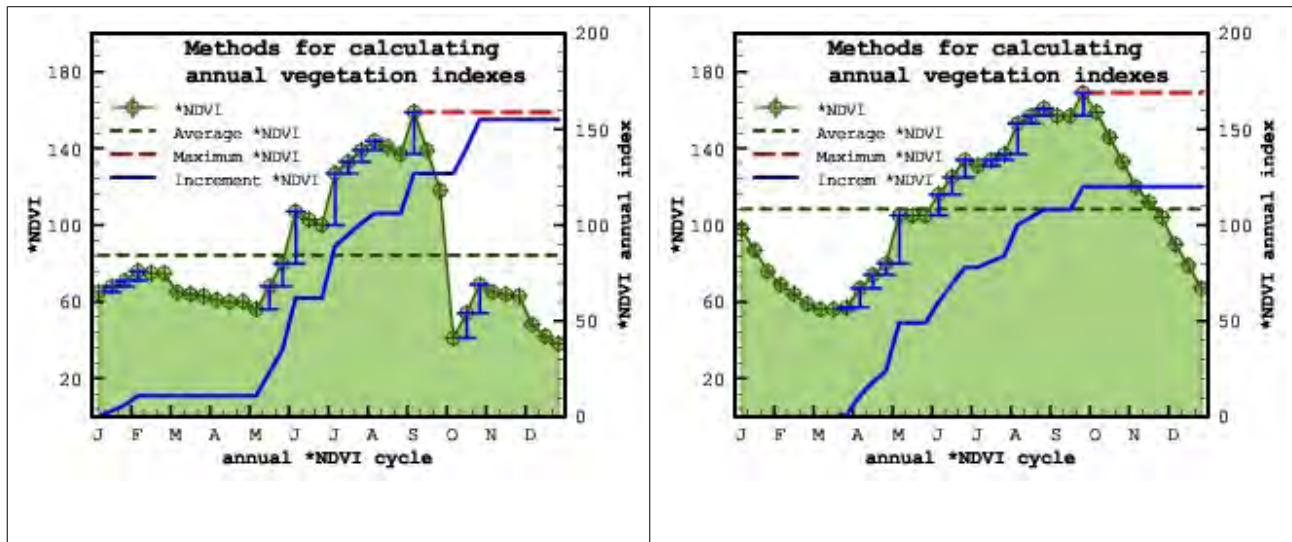
$$\text{NDVI} = (\text{near infrared} - \text{visible}) / (\text{near infrared} + \text{visible}) \quad (\text{Eq.1})$$

In this project we have used the AVHRR NDVI processed by the Global Inventory Monitoring and Modeling System (GIMMS) group at NASA Goddard Space Flight Center (GSFC). The GIMMS-NDVI is corrected for the different generation of sensors, for sensor degradation and for atmospheric disturbances. The global GIMMS dataset used here has a resolution of 8 * 8 km.

The AVHRR sensor was not designed for vegetation mapping, and NDVI is hampered by several shortcomings. To overcome the influence of soil signals at low vegetation densities affecting NDVI,

a previously published soil adjustment factor was applied. All the NDVI data used in this lesson, and all NDVI data available in the dataset are hence soil adjusted, sometimes denoted *NDVI.

The *NDVI data was recalculated to annual indexes of vegetation growth using three approaches: 1) annual average vegetation, 2) annual maximum vegetation and, 3) annual increments in vegetation. The annual indexes were calculated using the period January to December. The figure below illustrates the difference in the indexes (see section 5.7 in the main report).



Earlier studies have used either an annual (or growing season) NDVI average (or integral) for mapping vegetation growth. Studies restricted to cropland have tended to use the annual maximum NDVI, hypothesized to represent the standing crop before harvest. For this project we adopted a theoretically neutral annual vegetation index, defined as the summed annual increment in NDVI between each of the individual observations (i.e. difference in vegetation between current and previous observation if positive) over an annual NDVI cycle. The starting of each annual vegetation cycle was set to December the year before. By summing the increments we hypothesize that our index better captures the productivity of rangelands, and neutralizes the differences between rangelands, croplands and woodlands. It further has the advantage of neutralizing initial background effects (e.g. soil moisture conditions, influence of woody biomass – see the illustration above) at the start of an annual growing cycle. The developed index should hence be more suitable when comparing vegetation in landscapes with transient temporal and spatial changes, compared to either annual average or maximum vegetation used in most other studies. Each of the three indexes has its advantage and drawback: annual average vegetation can be misleading as it is sensitive to cloud contamination, and it overestimates vegetation growth in evergreen forests and shrub lands; maximum vegetation is not suitable for pastoral landscapes with continuous or intermittent grazing, where standing crop is low but growth can still be higher compared to fallow or croplands; the annual vegetation increment is sensitive to cloud contamination.

About grid data and data formats

In DIVA-GIS a grid layer is built of three separate files; a data file (with the extension *.gri), a header file that contains information on rows, columns, resolution, projection, data format and legend (with the extension *.grd) and an image file that is displayed in the Data View (usually a bitmap - *.bmp). Most GIS software only use two files for grids, and do not have a separate image file – the GIS software generate the display image when a grid is added. Here is a grid header file from DIVA-GIS, with some explanations:

[General]	# General information
Creator=Thomas Gumbricht	#
Created=Sun Oct 28 17:55:33 2007	#

Title=Afpopd90	#
[GeoReference]	# The Coordinate system of the grid
Projection=GEOGRAPHIC	# Either Geographic or Other (i.e. projected)
Datum=WGS84	# How the Earth roundness is set
Mapunits=people/km2	# The units in Other projections (see below)
Columns=216	# Nr of columns in grid
Rows=240	# Nr of rows in grid
MinX=33.000000	# Coordinate of left side
MaxX=42.000000	# Coordinate of right side
MinY=-5.000000	# Coordinate of bottom
MaxY=5.000000	# Coordinate of top
ResolutionX=0.041667	# Resolution in Grid X
ResolutionY=0.041667	# Resolution in Grid Y
[Data]	#
DataType=INT2BYTES	# Datatype can be byte, integer or real
MinValue=0	# Minimum data value in grid
MaxValue=10410	# Maximum data value in grid
NoDataValue=-9999	# Value denoting no data
Transparent=1	# if no data is transparent or not in the view
Units= people/km2	# Units in map (here = people/km2)
[Application]	#
Opt0= Imported from	# DIVA-GIS operations (entered by DIVA)
Hist0=Population data made	# Metadata put by the producer/user
Hist1=Produced by CIESIN....	# Any number of Hist entries allowed
[ContLegend]	# Default continuous legend (if no [Legend])
Count=2	# Nr of entries in default legend
Color1=16711680	# Color of entry 1 in default legend
Value1=0	# Value of entry in default legend (= min)
Label1=	# Text for entry 1 in default legend
Color2=255	# Color for label 2 in default legend
Value2=10410	# Value for label 2 in default legend (= max)
Label2=	# Color for label 2 in default legend
[Legend]	# If given overrides the [ContLegend]
Count=9	# Nr of entries in the Legend
Color1=13816530	# Color of entry 1 in Legend
Min1=0	# Min value for entry 1 in Legend
Max1=0	# Max value for entry 1 in Legend
Label1=0	# Label for entry 1 in Legend
.	.
.	.
.	.
Color9=65634	# Color of entry 9 in Legend
Min9=5001	# Min value for entry 9 in Legend
Max9=1000000	# Max value for entry 9 in Legend
Label9=>5000	# Label for entry 9 in Legend
Transparent=1	# If nodata is transparent (1) or not (0)
NoDataColor=16777210	# No data color (shown if Transparent = 0)
NoDataLabel=No Data	# Label for no data
isContinuous=0	# If legend is continuous or has gaps
SpacedByColor=0	# If Legend should be spaced by color

Most of the information given above can be retrieved from the Properties window of grid layers.

The georeference of a GIS data layer can either be Geographic (unprojected) or Projected (called Other in DIVA-GIS). A Geographic georeference is given in Longitude for X and Latitude for Y. DIVA-GIS can reproject layers from Geographic to many different Projections, but only version 7.2

can reproject an already projected layer. A Geographic georeference is defaulted to have Mapunits in decimal degrees (e.g. 33.500000 and not in degrees-minutes-seconds 33 30 0). For Projected layers Mapunits can be kilometers, meters, millimeters, miles, feet etc. The Datum describes the spherical shape of the Earth, as the Earth is not completely round but is flattened at the poles and thicker around the equator. WGS84 (World Geodetic System 1984) is the default for Geographic georeference. If you reproject data to a Projection, another Datum can be set (see [lesson 10](#)).

The DataType in GIS grid layers can have many formats, but is normally restricted to three formats:


BYTE (INT1BYTE) 0 .. 255
 INTEGER (INT2BYTES) -32768 ..32767
 REAL (FLT4BYTES) 1.5 x 10⁻⁴⁵ .. 3.4 x 10³⁸

In a Byte image each cell uses one byte (8 bits) for storing each cell value, and the data range can be a whole number between 0 and 255 (called unsigned byte by computer specialists as only positive numbers can be stored). In an Integer image each cell uses 2 bytes (16 bits) for storing each cell value, and the data range can be a whole number between -32768 and 32767 (signed small integer for computer specialists, as the data can have both positive and negative values). An image with Real values uses 4 bytes (32 bits) for storing each cell value, and the data range can be a ratio number (32 bit real is called Single by computer specialists). The trend images that you generated in lesson 4 contain Real data (FLT4BYTES in the DIVA-GIS header).

The image files for vegetation and rainfall that you will encounter in this lesson, where generated outside DIVA-GIS to make a more full use of color ramping. If you make changes to the legends in these files, DIVA-GIS will regenerate the image file and the pre-made color ramp in the images will disappear, You will not be able to recreate the symbolization using DIVA, so if you change the symbolization, then you can make use of the image files to get the original symbolization back by using an image program (photoshop, gimp etc) to convert the jpg image with the same name to a bitmap (bmp).

Framework data sets

Start a new project

Start DIVA-GIS and open a new project by clicking the New Project button, , or via the menu: Project - New. Save the new project, e.g. calling it lesson5.

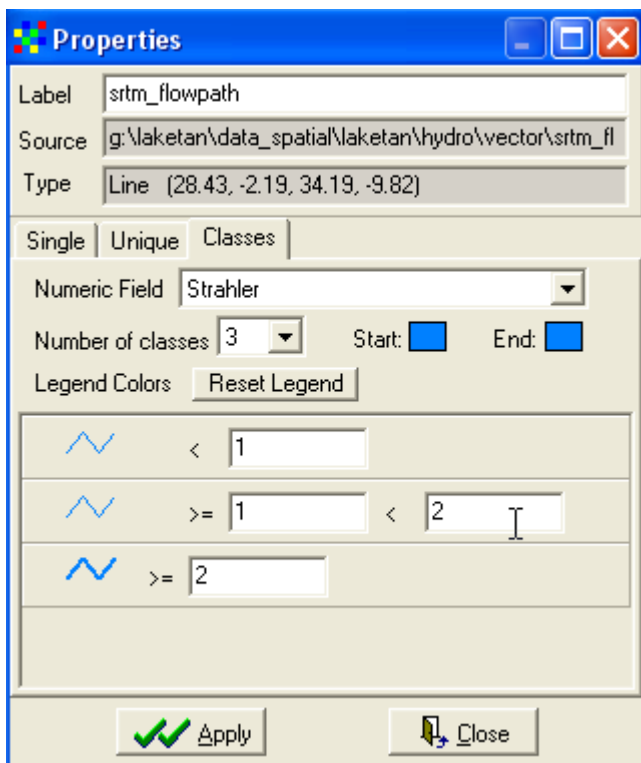
Add framework layers

In order to get some reference and background data to use for interpreting vegetation trends, start by adding the following framework (basic) data layers.

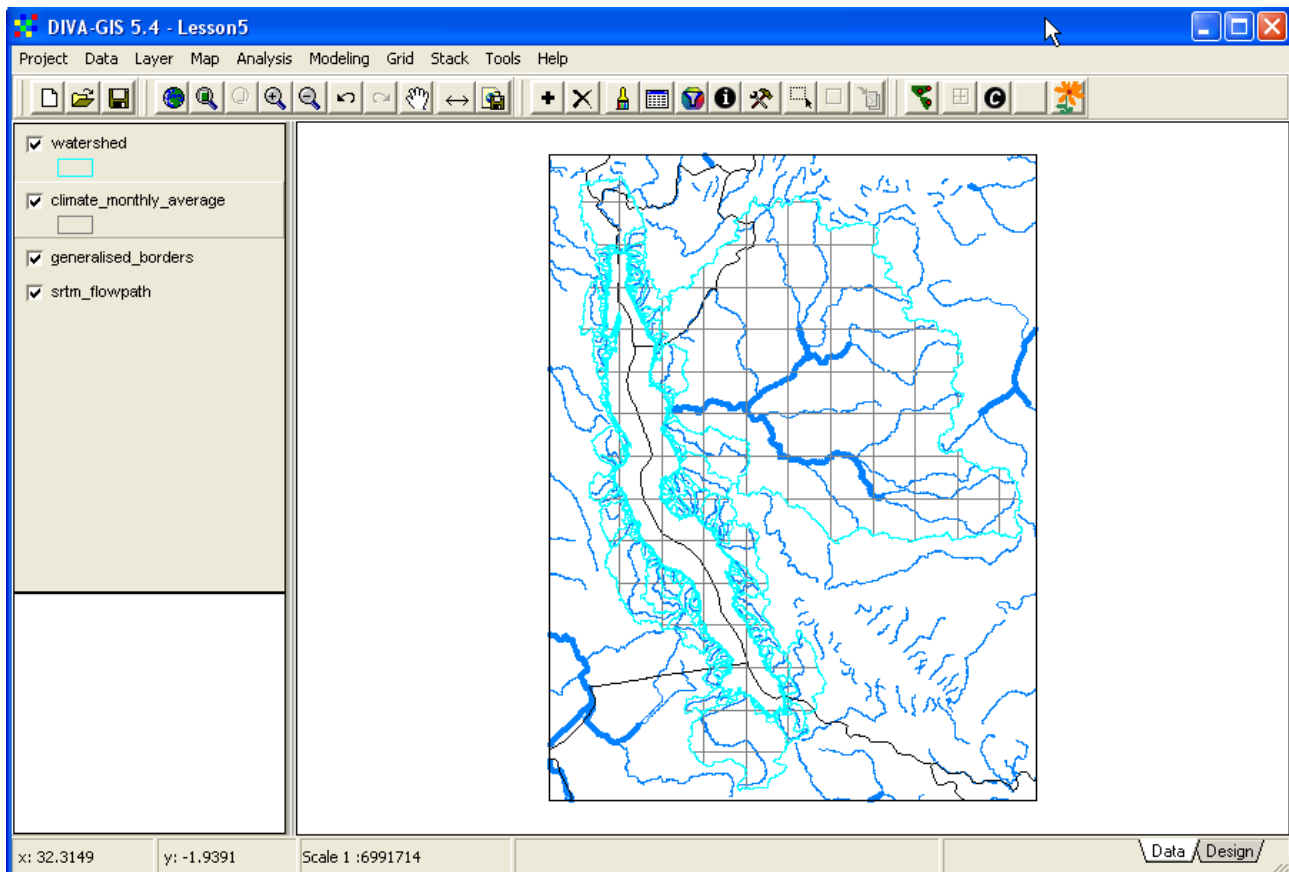
Data layer	Folder	File
Countries	\data_spatial\laketan\political\vector	generalised_borders
Rivers	\data_spatial\laketan\hydro\vector	srtm_flowpath
Basins	\data_spatial\laketan\hydro\vector	watershed
Climate	\data_spatial\laketan\climate	climate_monthly_average

Try to symbolize the layers. The country layer you can keep as it is (transparent fill with black boundaries). The river should be symbolized using the STRAHLER attribute that gives information on stream order (number of upstream divisions), for example as shown below. To generate stream order, the river data must have a correct topology (see [lesson 2](#)), that is to say that the streams must

be connected to each other with no gaps, and a new stream segment at each flow join or bifurcation (splitting of one stream to two). Stream order can be seen as ordinal data (see [lesson 2](#) for an explaining of ordinal data), and as this ordinal data is numerical you can actually use the classification option to symbolize the streams into small and large (as shown below).



The watershed layer (watershed) you can put any color to the border (magenta in the example below), but keep the transparent fill. The climate data you will not need until later in the lesson, and you can leave it for now. The illustration below shows the DIVA-GIS map view if you followed the symbolization suggested above.




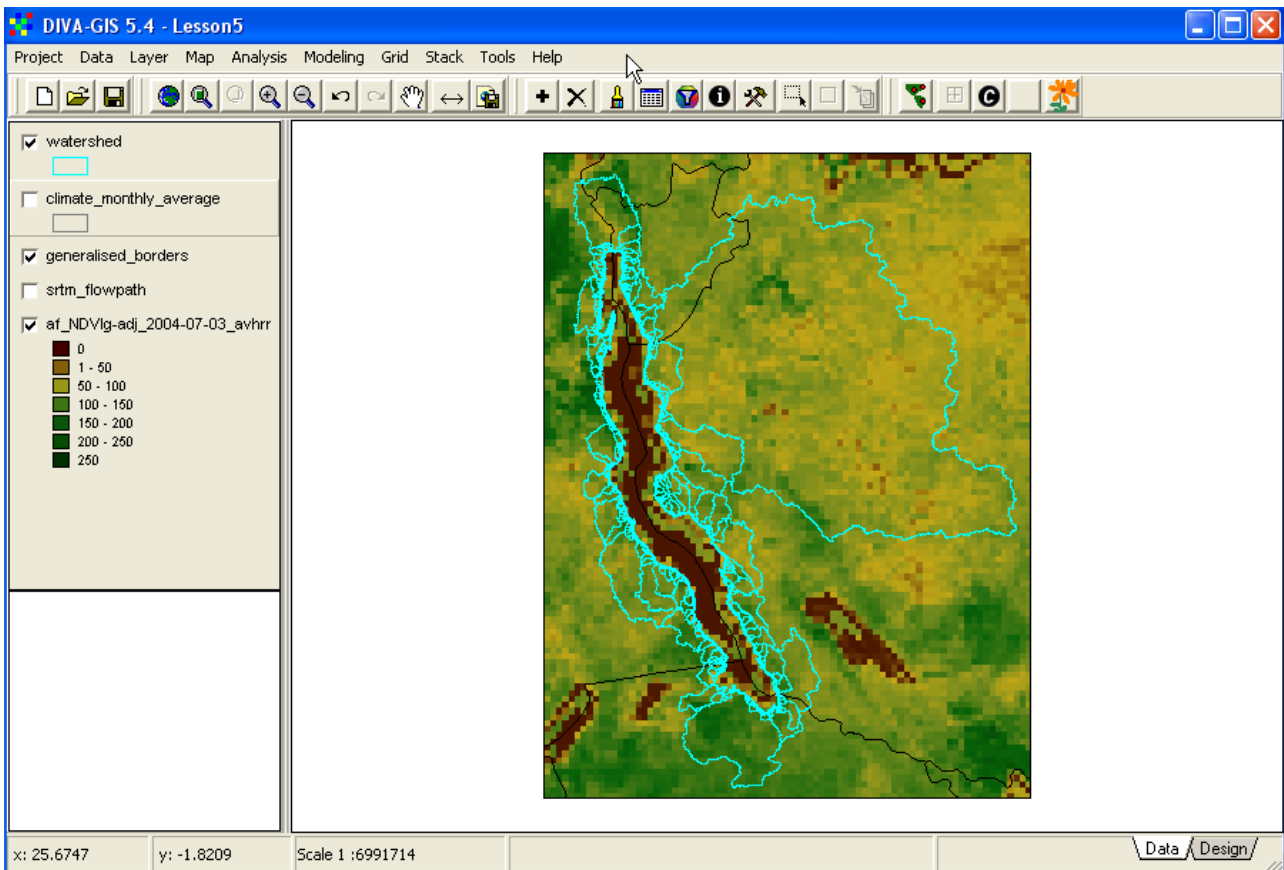
Vegetation grid data

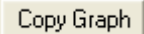
For the analysis of vegetation covering the whole of the Lake Tanganyika Basin, vegetation data from the AVHRR (Advanced Very High Resolution Radiometer) was used (see above). The AVHRR data is recorded daily, and then composed into 10-day values for vegetation using the maximum recorded values. The maximum value is preferred in many studies as clouds and atmospheric pollution tend to decrease the value of NDVI. All the NDVI vegetation data for Lake Tanganyika basin as a whole is under \data_spatial\laketan\rs\ndvi.

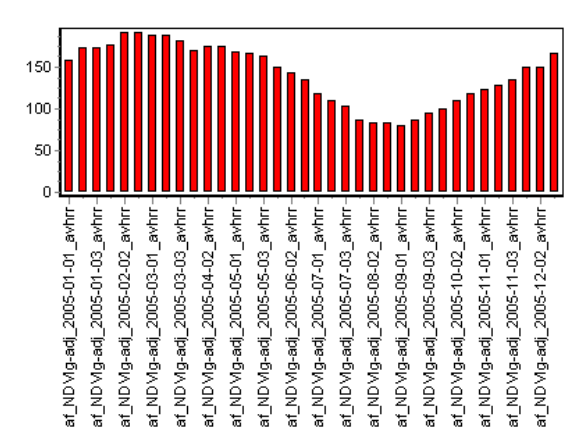
Annual vegetation cycles

The dataset includes 10day (sometimes called dekadal) Maximum Value Composites (MVC) of vegetation growth derived by finding the maximum NDVI value for 10 day cycles of satellite images. In this project all the original data was soil adjusted (see above) and the adjusted data for each year 1981 to 2009 is in separate folders under \data_spatial\laketan\rs\ndvig\GIMMS_NDVIg-10day_YYYY, where YYYY should be exchanged for an actual year (e.g. 2004). Use the Add

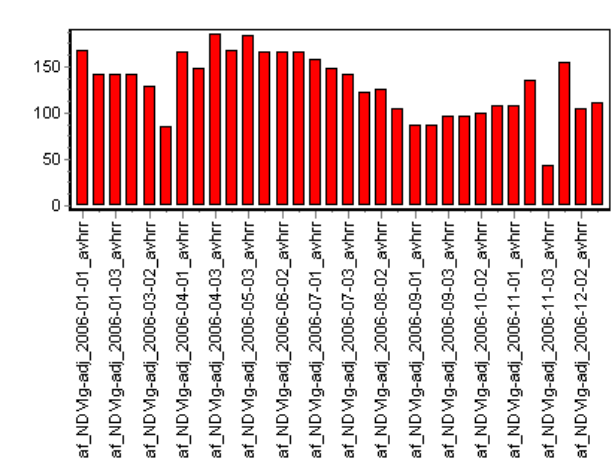
Layer tool, , to add any single vegetation image representing any year (3rd dekad of July 2004 in the example below). In the example below the flowpath layer has been turned off to allow the vegetation to be seen in more detail.



For each annual cycle of vegetation data, there is a pre-made stack (see lesson 4 to repeat stacks if necessary) available in the dataset. You can now use the stack plot function (available from the menu: Stack - Plot Stack) to view the annual vegetation cycle for any point in L.Tanganyika for any year 1981 to 2009. In the example below the annual vegetation cycles (*NDVI) for dekad data are shown for 2005 (dry) and 2006(wet) (Stacks are for Row 53, and Column 40). These plots were captured from DIVA-GIS by clicking the Copy Graph button, , in the Stack Plot window and then pasted and cut into the document (see lesson 4).



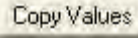
2005 (dry)

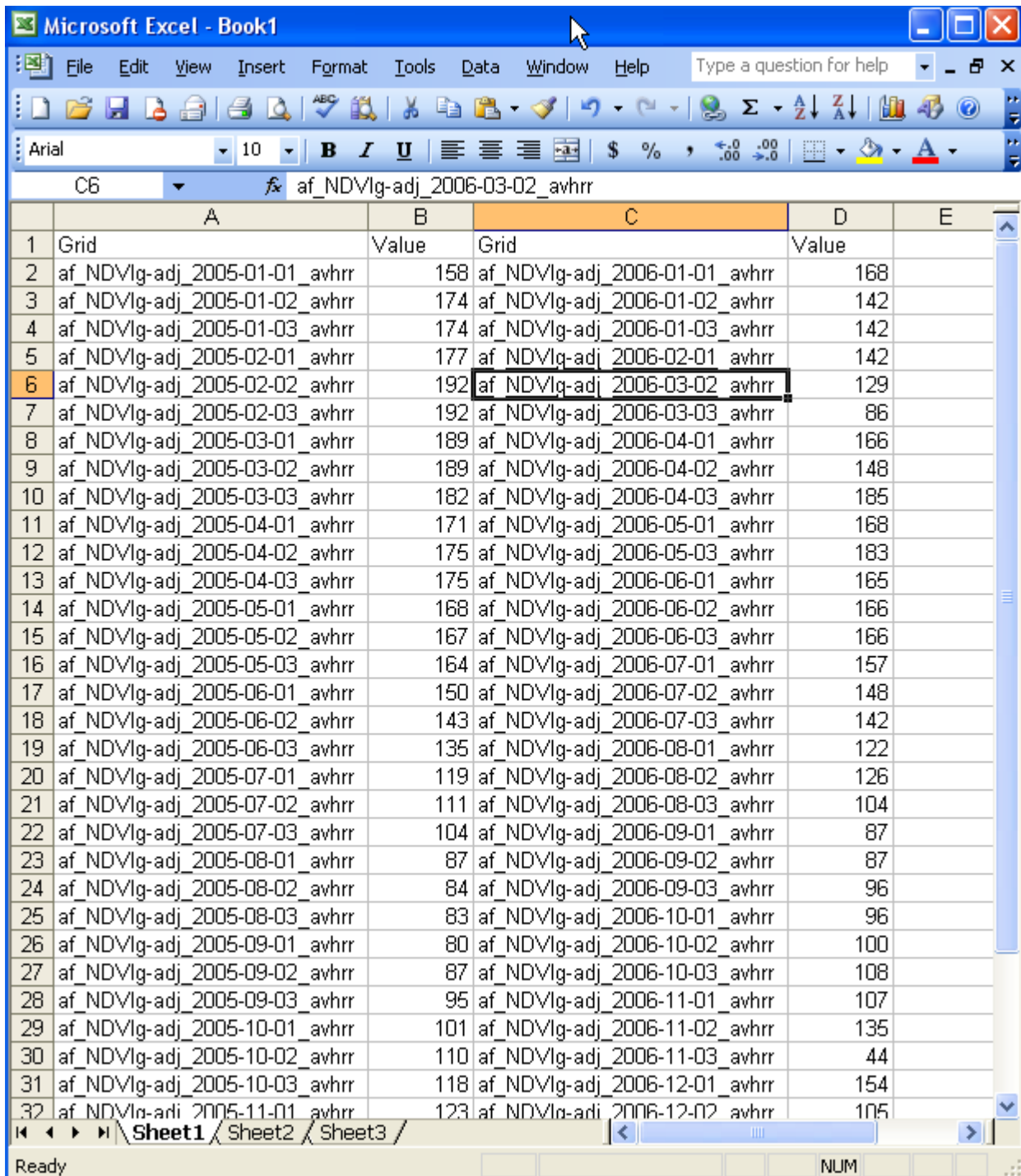


2004 (wet)


The plot images above clearly illustrate the variation in annual growing cycles under different rainfall conditions in the clicked position in the Lake Tanganyika Basin.

Using excel with DIVA-GIS

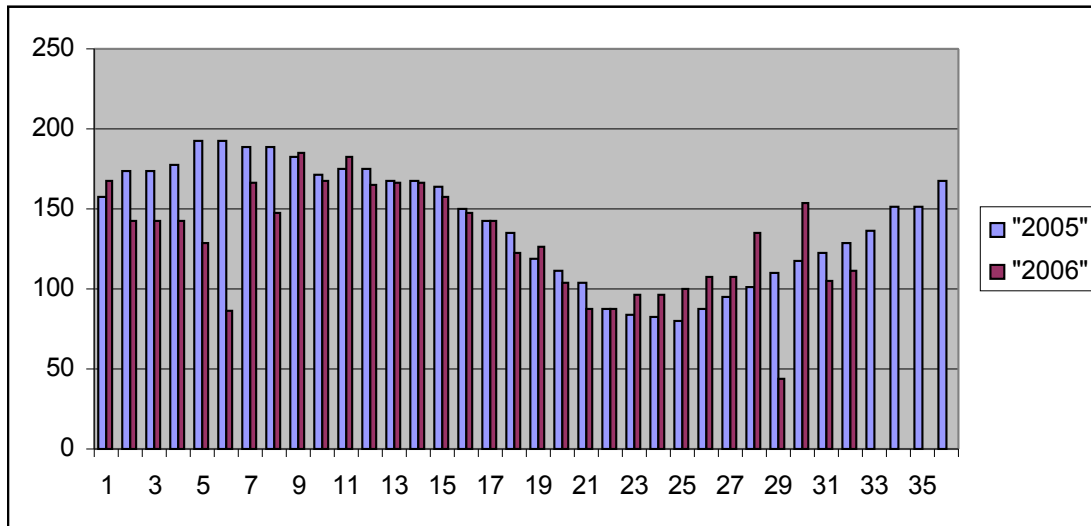
When you have made a stack plot you can use the Copy Values button,  to transfer the data of the plot to excel. After clicking the Copy Values button in the Stack Plot window of DIVA, change application to excel and position the cursor in the cell you want to paste the data to, then click ctrl-v in excel (or go via the menu in excel: Edit – Paste). In the example below the plot representing 2005 is pasted in cell A and the plot representing 2006 in the cell C.



	A	B	C	D	E
1	Grid	Value	Grid	Value	
2	af_NDVlg-adj_2005-01-01_avhrr	158	af_NDVlg-adj_2006-01-01_avhrr	168	
3	af_NDVlg-adj_2005-01-02_avhrr	174	af_NDVlg-adj_2006-01-02_avhrr	142	
4	af_NDVlg-adj_2005-01-03_avhrr	174	af_NDVlg-adj_2006-01-03_avhrr	142	
5	af_NDVlg-adj_2005-02-01_avhrr	177	af_NDVlg-adj_2006-02-01_avhrr	142	
6	af_NDVlg-adj_2005-02-02_avhrr	192	af_NDVlg-adj_2006-03-02_avhrr	129	
7	af_NDVlg-adj_2005-02-03_avhrr	192	af_NDVlg-adj_2006-03-03_avhrr	86	
8	af_NDVlg-adj_2005-03-01_avhrr	189	af_NDVlg-adj_2006-04-01_avhrr	166	
9	af_NDVlg-adj_2005-03-02_avhrr	189	af_NDVlg-adj_2006-04-02_avhrr	148	
10	af_NDVlg-adj_2005-03-03_avhrr	182	af_NDVlg-adj_2006-04-03_avhrr	185	
11	af_NDVlg-adj_2005-04-01_avhrr	171	af_NDVlg-adj_2006-05-01_avhrr	168	
12	af_NDVlg-adj_2005-04-02_avhrr	175	af_NDVlg-adj_2006-05-03_avhrr	183	
13	af_NDVlg-adj_2005-04-03_avhrr	175	af_NDVlg-adj_2006-06-01_avhrr	165	
14	af_NDVlg-adj_2005-05-01_avhrr	168	af_NDVlg-adj_2006-06-02_avhrr	166	
15	af_NDVlg-adj_2005-05-02_avhrr	167	af_NDVlg-adj_2006-06-03_avhrr	166	
16	af_NDVlg-adj_2005-05-03_avhrr	164	af_NDVlg-adj_2006-07-01_avhrr	157	
17	af_NDVlg-adj_2005-06-01_avhrr	150	af_NDVlg-adj_2006-07-02_avhrr	148	
18	af_NDVlg-adj_2005-06-02_avhrr	143	af_NDVlg-adj_2006-07-03_avhrr	142	
19	af_NDVlg-adj_2005-06-03_avhrr	135	af_NDVlg-adj_2006-08-01_avhrr	122	
20	af_NDVlg-adj_2005-07-01_avhrr	119	af_NDVlg-adj_2006-08-02_avhrr	126	
21	af_NDVlg-adj_2005-07-02_avhrr	111	af_NDVlg-adj_2006-08-03_avhrr	104	
22	af_NDVlg-adj_2005-07-03_avhrr	104	af_NDVlg-adj_2006-09-01_avhrr	87	
23	af_NDVlg-adj_2005-08-01_avhrr	87	af_NDVlg-adj_2006-09-02_avhrr	87	
24	af_NDVlg-adj_2005-08-02_avhrr	84	af_NDVlg-adj_2006-09-03_avhrr	96	
25	af_NDVlg-adj_2005-08-03_avhrr	83	af_NDVlg-adj_2006-10-01_avhrr	96	
26	af_NDVlg-adj_2005-09-01_avhrr	80	af_NDVlg-adj_2006-10-02_avhrr	100	
27	af_NDVlg-adj_2005-09-02_avhrr	87	af_NDVlg-adj_2006-10-03_avhrr	108	
28	af_NDVlg-adj_2005-09-03_avhrr	95	af_NDVlg-adj_2006-11-01_avhrr	107	
29	af_NDVlg-adj_2005-10-01_avhrr	101	af_NDVlg-adj_2006-11-02_avhrr	135	
30	af_NDVlg-adj_2005-10-02_avhrr	110	af_NDVlg-adj_2006-11-03_avhrr	44	
31	af_NDVlg-adj_2005-10-03_avhrr	118	af_NDVlg-adj_2006-12-01_avhrr	154	
32	af_NDVlg-adj_2005-11-01_avhrr	123	af_NDVlg-adj_2006-12-02_avhrr	105	

Mark Columns B and D by clicking in the gray area with “B” and “D”, hold down the Control-key while doing so. Then click the Chart Wizard, , and accept all the default settings that are suggested by the wizard. You will then get a graph that is only missing one thing, namely the coding of the legend. In the excel sheet change the values of B and D (the headings used in the chart) to

state the years of the data instead of “Value”. Your graph should then look something like the figure below.



You can use excel to design a better graph, including setting colors on the bars, putting labels on the axis, adding more years with data etc.

Annual indexes of vegetation production

Annual indexes of vegetation production were calculated using 12 months cycles starting in January and ending in December for each year 1981 to 2009. As stated above three different annual indexes were calculated (annual average *NDVI, annual maximum *NDVI and annual increment *NDVI), and they are all included on the project CD. The table below gives information on the different indexes and where to find the data in the dataset.

Annual vegetation production index	Folder
Annual average	\data_spatial\laketan\rs\ndvig\GIMMS_NDVIg-annual-average
Annual maximum	\data_spatial\laketan\rs\ndvig\GIMMS_NDVIg-annual-maximum
Annual increment	\data_spatial\laketan\rs\ndvig\GIMMS_NDVIg-annual-increment

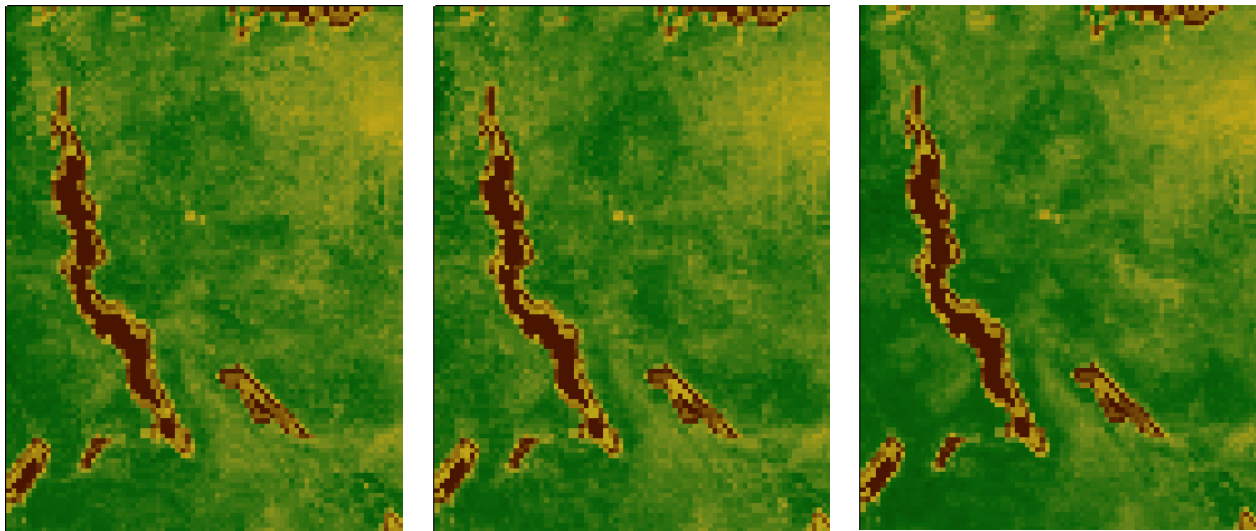
For each index there are layers representing the mean situation for the period 1982-2009, the normalized trend slope 1982-2009 based on z-scores (see below), the normalized trend slope coefficient of determination (regression r²) 1982-2009, and the significance (p) level of the normalized trend slopes.

As you are now going to add several grid files at the same time, it is better to uncheck the option for Layer is visible when added, in the Option window (found via the menu: Tools – Options; see [lesson 3](#) to set the Options).

When you add the vegetation layers to your DIVA-GIS project, only add the grid (*.grd) files, do not load any image (*.jpg) file. You do not need them, as the symbolization of the grid files is identical to the image files.

Start by adding at least three images showing the annual mean vegetation production going back to 1982. Use one image from the early 1980’s, one from the 1990’s and one recent image. In the example below, images for 1984, 1994, and 2004, were used.


Put the added vegetation images at the bottom of the Legend in the Data view, and turn on the layers showing the countries of LakeTanganyika Basin, then sequentially turn on each vegetation image. The years used in this tutorial show a remarkable difference in annual vegetation production. In the example below you can see the vegetation production in 1984, 1994 and 2004 using the increment *NDVI index as an example.

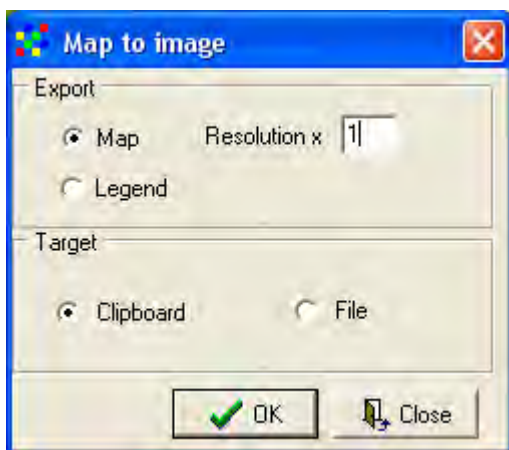


1984

1994

2004

The three small map-images above were generated by clicking the Map to Image tool,  in DIVA-GIS using the Clipboard as Target, and then pasted into this document (by clicking Ctrl-v), without being saved to any intermediate image file.



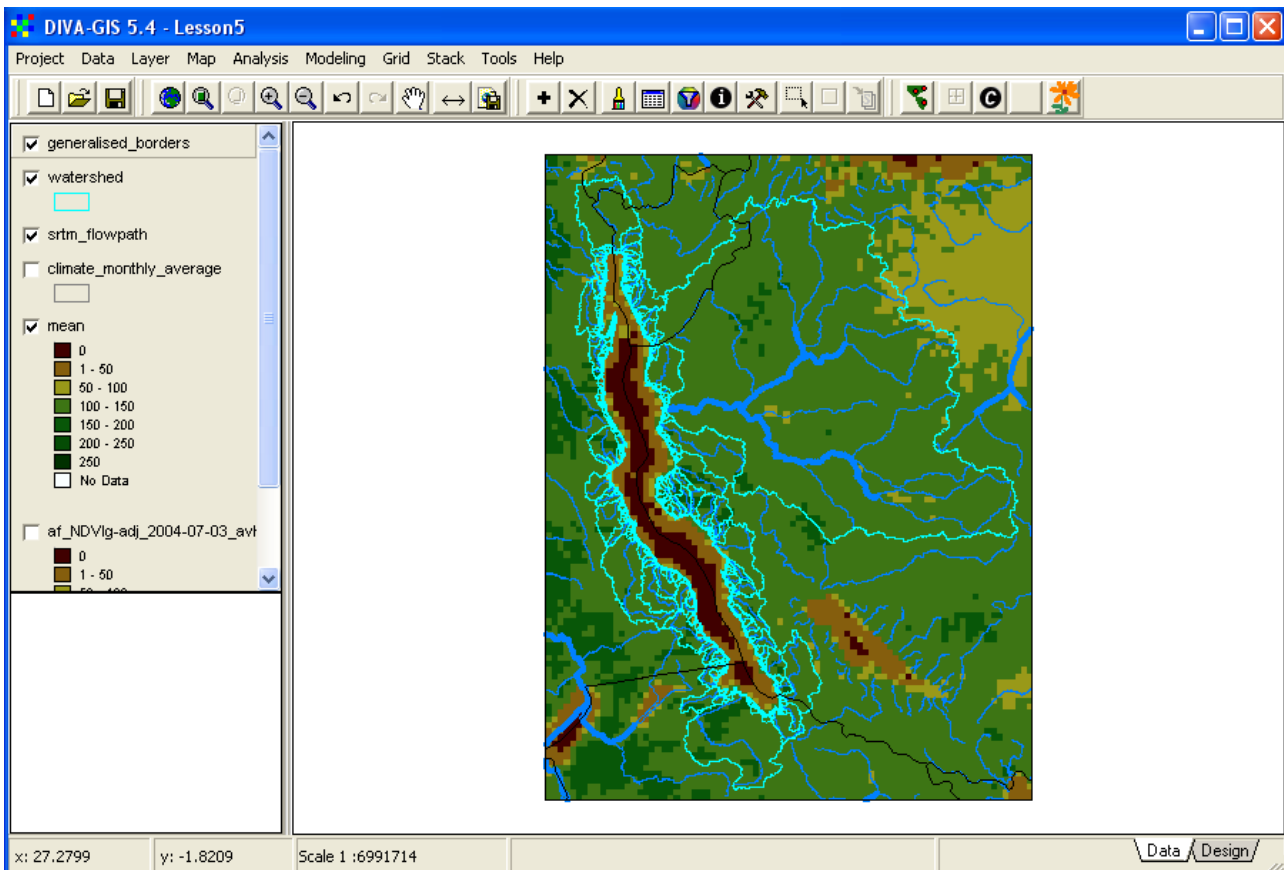
Vegetation production spatial dependency

You can use the framework dataset to do some simple visual interpretation of correlations between various spatial (geographic) phenomena and vegetation production.

Vegetation production and drainage

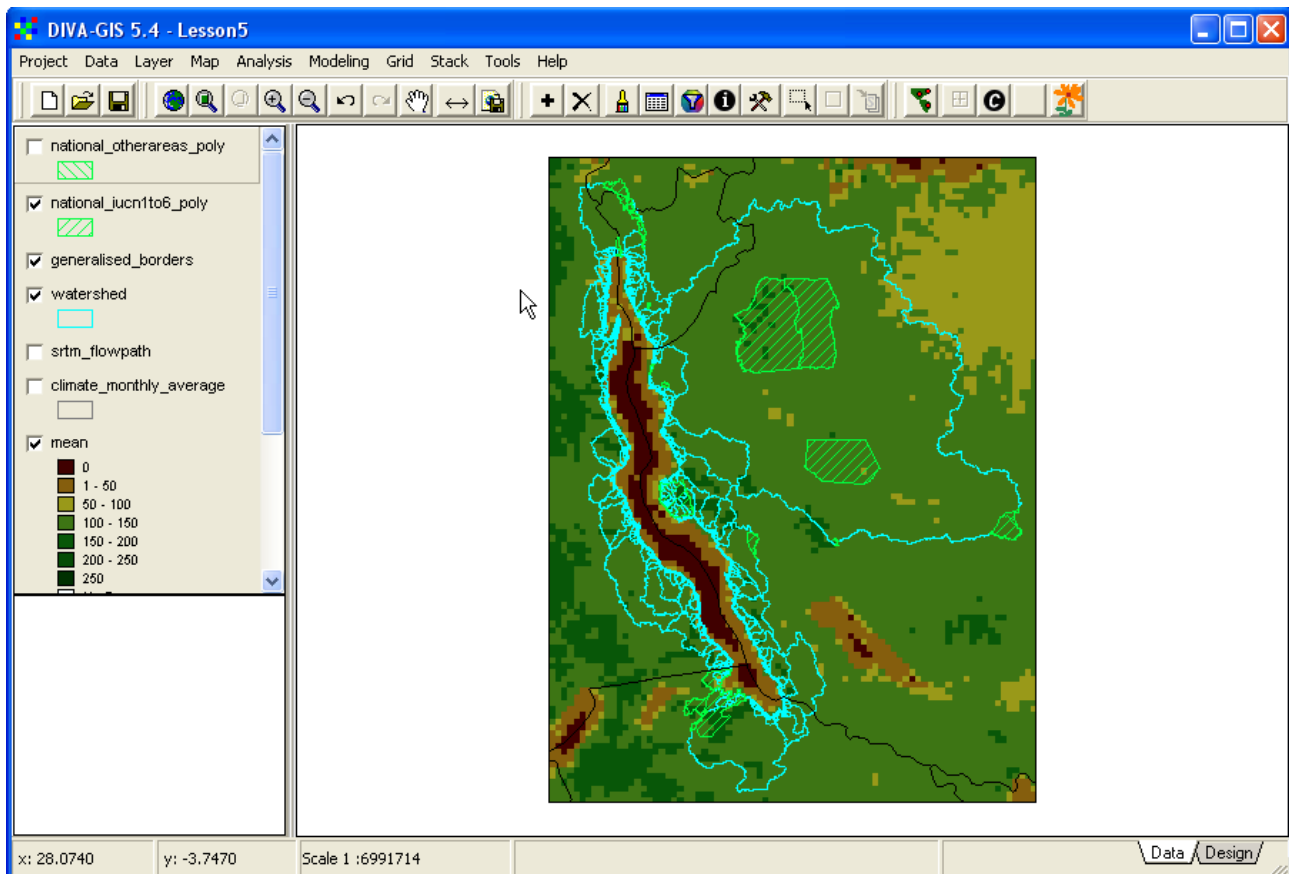
The framework dataset includes flow paths and watersheds, and you need to symbolize these to be able to visually interpret if they have any effect on the vegetation pattern. If you followed this tutorial, you should already have rivers and watersheds properly symbolized in your DIVA-GIS project, if not you need to go the Frameworks data set section above and do the symbolization now.

The image below shows drainage and average vegetation production (1982 to 2009) for the Tanzanian part of the Lake Tanganyika Basin, indicated in the overview map (see lesson 1 to repeat how to add an overview map to your DIVA-GIS project).




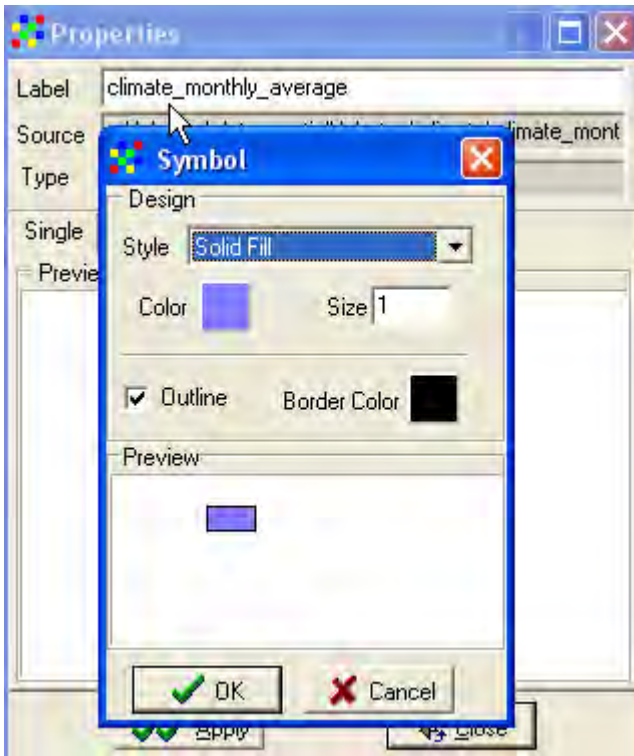
To make a more thorough analysis of topography and vegetation growth more detailed data is needed for both vegetation and topography.

In lesson 4 you used layers showing protected areas to visually analyze the tree cover and if tree cover differed inside and outside protected areas. You can now redo that analysis but instead of using tree cover you can use NDVI layers. In the example below IUCN protected national areas (under the folder `\data_spatila\laketan\mapdata\protected\vector`) are put on top of the average vegetation production layer, but there seems to be no significant difference in vegetation production between protected and un-protected areas. You can test it by using the Identify tool.

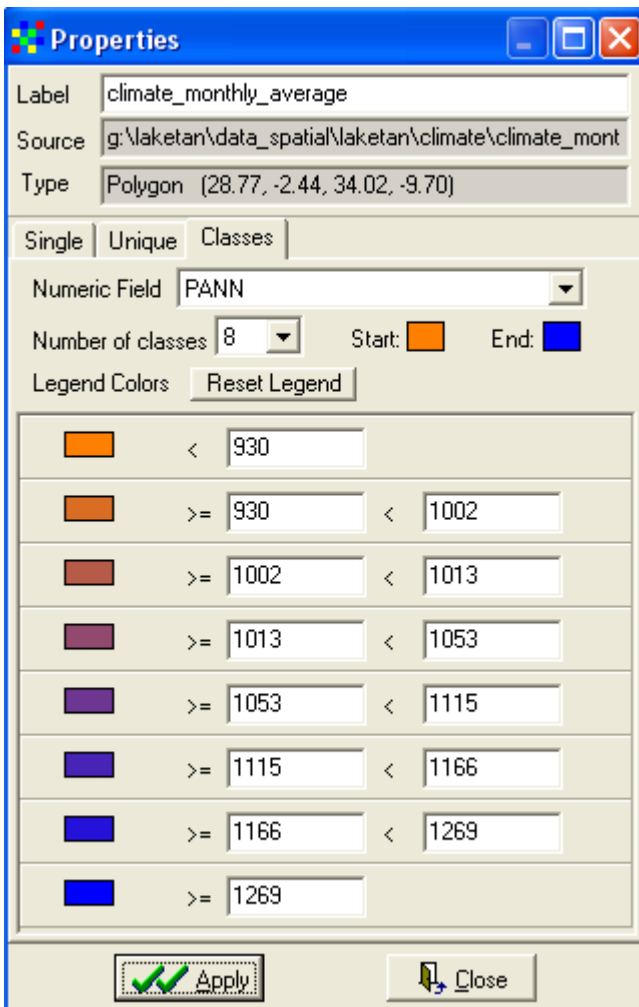


Vegetation production and rainfall

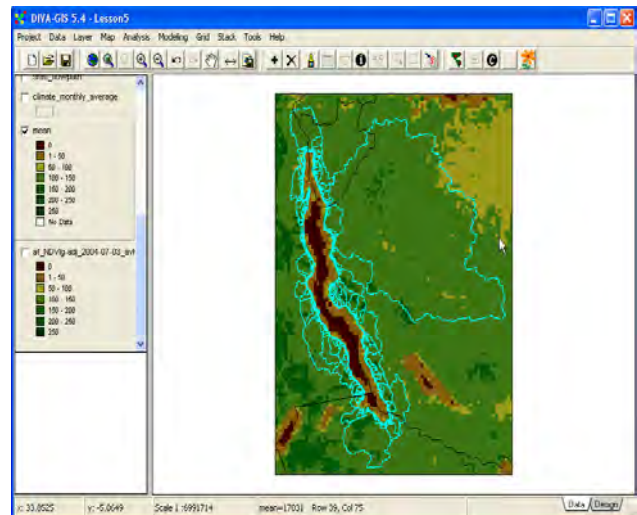
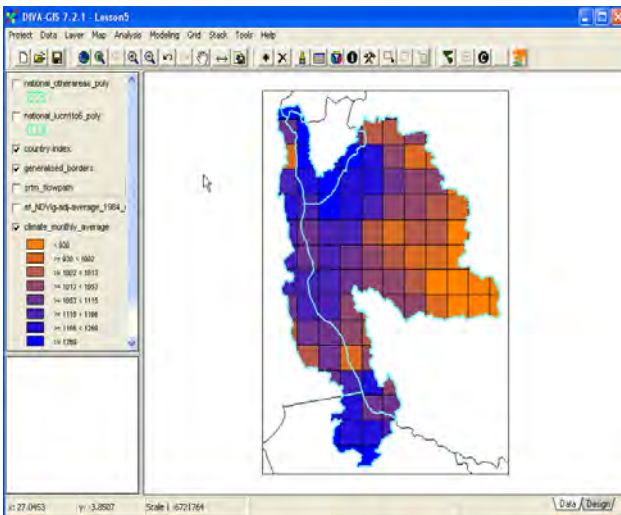
The framework dataset includes a vector layer with climate long term average data (\data_spatial\laketan\climate\climate_monthly_average.shp – if you followed this tutorial this vector layer should already be in your DIVA-GIS project). You can use that layer to see the general rainfall pattern over Lake Tanganyika Basin. Drag the climate vector layer to the top of the Legend in DIVA-GIS (see lesson 1 to repeat how to change the Legend and display order of layers). Open the Properties window for this layer by double clicking the theme in the DIVA-GIS legend, by making the layer active and clicking the Properties tool, , or via the menu: Layer- Properties. Then symbolize the climate layer using the attribute field PANN (precipitation annual). First you must use the Single tab in the Properties window to set the Fill type to solid and uncheck the Border, as shown below.



Change to the Classes tab in the Properties window and create a classification of the annual rainfall attribute (PANN) as shown below (precipitation is an attribute of the type ratio, and symbolization should be done using classes – click here to get to [lesson 2](#) explaining ratio data and other attribute data types). Change the label of the theme to Annual rainfall (mm) (or other relevant label), the label will then be changed in the Legend of the Data View.



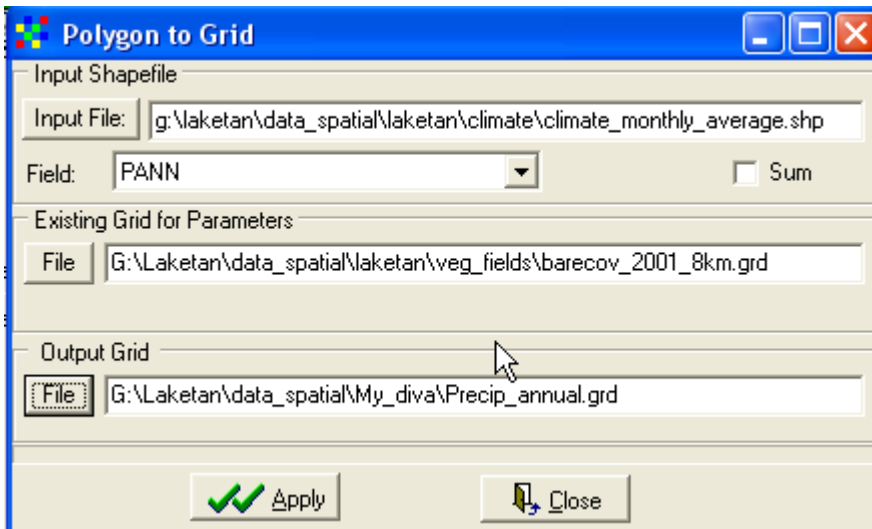
Toggle between viewing the annual rainfall layer and the average vegetation production layer by turning the rainfall layer on and off.



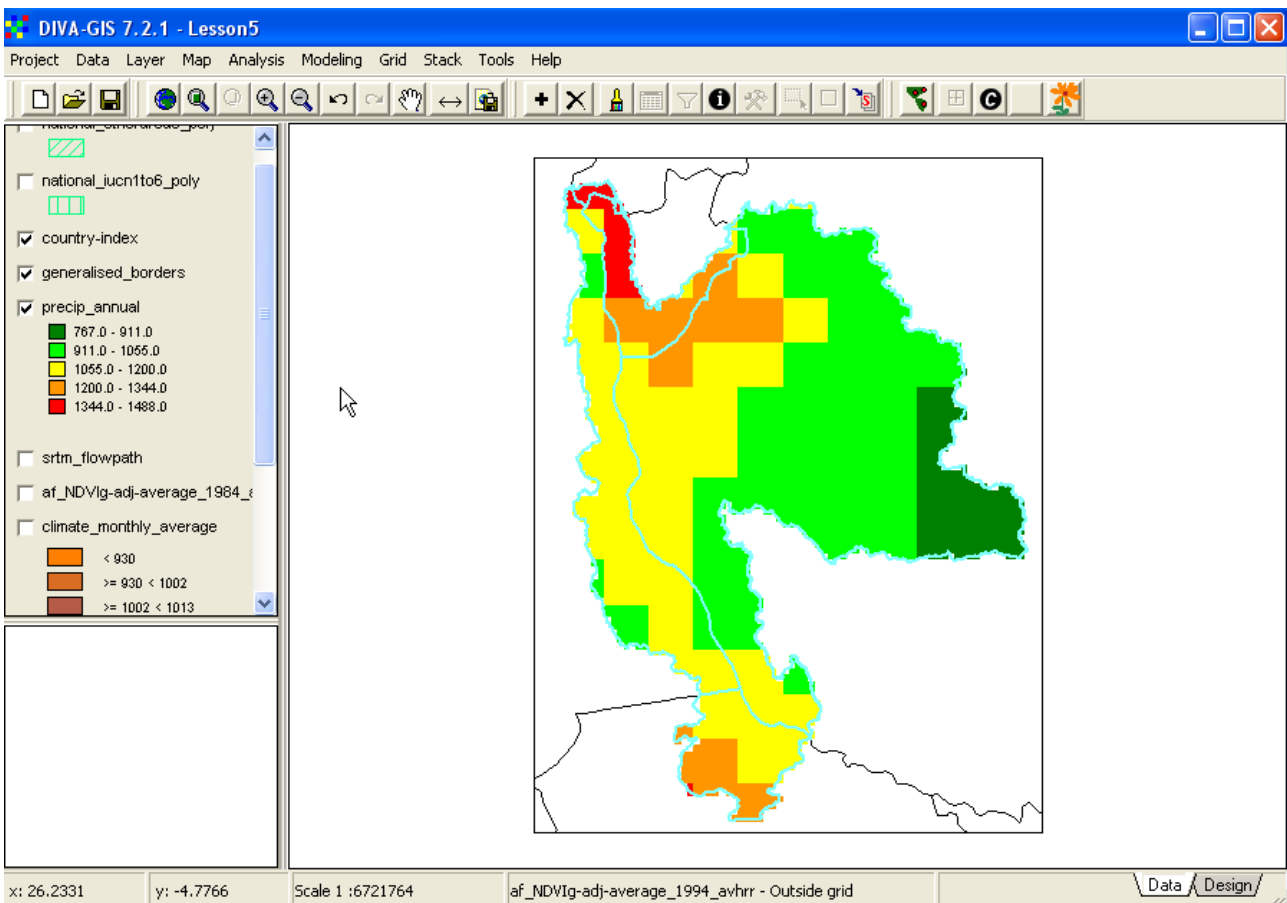
Vector to grid conversion

The climate statistical layer contains monthly data not only for precipitation, but also for temperature and soil water, and can be used for making further analysis. If you want to use the climate statistical data layer for calculations, DIVA-GIS can do that if you first convert the polygons to a grid, via the menu: Data – Polygon to Grid. In the Polygon to Grid window you can set the

parameters of the output grid to exactly fit the vegetation data. As a grid layer only contains one attribute, you need to select each attribute of the vector layer to generate a new grid file (that is why vector files consume less storage space in the memory). For now it is enough if you generate one grid file for the annual average rainfall pattern.

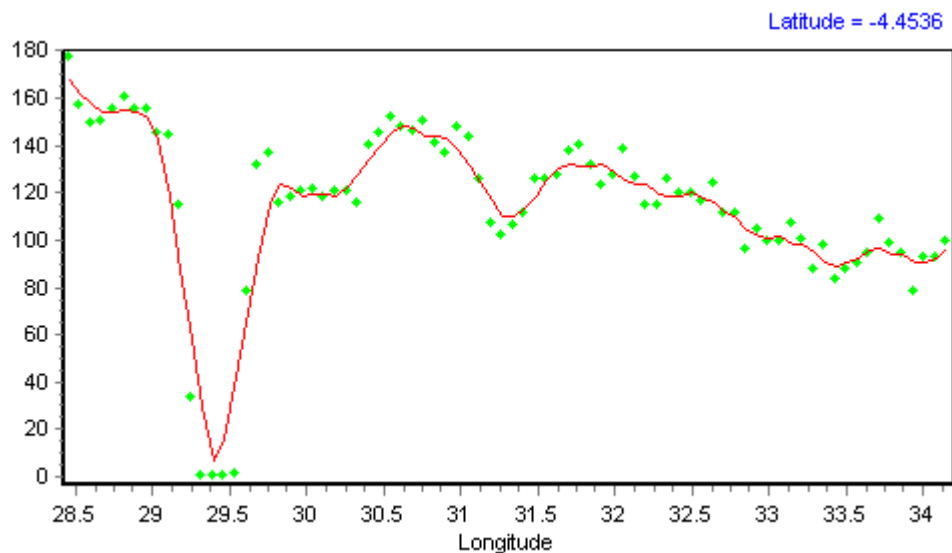


The image below shows the grid layer for annual rainfall generated by DIVA-GIS, and then symbolized using rainfall classes. You can use the Polygon to Grid function in DIVA GIS to convert any vector polygon (shape file) to a grid, including protected areas, countries, river basins etc. In lesson 6 you will use the country layer to create a grid mask in this way.



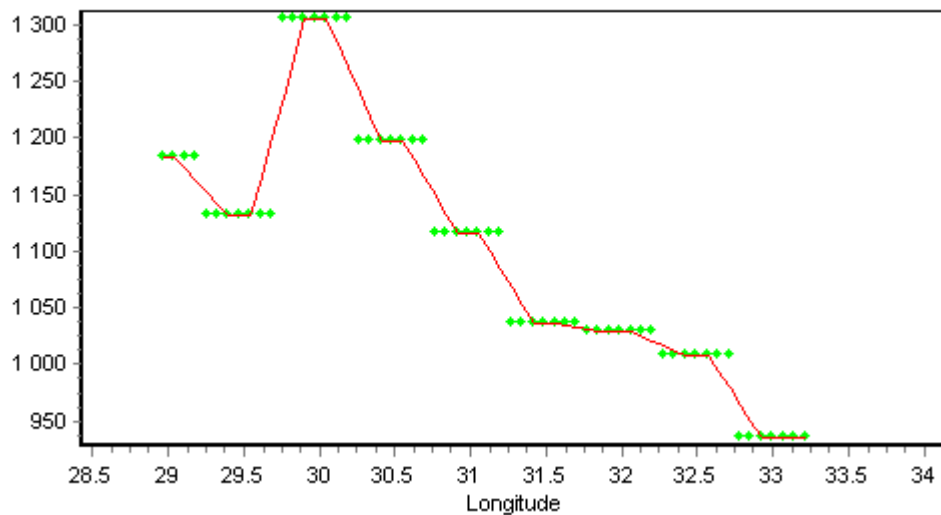
Vegetation production spatial trend

You can use the annual vegetation production layers (or any other vegetation layer) to analyze spatial trends by using the transect tool of DIVA-GIS introduced in lesson 4. The transect tool is available via the menu: Grid – Transect. If you remember from lesson 4 you can either analyze a spatial trend from East to West (default) or from South to North by changing from Row to Field in the drop down menu to the upper left in the Transect window. You can then either use the Click on map button, , and just click in the map to select the position of the transect, or you can manually enter a row or column (as shown below), and then click the OK button, . In the example below the average vegetation production for the period 1982-2009 is shown along the same South-North transect that was used in lesson 4 to analyze the spatial trend in tree cover.



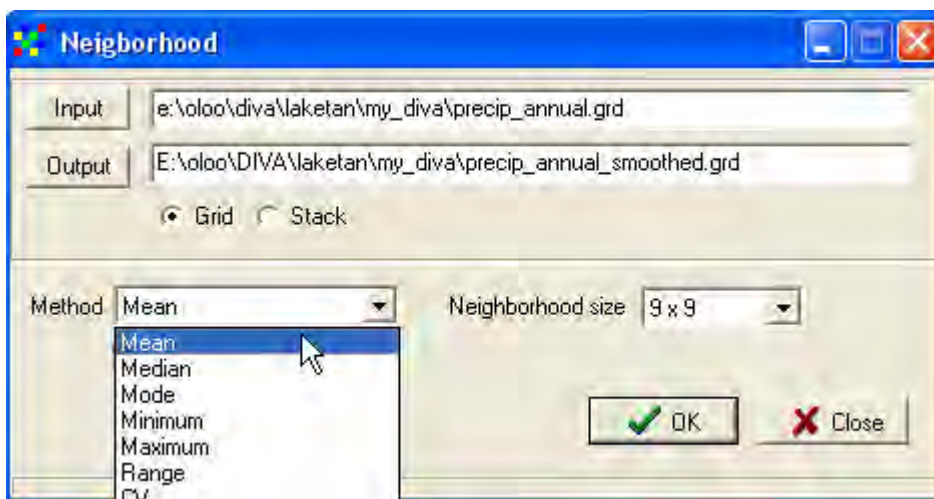
You can now change the Input file to the average annual rainfall grid map that you just created, and redo the transect for the same position (Longitude -4.45 in the example) by setting the Column position manually. In the example below the transect graph for annual rainfall was copied to the clipboard using the Graph to Clipboard button, , and then just pasted into this tutorial.. Note that the original resolution of the rainfall data polygon layer was 50 km, but when converted to a grid the grid resolution was set after the vegetation data in 8 km resolution – hence there will be about 8 neighboring cells (50 divided by 8 is approximately equal to 8) always having the same value. Because of this data values becomes clustered and the moving average aligned with individual data points.

Latitude = -4.4536



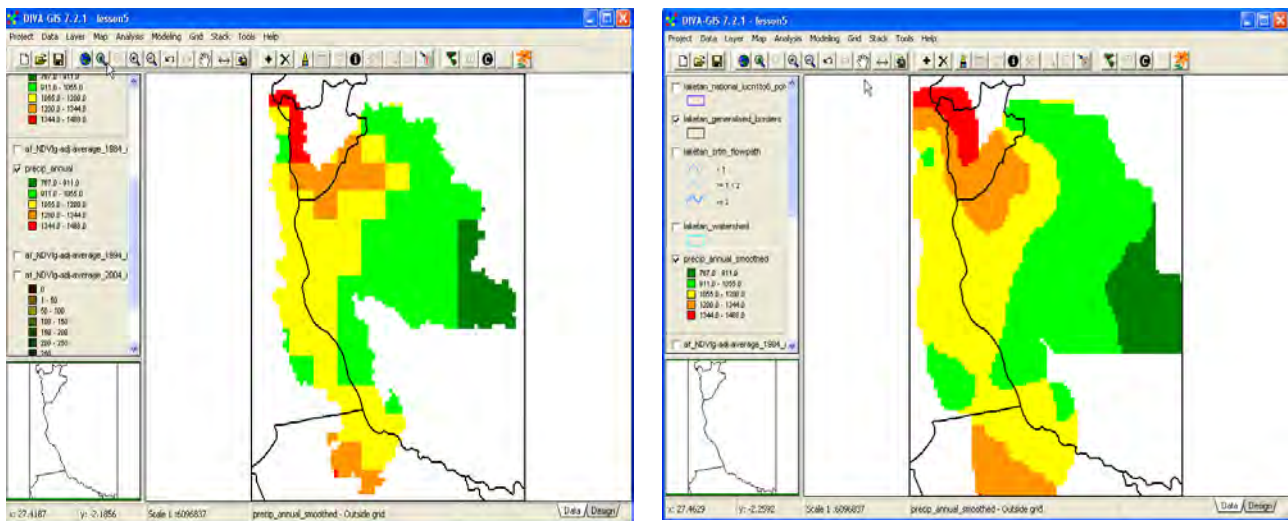
Filtering a grid

In lesson 4 you used map algebra to analyze the relations between cells in the same position, but in different maps. Comparing single cells over different grid layers is called map algebra or local analyzes in the GIS jargon. But you can also analyze a cell related to its neighbors, called focal (or neighborhood) analysis. The most common focal analysis is filtering. A filter is a small window (or kernel in GIS jargon) of cells (3x3, 5x5, 7x7 etc), where the central pixel in the kernel is assigned a value dependent on the values of its neighbors in the kernel. Filtering can be used for analyzing many spatial relations at a focal scale, including mean, median, mode, minimum and maximum values. But also more advanced functions can be used, including for analyzing variation and richness. Make the gridded rainfall map the active theme, and go via the menu: Grid – Neighborhood, to open the window for doing neighborhood analysis. In the dropdown menu for Method you can choose which kind of focal analysis you want to do.

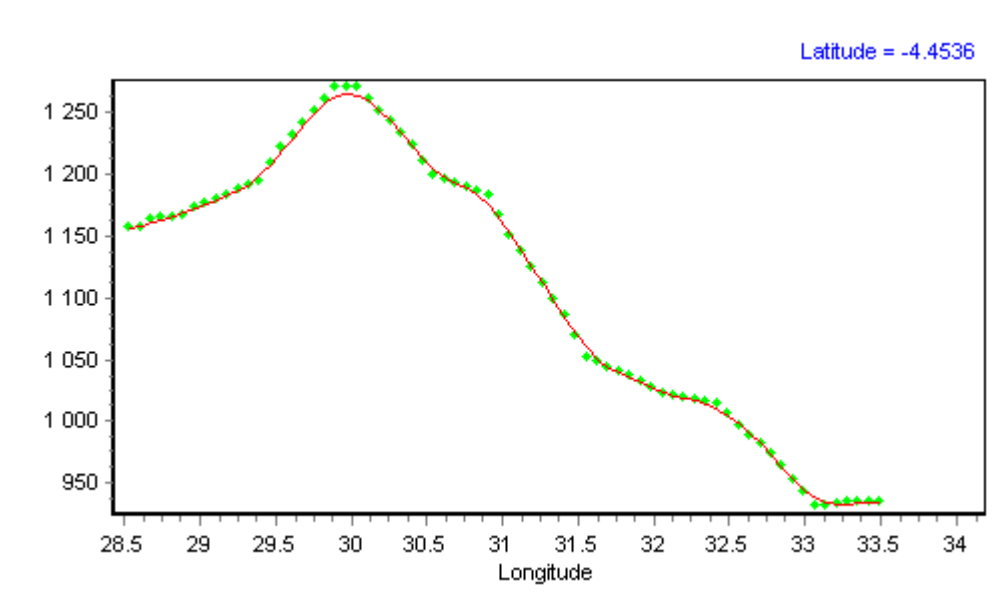


In this case you want to smooth the rainfall grid so that the clusters of pixels with similar values get a more smooth transition. Rainfall is a typical ratio attribute, and hence Mean is the best suited filtering method (if you want to filter nominal data you can not use Mean, but any of the other methods could be used). As the cluster of pixels with similar values are of the size 8x8 cells you should set the Neighborhood (filter or kernel) size to be 9 x 9 in the drop down menu as shown above. Then click OK to do the neighborhood analysis.

Open the properties window for the smoothed rainfall grid (precip_annual_smoothed in the tutorial) and load the legend from the original rainfall grid (see lesson 4 to repeat how to load a legend from an existing to grid). Arrange the display order of the rainfall grid maps and the layer showing the country borders as shown below, then toggle the top rainfall grid layer on and off and you will see the result of the Neighborhood analyses – the filtered version of the rainfall grid has a much more smooth appearance.



If you redo the transect analysis using the smoothed grid you will also see that the clusters of similar values have disappeared, and the values are smoothed.




Neighborhood analysis can also be used to calculate slope (steepness) and aspect (direction of slope). In DIVA-GIS this function is available under the menu: Modeling – Terrain modeling. It is intended to use with topographic data, and will not be used in this lesson.

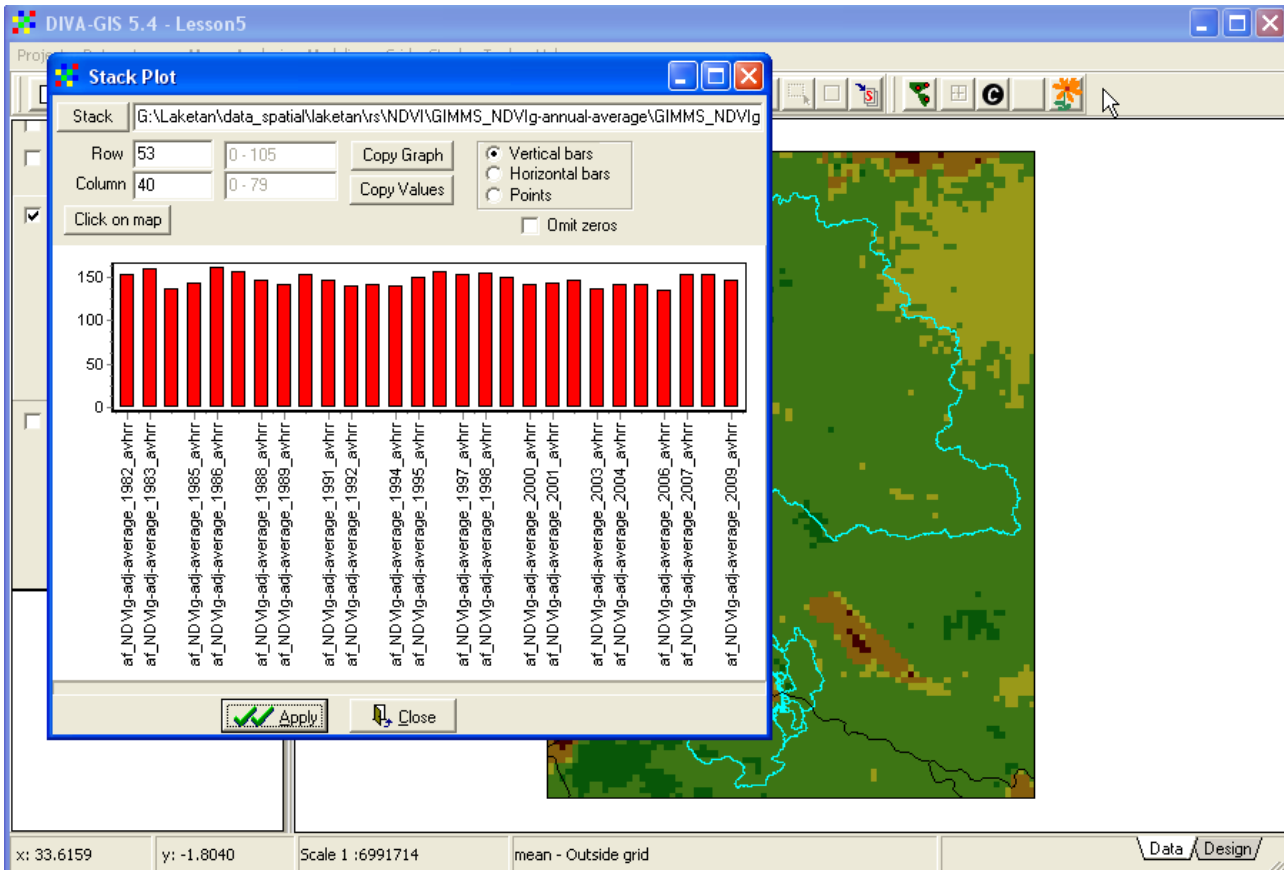
Time series analysis

For mapping out land degradation it is more interesting to analyze the temporal trends in vegetation than the vegetation cover in a particular year.

Annual vegetation production temporal trend

Above you used the DIVA-GIS stack plot function to analyze annual vegetation cycles, now you can also use the same function to analyze the trend in annual vegetation production 1982 to 2009.

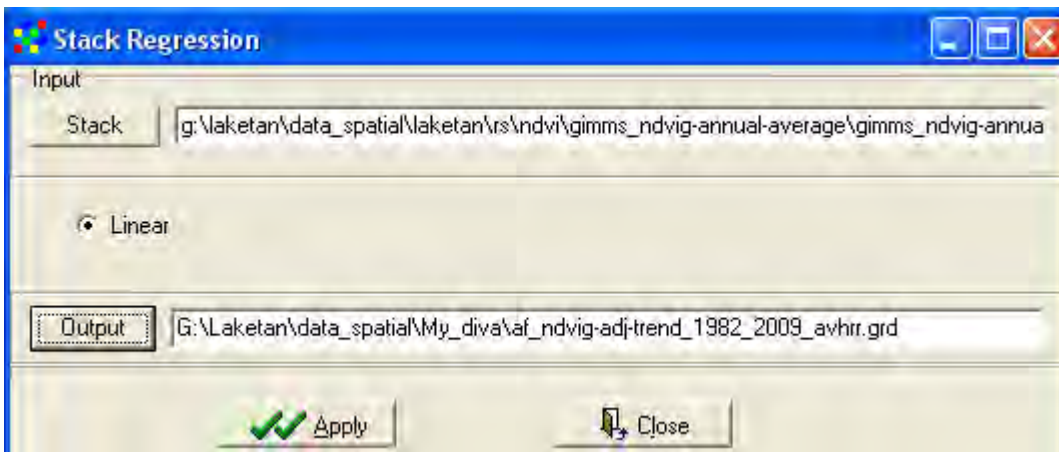
For each annual vegetation production index (average, maximum and increment *NDVI) there is a pre-made stack available for analyzing vegetation variation. The stack for each index is in the same folder as the grid data. Load the stack for the index you want to probe by opening the Stack Plot window from the menu: Stack –Plot. You can make the Stack Plot window larger by grabbing the blue border with the left mouse button, and dragging the mouse. Click the Click on Map button, , and then just click in the map and you will get the vegetation development of the click-point over the last 27 years. The example below shows the vegetation production a point within the Tanzanian side or Lake Tanganyika basin using the annual increment *NDVI index for the period 1982 to 2009 .



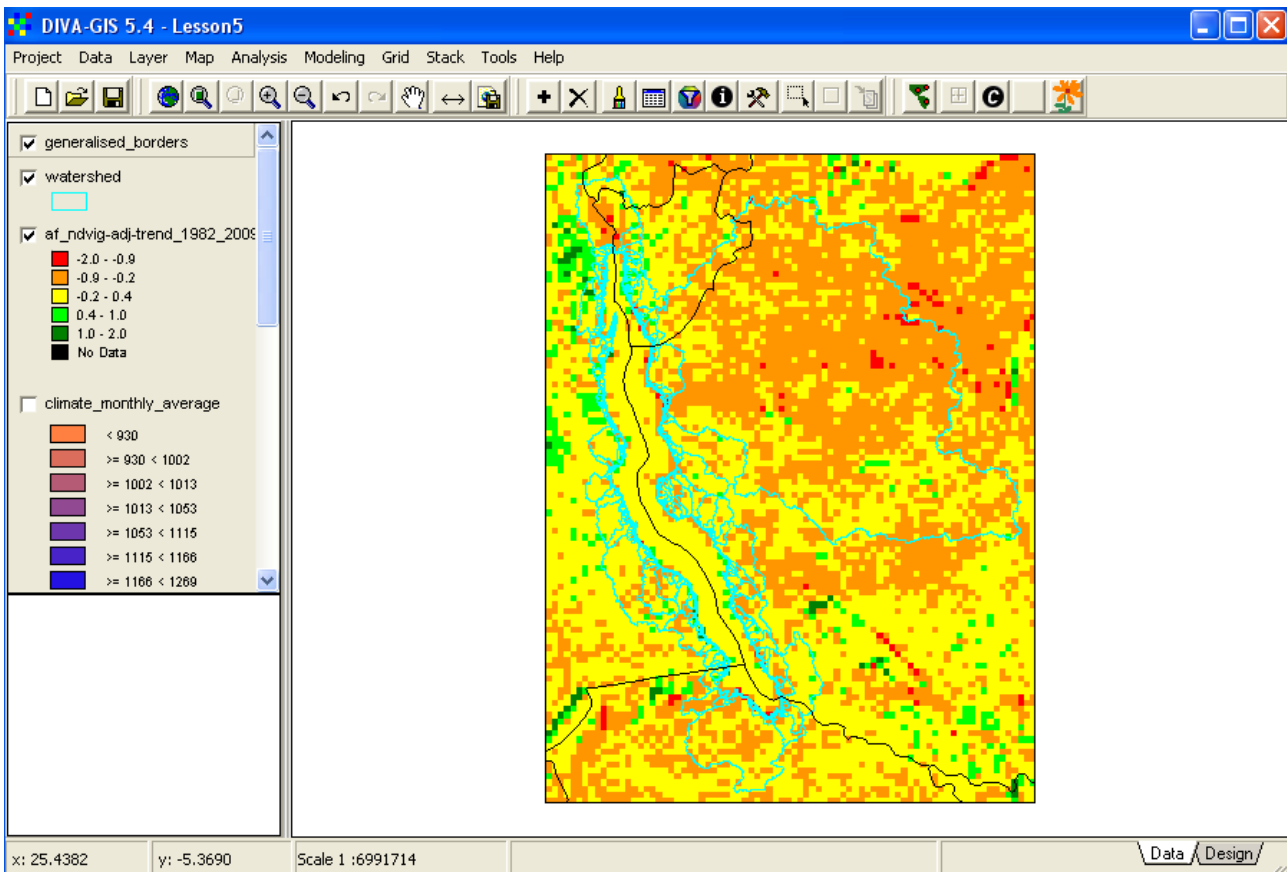
Instead of clicking in the map, you can also give values for which row and column to plot. After entering values in the text fields, just click Apply for DIVA-GIS to analyze your selected point. You can again copy the stack plot graph or the values behind it to the clipboard and just paste the graphics or numerical values into any other Windows application. An efficient way to add data to excel, make an image for a presentation, or for including in a report. Close the Stack Plot window.

Stack regression

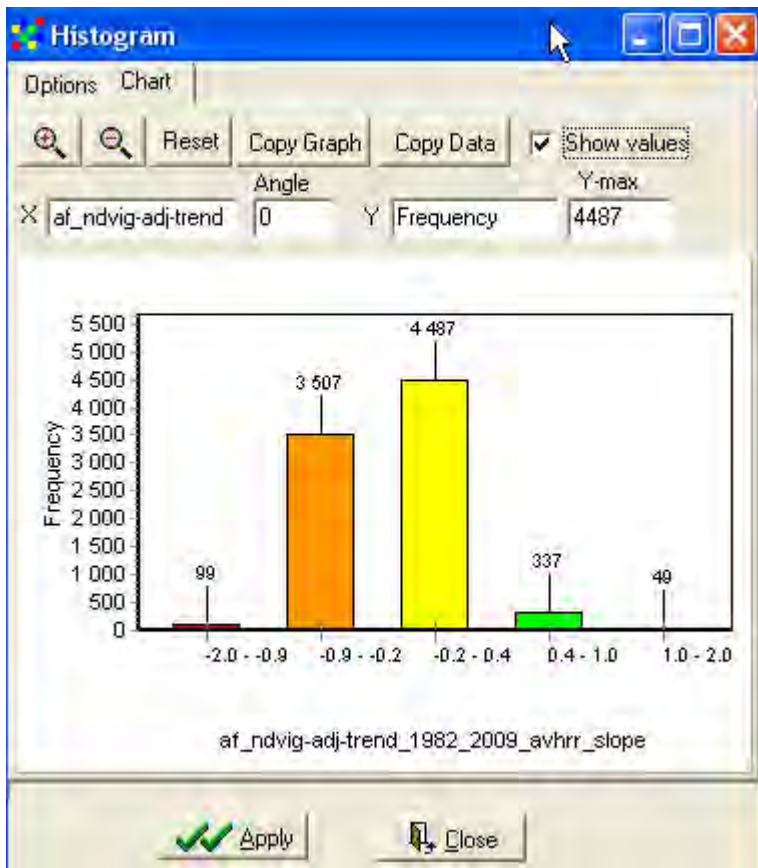
A stack can be used as input for map calculations in DIVA-GIS, introduced in lesson 4. One such option is to calculate a regression trend using a stack. DIVA-GIS supports calculating an ordinary least square (OLS) regression using a stack as input. We can hence use DIVA-GIS to find areas where vegetation has increased or decreased over the last 27 years. The regression option is reached via the menu: Stack – Regression. Select the same stack you used for the stack plot above, click the Output button and give a logical output name, and click Apply to perform the regression, illustrated below.



DIVA-GIS will generate two files from the regression, one with the suffix `_slope` (the strength of the slope) and a second with the suffix `_r2` (the correlation in the slope). Only the first file is automatically added to the project, the second must be added using the Add layer tool. Below you see an example of how the regression slope result has been symbolized using color ramping in 8 classes. It is very clear that vegetation production has reduced over most of Lake Tanganyika for the last 27 years.



Instead of analyzing the trend using the map, you can also use DIVA-GIS to plot a histogram of your results. Make sure the time-series regression layer is the active theme, and go via the menu: Analysis – Histogram. If you had the wrong theme as active layer, you can select which file to plot also directly from the Histogram window. You can set the Class width or the Number of classes you want for your histogram plot, but start with the default - Classes as defined in the legend of the grid layer. Just click Apply, and you will get a histogram where each bar is symbolized with the same color it has in the legend. Click the checkbox for Show values to see the actual numerical values (number of cells falling within each class).



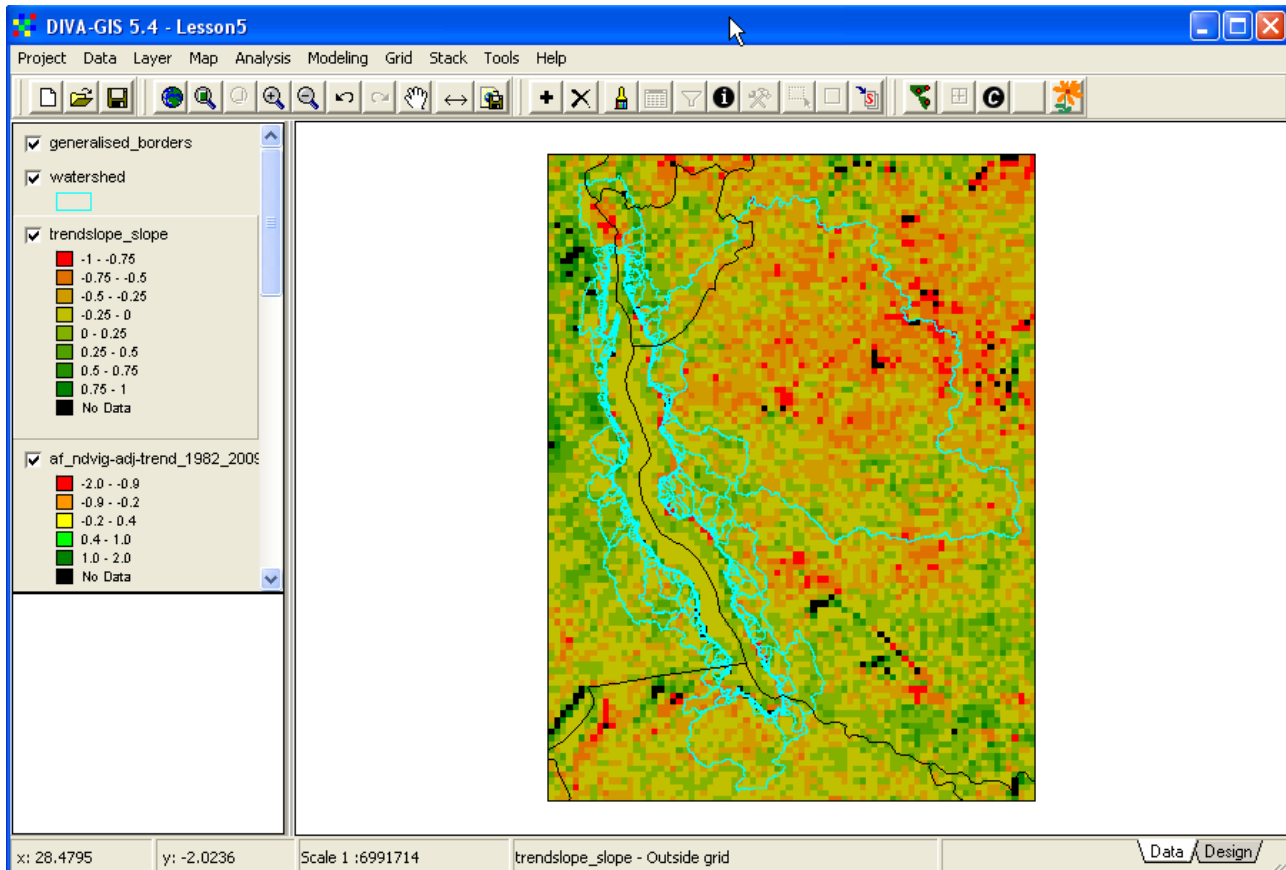
As you see above most cells fall in the class in the negative values indicating a decrease in vegetation cover for the last 27 years.

Again you can copy the graph and paste it into any other Windows document, or you can copy the values and paste them to Excel for further calculations or for building another chart type using the classified data. You can also zoom in and out in the chart to enlarge sections of the chart.

Normalized vegetation trends

The use of Ordinary Least Square (OLS) regression for identifying trends in time-series data with large variation is a bit questionable. For a first test it works but more advanced statistical methods should ideally be applied before any management or policy decision is taken. On the project CD there are ready-made trends using normalized vegetation data for each annual vegetation index included (the [section on normalization](#) at the end of this lesson to get more information on the normalization method). The normalized trends are also derived from OLS (and hence a bit questionable). The normalization (standardization) was done by calculating the average and variation (standard deviation) for each cell for the period 1982-2009. A value (called z-score) was then derived that measures the relation between the original NDVI value of each cell and year, and the time series average and standard deviation for that cell. In short this method is applied to remove the influence of extreme (high or low) values that otherwise have an un-proportional large influence on the regression outcome. You can now compare your regression using the original data, with the regression done using normalized data. The z-score regression is in the same folder as the original data, but with “z-slope” inserted. Below you see the normalized trend slope for increment

*NDVI 1982-2009.

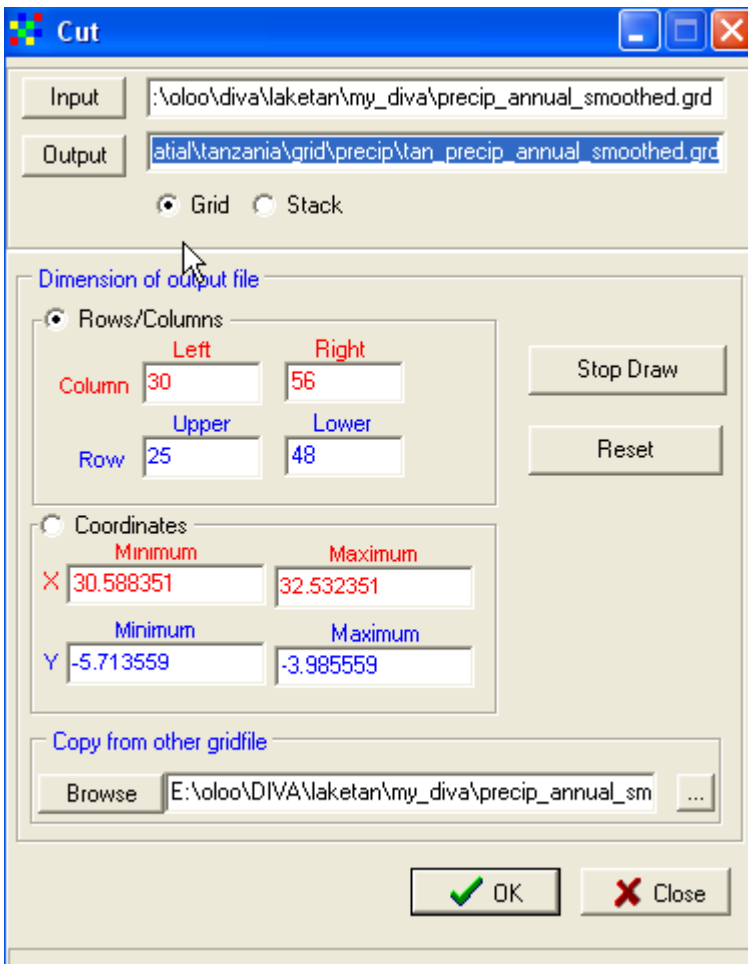


In DIVA-GIS you can also do an Ordinary Least Square (OLS) regression with either single or multiple independent variables to estimate the value of a dependent variable.

Cut Grid

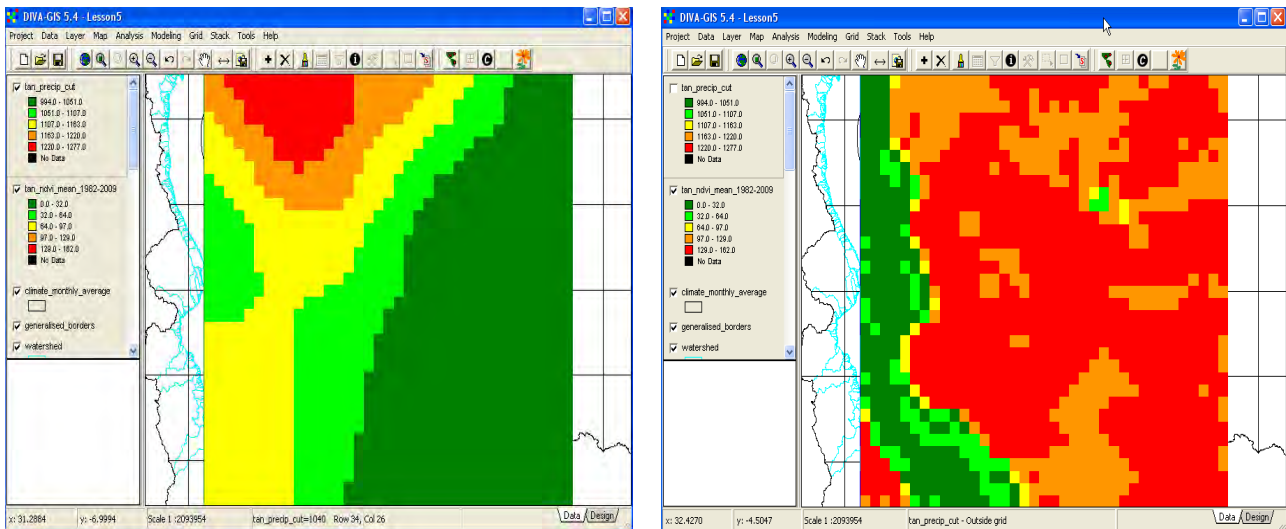
Make the smoothed rainfall grid you created above (precip_annual_smoothed) the active theme, and go via the menu: Grid – Cut. In the example below a new data site for Tanzania was created under the data_spatial directory, and then the naming principles used for other layers was obeyed, introducing a prefix “tan”. The easiest way to define the new grid size is to click the Draw

Rectangle button, , in the Cut window, and draw a rectangle defining the new grid size in the Data view. When you have given an Output name and defined a cut area click the OK button to cut the grid to a smaller size.



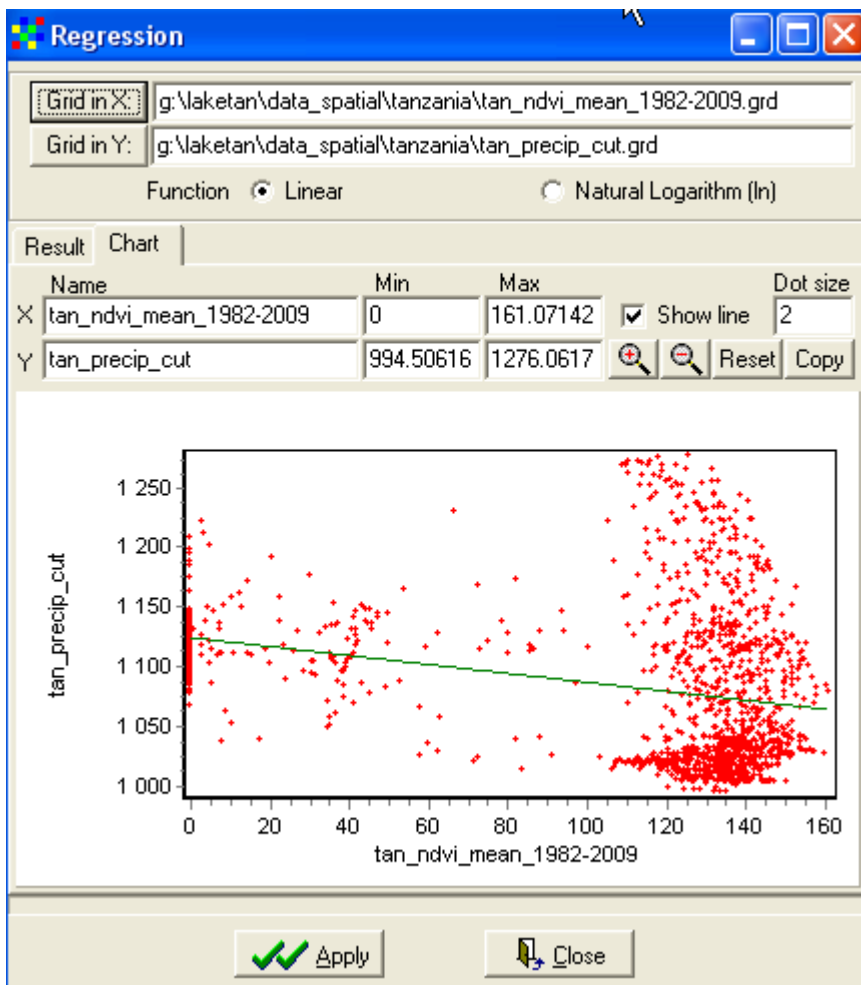
If you keep the Cut window open you can just change Input and Output layers to cut another grid to the same size used for the previous cutting. Otherwise you can also enter the dimensions of the output layer, either as Row/Columns positions or as Coordinates. You can also copy the extent of the cutting from an existing grid in smaller size (i.e. a grid that you already cut before). It is important that your grids covering a specific region have the same extent and resolution (see A note on grid data dimensions in lesson 4). Also cut one of the annual vegetation grids (mean for the whole period 1982-2009, or a grid representing a single year).

You can symbolize the cut grids (rainfall and vegetation production) by loading the legends from existing layers that are already symbolized. The two images below show the cut versions of the rainfall and averaget *NDVI (mean for 1982-2009) grid layers, symbolized using the original grid layers. Note how the colors are not as gradual in the cut *NDVI layer as in the original – the original layer was symbolized outside DIVA-GIS and can not easily be recreated using DIVA-GIS (you then need to define 250 entries in the Legend).



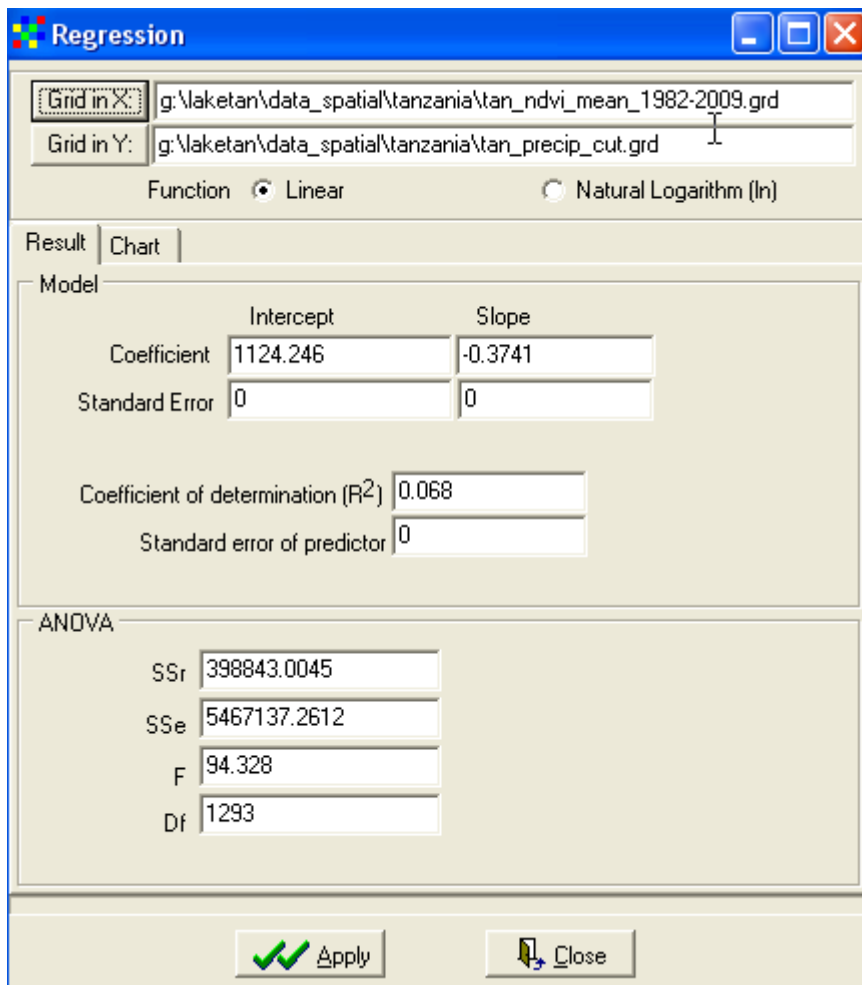
Regression with independent and dependent variables

With the two smaller grids shown in the image above, you can use DIVA-GIS to test the statistical dependency of vegetation production on rainfall. This regression function is under the menu: Analysis – Regression. In the Regression window set the cut vegetation grid to be the dependent variable (Y), and the rainfall to be the independent variable (X). Then click Apply to perform the regression analysis. When the regression is ready you can watch the results under the two tabs in the Regression window. The Chart tab opens automatically when the regression is done, and you see the regression as a plotted graph.



(If you look at the curve in the figure above, a logarithmic function should fit the data better than a linear model. In DIVA-GIS you can use the Natural Logarithm as a fitting function rather than a linear function, but the result is not as good as for the linear model. With another exponential function than the natural logarithm the fitting would be better.)

Under the result tab you can see some more details about the regression, including the regression intercept and slope.



In the example above the intercept (1124.246) and slope (-0.3741) leads to the following regression model for determining vegetation production in Lake Tanganyika basin:

$$\text{annual Average *NDVI} = 1124.246 - 0.3741 * \text{rainfall}$$

You can also see the coefficient of determination (R²) that is as high as 0.068 in the example above that can be interpreted as rainfall determining 6.8 % of the spatial variation in vegetation growth in Lake Tanganyika. There are also statistical results for ANOVA (analysis of variance) and about other statistical properties.

If you have more variables available that you think can explain vegetation growth, you can also include them by using a multiple regression, available in DIVA-GIS under the menu: Analysis – Multiple regression. Factors that could be tested together in a multiple regression apart from rainfall include: topography, protected area, population and rainfall the previous year.

Try it on your own – analyze another region

You should now be able to use what you learned above to redo the analysis of vegetation trends for your own country or district (districts for all the countries in L. Tanganyika basin are available in

the vector layer \data_spatial\laketan\mapdata\political\vector\admin_districts.shp). You first need to select a country/district to work with, and then cut the grid data fit that region, and then redo the analysis. Select a study site and use the Stack Plot tool to capture the vegetation change as recorded with the three different indexes (annual average *NDVI, annual maximum *NDVI and annual increment *NDVI). Transfer the results to excel and create a graph that shows the difference between the three indexes. Use the stack regression to analyze the rainfall relation of each index to see how they respond to rainfall Compile your results in a short report including a designed map, a histogram for each index, a comparing excel diagram, and a transect for the same position for each index.

Normalization (from wikipedia.org)

In statistics, the standard score, also called the z -score or normal score, is a dimensionless quantity derived by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation. This conversion process is called standardizing or normalizing.

The standard score indicates how many standard deviations an observation is above or below the mean. It allows comparison of observations from different normal distributions, which is done frequently in research.

The standard score is:
$$z = \frac{x - \mu}{\sigma}$$

where:
 x is a raw score to be standardized
 σ is the standard deviation of the population
 μ is the mean of the population.

The quantity z represents the distance between the raw score and the population mean in units of the standard deviation. z is negative when the raw score is below the mean, positive when above.