GeoCLIM Manual Version 1.2

FEWS NET

Feb-2017

The GeoCLIM Manual is intended to be a reference guide for all users of the climatological analysis tool, including climatologists, decision makers, researchers, etc. FEWS NET and the CHG are dedicated to providing tools to help mitigate or prevent humanitarian crises.

Famine Early Warning Systems Network

Table of Contents

Acknowledgements: GeoCLIM was developed by Tamuka Magadzire of the Famine Early Warning Systems Network (FEWS NET), and Karthik Vanumamalai, Cheryl Holen and Joshua Sickmeyer at the United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center in support of the US Agency for International Development's (USAID) Planning for Resilience in East Africa through Policy, Adaptation, Research and Economic Development (PREPARED) Project and Global Climate Change activities, in consultation with Chris Funk, Gilbert Ouma, Ismael Mulama Lutta and Gideon Galu. The documentation was supported by Diego Pedreros and Tamuka Magadzire with the additional support by Claudia J. Young and Libby White. Project oversight for the development of the GeoCLIM was provided by James Verdin, Chris Funk, and James Rowland.

Introduction

Summary

The GeoCLIM program is part of a set of agroclimatic analysis products developed by the FEWS NET/United States Geological Survey (USGS). The GeoCLIM is designed for climatological analysis of rainfall, temperature, and evapotranspiration data. The GeoCLIM provides an array of accessible analysis tools for climate-smart agricultural development. These user-friendly tools can be used to:

- Blend station information with satellite data to create improved datasets,
- Analyze seasonal trends and/or historical climate data,
- Analyze drought for a selected region by calculating the standardized precipitation index (SPI),
- Create visual representations of climate data, create scripts (batch files) to quickly and efficiently analyze large quantities of climate data,
- View and/or edit shapefiles and raster files, and extract statistics from raster datasets to create time series. (video- GeoCLIM overview)

Using the Manual

This manual presents examples and exercises to help users understand the different applications of the GeoCLIM tools. For video tutorials and to download practice exercises for the different application of the GeoCLIM go to the CHG website chg.geog.ucsb.edu/tools/geoclim.

Chapter 1: *Overview* provides a brief tour of the various functions available in the GeoCLIM.

Chapter 2: *Settings* will go through setting up the GeoCLIM and downloading data.

Chapter 3: *Data Types* provides a brief review of the different data types used in the GeoCLIM.

Chapter 4: *Spatial Data Viewer* gives a tutorial for the viewing, editing, and creation of shapefiles and rasters on the GeoCLIM.

Chapter 5: *Climatological Analysis* explains how to calculate statistics, trends, SPI, and other functions for a set period of time.

Chapter 6: *Rainfall Summaries* shows how to calculate totals, averages, and anomalies for a period of time.

Chapter 7: *Climate Composites* allows for seasonal analysis among a group or two groups of non-consecutive years.

Chapter 8: *Contour Tool* briefly reviews how to visualize spatial rainfall distribution based on contour lines.

Chapter 9: [Calculate Difference in Averages] shows another way of estimating trends by comparing changes in averages for two periods.

Chapter 10: [BASIICS] is a walkthrough of the process of blending station and raster data.

Chapter 11: [Extract Statistics] explains how to create summaries of historical rainfall for a given region.

Chapter 12: [Create archives] explains how to create climate data archives that could be shared with other GeoCLIM users.

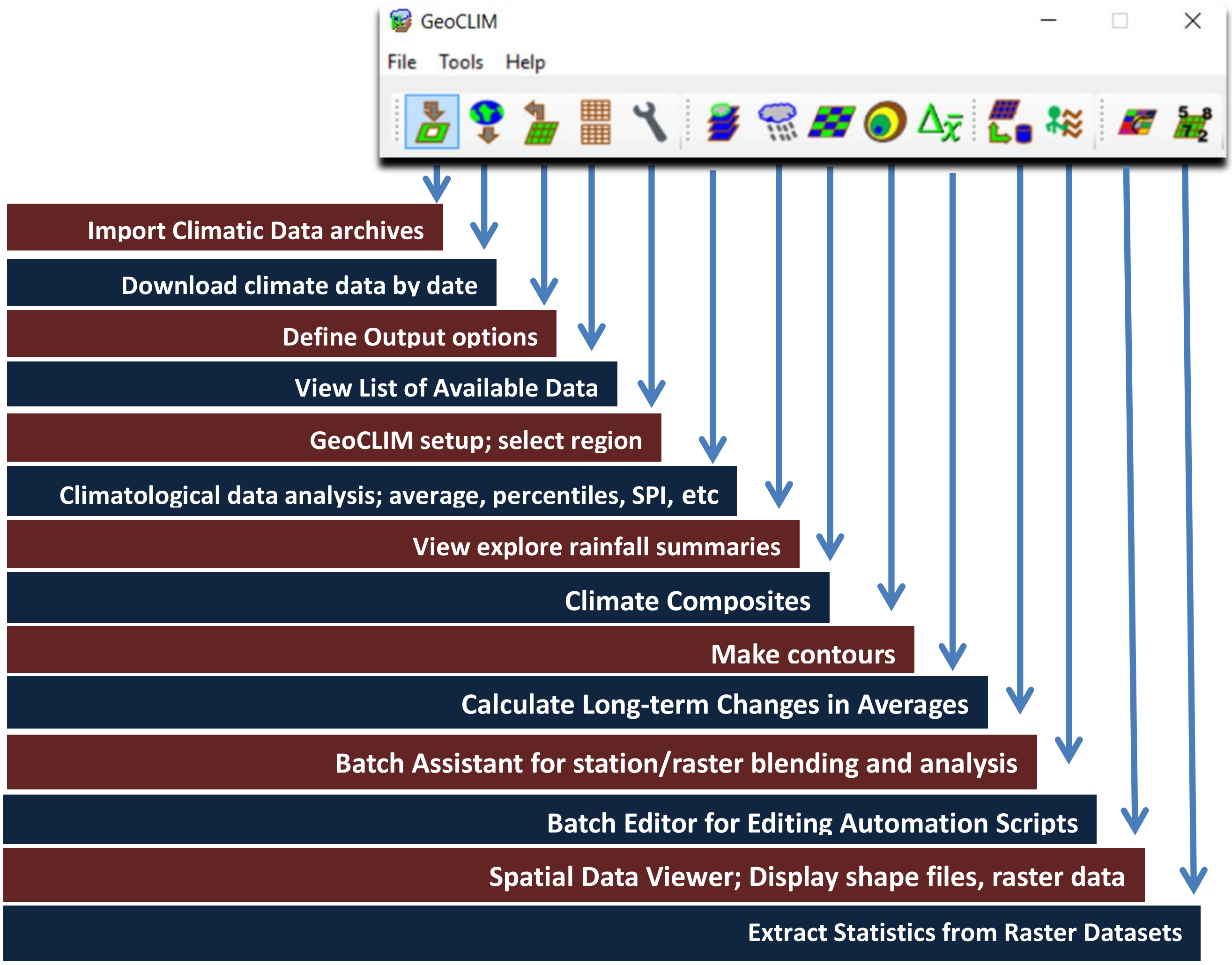
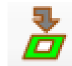


Figure 1.1 *GeoCLIM Main Toolbar.*

Summary

Figure 1.1 shows the main tools available from the GeoCLIM toolbar. These tools consist of settings, data management, and analysis tools. This chapter briefly explores these main tools. In the following chapters, each tool will be described in more detail with example and exercises will be available at the end of each chapter.

1.1. Import GeoCLIM Climate Archives 

A GeoCLIM archive is a compressed file containing data for a given climate variable and specific information so that it could be imported into the GeoCLIM. The *Import Climate Archives* tool (Figure 1.2) is used to make datasets available in GeoCLIM for analysis. These archive files are useful for sharing data among GeoCLIM users. For information on creating data archives, see chapter 12.

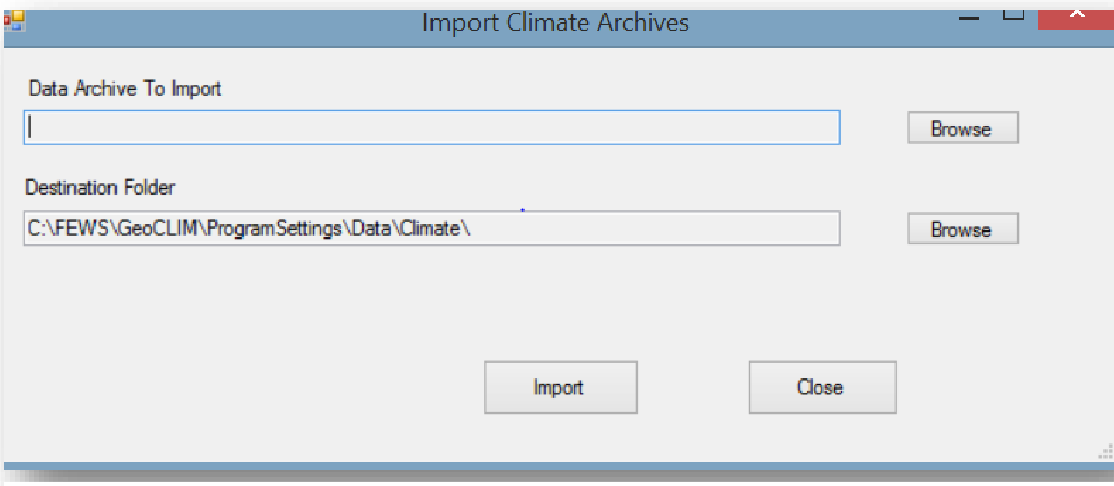


Figure 1.2 *Climate data can be made available in the GeoCLIM through downloading archives.*

1.2. Download Climate Data 

The *Download Selection* tool (Figure 1.3) facilitates bulk downloads of available climate data via FTP or HTTP from different sources (e.g. UCSB, USGS, etc.). See chapter 2 for more information on selecting datasets, including changing the default dataset.

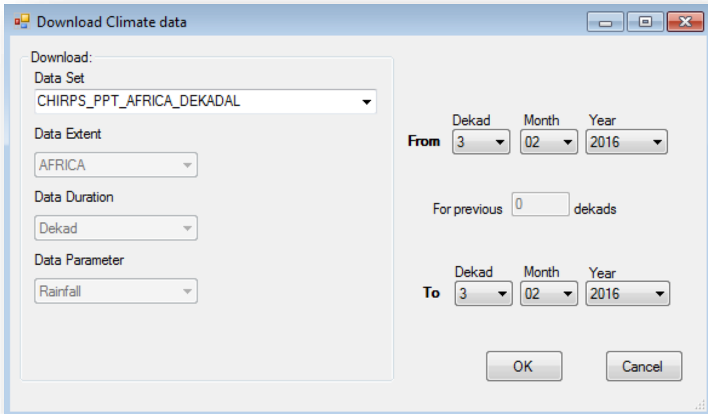
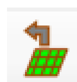


Figure 1.3 Rainfall, Temperature, or Evapotranspiration data can be dowloaded directly from an FTP site.

1.3. Define Output Options 

The *Define Output Options* tool enables the user to specify how the GeoCLIM outputs should be generated and saved including, see Figure 1.4, **(1)** the windows size of the GeoCLIM main toolbar, **(2)** the title fonts for the output graphics, **(3)** whether to format outputs for ArcGIS, **(4)** the file size limits of the temporary directory, **(5)** the output directory, and **(6)** the prefix for the output files from the analysis tools (Figure 1.4). The default values are as shown in Figure 1.4.

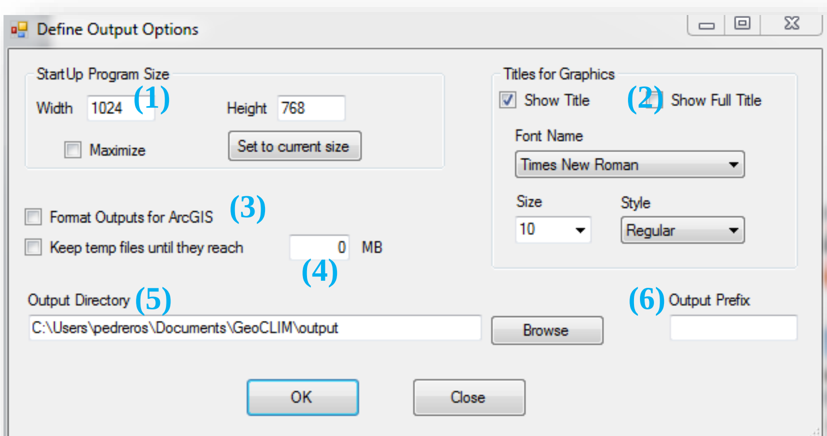


Figure 1. 4 The tool allows the user to define output settings.

1.4. View Available Data

The *View Available Data* tool shows a list of the data available for analysis based on the climate dataset selection (rainfall, temperature, or evapotranspiration). Figure 1.5 shows an example of a list of dekadal (10-day totals) rainfall starting in January 198101. For this example, the selected rainfall dataset by the user is `CHIRPS_PPT_AFRICA_DEKADAL`. The **List Missing Data** button provides a list of the missing files in the time series between the first and last date of available data. The **Export** button can be used to export data from the selected climate dataset in different formats (single BIL or NetCDF files, or as a GeoCLIM Archive) for sharing or backup. See chapter 2 for more information.

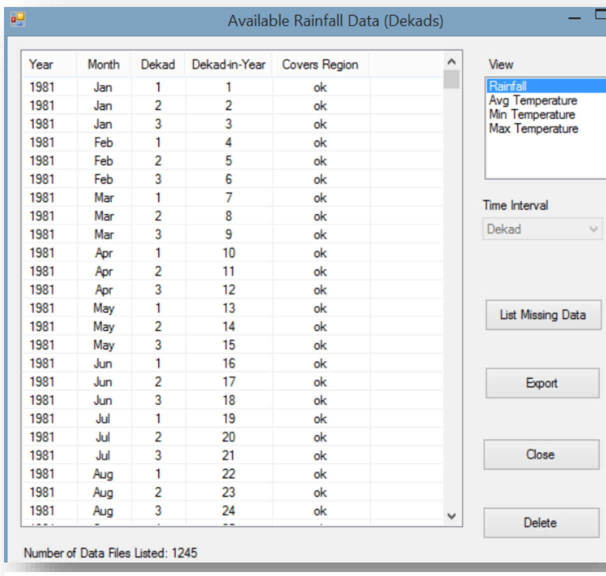


Figure 1.5 *The user could list the available data from the different climate variables and their spatial coverage.*

1.5. GeoCLIM Settings.

The *GeoCLIM Settings* tool contains three main tabs (see Figure 1.6):

1.5.1. **Region -** Select a pre-defined region based on a geographical area (e.g., country, county, city, or pre-defined region).

1.5.2. **Mask** - Define a specific mask for the selected region for computation and display (e.g., land masses, non-desert regions).

1.5.3. **Data** – Select a dataset for each of the available climate variables in the program (precipitation or rainfall, average temperature, minimum temperature, maximum temperature, and evapotranspiration). Click on the **Select Dataset** button to change the data to be used for analysis. For more information see chapter 2.

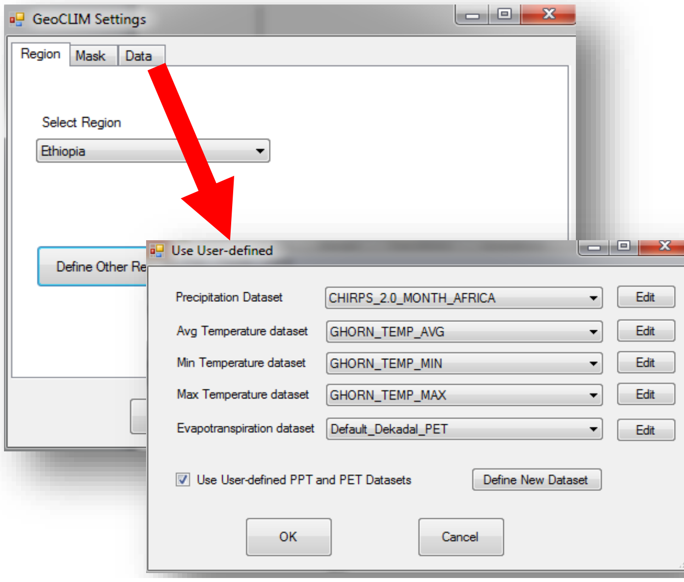
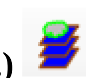


Figure 1.6 *The settings tool allows the user to select the climate data and the region for analysis.*

1.6. Climatological Data Analysis (mean, trend, percentile, etc.) 

The *Climatological Analysis of Climatic Variables* tool (Figure 1.7) is designed to analyze and display the statistical characteristics of rainfall, evapotranspiration, and temperature data based on the time series of the data. The tool displays all the years and periods (months, dekads, or pentads) available for a selected climate dataset. The user may then select a period or a group of periods, for example: months (**Select [Dekads/Months/Pentads] To Process**) **(1)** and years (**Select Years to Analyze**) **(2)**. In addition to specifying the period and years to analyze, the user may choose an Analysis Method **(3)** such as Average, Trend, SPI, among others; and a Parameter to Analyze. Finally, the user can also specify the output directory. Clicking on the **Analyze** button will produce a map using the selected input parameters. See chapter 5 for a more in-depth discussion of this tool.

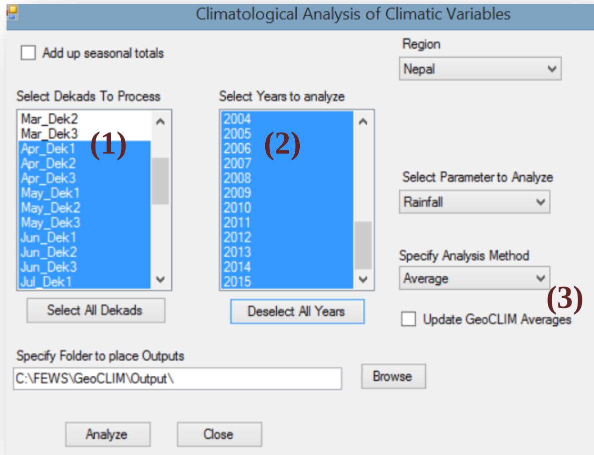


Figure 1.7 *The user could calculate statistics, trends, SPI among other functions for specific seasons.*

1.7. Rainfall Summaries 

The *Rainfall Summaries* tool (Figure 1.8) calculates the total rainfall for a selected time period and region, and expresses this as the actual total, the difference from the long term average, or as a percent of the long-term average. More details on this tool can be found in chapter 6.

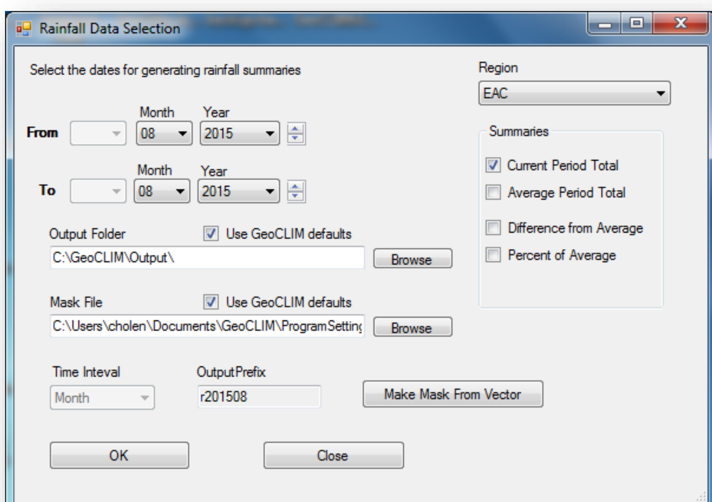


Figure 1.8 *The rainfall summaries tool facilitates the calcualtion of rainfall total and anomalies for a selected period of time.*

1.8. Climate Composites

The Climate Composites facilitates the seasonal analysis among a group or two of non-consecutive years. The tool calculates seasonal average from a group of years or compares the seasonal rainfall performance among two groups of years. See Figure 1.9. See chapter 7 for more details.

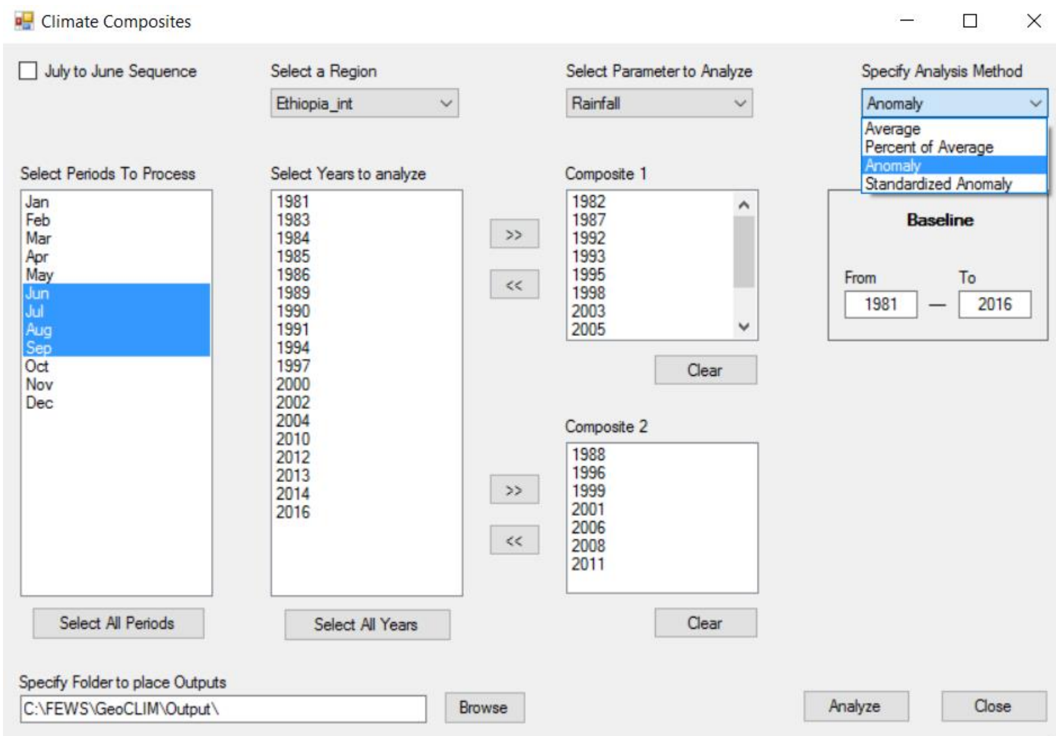
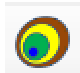


Figure 1. 9 *The Climate Composites tool facilitates the seasonal analysis for one or among two groups of years*.

1.9. Make Contours 

The *Make Gridded Contours* tool (Figure 1.10) allows the display of smoothed contours for a specified interval based on a raster BIL file. This tool is useful for displaying the change of area covered by a contour (of a climate variable) from one period to another. See more about making contours in chapter 10. The contours displayed in this tool are color-coded for ease of display.

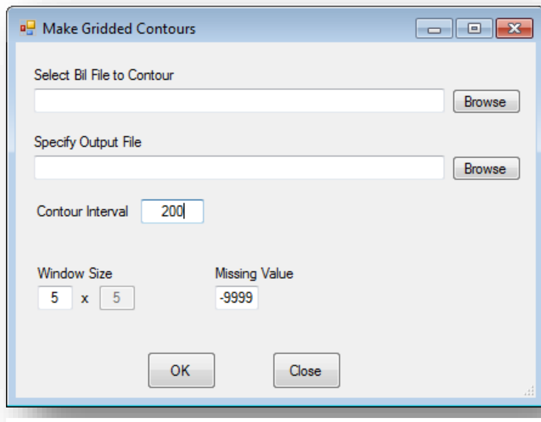


Figure 1.10 *Display rainfall data based on contour intervals.*

1.10. Calculate Long-Term Changes in Average 

The *Climate Trends - Changes in Averages* tool (Figure 1.11) analyzes how the average rainfall, temperature or evapotranspiration for a given season (a set of months) changes between two historical time periods (denoted as *Series 1* and *Series 2* in Figure 1.11). For example, the user can select a set of months and choose range of years (**From** and **To**) for **Series 1** and **Series 2** to ascertain how much the average rainfall or temperature has shifted over those two time periods. See chapter 9 for more details.

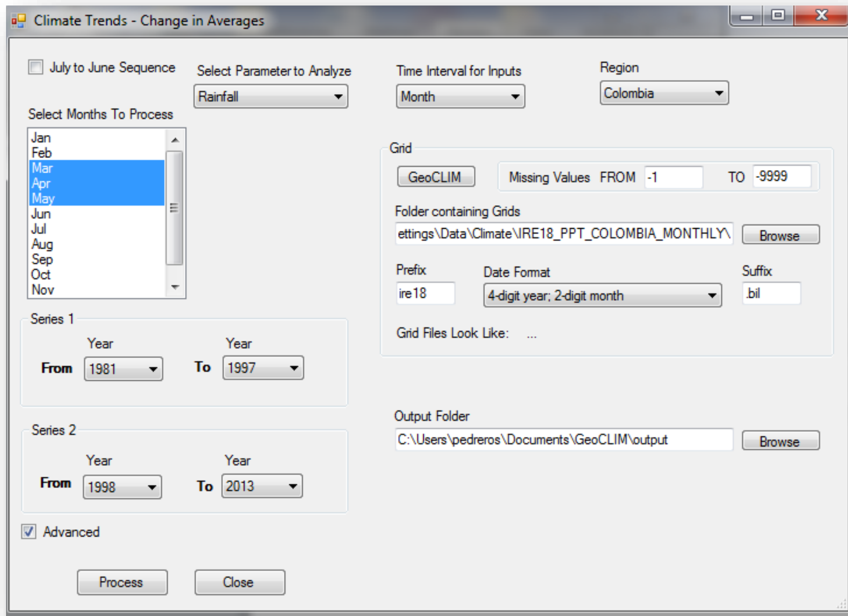
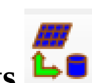


Figure 1.11 *The change in average tool compares the average rainfall for two periods of time, identifying trends.*

1.11. Batch Assistant for Easily Developing Automation Scripts 

The *Batch Assistant* tool, see Figure 1.12, helps to create automated scripts for frequently run processes. The tool contains functions that enable the generation of climate surfaces by interpolating station data, as well as analysis of the relationship between station data and raster datasets. The tool contains the following modules: for more information see chapter 10.

- **Blend rasters/grids with stations**: This function blends raster (e.g., satellite data, etc.) with stations available for a specific period to create a new, improved dataset using the **B**ackground **A**ssisted **S**tation **I**nterpolation for **I**mproved **C**limate **S**urfaces (BASIICS) module
- **Validate Satellite Rainfall**: Validates a raster dataset using station data by comparing the point-to-pixel value for each station. This indicates how much the station and the raster values differ.
- **Interpolate just stations**: This function uses a modified inverse distance weighting process to interpolate the stations available. In this process there is no raster data involved. See chapter 10 for more information.

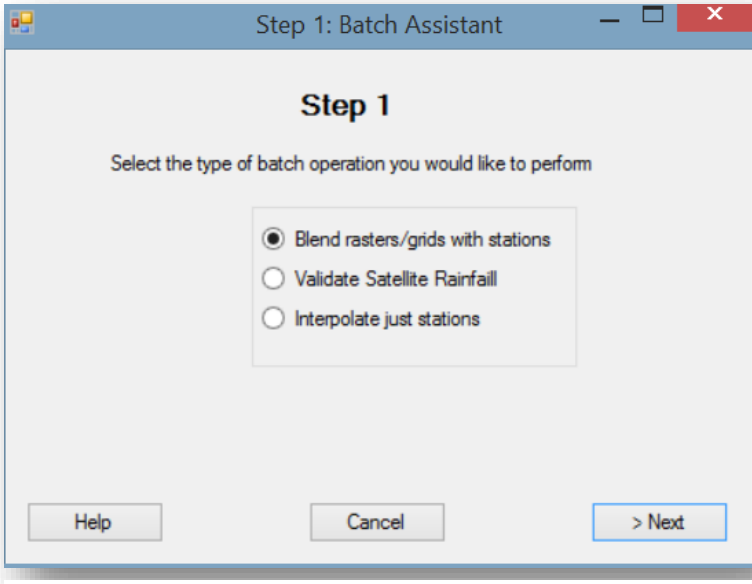


Figure 1.12 The Batch Assistant tool has functions that validate satellite estimated data using station values, blends station data with raster (BASIICS) and interpolates station data.

1.12. Batch Editor for Editing Automation Scripts

The *Batch Editor* allows the user to manually change the code that has been developed using the Batch Assistant (section 1.10) in the batch file to run a specific funtion. See Figure 1.13. See chapter 10 for more information.

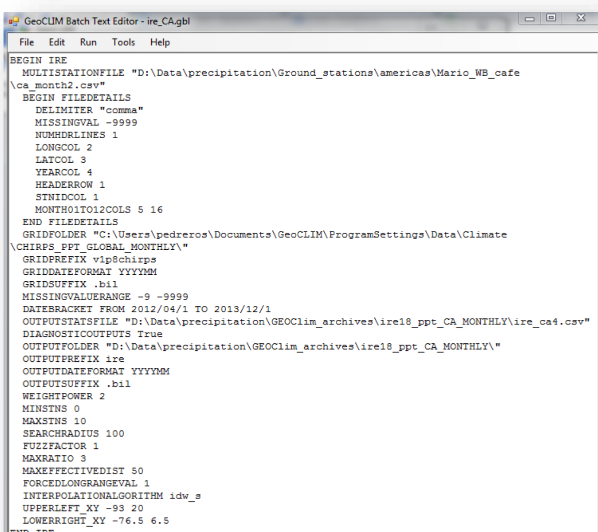


Figure 1.13 The BASIICS process could be re-run using the batch assitant tool.

1.13. Spatial Data Viewer 

The *Spatial Data Viewer* facilitates the display of raster and vector data, editing of legends, digitizing regions, and editing raster data by "painting" over the raster with a newly defined value. See Figure 1.14. See the [Spatial Data Viewer] section for more details on this tool.

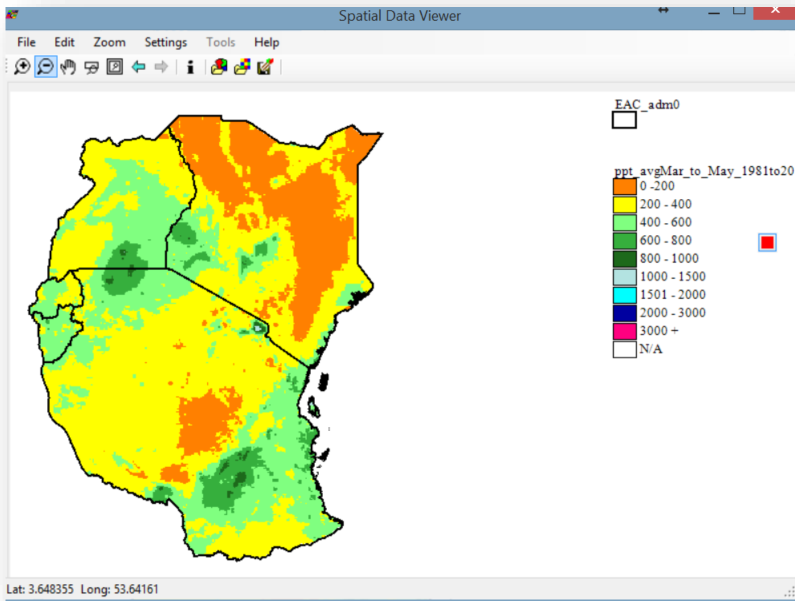


Figure 1.14 *The spatial data viewer allows the display and editing of raster data and shape files.*

1.14. Extract Statistics from Raster Datasets 

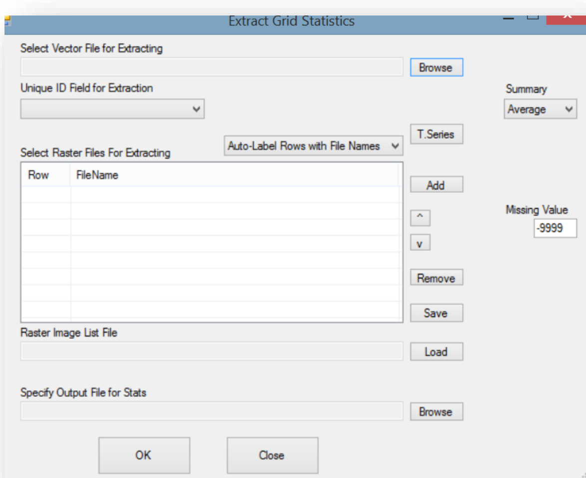


Figure 1.15 *The user could extract spatially statistics based on polygons.*

The *Extract Grid Statistics* tool produces a table with areal statistics such as averages, totals, or max/min for raster data over a set of polygons. See Figure 1.15. For example, the areal average rainfall for each district during the main growing season can be calculated using this tool. For more on this tool, see Chapter 11 for more information.

Chapter 2: Settings

Summary

This chapter focuses on the GeoCLIM settings such as adding new climatological data, changing the workspace, and creating a new area of interest (region) to use with the different analysis tools.

2.1. GeoCLIM Settings 

Once the GeoCLIM program is installed, the user can change settings such as the region of work or the selected climate datasets, by clicking on the  icon from the GeoCLIM toolbar. See Figure 2.1.

2.1.1. Select the **Region** of interest. If the desired region is not in the predefined list, see section 2.6 to learn how to create a new region.

2.1.2. The **Mask** tab allows the user to select a new mask or edit an existing one. Masks are raster maps that are used to include the desired area of work (region) in the analysis, and ignore anything else. Masks typically have a value of 1 over the area of work and values of 0 elsewhere, which helps to make calculations run faster and to provide a map display cleaner.

2.1.3. The **Data** tab, as shown in Figure 2.1, facilitates the selection of available datasets (rainfall, temperature and evapotranspiration) for analysis. The Data tab also allows the user to add new or edit existing climate datasets (see the next two sections).

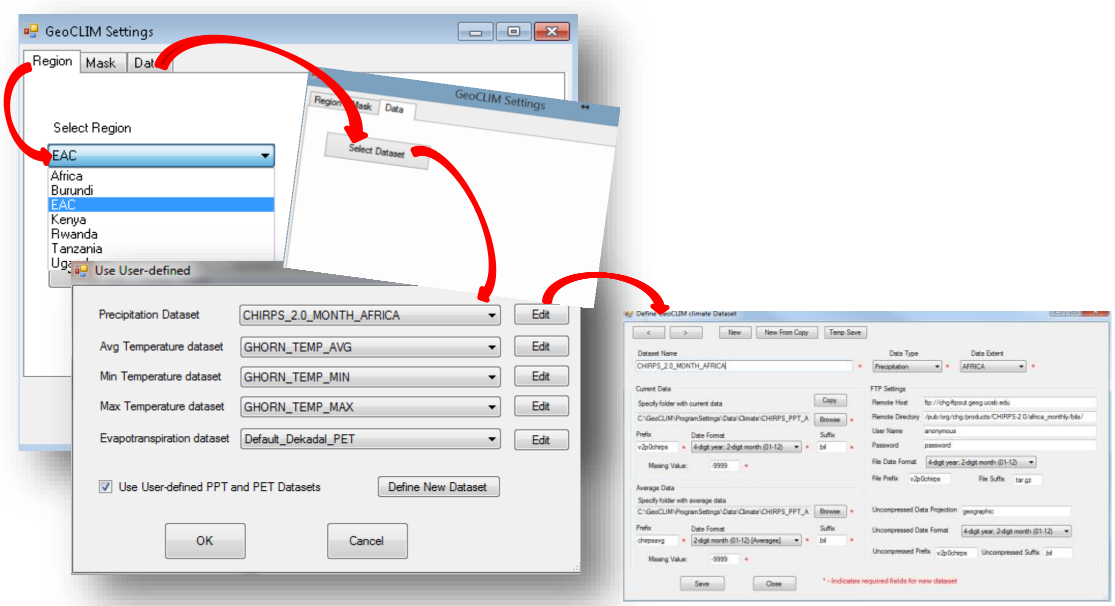


Figure 2.1 The GeoCLIM Settings allow the user to select a new region, a mask and/or climate datasets or/and edit the parameters such as ftp information.

2.2. Making data available for the GeoCLIM

The GeoCLIM uses climate datasets in raster format. Users can use existing datasets (eg. CHIRPS, TRMM, RFE among others) or create their own to use with the GeoCLIM's analysis tools. For example, a new dekadal rainfall dataset could be created in-house by blending gridded satellite rainfall estimates and rainfall station values using the BASIICS tool. For the new dataset to be usable in the GeoCLIM, the dataset's location and file-naming convention need to be defined so that the GeoCLIM can find and read the data.

NOTE: The GeoCLIM directory could be in a location defined by the user. The path to climate data within the GeoCLIM directory is fixed `\GeoCLIM\ProgramSettings\Data\Climate`.

2.2.1. Define climate data filename

The file name in a climate dataset must adhere to the following format in order for it to be readable by GeoCLIM:

<prefix><date-format><suffix>

where

<prefix> is a set of characters before the year (see 1 on Figure 2.2) that could be associated with the dataset name, descriptor or source.

<date-format> the date is composed of the <year><period of time(pentad, dekad or month)>. The GeoCLIM program takes a variety of formats for the date of the data, for example: yyyymm means 4 digits year followed by 2 digit month (01, 02, 03…12). More formats are available for pentadal, decadal, and monthly datasets under the 'define GeoCLIM climate Dataset'. (See 2 on Figure 2.2).

<suffix> is any character after the date, including the extension of the file (e.g., .bil). (See 3 on Figure 2.2)

For example, to name the rainfall total for the 36th dekad of 1991 from CHIRPS 2.0, the recommended name of the BIL file could be *v2p0chirps199136.bil*. In this case, the <prefix> is v2p0chirps, to indicate that it is CHIRPS 2.0 data, the <date-format> comprises of a 4-digit year (1991) and a 2-digit dekad (36), and the <suffix> is the extension of BIL files including the '.' (.bil). To learn more about the data formats used on the GeoCLIM, see Chapter 3

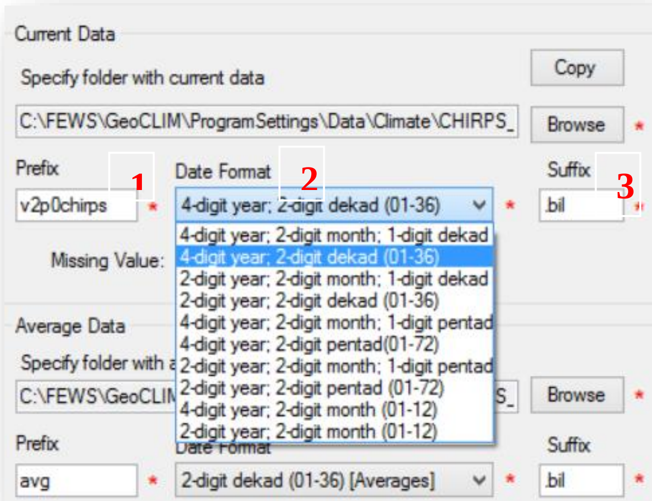


Figure 2.2 *The name of the data file must be composed of a prefix a date (year and period of time) and a suffix.*

2.2.2. Define a new dataset in the GeoCLIM

The climate dataset definition specifies the location of the folder containing the data, the file name (file prefix, date format and file suffix), the missing data value, and where applicable, the download information for the dataset. The climate dataset is defined through the following steps:

1) Create a new sub-directory and copy the BIL raster dataset (*.bil and *.hdr files). This new subdirectory should be under the `\GeoCLIM\ProgramSettings\Data\Climate` directory. In this example, the new subdirectory will called `NEW_PPT_CHIRPS_DEKADAL` and the complete path would be `\GeoCLIM\ProgramSettings\Data\Climate\NEW_PPT_CHIRPS_DEKADAL`

   Once the dataset is ready, the next step is to define it in the GeoCLIM, so that the tool could read the data.

2) Specify the dataset information in the dataset definition form. There are three ways to get to the dataset definition form,

   - Option 1: From the GeoCLIM menu, go to **File** > **New** > **Dataset definition** to create a new GeoCLIM climate dataset definition, see (1) on Figure 2.3.

   - Option 1b: On the GeoCLIM toolbar, click the **Settings** spanner button, then choose the **Data** tab **>** click the **Select Dataset** button **>** and click on **Define new Dataset**, see (1b) on Figure 2.3.

- Option 1c: Open the definition of an existing dataset by clicking on the ‘edit’ button see (1c) on Figure 2.3 and make the changes to reflect the new dataset. Then save the form with a new name.

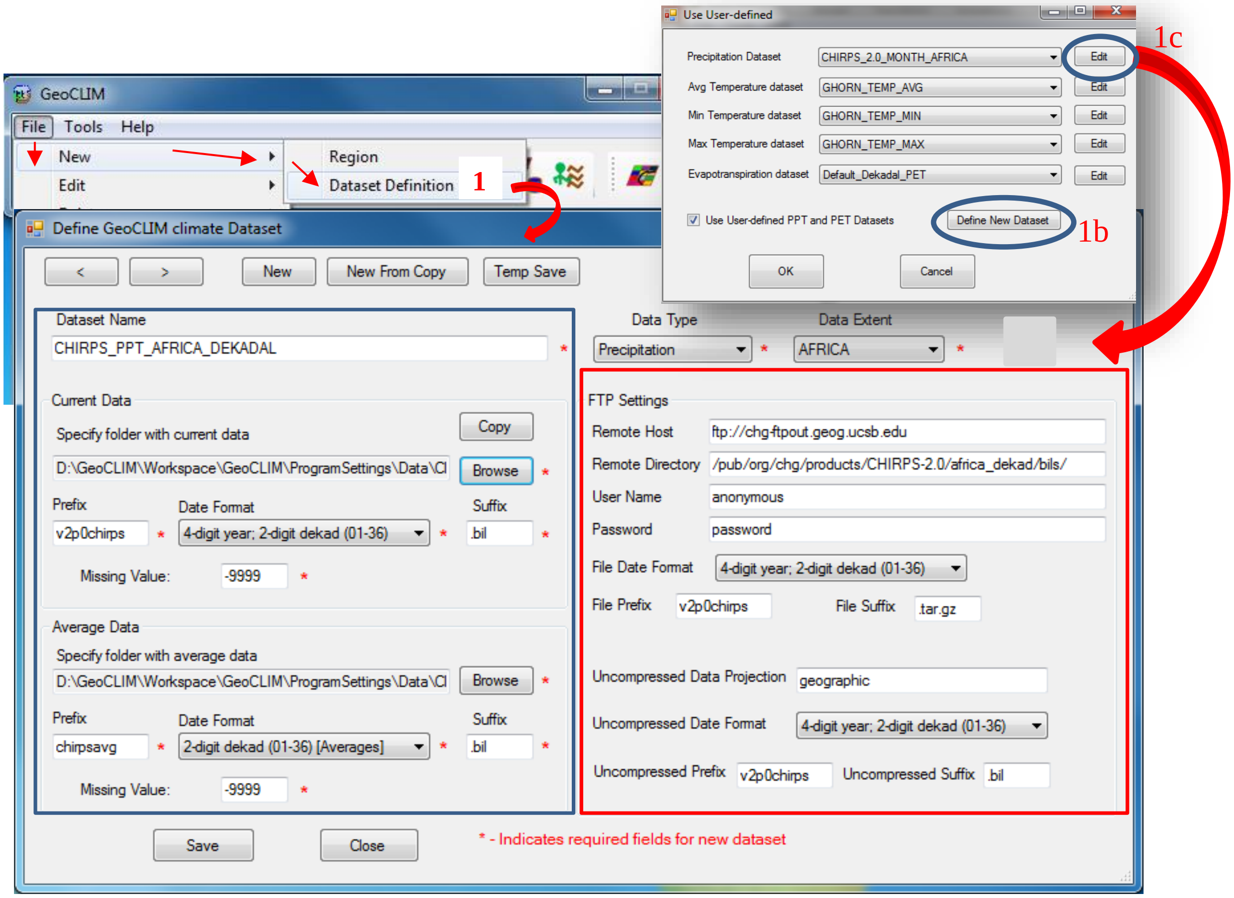


__Figure 2.3__ There are different ways to get to the Data definition form in the GeoCLIM: 1) File>New> Data Definition, or 1b) Settings>Data>Select Dataset>Define New Dataset.

Figure 2.3 shows a completed data definition form. The left side of the form, in blue box, defines the local aspects of the dataset: its name, the path to the data directory, the file-naming convention (prefix, date format and suffix), the missing (NoData) value and the location and format of the dataset long-term averages. It is recommended that the average files be located in the same directory with the data. The right side of the form, in red box, defines the ftp settings that enable data updates to be downloaded for the dataset, where appropriate.

3) **Dataset Name:** type the name of the dataset – it should be the same as the name of the new directory containing the data, `NEW_PPT_CHIRPS_DEKADAL`, as describes on (2).

4) **Select the data Type:** select the type according to the data; Precipitation, Avg Temperature, Min Temperature, Max Temperature, or Evapotranspiration**.**

5) **Select the Data Extent:** there are only three data extents on the current GeoCLIM version: Africa, Central America and Global. If data is not for a location in Africa or Central America, select *Global*

6) **Specify folder with current data:** Browse to the new directory `\GeoCLIM\Programsettings\Data\Climate\NEW_PPT_CHIRPS_DEKADAL`

7) **Prefix:** Fill out the prefix as defined above.

8) **Select the date-format.** For this example, select *4-digit year; 2-digit dekad (01-36)*

9) Fill out the **suffix** *.bil*. NOTE: do not forget the '.'

10) Fill out the **missing value**, for example for CHIRPS data the missing data value used is *-9999.*

11) Click on the **Copy** button below the Dataset name - this copies the data directory path onto the average section to ensure that the averages are located in the same directory as the data files.

    Next, fill the information for the average data. This information is used when calculating GeoCLIM averages or comparing actual data with long term averages.

12) Give a **prefix** to the average files, for example: *avgppt*.

13) Fill out the **date-format.** For this example, select *2-digit dekad (01-36) [Averages]*

14) The **suffix** is the same *.bil*.

15) The **missing value** should be the same as defined for the current data.

16) **FTP Settings** contains the necessary information that enable data updates to be downloaded for the dataset. If the data does not have FTP information, this section can be empty.

Once the form is completed, click **Save** and select the correct choice. In this example, since a new dataset is created, select **YES**, and then **Yes** to confirm.

To start using the new dataset, go to the **GeoCLIM settings** icon > **Data** > **Select Dataset** and select the new dataset from the Precipitation Dataset list. Then, click **OK** to close all the windows. Once a dataset is selected on the settings, all the GeoCLIM functions will use it as default.

2.3. Check availability of the data and compatibility with the selected region

After selecting a dataset, click on the *View Available Data* icon  to make sure that the data covers the region selected. See Figure 2.4. If the **Covers Region** column shows **ok**, the user can

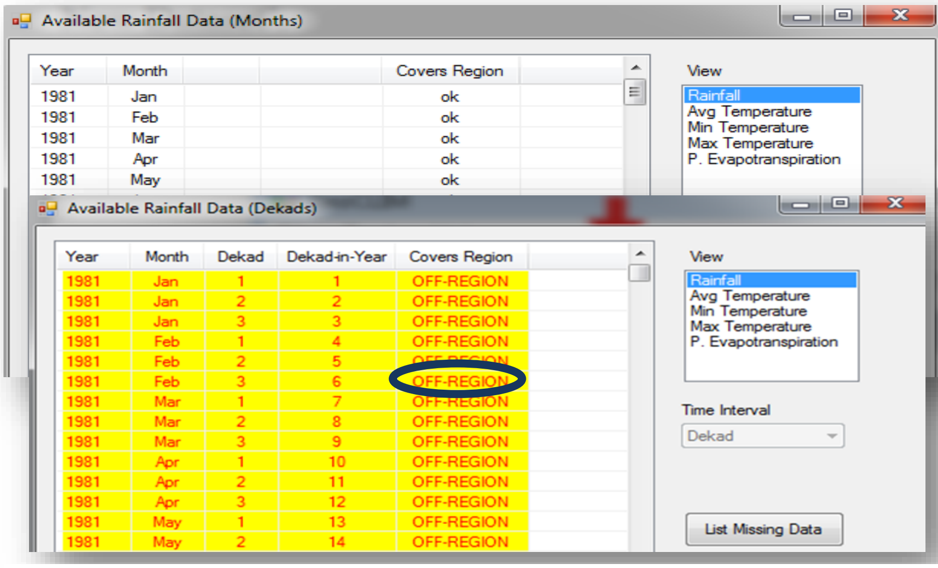


Figure 2.4 The available Data window shows the length of the time series and how well the data extent covers the region of work.

proceed with the analysis functions. If the column shows as **OFF-REGION**, then the data does not cover the region completely and the region will need to be modified, or a different region will need to be selected. To change the region settings, from the GeoCLIM menu, go to **File** > **Edit** > **Region** and make sure that the coordinates of the region are within the domain of the dataset (see section 2.6. Create a new region in the GeoCLIM to learn how to create/edit a region).

The **Available Rainfall Data** tool also allows for the identification of temporal gaps ("missing data") in the dataset between the first and the last date available for the selected climate datasets. Click on the **List Missing Data** button to get a complete list of missing data.

2.4. Review the GeoCLIM Directory Structure

Now let's review the directory structure and data paths in the GeoCLIM. The default directory (in Windows Vista, 7, and 10) is

```
C:\Users\<USER>\Documents\GeoCLIM
```

Where `<USER>` is the Windows username. If the path to workspace was changed, go to new path, see section 2.5 **Change Workspace**. There are two subdirectories in the GeoCLIM folder: `Output` and `ProgramSettings`. The `Output` directory is where all the analysis results will be saved. This default location can be changed in the **Output Options** tool. An outline of the contents of `ProgramSettings` **(1)** is shown on Figure 2.5 and described below:

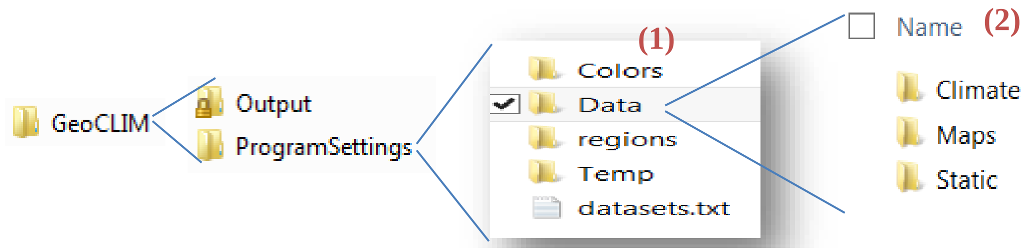


Figure 2.5 The GeoCLIM directory contains two folders; the 'output' where all the results go by default, and the 'ProgramSettings' that contains the 'Data' directory among others.

- **Colors:** Default color files for legends and maps produced by the GeoCLIM.
- **Data:** see (2) on Figure 2.5
  - ***Climate*** - All downloaded and imported data are stored here by default. See the import data section above for information on how to import a new dataset.
  - ***Maps*** - Contains all the shapefiles for the maps of the regions and countries required. Additional shapefiles/maps can be added to this directory.
  - ***Static*** - This directory contains the masks for the different regions. Masks are maps in raster format that are used to define the area of work (region) and ignore the rest of the data. For example the GeoCLIM contains rainfall data for the entire continent of Africa but the analysis is done only for the EAC region. The masks are raster files in BIL format with a value of 1 in the area of interest (e.g. land areas in Africa) and a value of 0 (zero) outside the area of interest.
- **Regions**: This directory contains GeoCLIM region files that define the area to analyze/display. The region files specify the min and max longitude of the area to analyze, the pixel size, the mask file to use for isolating areas of interest, and the shapefile to use when displaying analytical outputs. GeoCLIM Regions are typically countries or regional groupings. New regions can be created as needed. See how to create a new region in section 2.6.
- **Temp:** This directory stores temporary files, such as the downloaded .tar.gz files.

NOTE: *It is strongly recommended that the user becomes familiar with the structure of the* `ProgramSettings` *directory.*

2.5. Change workspace

As was shown above, the default workspace is installed on the `C:\` drive. Sometimes the `C:\` drive is too small to hold all the data and outputs from the GeoCLIM program so it may be necessary to change the workspace. Another benefit of having the workspace on a different path is that if a new version of the GeoCLIM is installed, the workspace can be reused. When the workspace remains on the default path, it gets replaced upon re-installation of the GeoCLIM.

To change the default workspace:

- From the GeoCLIM menu, go to **File** > **Workspace Settings**, see Figure 2.6 **(1a)**.

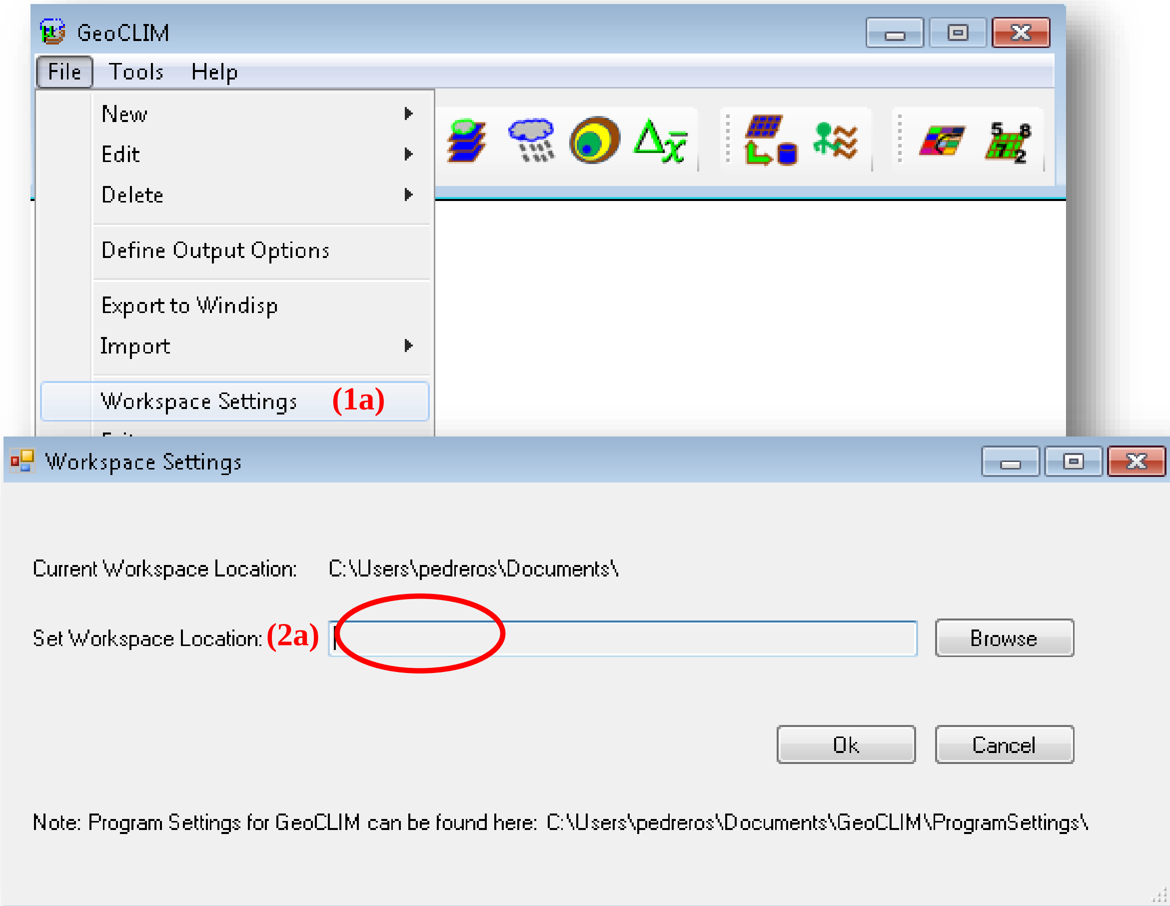


Figure 2.6 The GeoCLIM allows the user to change the workspace directory from the default to a new location.

- Change the path in the **Set Workspace Location** field, see Figure 2.6 **(2a)**.

2.6. Create a new region in the GeoCLIM

There are two ways to create a user-defined region in the GeoCLIM program. The first way is to go to **File** > **New** > **Region** and fill out all the fields. The other way, is to modify an existing region. The second option is outlined below.

1) Open an existing region by clicking the **File** menu item, and navigating to **Edit** > **Region**. See Figure 2.7

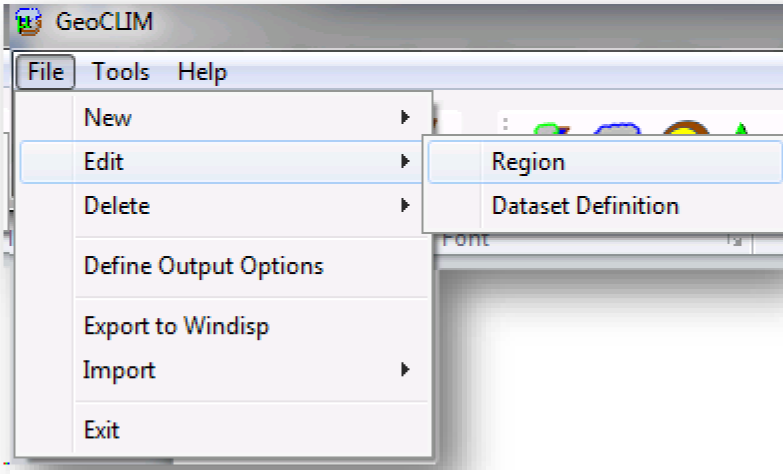


Figure 2.7 *To open an existing region, go to File >Edit>Region*

2) Select an existing region and click **OK**, see Figure 2.8.

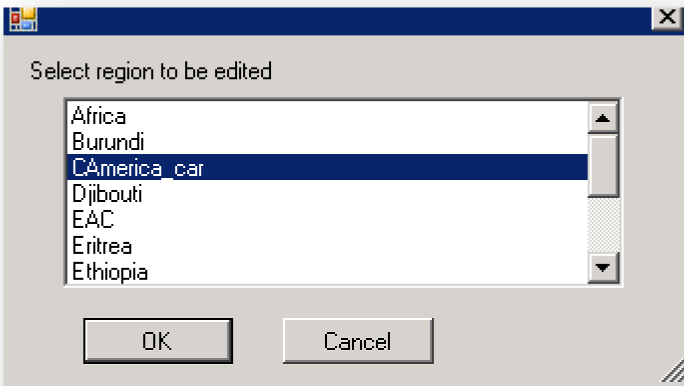


Figure 2.8 *The GeoCLIM contains several regions that could be used as example to create a new one.*

3) Alter the parameters of the existing region to align with the parameters of the new area of interest and click **Save As** to save it as a new region. See Figure 2.9.

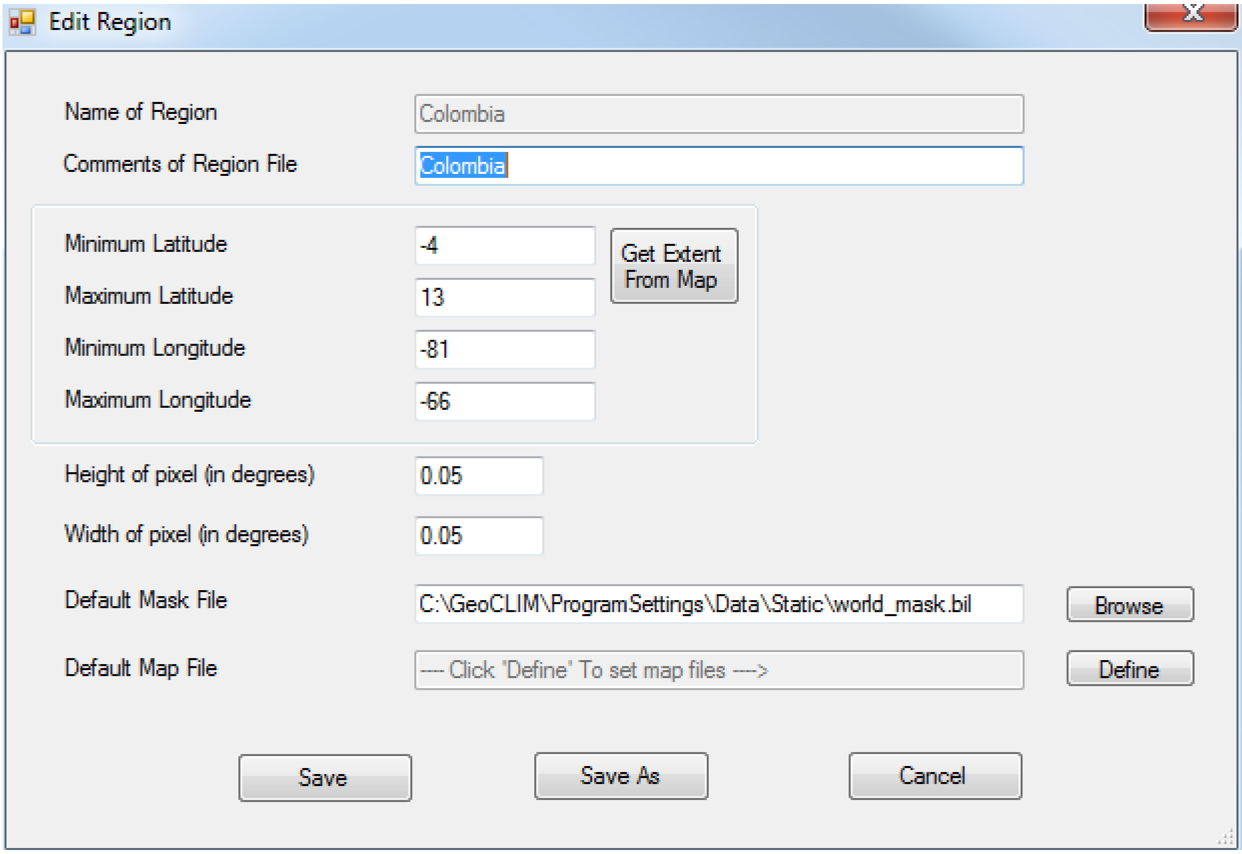


Figure 2.9 *The 'Edit Region' form describes the geographic boundaries of the region, the pixel size for the outputs, the mask used and the outline shape file.*

4) The fields **Height of pixel** and **Width of pixel** refer to the pixel size that will be used in analyzing the data – ideally this should match the pixel size of the source climate data. Create a mask for the new region. A mask is a raster dataset with a '0' value for outside the region of interest and a '1' for the area within the region. In the example on Figure 2.9, a global mask is used but the geographic boundaries of the region limit the output area.

5) Finally, specify a shapefile for the new region in the **Default Map File** field. Click the **Define** button shown in Figure 2.9, then click **Add** as shown in Figure 2.10. This map is used as outline when displaying the results.

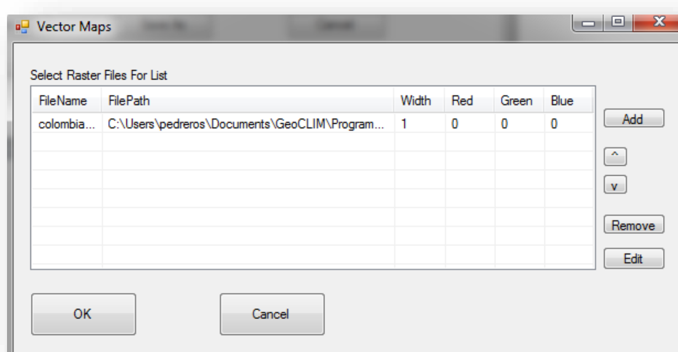


Figure 2.10 *A shape file that serves as the outline on the output products.*

NOTE: One possible problem when creating a new region is that the coordinates could show up as a long number. This issue happens in countries that use ‘,’ to separate decimals. To fix this problem, from Windows, go to **Control panel > Clock, Language, and Regions** and change date, time, or number in the **Formats** tab, then click **Additional Settings**. Make sure that the decimal symbol is ‘.’. See Figure 2.11.

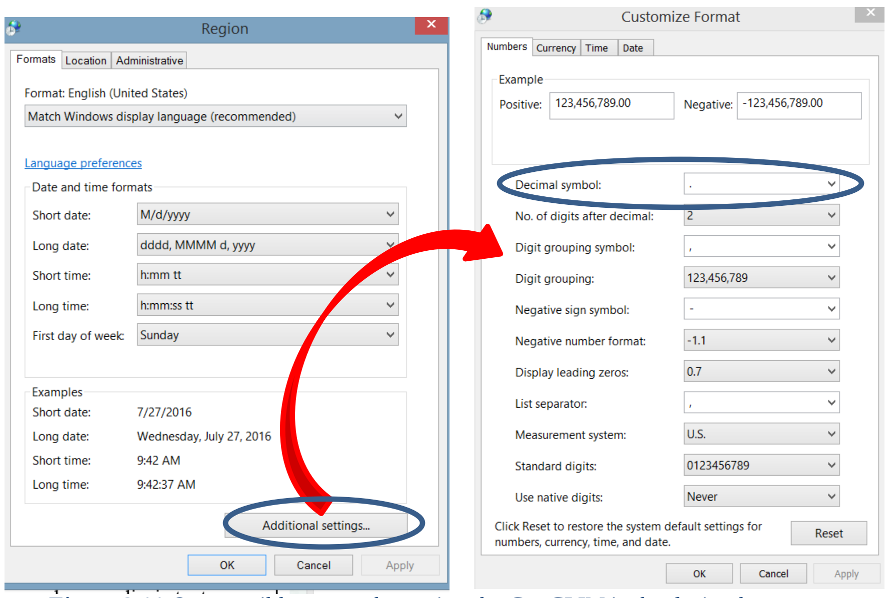


Figure 2.11 *One possible error when using the GeoCLIM is the decimal separator. Go to the Additional settings and change the ‘Decimal symbol’*

Chapter 3: Data Types in the GeoCLIM

Summary

This chapter examines the types of data used in the GeoCLIM program. The GeoCLIM uses three main data types: raster data in BIL format (*.bil), vector data in shapefile format (*.shp), and tables in comma delimited format (*.csv). The Spatial Data Viewer will be used to explore the different types of data.

3.1. Characteristics of the Raster dataset

A band interleaved by line (BIL) dataset contains two main files: a (*.bil) file and a header file (*.hdr). The .bil file contains the data (e.g. rainfall, temperature), while the HDR file has the characteristics of the dataset (geographic information as well as pixel size and depth which determines the range of values a raster dataset can store), see an example of the *.hdr file on Figure 3.1.

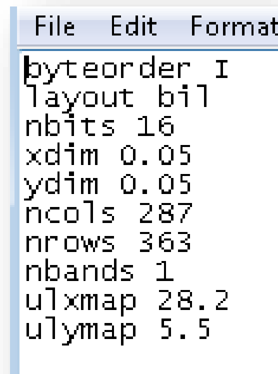


***Figure 3.1** Example of a HDR file.*

To take a closer look, open the HDR file (*.hdr) in a text editor, such as Notepad, see Figure 3.1. The HDR file is a simple file that contains key information such as the number of columns, rows, bits, size of pixel, etc. The `ulxmap` and `ulymap` values indicate the coordinate of the center of the upper left corner pixel of the dataset. In this example, the `xdim` and `ydim` values correspond to horizontal and vertical dimension of a pixel (0.05 degrees is about five kilometers). There are additional keywords that the header could have, see Figure 3.2 (ESRI, help 20010). Sometimes it is important to modify the HDR file so that the data is read correctly by the program.

Keyword	Acceptable Value	Default
nrows	Any integer > 0	None
ncols	Any integer > 0	None
nbands	Any integer > 0	1
nbits	1, 4, 8, 16, 32	8
pixeltype	SIGNEDINT	Unsigned Integer
byteorder	I = Intel; M = Motorola	Same as host machine
layout	bil, bip, bsq	bil
skipbytes	Any integer ≥ 0	0
ulxmap	Any real number	0
ulymap	Any real number	nrows - 1
xdim	Any real number	1
ydim	Any real number	1
bandrowbytes	Any integer > 0	Smallest integer ≥ (ncols x nbits) / 8
totalrowbytes	Any integer > 0	For bil: nbands x bandrowbytes; for bip: smallest integer ≥ (ncols x nbands x nbits) / 8
bandgapbytes	Any integer ≥ 0	0

Figure 3.2 *The HDR file is composed of a series of key words and their respective accepted values.*

One important keyword on the HDR file is the `pixeltype` since it defines the type of value (+ or + and -) a pixel could take. For example, rainfall could only take unsigned (+) values, while temperature could take + and – values. Another important keyword is the `nbits` because it indicates the number of bits per pixel in the dataset (nbits=10 means 10 bits per pixel) Figure 3.3 (ESRI, Support 2016) shows a list of values a raster data set could take depending on the pixel depth or `nbits` value.

unsigned 1 bit = 0 to 1
unsigned 2 bit = 0 to 4
unsigned 4 bit = 0 to 16
unsigned 8 bit = 0 to 255
signed 8 bit = -128 to 127
unsigned 16 bit = 0 to 65535
signed 16 bit = -32768 to 32767
unsigned 32 bit = 0 to 4294967295
signed 32 bit = -2147483648 to 2147483647
floating point 32 bit = -3.402823466e+38 to 3.402823466e+38

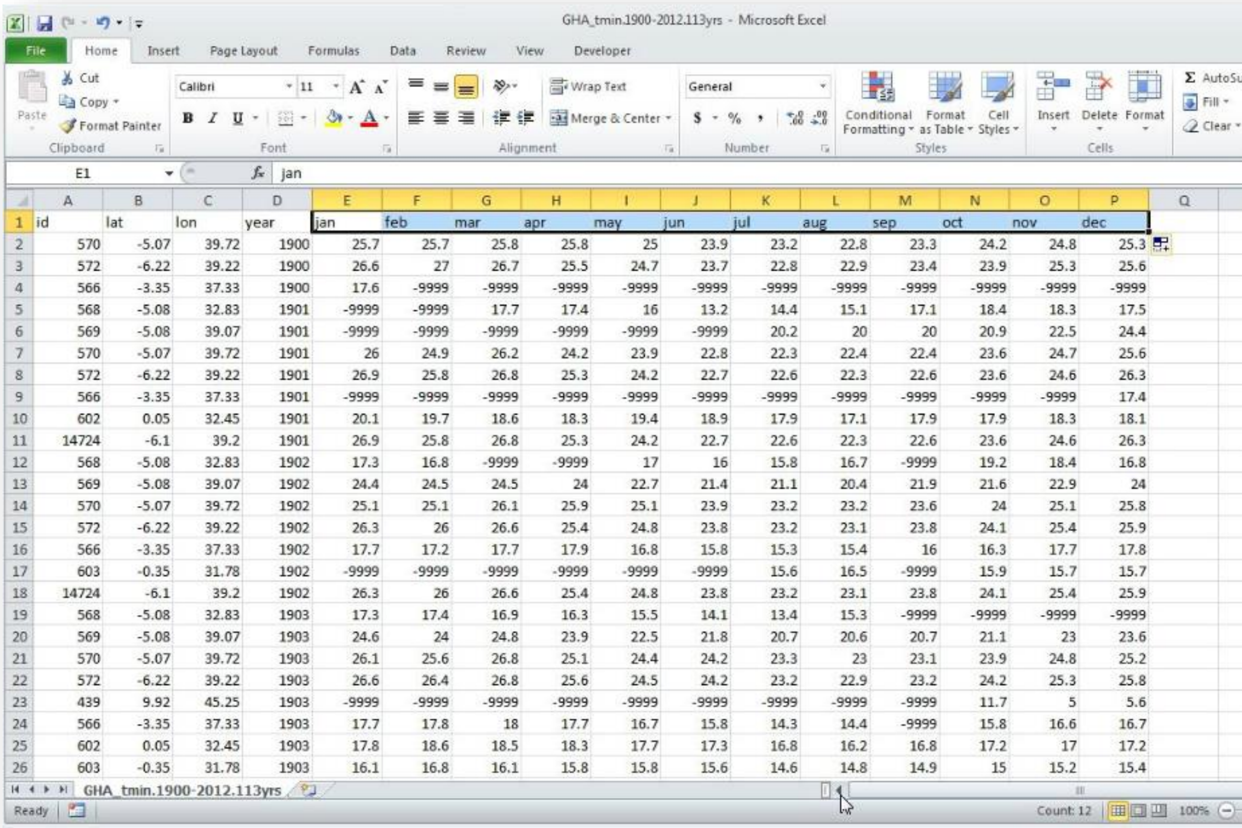
Figure 3.3 *The range of values a dataset could store depends on the nbits*

3.2. Vector data: shapefiles (*.shp)

Another type of data used in GeoCLIM are vector data in shapefile format (*.shp). The current GeoCLIM version (1.2) only allows access to polygons. To open a shapefile file follow the steps below, to create or edit shapefiles go to chapter 4.

3.3. Tables (*.csv)

The GeoCLIM program also uses tables in comma delimited format (*.csv). These tables are used in the process of blending raster data with station values or to validate raster data. For the blending process, the CSV files are required to have columns for *ID*, *lat*, *long*, *year*, *and time period* (pentads, dekads or months) – such as the months of January-December in the Figure 3.4. The metadata columns do not have to be in any particular order and additional columns are permitted. However, the time-period columns need to be consecutive as shown on Figure 3.4.



id	lat	lon	year	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
570	-5.07	39.72	1900	25.7	25.7	25.8	25.8	25	23.9	23.2	22.8	23.3	24.2	24.8	25.3
572	-6.22	39.22	1900	26.6	27	26.7	25.5	24.7	23.7	22.8	22.9	23.4	23.9	25.3	25.6
566	-3.35	37.33	1900	17.6	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999
568	-5.08	32.83	1901	-9999	-9999	17.7	17.4	16	13.2	14.4	15.1	17.1	18.4	18.3	17.5
569	-5.08	39.07	1901	-9999	-9999	-9999	-9999	-9999	-9999	20.2	20	20	20.9	22.5	24.4
570	-5.07	39.72	1901	26	24.9	26.2	24.2	23.9	22.8	22.3	22.4	22.4	23.6	24.7	25.6
572	-6.22	39.22	1901	26.9	25.8	26.8	25.3	24.2	22.7	22.6	22.3	22.6	23.6	24.6	26.3
566	-3.35	37.33	1901	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	17.4
602	0.05	32.45	1901	20.1	19.7	18.6	18.3	19.4	18.9	17.9	17.1	17.9	17.9	18.3	18.1
14724	-6.1	39.2	1901	26.9	25.8	26.8	25.3	24.2	22.7	22.6	22.3	22.6	23.6	24.6	26.3
568	-5.08	32.83	1902	17.3	16.8	-9999	-9999	17	16	15.8	16.7	-9999	19.2	18.4	16.8
569	-5.08	39.07	1902	24.4	24.5	24.5	24	22.7	21.4	21.1	20.4	21.9	21.6	22.9	24
570	-5.07	39.72	1902	25.1	25.1	26.1	25.9	25.1	23.9	23.2	23.2	23.6	24	25.1	25.8
572	-6.22	39.22	1902	26.3	26	26.6	25.4	24.8	23.8	23.2	23.1	23.8	24.1	25.4	25.9
566	-3.35	37.33	1902	17.7	17.2	17.7	17.9	16.8	15.8	15.3	15.4	16	16.3	17.7	17.8
603	-0.35	31.78	1902	-9999	-9999	-9999	-9999	-9999	-9999	15.6	16.5	-9999	15.9	15.7	15.7
14724	-6.1	39.2	1902	26.3	26	26.6	25.4	24.8	23.8	23.2	23.1	23.8	24.1	25.4	25.9
568	-5.08	32.83	1903	17.3	17.4	16.9	16.3	15.5	14.1	13.4	15.3	-9999	-9999	-9999	-9999
569	-5.08	39.07	1903	24.6	24	24.8	23.9	22.5	21.8	20.7	20.6	20.7	21.1	23	23.6
570	-5.07	39.72	1903	26.1	25.6	26.8	25.1	24.4	24.2	23.3	23	23.1	23.9	24.8	25.2
572	-6.22	39.22	1903	26.6	26.4	26.8	25.6	24.5	24.2	23.2	22.9	23.2	24.2	25.3	25.8
439	9.92	45.25	1903	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	11.7	5	5.6
566	-3.35	37.33	1903	17.7	17.8	18	17.7	16.7	15.8	14.3	14.4	-9999	15.8	16.6	16.7
602	0.05	32.45	1903	17.8	18.6	18.5	18.3	17.7	17.3	16.8	16.2	16.8	17.2	17	17.2
603	-0.35	31.78	1903	16.1	16.8	16.1	15.8	15.8	15.6	14.6	14.8	14.9	15	15.2	15.4

Figure 3.4 Tables of station data values are used in Comma delimited tables (CSV).

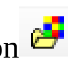
Summary

The spatial data viewer facilitates the visualization of both raster and vector data.

4.1. Displaying a Raster dataset using the GeoCLIM

1) Click on  to open the **Spatial Data Viewer** (Figure 4.1 **(1)**).

Figure 4.1 *The GeoCLIM main tools*

2) Click on the **Open Raster Map** button , as indicated in Figure 4.2 **(2)**, and then select a dataset using the **Select Raster Image** window.

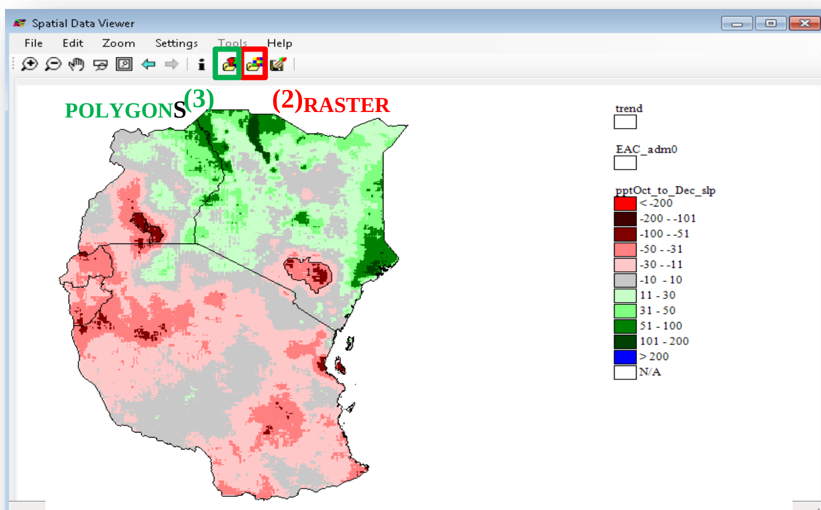


Figure 4.2 *The GeoCLIM allows the user to work with both raster (.bil format) and shape files.*

3) Once in the ***Select Raster Image*** window, click the 'Browse' button next to the **Select Raster File for Display** box to select a raster file (.bil). The file will open in the Spatial Data Viewer. See Figure 4.3a. This raster file shows a value of `-9999` as the 'No Data' value when clicking on the raster legend. When the red square appears, move the mouse over the image. The value for the pixel (i.e., where the mouse is) will show in the lower left corner of the *Spatial Data Viewer*. (Figure 4.3 b)

4) Right click on the legend and select change legend. Navigate to `Documents\GeoCLIM\ProgramSettings\Colors` and select RFE.clr. Once a legend is added, the image displays the ranges of values, see Figure 4.3b.

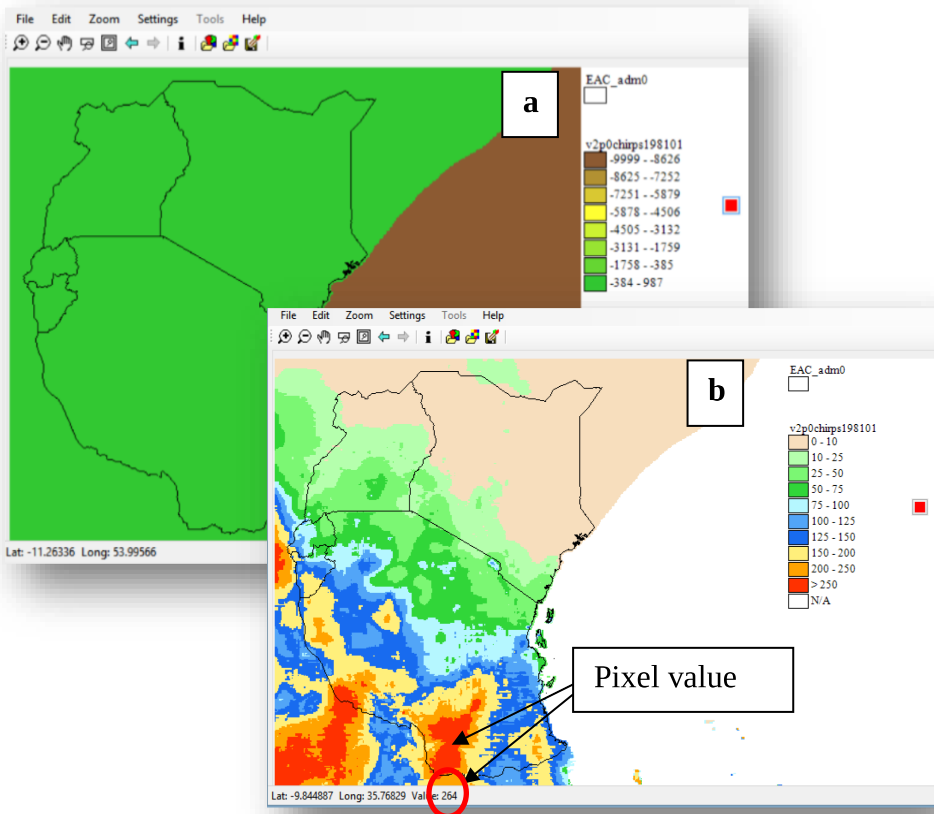


Figure 4.3 *The GeoCLIM allows the use of raster data in .bil format. (a) Shows a raster dataset using the default legend (b) Shows the same raster file after assigning a different legend, and also shows how to view actual pixel values.*

Chapter 5: Climatological Analysis

Summary

The **Climatological Analysis** tool facilitates the calculation of statistics, trends, and frequencies (among others) for rainfall, temperature, and evapotranspiration. The tool uses data that have already been downloaded or imported into GeoCLIM (see chapter 2 for how to make data available in the GeoCLIM) and for regions that are completely covered by the available data. The analysis can be done for an entire time series or just a selected climate subset, such as the March-April-May season for El Niño years.

For example: the user could calculate the annual average rainfall for the period 1981-2015 using the CHIRPS dataset. In addition, the standard deviation of total rainfall over the March-to-May season, from 1981 to 2000, can be also calculated. When dealing with temperature datasets, seasonal averages are calculated rather than seasonal totals. Users can also perform analysis on seasons that start in one year and end in the following year (e.g., the frequency of receiving greater than 500 mm of rainfall for the Dec-Jan-Feb season between the 1991/92 and the 2009/2010 seasons). In cases where a season crosses years, the season is referred to by the year in which the first month of the season falls (for example, Dec-Jan-Feb 2014/2015 is referred to as Dec-Jan-Feb 2014 in GeoCLIM)

The **Climatological Analyses** tools include the following analysis methods:

- Average
- Median
- Standard deviation
- Count
- Coefficient of variation
- Trend
- Percentiles
- Frequency
- Standardized Precipitation Index (SPI)

5.1. Climatological Analysis

1) To access the climatological analysis tool (Figure 5.1), use any of these two options: 1) click the Rainfall Climatological Analysis  icon from the tool bar, or 2) click on the **Tools** dropdown menu from the main GeoCLIM toolbar and select **Rainfall Climatological Analysis**,
2) Make sure that the region of interest is selected from the Region dropdown list, see **(1)** on Figure 5.1. The default region selected is the region set up during first-time run or from the GeoCLIM Settings tool.

3) Select the periods comprising a season of interest on the left panel. The data period (pentads, dekads or months) is based on the selected climate dataset. See **(2)** on Figure 5.1, in this case, the data period is in *Months*.
4) Select the years of interest on the right panel, see **(3)** on Figure 5.1.
5) Check the **Add up seasonal totals** box in the upper left corner see **(4)** on Figure 5.1, to calculate the total of the selected season, for each year.
6) If the season to analyze goes from one year to another, for example Oct to Jan, check the **July to June Sequence** checkbox. See **(5)** on Figure 5.1.
7) See section 5.2 to calculate averages for each period.

NOTE: Make sure the last year selected contains a complete season otherwise there will be a missing data error message and the tool will not run.

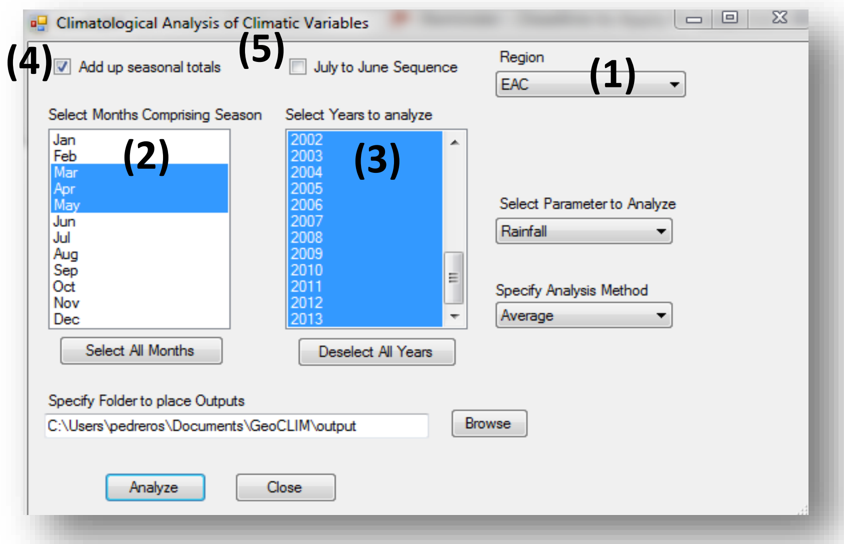


Figure 5.1 *The Climatological Analysis tool facilitates the calculation of statistics, trends, SPI among others, using the complete or part of the time series.*

The outputs from the analysis will be located in the folder assigned in the **Specify Folder to place Outputs** field. There are two products. The main output from this analysis is displayed on the Spatial Data Viewer (see Figure 5.2). This result is also saved in the output folder as a raster image (BIL dataset). Another product is the seasonal total for each year specified in the analysis (see Figure 5.6). This set of seasonal totals could be used to create a time series for any given region of interest. See Chapter 9 for an example of how to use seasonal totals.

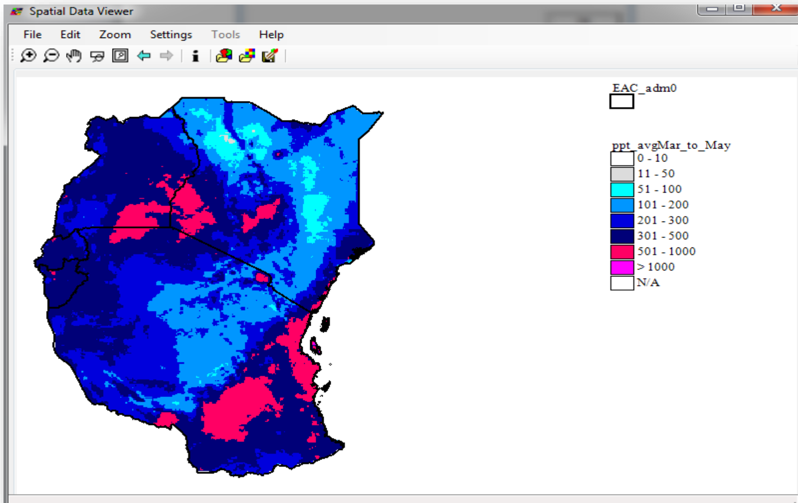


Figure 5.2 *Average rainfall for the period March-May 1981-2013*

When the viewer window is closed, a window message is displayed to provide information of the location of a log file (a text file), see Figure 5.3. This log file is useful as reference of the outputs obtained in the analysis.

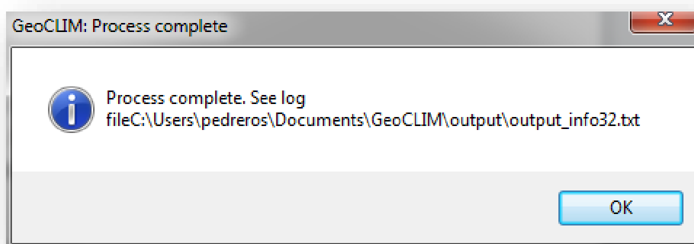


Figure 5.3 *A log file is created after each process.*

NOTE: If multiple periods are selected (March-April-May) and the **Add up seasonal totals** is not selected, the process will be calculated for each individual period (pentad, dekad or month) and there will not be a map display.

5.2. Updating GeoCLIM averages

In order to calculate anomalies, it is necessary to have the long-term average for each period. The GeoCLIM calculates the average for every pentad, dekad, or month when checking the **Update GeoCLIM Averages** box, see (1) on Figure 5.4. The results will be saved in the directory defined from the climate dataset definition (see GeoCLIM Settings chapter 2). Before calculating these averages, make sure the **Add up seasonal totals** box is NOT checked, see (2) on Figure 5.4, so the tool calculates the long term average for every period, Jan_Dek1, Jan_Dek2, and so forth.

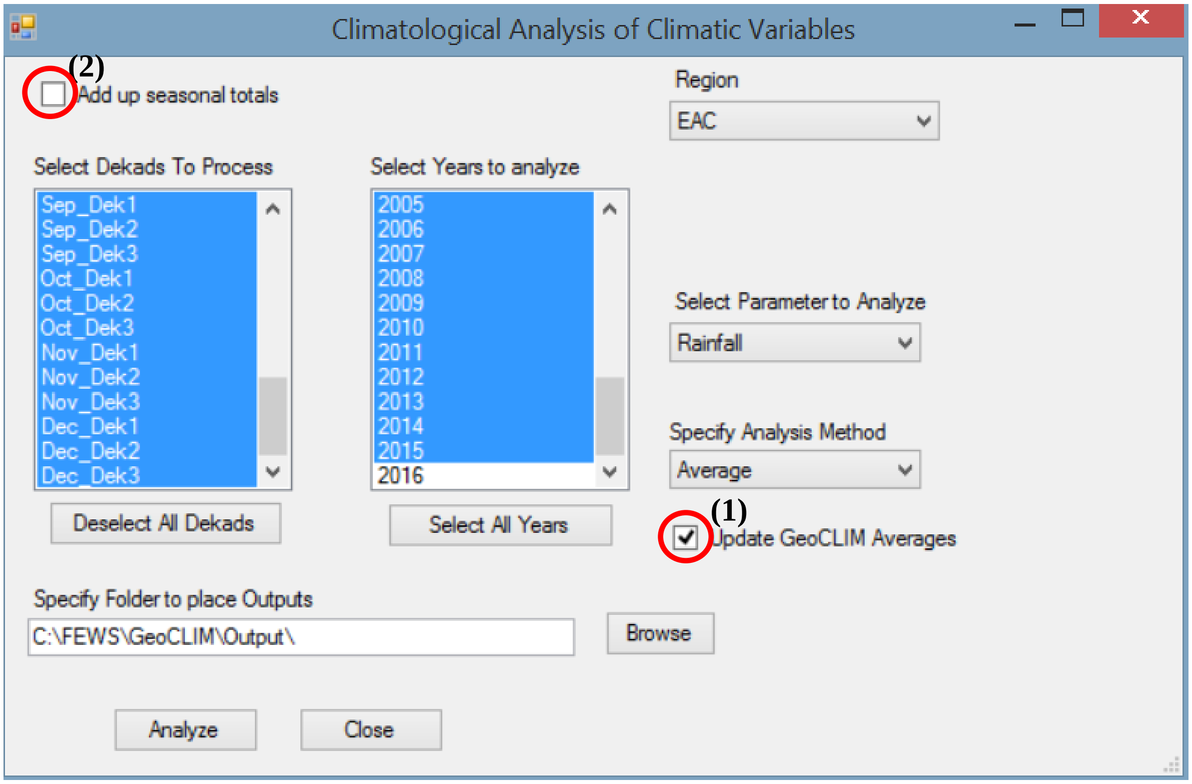


Figure 5.4 *Calculate average rainfall for each pentad, dekad or month using the Climatological analysis tool.*

NOTE: The **Update GeoCLIM Averages** option creates the average for just the selected region extent; it does not calculate it for the extent of the selected climate dataset. For example: if the extent of the climate dataset is Africa, but the region selected in the tool is Kenya, the extent of the average would be for Kenya only. If the region is changed, it would be necessary to calculate these averages again to have the data in the region of interest. It is ideal to select a region that covers the entire dataset before running the **Update GeoCLIM Averages** option.

5.3. Analysis Methods

5.3.1. Average

The ***Average*** analysis method calculates the statistical average value for each pixel for the season or group of periods using all the years selected (see Figure 5.5). Another by-product of this process is the seasonal total, in raster format (.bil) for each year (see Figure 5.6).

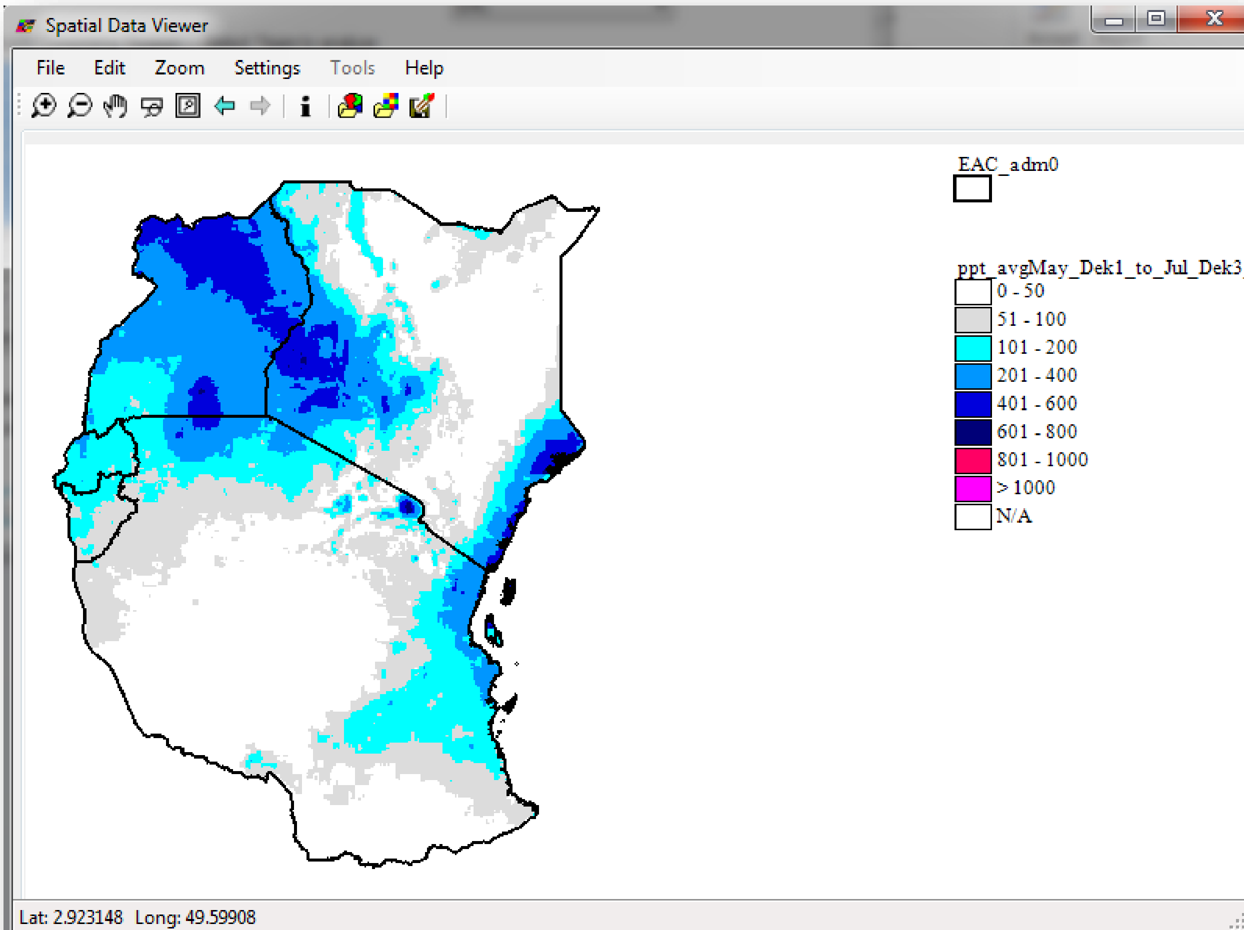


Figure 5.5 *Average rainfall for the May-July season for the years 1981-2013.*

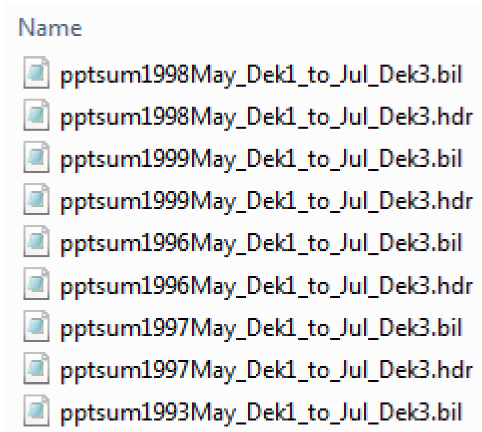


Figure 5.6 *Seasonal totals for each year for the selected period (May dek1 to July dek3) for 1981-2013.*

5.3.2. Median

The ***Median*** analysis method in the tool calculates the midpoint value of a frequency distribution for the selected climate variable. See Figure 5.7 for an example of Median calculation.

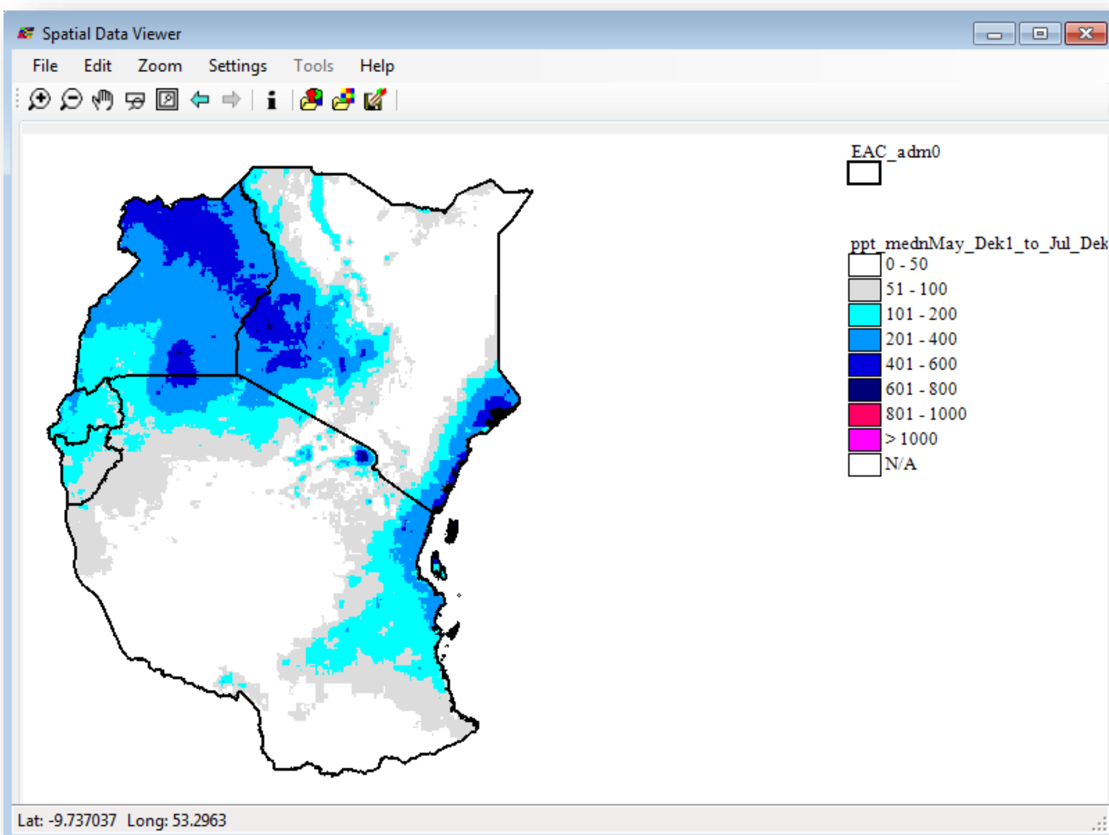


Figure 5.7 *Median for the season May-July for the years 1981-2013.*

5.3.3. Measuring variability with Standard deviation (SD) and Coefficient of variation (CV)

The GeoCLIM provides different methods of estimating variability. The standard deviation shows the variability within the time series, see Figure 5.8 and 5.9, for each pixel while the coefficient of variation shows the SD as percent of average, facilitating the comparison of variability among regions. See Figure 5.10 (a) (b).

5.3.3.1. Standard deviation

Standard deviation (SD) is a measure of variation, or how spread out the data are from the mean (see Figure 5.9).

$$S=\sqrt{\frac{\sum(X-M)^2}{n-1}}$$

© easycalculation.com

Figure 5.8 the standard deviation of a sample dataset.

SD can be used to determine how likely a given deviation will occur in any given year. SD is based on the concept of the bell curve (see Figure 5.9).

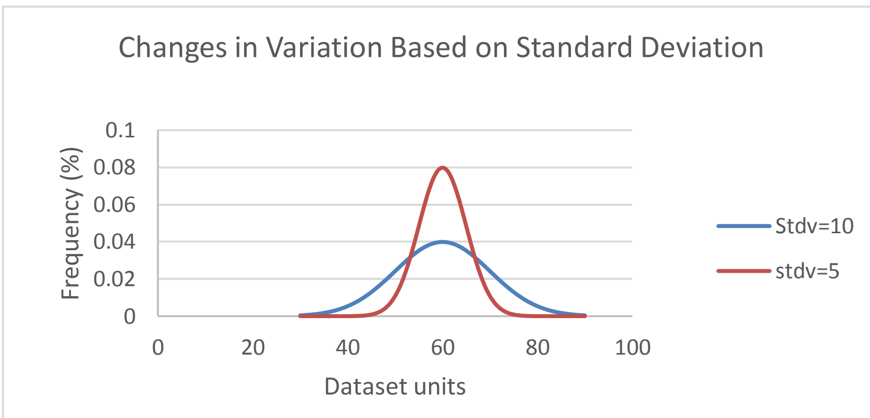


Figure 5.9 The distribution of two datasets with same mean and different SD. The red line shows a low SD indicating low variability within the data; values are closer to the mean. The blue line shows the distribution of a more variable data set.

NOTE: In doing analysis based on the average, it is important to include the SD to indicate how much the values could go above or below the mean.

5.3.3.2 Coefficient of variation.

The CV is the ratio of the SD to the average

CV= (SD/average) *100

SD	Mean	CV
171mm	721mm	24%

CV allows for the comparison among different magnitudes of variation or between regions with different means. See figure 5.10 (b).

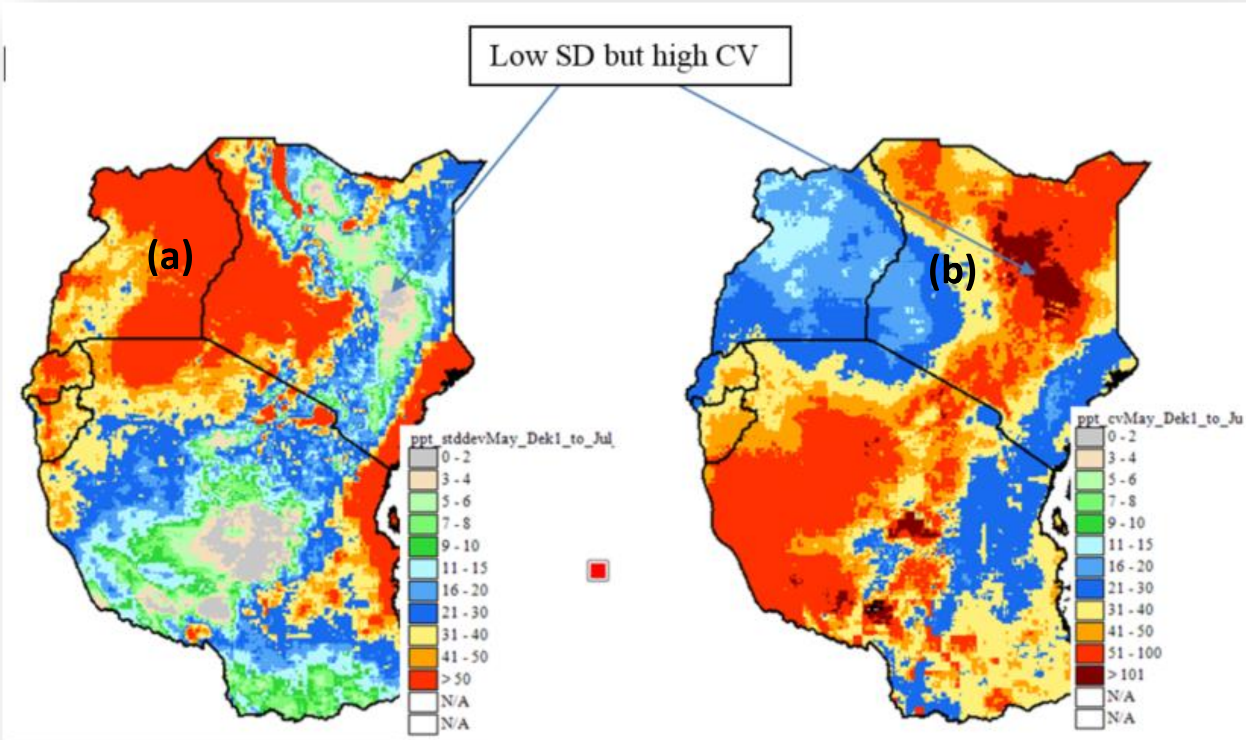


Figure 5.10 *Measuring variation; the SD (a) shows the variation of rainfall in mm, while the CV (b) presents the SD as percent of the average allowing the comparison among regions. The SD of the area pointed by arrow shows a low value but the average is so low that it is a highly variable region, relative to expected rainfall, as shown by the CV.*

5.3.4. Count

The count function on the Climatological Rainfall Analysis tool gives out the number of pixels, in the time series, with value different than the *missing value*, which is typically `-9999` For the example in Figure 5.11, the count is 35 for all pixels since there are no missing data in the time series from the selected climate dataset used in the analysis.

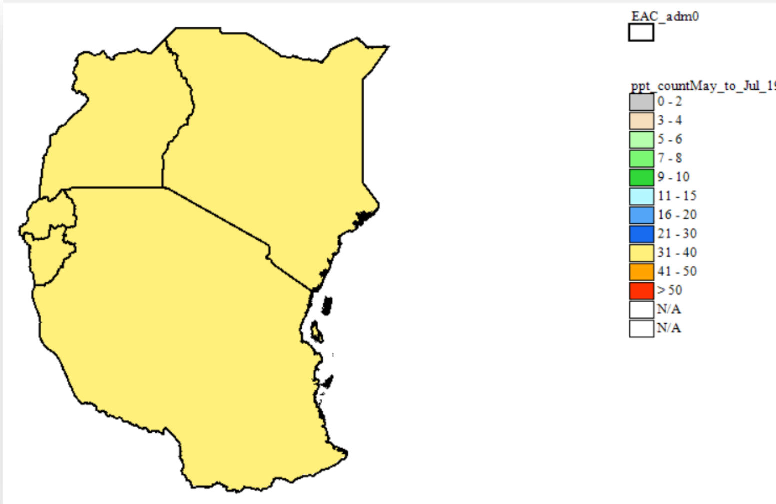


Figure 5.11 The count function looks at the time series for each pixel and counts the number of times there is a valid, non-missing value.

5.3.5. Trend

The trend analysis function calculates the seasonal total rainfall (or seasonal average, in the case of temperature) for each selected year, and then calculates a linear trend line using a regression analysis of the seasonal values.

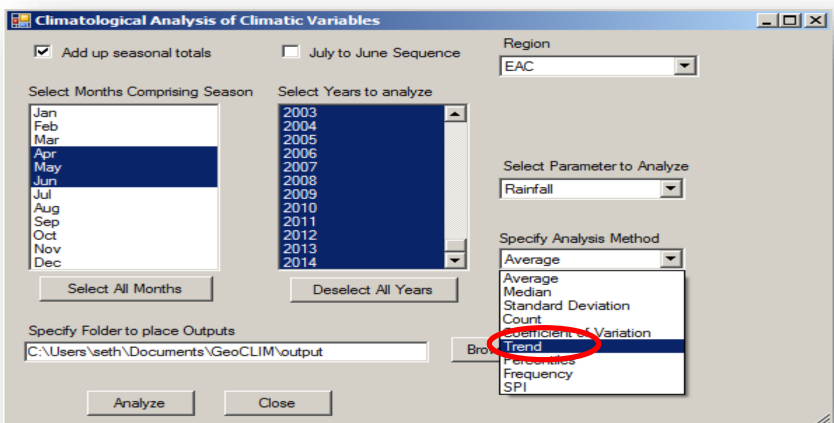


*Figure 5.12 To calculate the trend for a climate variable, select the season, make sure that the **Add up seasonal totals** is checked, and select the years to be used in the calculation.*

To calculate trends:
1) Start the **Climatological Analysis** tool as described above.
2) Select the data to be analyzed. Select the season on the left panel and the years of the time series to be used on the right panel on Figure 5.12.
3) Select *Trend* from the analysis methods menu, as shown in Figure 5.12.

Figure 5.13 shows the annual trend for Ethiopia from 1981-2016. The trend function in the GeoCLIM produces two maps, a map showing the slope of the regression and another one showing the r^2. Figure 5.13(1) shows the slope of the regression line or mm of increasing (green – blue) or decreasing (pink-red) rainfall totals, the legend shows the results per decade (10 years) on average. Figure 5.13(2) shows the coefficient of determination (r-squared, or r^2) of the regression as an indication of the strength of the trend. It is important to use both maps to develop a conclusion about a region. Points a, b and c on Figure 5.13(1), show three trend areas with different r^2. Figure 5.14 presents the scatter plot of the time series with the regression line for each of the three points. Site (a) has a -60mm lost per dacade on Figure 5.13(1) with r^2= 4%, point (b) shows a 121mm gained per decade on Figure 5.13(1) with a r^2 = 15% while site (c) shows -224mm lost per decade with r^2 = 48%.

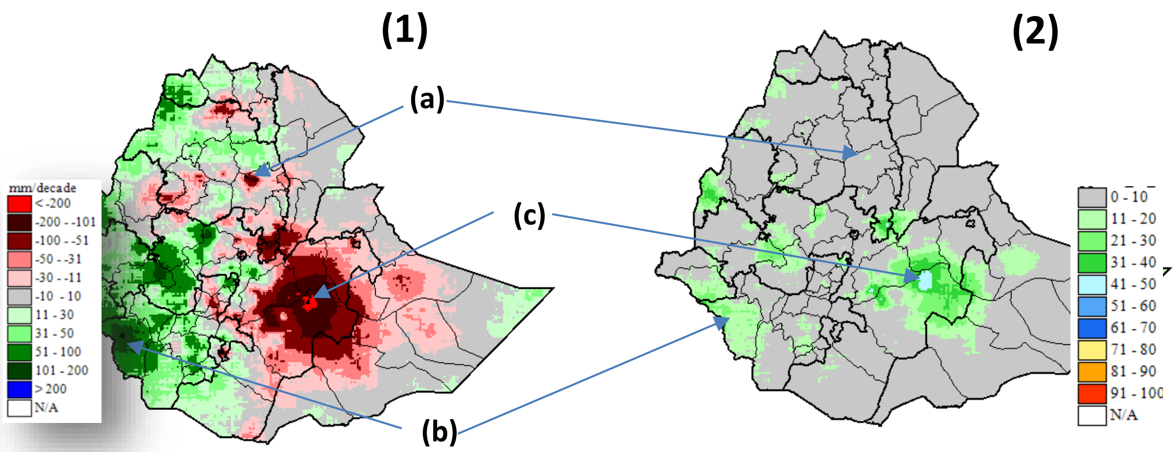


Figure 5.13 *The trend function in the GeoCLIM produces two maps. The map on the left (1) shows the slope of the regression in mm per decade lost (- on red) or gained (+ green-blue). The map on the right (2) shows the r^2 of the regression.*

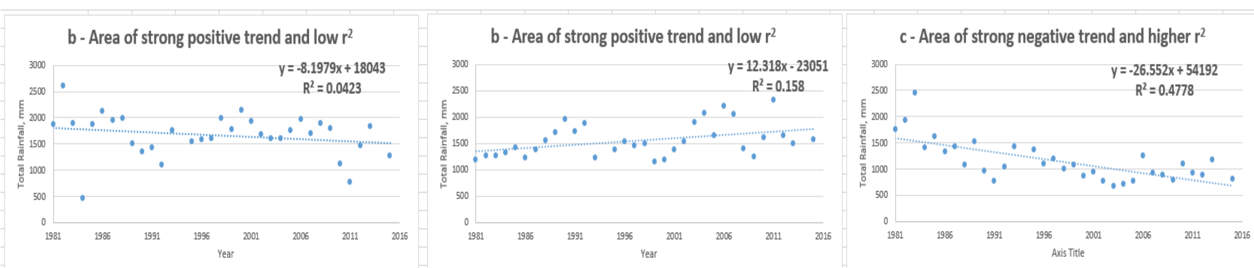


Figure 5.14 It is important to evaluate the strength of the relationship (r^2) before making conclusions about the trend for a region. Plots show three regions that present strong trend on Figure 5.13(1) with different r^2.

5.3.6. Percentiles

Percentiles, in the GeoCLIM, are calculated by expressing a specific value as a percent rank of all the values in the selected series at a given pixel location. If the sample size is large enough (>~30), and there are no extreme values, this method will approximate the true distribution of the rainfall better than the normal distribution, and with a potentially comparable accuracy to the gamma distribution. The geotools use a percentrank scheme of 0 to 100, with the smallest value data point in the series having a percentrank of 0, while the largest value has a percentrank of 100. This is a similar concept to the PERCENTRANK.INC function in Excel.
The Percentile method in the geotools uses a linear interpolation between known points. So for example, if the driest 3 records in a 36 year history are 115mm, 174mm and 178mm, then 115mm is a percentile of 0; 174 is a percentile of 2.9, 150 would be a percentile of 1.7 and 170 would be a percentile of 2.7.

Figure 5.15 presents an example map of rainfall accumulations (mm) corresponding to the 15th percentile rank. This percentile rank defines a set of very low frequency, extreme dry events. A map of 90th percentile accumulations would indicate the values (mm) below which 90% of historical events fall; this percentile rank provides the threshold defining a set of extreme wet event.

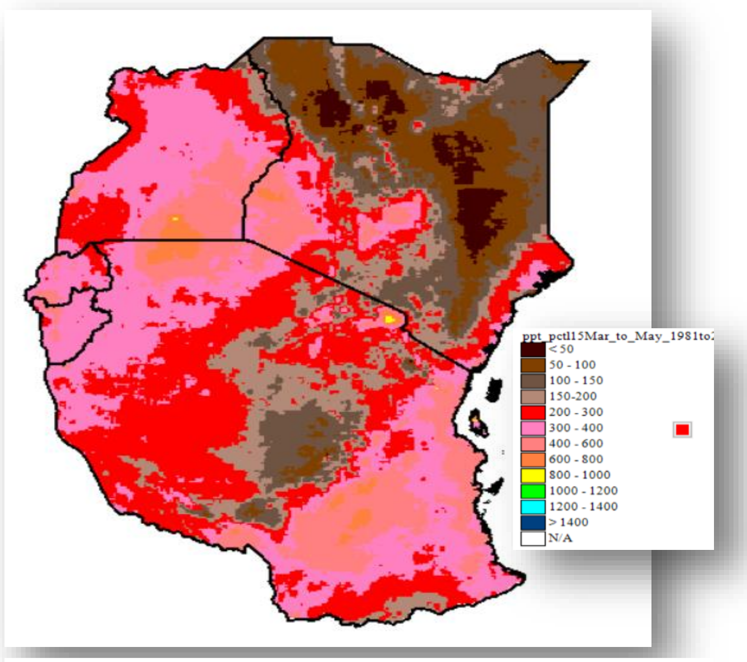


Figure 5.15 *The map shows, by pixel, the rainfall value that is lower or equal to 15% of the values in the frequency distribution. These values represent extreme dry events.*

5.3.7. Frequency

The Frequency analysis method in the GeoCLIM Climatological Analysis module (Figure 5.17) calculates the number of times a range of values has occurred in the time series. The legend of the map on Figure 5.18 is given for every 10 years. The Frequency method helps answer to questions such as "How many times has the total seasonal rainfall been less than 400 mm?" The answer may translate into failure for many crop types.

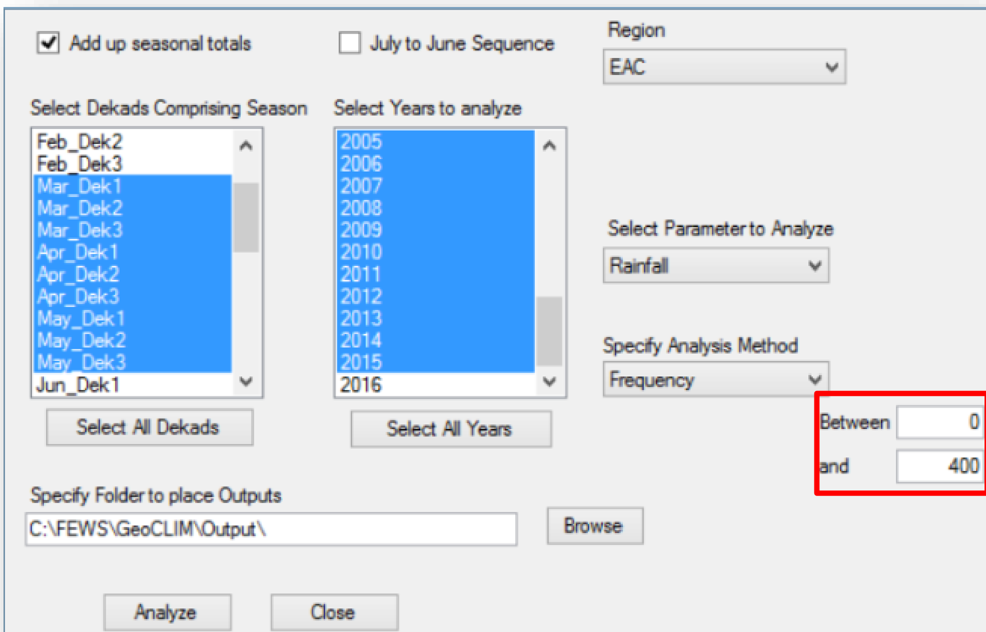


Figure 5.17 *The Frequency function allows for the selection of a range of values (red box) and identify the number of times this range has occurred in the times series.*

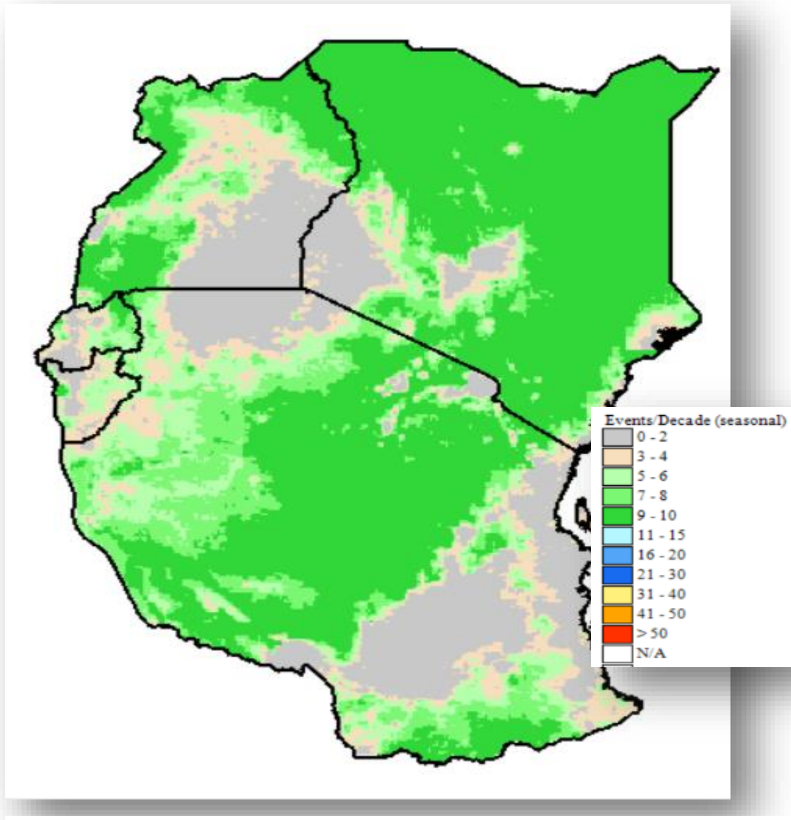


Figure 5.18 *The tool calculates the number of times the selected range of values took place during the time series selected.*

5.3.8. Standardized Precipitation Index (SPI)

The Standardized Precipitation Index presents a rainfall anomaly as a normalized variable that conveys the probabilistic significance of the observed/estimated rainfall (McKee, 1993). By expressing anomalies in terms of their likelihood of occurrence it is easier to evaluate the rarity of the observed event, in the absence of a nuanced understanding of the rainfall regime at a location. This method offers a different, and complementary, perspective compared to anomalies (which can be relatively large, but not very significant in areas with highly variable rainfall) or percent of normal (which can be extreme, but not very significant in dry locations).

To evaluate the likelihood of occurrence, probability distribution functions (PDFs) are fit at each pixel for each accumulation interval. These PDFs are fit to a historical dataset such as CHIRPS (Funk et al., 2016), which provides a 35-year timeseries with which to estimate gamma distribution parameters. The CHIRPS data establishes the shape of the distribution, as well as an estimate of the variance.

SPI values greater than zero indicate conditions wetter than the median, while negative SPI indicate drier than median conditions. For drought analysis, a SPI less than -1.0 indicates that the observation is roughly a one-in-six dry event, and is termed "moderate". A SPI less than -1.5 indicates a one-in-fifteen dry event, and is termed "severe". Values less than -2.0 are typically referred to as "extreme", indicating the event is in the driest 2% of all events (see Figure 5.19). (Early warning USGS-FEWS NET product documentation website)

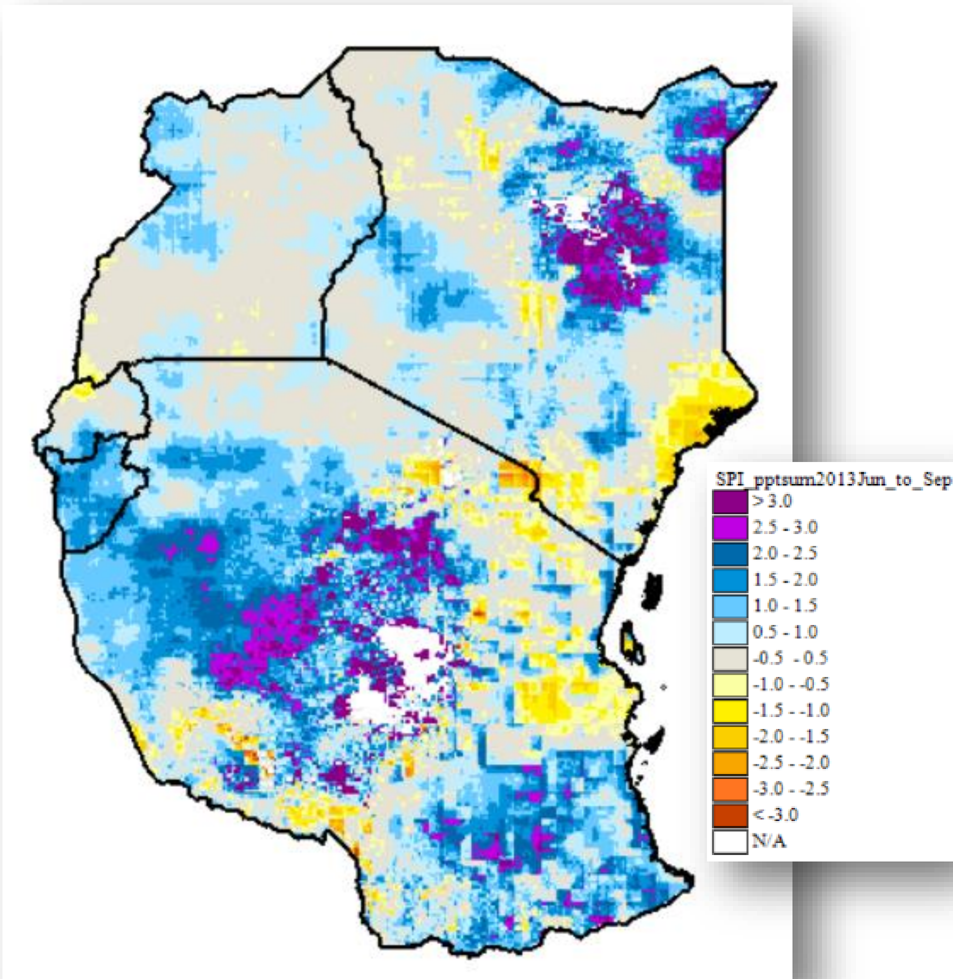


Figure 5.19 *The product of the SPI function in the GeoCLIM, is in units of [SPI * 100], but the legend shows actual SPI values.*

Chapter 6: View and Explore Rainfall Summaries

Summary

The **Rainfall Summaries** tool (Figure 6.2) facilitates the calculation of seasonal totals and anomalies (difference and percent) for any period of time, providing answers to questions like "How different was June-Sep 2012 from average?"

6.1. Requirements

In order to undertake analysis using the **Rainfall Summaries** tool, the dataset must have the GeoCLIM averages available, and the averages should cover the selected region. In case the averages do not exist, an error message will be displayed and the Climatological Analysis of Climatic Variables tool (see Figure 6.1) will open.

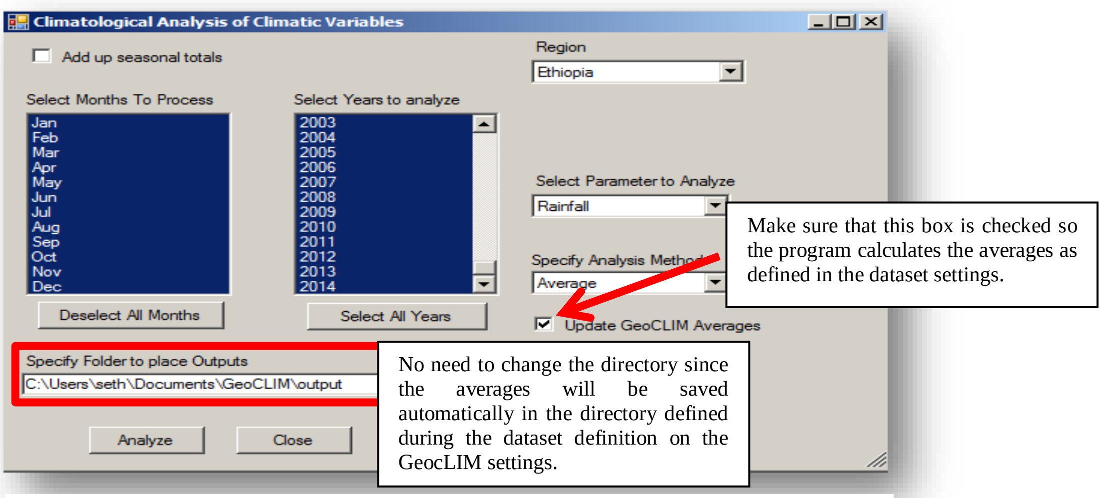


***Figure 6.1** In order to run the rainfall summaries tool, it is necessary to calculate the average for each pentad, dekad or month.*

6.2. Calculate seasonal total and anomalies

Once the averages are calculated, close the Climatological Analysis tool and go back to the **Rainfall Summaries** tool (see Figure 6.2).

1) Select the period of analysis (defined by the period between the **From** date and the **To** date

2) Select the type of summary

   2.1. **Current Period Total:** total rainfall for the selected period

   2.2. **Average Period Total:** Long term average for the selected period

   2.3. **Difference from Average:** (current minus average)

2.4. **Percent of Average:** (current / average)*100

NOTE: to save outputs on a different directory, uncheck the box *Use GeoCLIM defaults* and browse to the new directory.

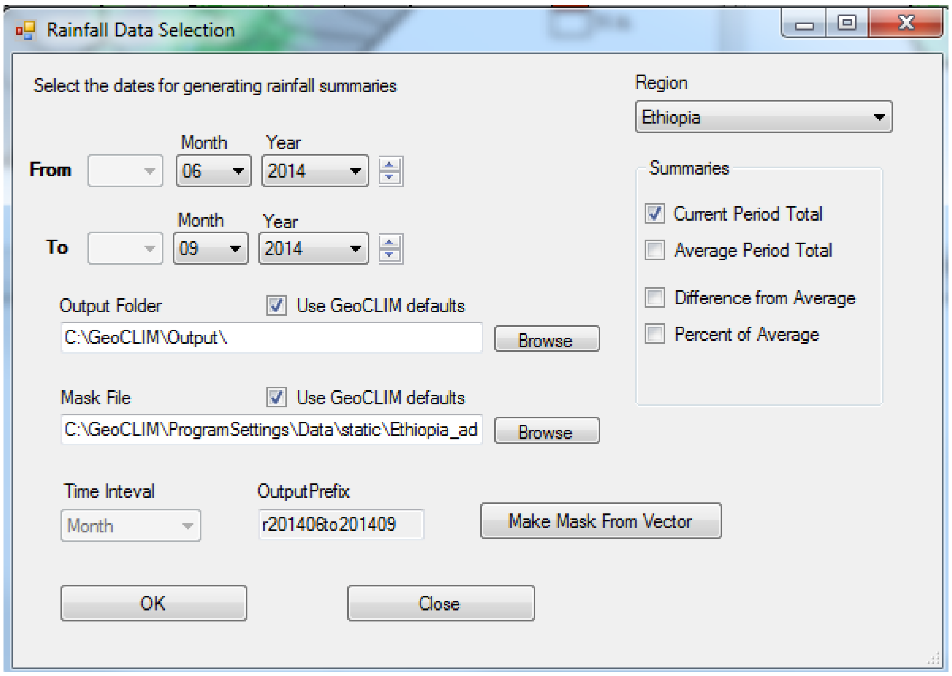


***Figure 6.2** The rainfall summaries tool enables the calculation of seasonal total, averages and anomalies for a specific time period.*

Chapter 7: Climate Composites

Introduction

Sometimes it is important to be able to analyze a season among a group of nonconsecutive years or compare the seasonal rainfall performance among two groups of years. For example: compare the difference in rainfall condition of the MJJ season during El Niño and La Nina years in Central America. The **Climate Composite** tool calculates the seasonal average for a group of years, the percent of average, as well as anomalies or standardized anomalies for one or two groups of years using a baseline average defined by the user.

7.1. Average

Calculate the seasonal average from a subset of years. In this chapter we will be comparing the El Nino/La Nina years from moderate to very strong as defined by the Oceanic Nino Index (ONI). El Nino (1982-83, 1986-87, 1987-88, 1991-92, 1997-1998, 2002-03, 2009-10, 2015-16) and la Nina (1988-89, 1998-99, 1999-00, 2007-08, 2010,11).

1. Select the region and season to be analyze
2. Select the years to be included for composite 1
3. Move the selected years to the *composite 1* box by clicking on the >>

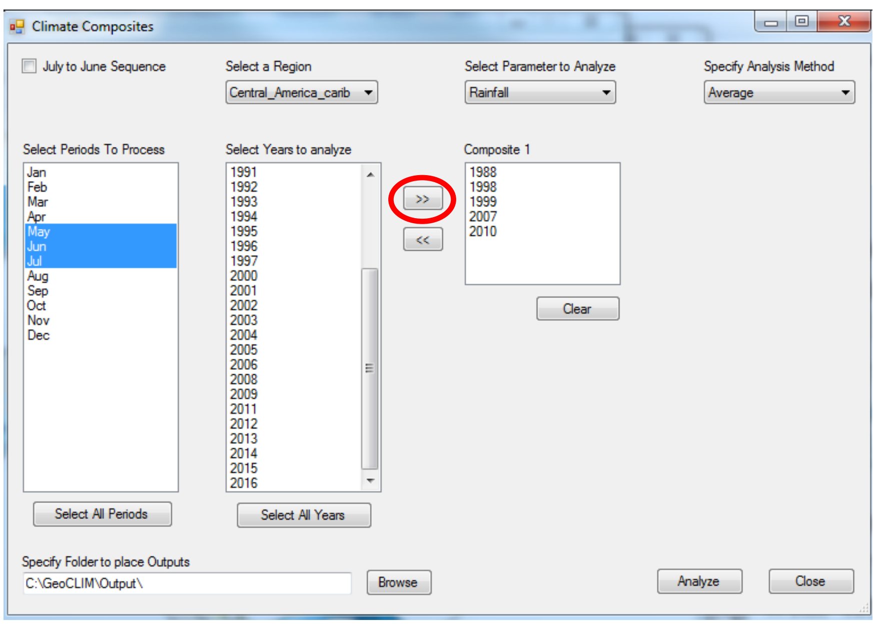


Figure 7. 1 *The tool calculates the seasonal average from a group of years and the output is displayed on the Spatial Data Viewer, and a PNG file is created*.

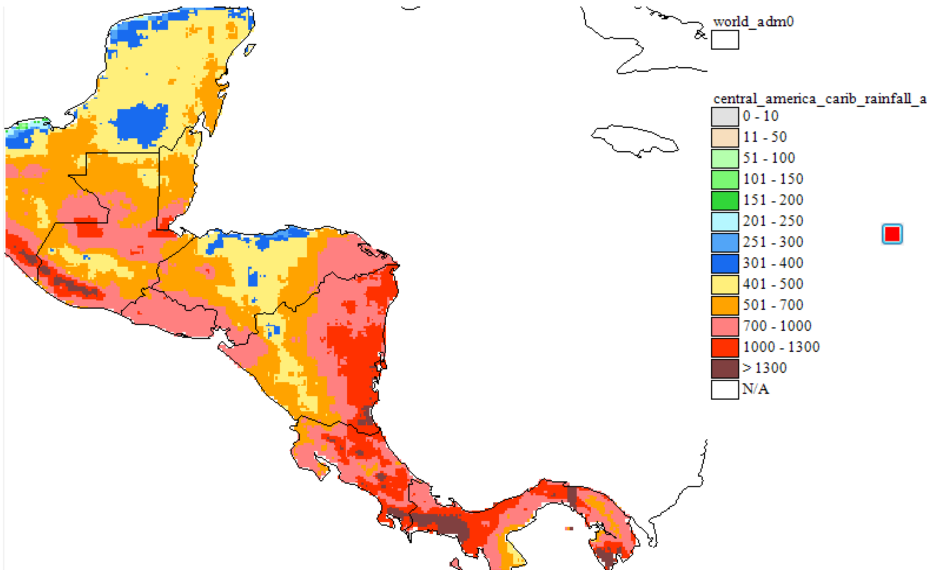


Figure 7. 2 *Average rainfall for La Nina years in Central America, composite1. The default legend was modified based on the range of values.*

7.2. Percent of Average: (apply for composite 1 and composite2)

Calculation of percent of average for composite 1:

$$\text{pct\_comp1} = \left(\frac{\text{avg}_{\text{comp1}}}{\text{average}_{\text{baseline}}}\right) * 100 \qquad (7.1)$$

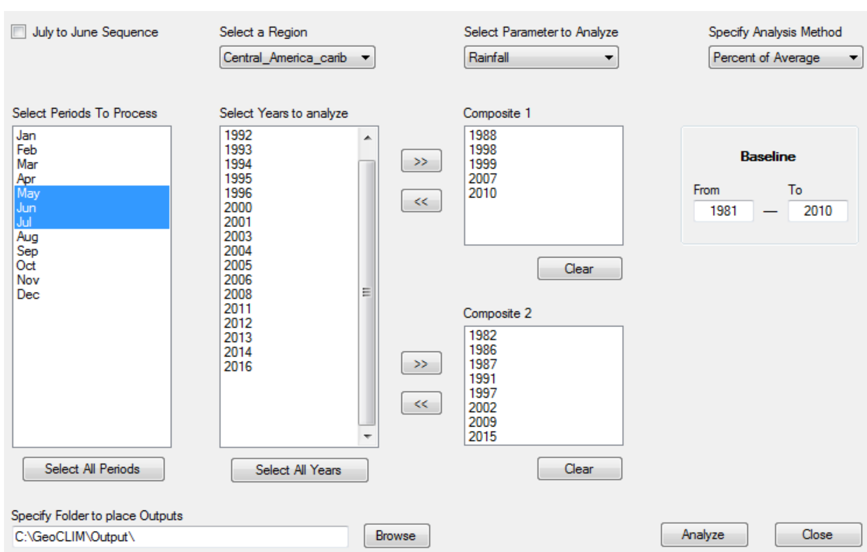


Figure 7.3 *The composites function allows for the calculation of percent of average for a single group of year or comparing two groups of years.*

- If composite 2 is empty, the program saves pct_comp1 as final output and displays it in the Spatial Data Viewer.

- If Composite 2 is **not** empty, the program calculates the difference between the two composites as follows:

$$\text{pct\_comp} = \left(\frac{\text{avg}_{\text{comp1}} - \text{avg}_{\text{comp2}}}{\text{average}_{\text{baseline}} + 0.1}\right) * 100 \qquad (7.2)$$

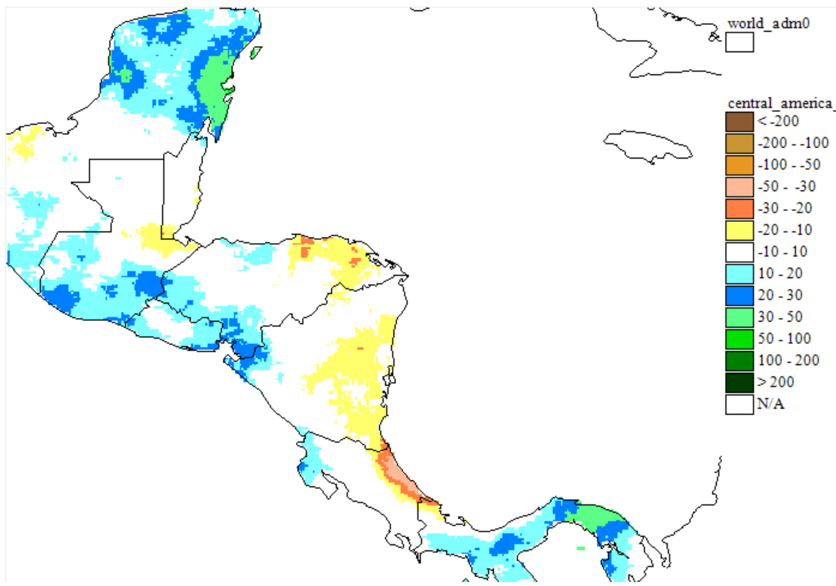


Figure 7.4 *Percent average for composites 1 (la Niña) and 2 (El Niño). In this example the positive values indicate that precipitation during la Niña years is higher, on average, than during El Niño years. The default legend was modified based on the range of values.*

7.3 Anomaly: (apply for composite 1 and composite 2)

This function calculates averages for each composite and the baseline, then it calculates the anomaly for each composite as follows:

$$\text{anom}_{\text{compN}} = \text{average\_compN} - \text{average\_baseline} \qquad (7.3)$$

- If composite 2 is empty, the anom_compN is saved as final output and display it in the Spatial Data Viewer.
- If Composite 2 is **not** empty, the program calculates the difference between the anomalies of the two composites as follows:

$$\text{anom\_comp} = \text{anom\_comp1} - \text{anom\_comp2} \qquad (7.4)$$

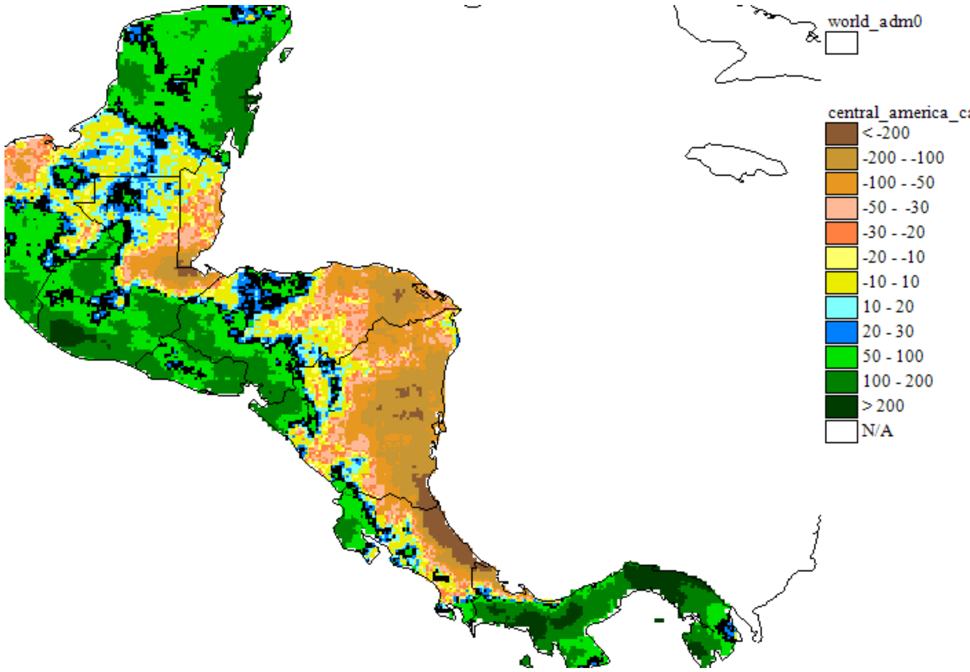


Figure 7.5 *The + anomalies show areas were, on average, la Niña years have higher values than el Niño. The results are shown in mm. The default legend was modified based on the range of values.*

7.4. Standardized Anomaly: (apply for composite 1 and composite 2)

This function calculates the difference anomaly of the average precipitation for a group of years and expresses it in terms of standard deviations away from the mean.

The function:

1) Validates if data exists for the selected years for composites and baseline.
2) Calculates standard deviation for the complete time series
3) Calculates average for the composites and baseline
4) Calculates anomaly for each composite
5) Calculates the standardized anomaly for each composite, see equation 7.5

$$\text{stdanom\_compN} = \left(\frac{(\text{avg\_compN} - \text{avg\_baseline}) + 0.1}{\text{stdev\_baseline} + 0.1}\right) * 100 \qquad (7.5)$$

- If composite 2 is empty, the function saves stdanom_compN as final output and displays it in the Spatial Data Viewer.
- If Composite 2 is **not** empty, the function calculates the difference between the two composites as follows:

$$\text{stdadnom\_comp} = \text{stdanom\_comp1} - \text{stdanom\_comp2} \qquad (7.6)$$

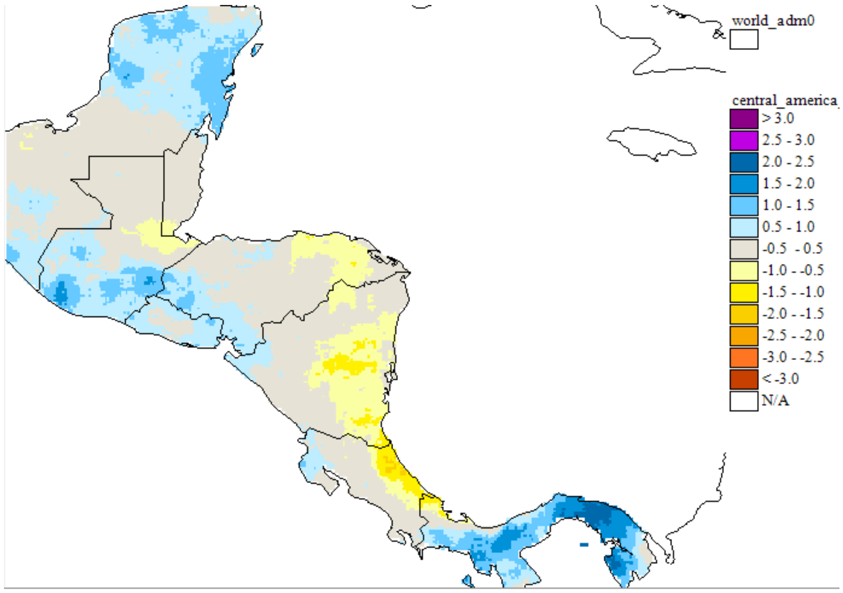
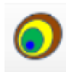


Figure 7.6 *this function calculates the difference anomaly of the average precipitation for a group of years and expresses it in terms of standard deviations away from the mean.*

NOTE: Event thought the values on the map are whole numbers the legend shows the results in number of standard deviations away from the mean.

Chapter 8: Contour Tool

Summary

The **Make Contours** tool  can be used to delineate areas within a defined interval of rainfall, evapotranspiration, or temperature. Analyzing contours from different intervals helps identify change on a given variable within a region.

8.1. Making Contours

1) Open the **Make Contours** tool from the GeoCLIM main toolbar. See Figure 8.1.
2) Specify the BIL input dataset. The tool automatically specifies the output file.
3) Select a contour interval value. In this case, 400 for an interval of rainfall of 400 mm.

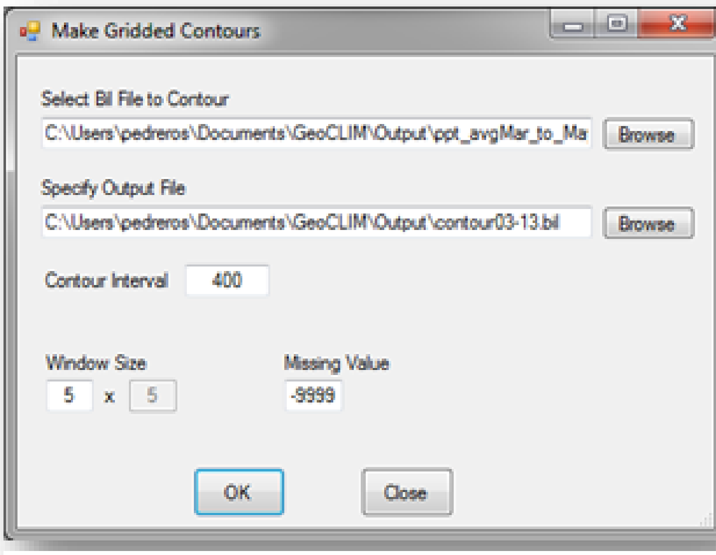


Figure 8.1 Select a rainfall dataset and the contour interval.

The contour tool can be used to analyze changes in average rainfall for different periods of time. For example, Figure 8.2 shows the changes in average for the years 19101-1992, 1993-2004, and 2005-2016 for the East Africa Countries (EAC) for the March-April-May season.

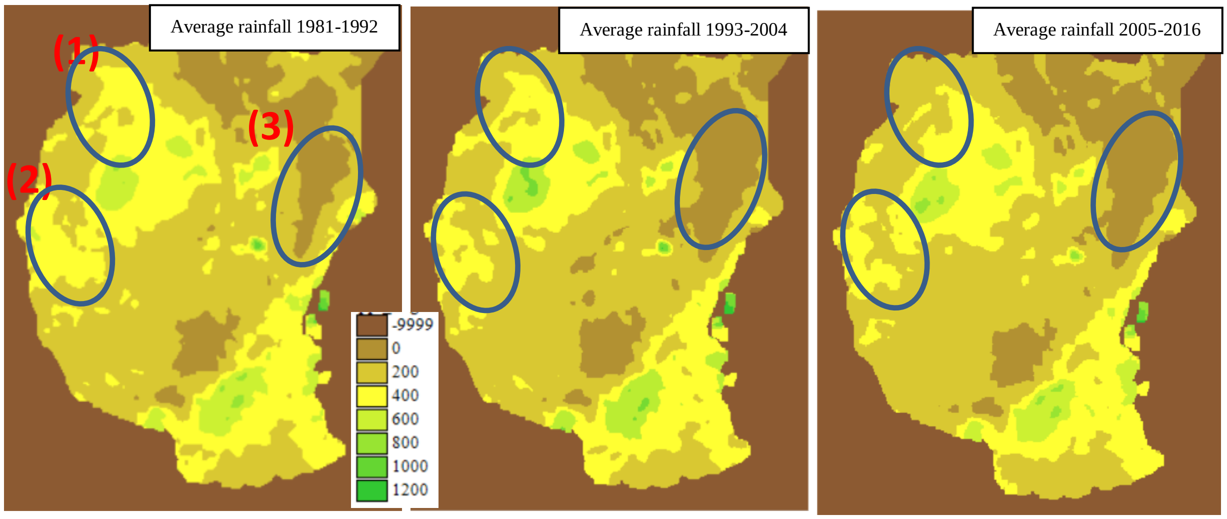


Figure 8.2 *The 200 mm interval for the average rainfall of the season March-April-May for the years 1981-1992, 1993-2004, and 2005-2016 show that areas of 400mm in the western part of the region (polygons 1 and 2) are changing into 200mm. Also, areas of 200mm in Kenya are changing into the zero interval (polygon 3). These polygons show that rainfall is decreasing in these regions.*

Chapter 9: Calculate Long-Term Changes in Average

Summary

Another way to estimate trends is by comparing the changes in average between two periods on a time series. The ***Calculate Long-Term Changes in Average*** tool allows for the selection of two different sets of years and calculate the difference between their long-term averages.

9.1 Estimate trends using difference in averages tool

1) Open the tool from the main bar menu (see Figure 9.1).
2) Select the season to be analyzed.
3) On the series 1, select the first period of time.
4) On series 2, select the second period.

The right side of the form is completed automatically with the information from the default rainfall data. Click **Process** to finish.

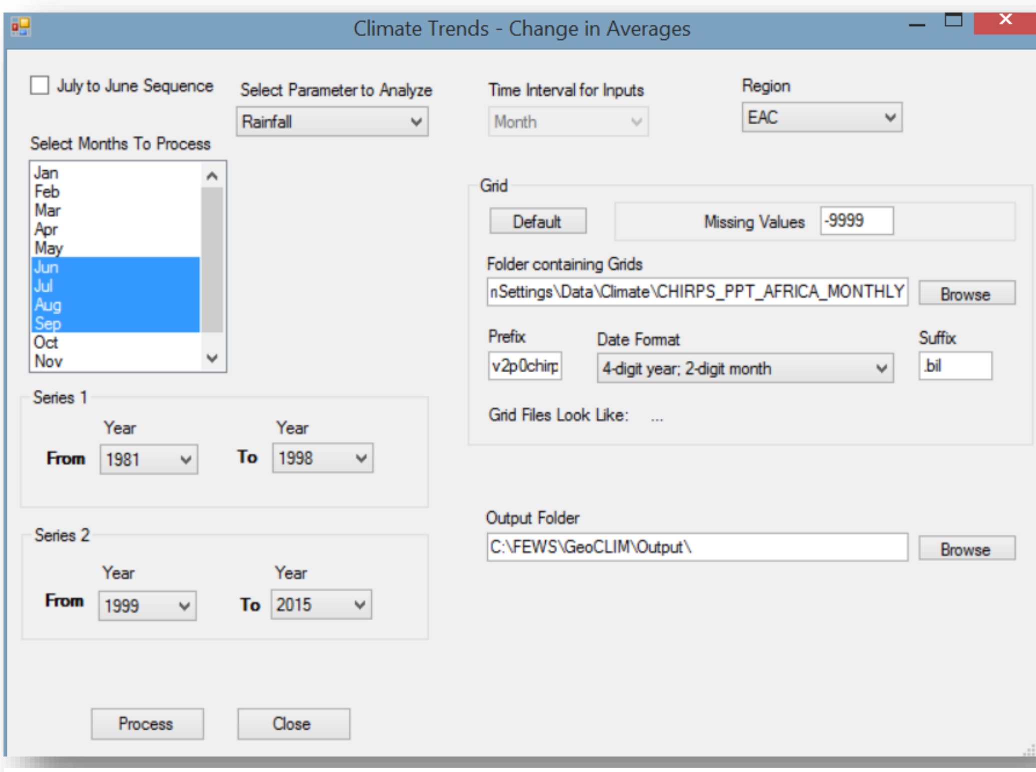


Figure 9.1 *The **Calculate Change in Average** tool is another way of estimating trends by comparing changes in averages between two periods.*

For example: Figure 9.2 shows the calcualtion of the change in average from the 1981-1998 period to the 1999-2015 period, for the June-September season. Figure 9.2 shows that there is an increasing trend on the blue polygon while there is a decreasing trend on the red polygon.

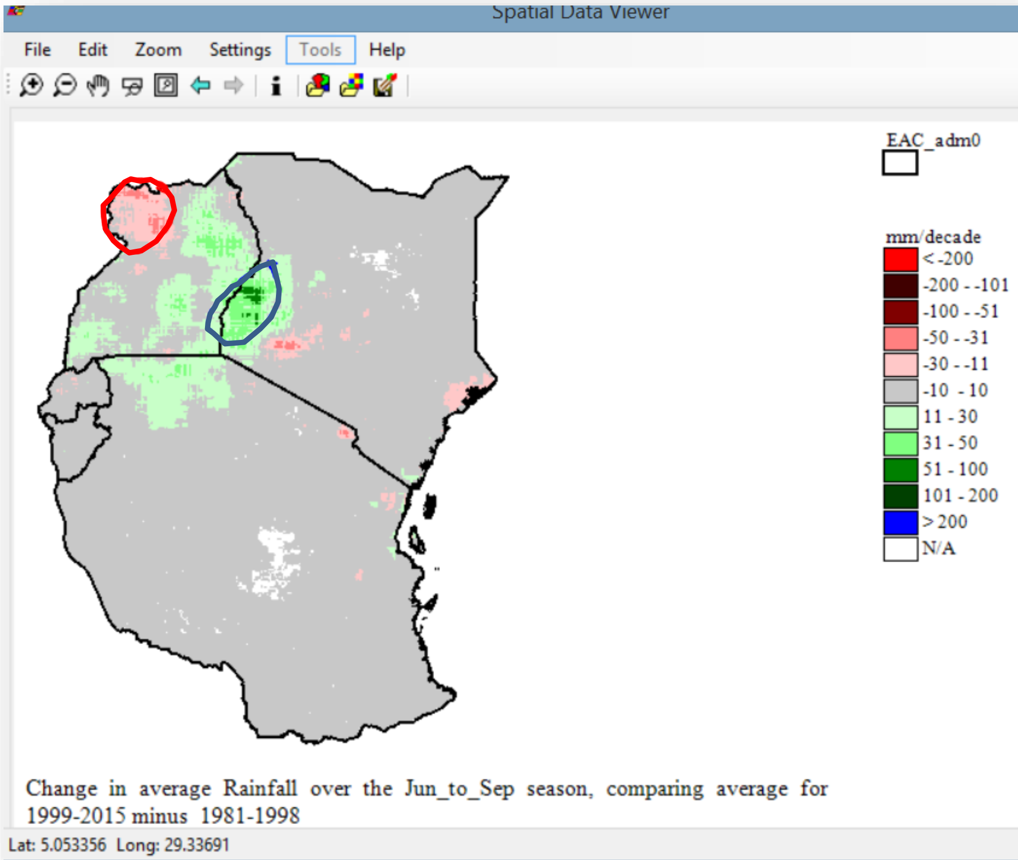


Figure 9.2 *Difference in averages; green colors show an increase in the latest period while red colors show a decrease in rainfall in the last period.*

Chapter 10: BASIICS

Summary

Satellite data provide useful information on rainfall, evapotranspiration, or temperature patterns. But, sometimes this data contains biases and inaccuracies due to incorrect or limited ground data used during calibration. Some raster data also have low resolution, which means the size of the pixel is too large for the area of interest. This data could be improved by combining them with ground station information. This chapter will review two processes: (1) validate the accuracy of rainfall estimates, in raster format, against ground stations data, and (2) improve rainfall estimates by blending them with ground station data using the BASIICS algorithm. These are two of the processes available through the GeoCLIM's Batch Assistant and Batch Text Editor operations.

10.1. BASIICS 

The **B**ackground-**A**ssisted **S**tation **I**nterpolation for **I**mproved **C**limate **S**urfaces (BASIICS) tool blends raster/grid datasets, such as satellite data, with point datasets, particularly station data. This blending is done using a modified Inverse Distance Weighting (IDW) approach that borrows some concepts from kriging, particularly simple and ordinary kriging. The basic process uses two datasets: (1) a point dataset with values at discrete points in space (e.g., rain gages) and (2) a grid/raster dataset with values varying continuously over space (e.g., a satellite-based rainfall estimate grid or a climatic average). To use the approach effectively, the two datasets need to be correlated. The first step in the process is to extract values from the grid at all locations where the point data have valid values (missing values can be specified by the user). This produces a comparable dataset of grid values that can be directly compared to the point values. The process can also carry out a least squares regression between the collocated point and extracted grid values and output the R-squared value in a statistical diagnostic file.

The following four-step process is recommended to produce the improved gridded datasets:

1. Download or import the raster datasets to be improved into the GeoCLIM, see chapter 2.
2. Validate the rainfall estimates.
3. Improve rainfall estimates by blending them with stations values.
4. Use batch files to run the algorithm for subsequent periods.

10.2 Validate Satellite Rainfall

One of the batch operations of the GeoCLIM is the **Validate Satellite Rainfall** operation, see Figure 10.1, which allows the comparison between satellite-estimated rainfall and station data.

To use this function, follow the three steps below:

Step 1

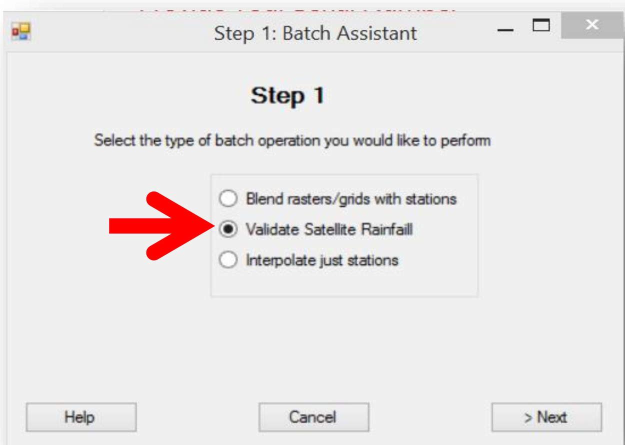


Figure 10.1 *There are three functions; Blend stations and raster data, Validate Satellite Rainfall and Interpolate Just Stations.*

- Click on the **BASIICS** icon from the GeoCLIM toolbar to open the Batch Assistant dialog box (**Step 1)**, see Figure 10.1.
- Select the **Validate Satellite Rainfall** option.
- Click on the > **Next** button to proceed to Step 2.

Step 2

- Select the time period of the rainfall estimate to validate. The time period and time interval as based on the selected climate dataset definition.
- Click on > **Next** button to proceed to Step 3. See Figure 10.2.

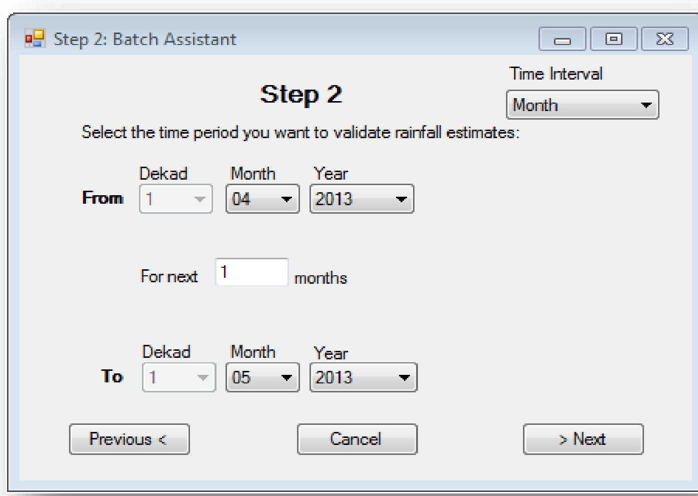


Figure 10.2 *Step 2 requires the dates* ***From*** *and* ***To*** *of the data to be analyzed.*

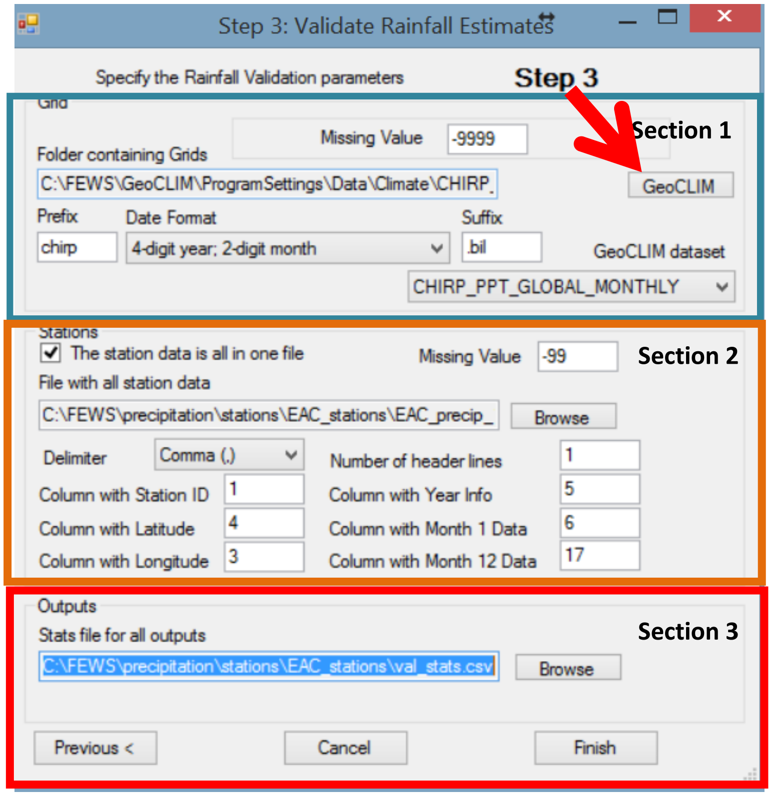


Figure 10.3 *To validate raster data using station values, it is required to have the path to the raster data and a table with the station values.*

Step 3: Step with three sections.

Section 1 (input grids) in Figure 10.3: This section relates to the raster/grid input parameters. Here, the user will need to indicate the file path to the gridded data to be validated. If the data to validate is the selected climate dataset, click on the **GeoCLIM** button to populate all the fields in this section automatically with the information of the raster data selected in **Program Settings**. Otherwise, fill all the information manually.

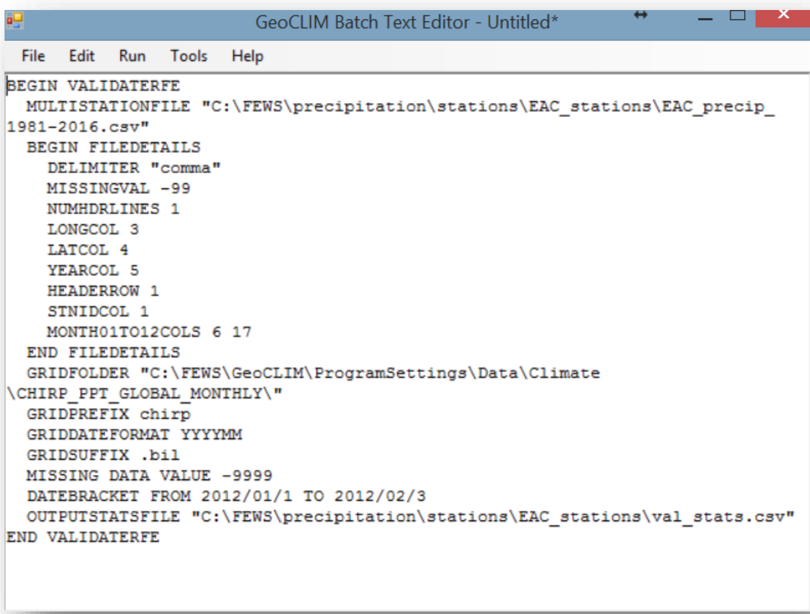
Section 2 (Stations) in Figure 10.3: This sections relates to the station input parameters.

1) Check on **The station data is all in one file** box, then select the file which contains the station data. The file must be in CSV file type, and can be prepared in Excel. See the Data Types chapter for more information on the format of the table and other file types in the GeoCLIM. If the station data are in different files, leave the box unchecked.
2) After selecting the stations file, the **Define Delimited Data Text File** dialog box will open where the format of the station file is defined - the header row (usually row 1), the first row that contains actual data (usually row 2), and the delimiter (usually a comma). Make any necessary changes for the correct specifications. Click **OK** when all the specifications are defined.

3) Next, make sure that the columns with **Station ID**, **Latitude**, **Longitude**, **Year Info**, and the first and last period (the period could be pentad, dekad, or month), as well as the missing value, have all been specified.

Section 3 (Outputs) see Figure 10.3:

1) Specify the file location where the statistical outputs will be written.
2) Click on **Finish**. Next, a batch file is generated and displayed, see Figure 10.4. This batch file can be saved for future reference.
3) Go to **Run** and select **Run Batch File** (Figure 10.4)




```
BEGIN VALIDATERFE
  MULTISTATIONFILE "C:\FEWS\precipitation\stations\EAC_stations\EAC_precip_
1981-2016.csv"
  BEGIN FILEDETAILS
    DELIMITER "comma"
    MISSINGVAL -99
    NUMHDRLINES 1
    LONGCOL 3
    LATCOL 4
    YEARCOL 5
    HEADERROW 1
    STNIDCOL 1
    MONTH01TO12COLS 6 17
  END FILEDETAILS
  GRIDFOLDER "C:\FEWS\GeoCLIM\ProgramSettings\Data\Climate
\CHIRP_PPT_GLOBAL_MONTHLY\"
  GRIDPREFIX chirp
  GRIDDATEFORMAT YYYYMM
  GRIDSUFFIX .bil
  MISSING DATA VALUE -9999
  DATEBRACKET FROM 2012/01/1 TO 2012/02/3
  OUTPUTSTATSFILE "C:\FEWS\precipitation\stations\EAC_stations\val_stats.csv"
END VALIDATERFE
```

Figure 10.4 *The Batch File is a text file with a list of commands from the step 3 form.*

This batch operation will create the following four outputs: (1) a map graphic (PNG) showing the stations used for validation overlaid on the rainfall field (Figure 10.5), (2) a scatterplot showing the rainfall field values against the stations (Figure 10.6), (3) a CSV file with a column with the station value, a column with the raster value for the point where the station falls, and a column with the value of an interpolated field using stations. The file includes some statistics showing how well the rainfall field and station data are related (Figure 10.7). And (4) a shape file containing all the stations that were used in the process. These outputs give the basis to decide if it is appropriate to blend the stations and the raster datasets.

NOTE: A map output will only display if a single period is selected. If multiple periods are selected (e.g. three months), the map graphics will not be displayed but they can be found as PNG files in the GeoCLIM output folder.

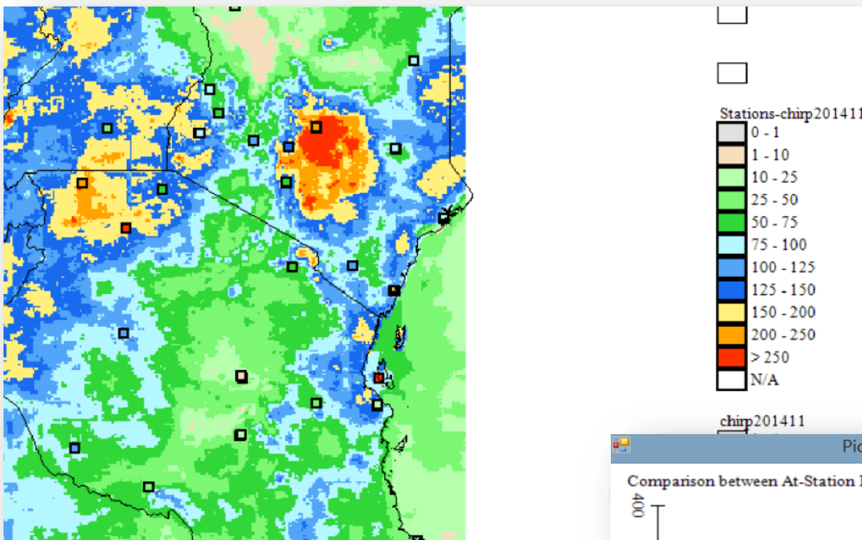


__Figure 10.5__ a map of the stations that took part in the process of validating CHIRP.

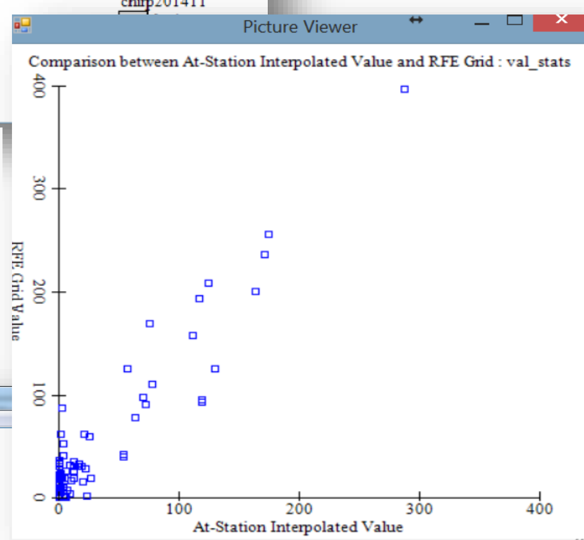
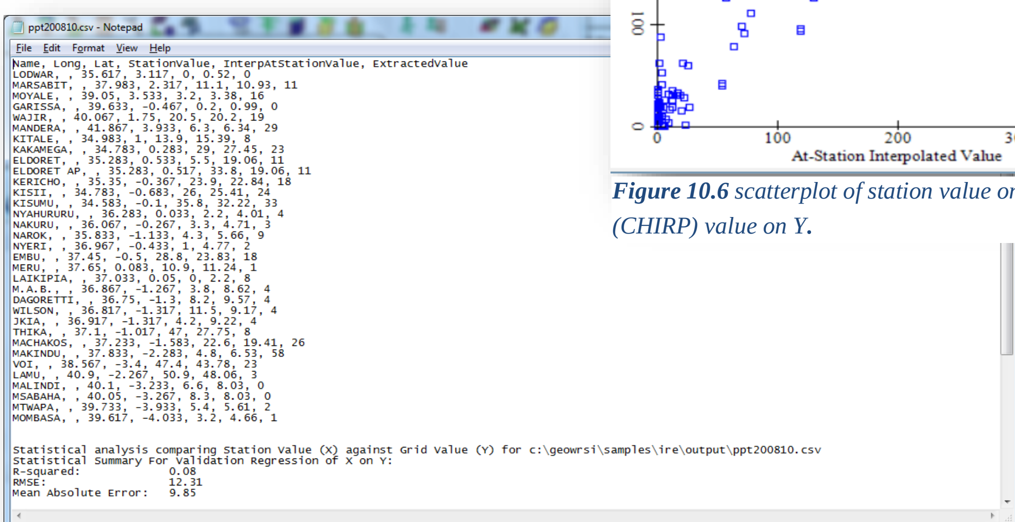


__Figure 10.6__ scatterplot of station value on X and raster (CHIRP) value on Y.



ppt200810.csv - Notepad

File Edit Format View Help

```
Name, Long, Lat, StationValue, InterpAtStationValue, ExtractedValue
LODWAR, , 35.617, 3.117, 0, 0.52, 0
MARSABIT, , 37.983, 2.317, 11.1, 10.93, 11
MOYALE, , 39.05, 3.533, 3.2, 3.38, 16
GARISSA, , 39.633, -0.467, 0.2, 0.99, 0
WAJIR, , 40.067, 1.75, 20.5, 20.2, 19
MANDERA, , 41.867, 3.933, 6.3, 6.34, 29
KITALE, , 34.983, 1, 13.9, 15.39, 8
KAKAMEGA, , 34.783, 0.283, 29, 27.45, 23
ELDORET, , 35.283, 0.533, 5.5, 19.06, 11
ELDORET AP, , 35.283, 0.517, 33.8, 19.06, 11
KERICHO, , 35.35, -0.367, 23.9, 22.84, 18
KISII, , 34.783, -0.683, 26, 25.41, 24
KISUMU, , 34.583, -0.1, 35.8, 32.22, 33
NYAHURURU, , 36.283, 0.033, 2.2, 4.01, 4
NAKURU, , 36.067, -0.267, 3.3, 4.71, 3
NAROK, , 35.833, -1.133, 4.3, 5.66, 9
NYERI, , 36.967, -0.433, 1, 4.77, 2
EMBU, , 37.45, -0.5, 28.8, 23.83, 18
MERU, , 37.65, 0.083, 10.9, 11.24, 1
LAIKIPIA, , 37.033, 0.05, 0, 2.2, 8
M.A.B., , 36.867, -1.267, 3.8, 8.62, 4
DAGORETTI, , 36.75, -1.3, 8.2, 9.57, 4
WILSON, , 36.817, -1.317, 11.5, 9.17, 4
JKIA, , 36.917, -1.317, 4.2, 9.22, 4
THIKA, , 37.1, -1.017, 47, 27.75, 8
MACHAKOS, , 37.233, -1.583, 22.6, 19.41, 26
MAKINDU, , 37.833, -2.283, 4.8, 6.53, 58
VOI, , 38.567, -3.4, 47.4, 43.78, 23
LAMU, , 40.9, -2.267, 50.9, 48.06, 3
MALINDI, , 40.1, -3.233, 6.6, 8.03, 0
MSABAHA, , 40.05, -3.267, 8.3, 8.03, 0
MTWAPA, , 39.733, -3.933, 5.4, 5.61, 2
MOMBASA, , 39.617, -4.033, 3.2, 4.66, 1


Statistical analysis comparing Station Value (X) against Grid Value (Y) for c:\geowrsi\samples\ire\output\ppt200810.csv
Statistical Summary For Validation Regression of X on Y:
R-squared:              0.08
RMSE:                   12.31
Mean Absolute Error:    9.85
```

__Figure 10.7__ text file that includes a list of the station value and its corresponding raster value together with statistics for the two groups.

10.3 Improve rasters with stations using the blending algorithm

To create improved rainfall estimates, follow the steps below:

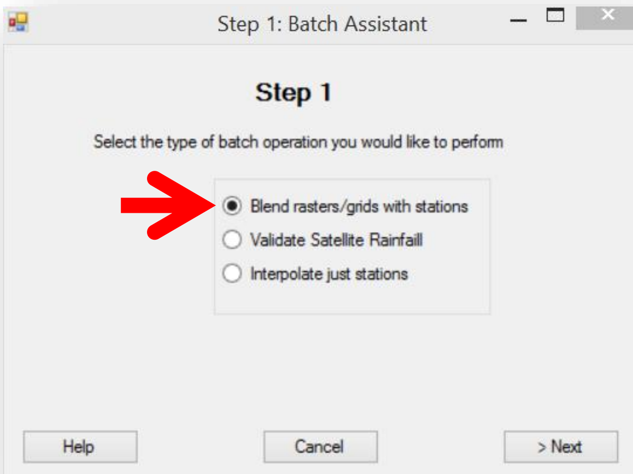


Figure 10.8 *Select the Blend raster/grids with stations option.*

Step 1

1) Click on the **BASIICS** button on the toolbar (Figure 10.10).
2) This will open the **Step 1** window, select the **Blend rasters/grids with stations** batch operation, and click on the > **Next** button. See Figure 10.8.

Step 2

In this step, select the **Time Interval** and the periods to be improved. Make sure that there are stations available for the same period to be improved. Click on **>Next** to continue. See Figure 10.9.

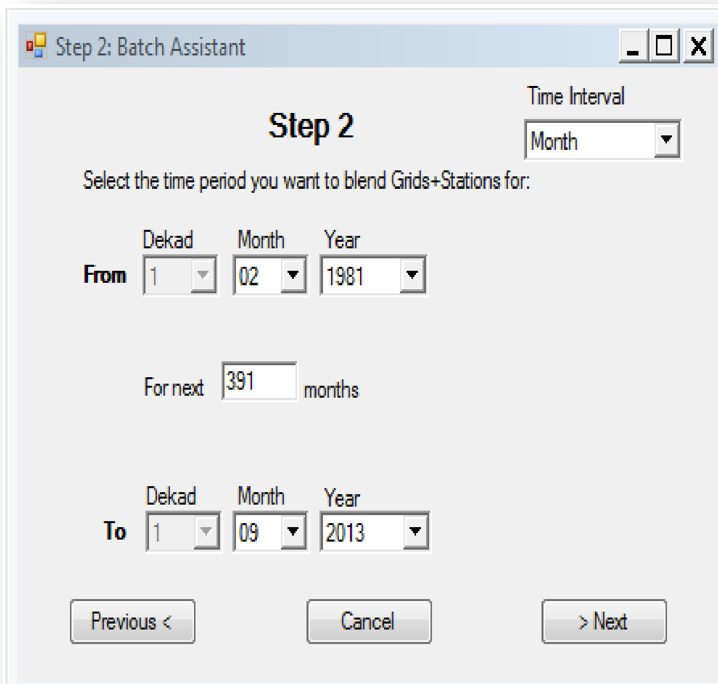


Figure 10.9 *Select the starting and ending dates that would be part of the process.*

Step 3

This step consists in four input sections: (1) grid (raster) information, (2) station data, (3) output information and (4) advanced options.

(1) Describing Grid data (Figure 10.10).

In this section, to use the default climate dataset, simply click on the **GeoCLIM** button to retrieve automatically all the parameters needed for the process; otherwise add all the inputs manually. The *Missing Value* field must be manually typed.

(2) Describing station data (Figure 10.10). In this section, specify the file containing the station data, the missing value in the rain gauge data, and the column numbers for each of the required input information (Station ID, latitude, longitude, etc.). The tool will try to retrieve all these information automatically, but it is important to verify that the information added by the tool was loaded correctly.

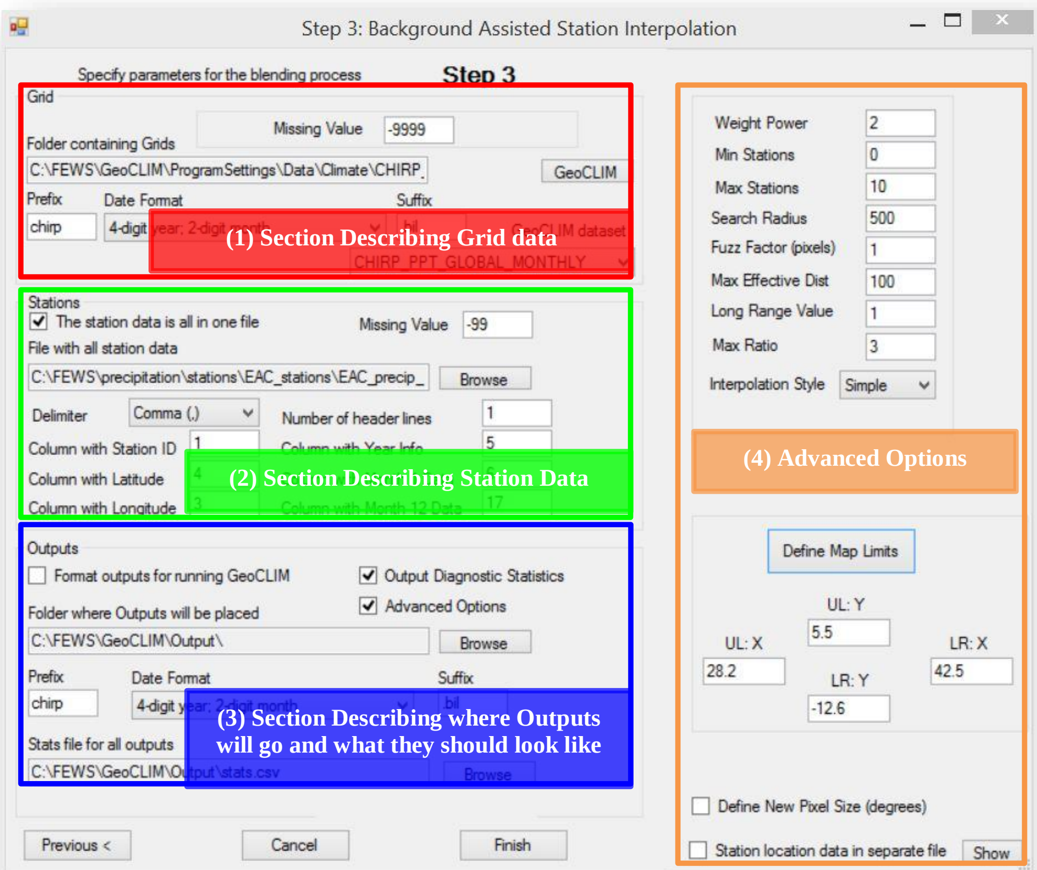


Figure 10.10 Step 3 requires information about the raster data to be improved, the stations with station metadata, and rainfall values.

The station data should be a CSV file and be in the format described in Figure 10.11. The order of the columns is not important, but must include:

- A unique station identifier ID, column 1
- A longitude column(*lon* in the example)
- A latitude column (*lat* in the example)
- Year, a column depicting the year value
- A series of consecutive columns for the number of periods, 72 for pentads, 36 for dekads or 12 for months.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
ID	FEWSID	lon	lat	year	month01	month02	month03	month04	month05	month06	month07	month08	month09	month10	month11	month12
14741	63932002	33.6	-9.3	1985	115	269	258	915	250	82	46	18	30	68	234	90
14688	63844007	38.4	-5.1	1984	72.8	19.2	172.7	913	306.9	295.4	114	40.1	98.8	202.8	165.1	186.5
14658	63790003	37.6	-3.2	1990	128.8	172.1	514.3	878.9	150.1	41.7	35	23.3	6.2	96.6	222.3	187.9
14661	63791001	37.1	-3.2	1984	23.4	37.1	64.1	821.1	275.2	158.9	87.3	21	23.8	49.4	117.5	101.1
14691	63844010	38.6	-5.2	1984	101.3	64.1	151.8	788.7	279.3	246.8	133.7	37.6	96.5	230.7	457.7	320.3
14657	63790002	37.3	-3.3	1988	93.5	8	166	756.7	371.5	217	26.5	38.5	54	3.5	111.5	72.5
14728	63887003	36.7	-8.6	1993	280	161	607	752	293	27	36	6	0	0	111	100

Figure 10.11 the CSV table with station data must contain a statin ID, lon, lat, year and a column for each pentad, dekad or month.

(3) Describing the outputs, (Figure 10.10)

In the third section, specify where the outputs should go and what format they should be in. If **Output Diagnostic Statistics** is selected (highly recommended), then the last part of this section will ask for the path where the statistics file should be saved. This is a very important file since it gives an evaluation of how well the blending operation performed.

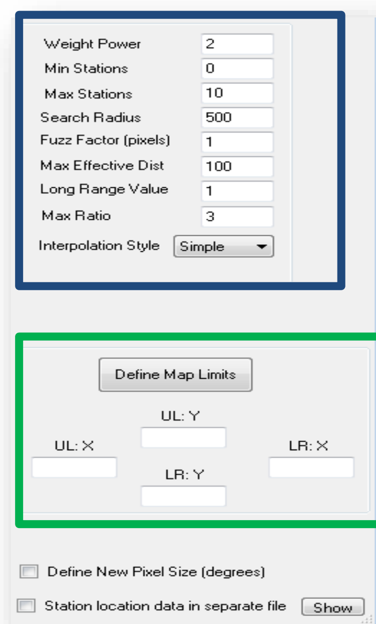


Figure 10.12 The advance option panel allows the user to modify the blending process.

(4) Advanced Options (the Blending algorithm)

When the **Advanced Options** checkbox is checked, this will open a set of options to the right of the window. The upper section describes parameters for the blending algorithm, see Figure 10.12 blue box.

The IRE algorithm is a methodology that was designed for combining rain gage data with satellite-based rainfall estimates to produce a more accurate gridded dataset of rainfall estimates. It has been implemented within the GeoCLIM program to allow for the generation of enhanced accuracy rainfall grids that can be used in soil water balance modeling, as well as for other purposes. Generically, the algorithm combines spatially discrete point data with spatially continuous grid data by interpolating ratios between the point and the grid where these two data are collocated, then multiplying the ratio by the grid. The technique is similar in principle to the SEDI technique that was originated for the Southern African Development Community (SADC)/FAO Regional Remote Sensing Project and developed by Peter Hoefsloot. The interpolation technique is done using the Inverse Distance Weighting (IDW) approach but borrows some concepts from the kriging approach, particularly the use of the concepts of simple and ordinary kriging. The equivalent concepts of "simple" and "ordinary" IDW are explained at a later stage. The basic concept is the use of two datasets: (1) a point dataset with values at discrete points in space (for example, rain gages) and (2) a grid dataset with values varying continuously over space (for example, a satellite-based rainfall estimate grid or a climatic average). For the algorithm to be used effectively, the two datasets need to be correlated. The first step in the algorithm is to extract values from the grid at all locations where the point data have valid values (missing values can be specified by the user). This produces a comparable dataset of grid values that can be directly compared to the point values. As one option, the user can instruct the program to carry out a least squares regression between the collocated point and extracted grid values and output the r^2 value in a statistical diagnostic file. This allows the user to determine whether the relationship between the grid and the point dataset is worth using to interpolate the data.

MaxRatio

When the values for the grid at the point locations have been extracted, the ratio between the grid and the point value at each point location are calculated to produce a dataset of station/grid ratios. One of the options in the program at this point is that user can specify that the station/grid ratios should be capped to avoid "run-away" values during the interpolation and grid multiplication, which follows after the ratio calculation. For example, assume that rain gage data were being interpolated with a rainfall estimate grid. At a specific point, assume that the point value recorded for the station was 10 mm, while the gridded rainfall estimate had a value of 1 mm. Although the absolute difference between the two estimates is only 9 mm, the station/grid ratio is 10, or 1,000 percent. This ratio will be interpolated, then multiplied by original grid. Assume that 50 km away from that point, the grid had a value of 30 mm. This 30 mm will be multiplied by a value close to 10, depending on the surrounding ratios in the interpolation, and the resultant value may be close to 300 mm. This error can be limited by capping the ratio and instructing the program not to cut off any ratios greater than a certain value. A cut-off ratio of 3 is used by default in the algorithm, meaning that any ratio greater than 3 is reset to 3 (in the example above, the ratio would be 3 instead of 10). However, this ratio can be set to any value by the user (a very large cut-off can be used, for example, 100,000) if the user does not want to have the ratios capped. This maximum, or cut-off, ratio is referred to in the IRE as the MaxRatio.

Search Radius, and Minimum and Maximum Stations `(SEARCHRADIUS)`, `(MINSTNS)`, `(MAXSTNS)`

These ratios are interpolated to produce a continuous raster dataset of station/grid ratios that, by default, has the same dimensions as the original grid with which the analysis is being done. Optionally, the user can specify latitude/longitude coordinates that define the spatial extent of the output grid, either directly or using a vector file such as a shapefile to define the extent. The interpolation needs to specify how many points should be used in the interpolation; both the minimum and maximum number of points should be used in the interpolation, as well as the maximum search radius within which to search for points, at any given location. At every location where a value is to be produced, the algorithm will search for the nearest stations that it can find within the allowable search radius from that location and use these in the interpolation. For example, assume that the user specifies that between 3 (minimum) and 10 (maximum) points are to be used within a radius of 200 km. The algorithm will search for the nearest 10 stations within a radius of 200 km. If fewer than 10 stations are found, for example 7, then those 7 stations will be used. However, if the number of stations found is less than the minimum allowable specified by the user, then that location will have a missing value. Use a value of *0* to produce a value everywhere in the output.

Weight Power **(WEIGHTPOWER)**: The power to which the inverse distance is raised in calculating the weight. For example, a weight of *2* means that the inverse distance to each station will be squared (power of 2) to calculate the corresponding weight.

Fuzz Factor (pixels) (**FUZZFACTOR):** In order to avoid reverse engineering the pixel value, the fuzz factor allows to hide the location of the station by a number of pixels.

Max Effective Dist **(MAXEFFECTIVEDIST)**: This parameter only works with the Simple interpolation Style (idw_s, see the interpolation style section below). It is the maximum distance to which the stations would have an influence. It is very important to understand the local characteristics of the region being interpolated to choose a proper value.

Max Ratio (**MAXRATIO)**: The maximum station/grid value that is allowed in the interpolation. When the station/grid ratio is greater than this parameter value, then this value is used.

Interpolation Style (**INTERPOLATIONALGORITHM**): The type of interpolation algorithm could be used in the process, *Simple* (idw_s) or *Ordinary* (idw_s) inverse distance weighting (IDW). In the ordinary IDW, the interpolation weights are dependent only on the surrounding stations. In the Simple IDW, the background grid also contributes a weight to the interpolation routine, and the relative weight of the background grid increases with increasing distance to surrounding stations.

The **Define Map Limits** option, see green box on Figure 10.12, allows the user to define a smaller area for the interpolation than the complete gridded dataset. This area can be defined by using the extent of an existing GeoCLIM Region (Figure 10.13) or other spatial data (raster or vector). This option helps to speed up the interpolation process considerably.

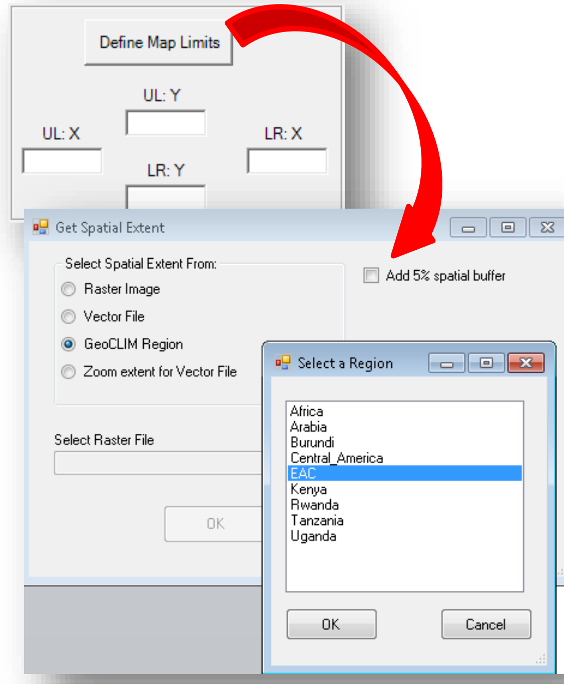


Figure 10.13 *Defining the output extend could be done by selecting an existing raster, region or shape file.*

For example to run the blending process for the EAC region (Figure 10.13), follow these steps:

1) From the **Get Spatial Extent** window, select the **GeoCLIM Region** option from the **Select Spatial Extent From** section. This will bring up the **Select a Region** window.

2) Choose the *EAC* region from the list. The selected region will define the extent of the outputs (map extent, points to analyze, etc).

3) Click on **OK**

4) Then click **OK** on the **Get Spatial Extent** window as well.

5) After all the fields are completed, the Step 3 form should look like Figure 10.10.

6) Next, click on the **Finish** button at the bottom of the Step 3 window. This will generate a batch file which can be run as needed.

Before running the batch file, it is strongly recommended to save it. To do this,

- Press **CTRL+S** or,

- Go to **File** > **Save As** from the **Batch Text Editor** menu, see Figure 10.14. This will enable the user to access and edit the saved file later.

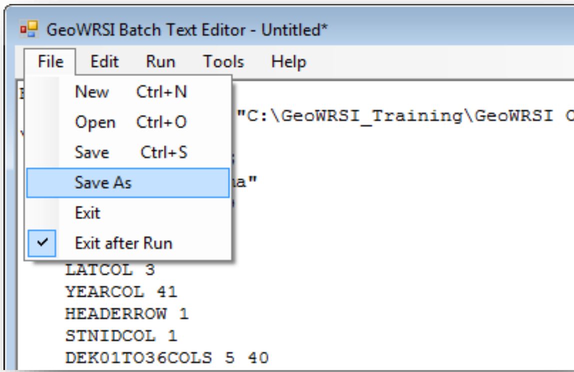
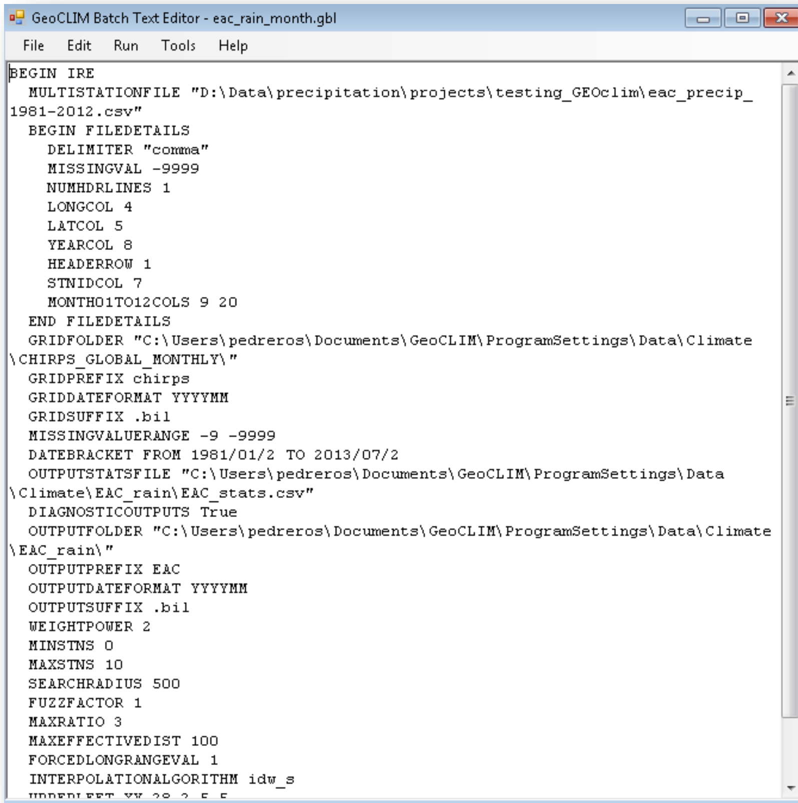


Figure 10.14 The batch file could be saved and re-ran later.

To run the blending batch file, either (a) Press the **F5** key on the keyboard, or (b) go to **Run** > **Run Batch File** from the **Batch Text Editor** menu.

The batch file contains all the settings specified in step three of the blending process. The commands are self-descriptive, so users could understand the majority of the command lines. See Figure 10.15.




```
BEGIN IRE
  MULTISTATIONFILE "D:\Data\precipitation\projects\testing_GEOclim\eac_precip_
1981-2012.csv"
  BEGIN FILEDETAILS
    DELIMITER "comma"
    MISSINGVAL -9999
    NUMHDRLINES 1
    LONGCOL 4
    LATCOL 5
    YEARCOL 8
    HEADERROW 1
    STNIDCOL 7
    MONTH01TO12COLS 9 20
  END FILEDETAILS
  GRIDFOLDER "C:\Users\pedreros\Documents\GeoCLIM\ProgramSettings\Data\Climate
\CHIRPS_GLOBAL_MONTHLY\"
  GRIDPREFIX chirps
  GRIDDATEFORMAT YYYYMM
  GRIDSUFFIX .bil
  MISSINGVALUERANGE -9 -9999
  DATEBRACKET FROM 1981/01/2 TO 2013/07/2
  OUTPUTSTATSFILE "C:\Users\pedreros\Documents\GeoCLIM\ProgramSettings\Data
\Climate\EAC_rain\EAC_stats.csv"
  DIAGNOSTICOUTPUTS True
  OUTPUTFOLDER "C:\Users\pedreros\Documents\GeoCLIM\ProgramSettings\Data\Climate
\EAC_rain\"
  OUTPUTPREFIX EAC
  OUTPUTDATEFORMAT YYYYMM
  OUTPUTSUFFIX .bil
  WEIGHTPOWER 2
  MINSTNS 0
  MAXSTNS 10
  SEARCHRADIUS 500
  FUZZFACTOR 1
  MAXRATIO 3
  MAXEFFECTIVEDIST 100
  FORCEDLONGRANGEVAL 1
  INTERPOLATIONALGORITHM idw_s
```

__Figure 10.15__ The batch file describes, in command form, all the fields from the step 3. The user could edit the batch file and re-ran the process.

NOTE: The batch file can be saved for later use. The file can be edited and re-run for different periods.

Chapter 11: Extracting Raster Statistics and Time Series

Summary

The **Extract Grid Statistics** tool calculates summary statistics for a polygon or a set of polygons from a shapefile (SHP) file over a raster dataset or a group of rasters. For example, the average rainfall of each district for each January from 19101 to 2014 can be calculated.

11.1. Extract Statistics

1) Open the **Extract Raster Statistics** tool  from the menu bar (see Figure 11.1).
2) Select a shapefile containing the polygons of interest (e.g. districts) and select a unique ID (a data field in the shapefile that will uniquely identify each polygon, such as district names) see 2 on Figure 11.2.
3) Select the type of summary for the pixels within the polygon, see the **Summary** pulldown menu (3 on Figure 11.1).
    a. Select the raster file(s) to be used. There are two ways to populate the data fields, (see 4 on Figure 11.2): (a) using the **T-series** (Time-series) button and (b) using the

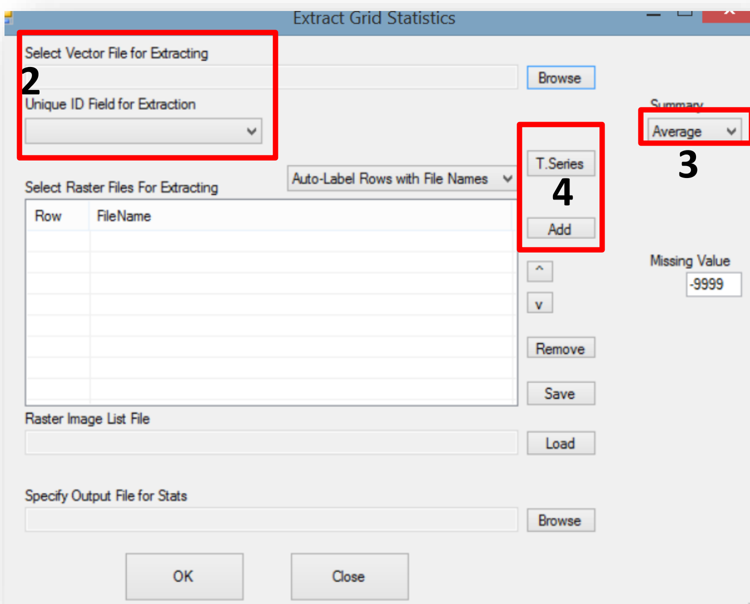


__Figure 11.1__ The tool calculates the spatial statistics for each raster dataset, based on each polygon on the shp file. The product is a table of a row for every polygon containing a single value (statistics selected) a column based on each raster.

Add button to select the raster datasets in BIL format. The **T-Series** button allows the generation of a list of raster files to extract from based on a time series, while the **Add** button allows the adding of individual files. Click on the **T-series** button.

This option will open a new window to define the climate dataset and time period to extract from.

i. Click on the **Use GeoCLIM Data** button to retrieve all the information from the default dataset. Or browse to the directory containing the data to be used.

ii. Select the period and years to add (Figure 11.2).

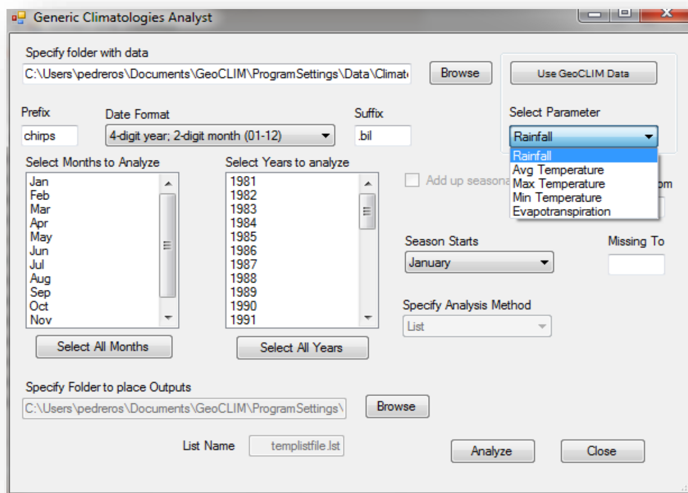


Figure 11.2 Select a T-series from the default data or point the path for a new data source.

iii. Once the periods and years are selected, click on the **Analyze** to create a temporal list file. The inputs defined previously are compiled in a list file that populates the fields in the previous window. A window message will confirm the creation of a list file.

iv. Click **OK** to close the window message and then click on the **Close** button to close the T-series window.

v. Back in the **Extract Grid Statistics** tool, the **Select Raster Files for Extracting** fields are now populated.

vi. Specify the output file where the output file with the results would be saved.

vii. Click on **OK** to run the process.

b. The second method of populating the **Select Raster Files for Extracting** fields is by clicking on the **Add** button and browsing to the directory where the files are.

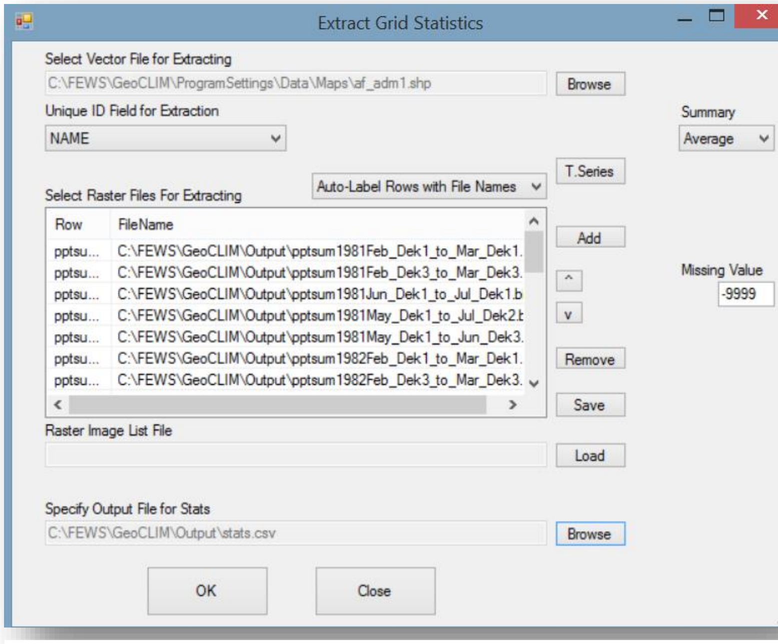
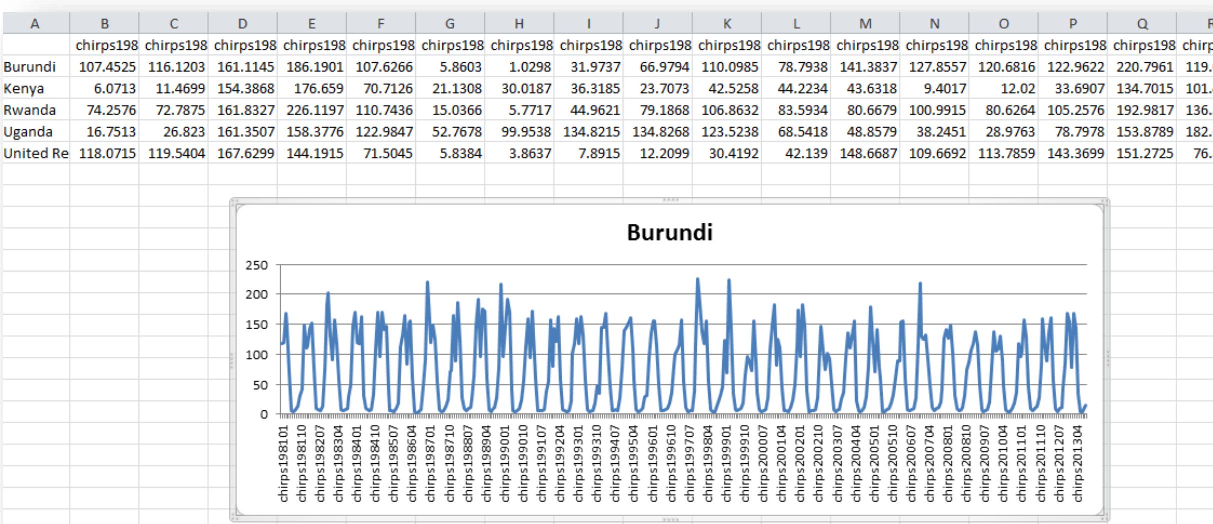


Figure 11.3 *Once all the data are selected and the shape files are in place, the form should look like the above.*

11.2. The Results

The **Extract Statistics** tool produces a CSV file with rows corresponding to the polygons from the input shapefile. Each column contains the summary value for each raster file selected. Figure 11.4 shows the spatially averaged rainfall using CHIRPS dekads, for each of the countries in the EAC region. To do additional analysis of the results such as production of time series graphs, open the CSV file in Microsoft Excel or similar program.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps198	chirps
Burundi	107.4525	116.1203	161.1145	186.1901	107.6266	5.8603	1.0298	31.9737	66.9794	110.0985	78.7938	141.3837	127.8557	120.6816	122.9622	220.7961	119.0
Kenya	6.0713	11.4699	154.3868	176.659	70.7126	21.1308	30.0187	36.3185	23.7073	42.5258	44.2234	43.6318	9.4017	12.02	33.6907	134.7015	101.4
Rwanda	74.2576	72.7875	161.8327	226.1197	110.7436	15.0366	5.7717	44.9621	79.1868	106.8632	83.5934	80.6679	100.9915	80.6264	105.2576	192.9817	136.5
Uganda	16.7513	26.823	161.3507	158.3776	122.9847	52.7678	99.9538	134.8215	134.8268	123.5238	68.5418	48.8579	38.2451	28.9763	78.7978	153.8789	182.6
United Re	118.0715	119.5404	167.6299	144.1915	71.5045	5.8384	3.8637	7.8915	12.2099	30.4192	42.139	148.6687	109.6692	113.7859	143.3699	151.2725	76.8

Figure 11.4 *The resulting table has a row for every polygon and every column represents the summary value for each raster.*

Chapter 12: Creating Archives

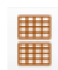
Summary

As mentioned in Chapter 2, there are different ways to make data available in the GeoCLIM program. One way is reading data that is already in the GeoCLIM workspace (`GeoCLIM\ProgramSettings\Data\Climate` directory). Another way to make data available in the program is importing a GeoCLIM archive. These archives are groups of raster files compressed into a single file that contains specification such as name format (e.g., 'v2p0chirpsyyyymm' - prefix, year, month) or time period (pentad, dekad, month) so that the GeoCLIM can read the data automatically. GeoCLIM archives are a great way to share data among colleagues. This chapter will go over how to create a data archive.

For a brief review of GeoCLIM data settings, go to chapter 2.

12.1. Create an Archive

To create an archive, follow these steps.

1) Open the **View Available Data** tool .
2) Make sure that the correct climate variable is selected
3) Select the part of the time series to be made into an archive
4) Click export (see Figure 12.1).
5) Select **GeoCLIM Archive** and click ok, a message ensuring the correct name of the new archive will pop up.
6) Click ok and name the new archive.

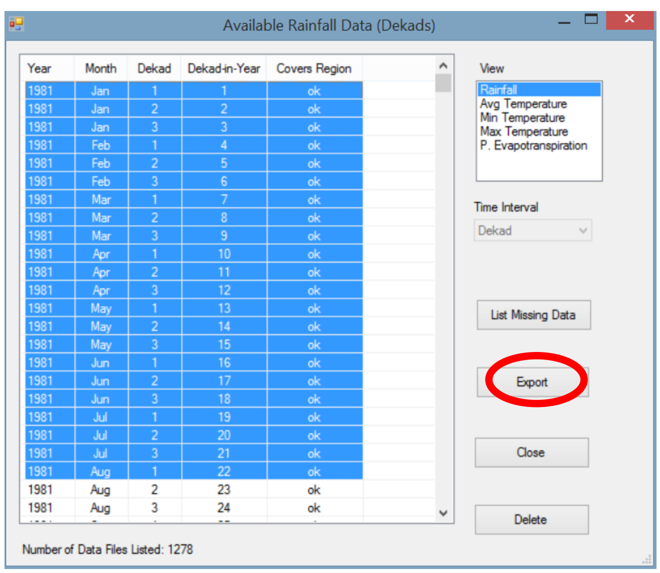


***Figure 12.1** Exporting a time series from the Available Rainfall Data tool will create an archive.*

A new file with the .climdata extension will show up in the assigned directory. This file contains all the information necessary so the GeoCLIM could read the data.

12.2. Importing archives

To import an archive follow these steps:

1) Open the Import Climate Data Archives  tool.
2) Select the archive to be imported and Click import, see Figure 12.3.

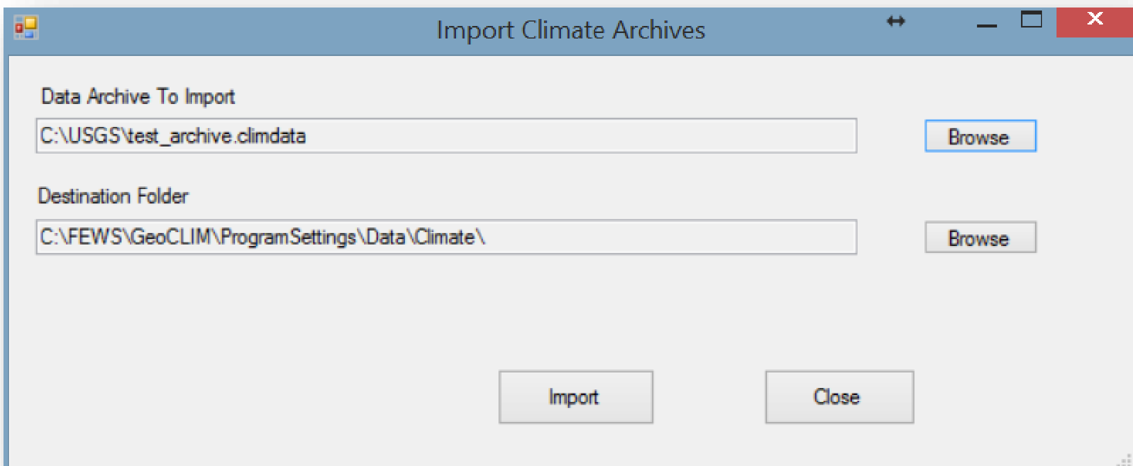


Figure 12.3 *The Import Climate Archives tool.*

3) Close the console (figure 12.5) once it is done. And make the new dataset the default.

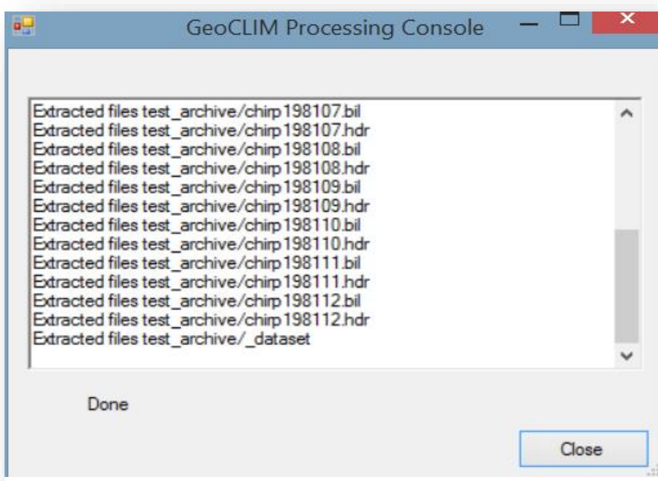


Figure 12.4 *The GeoCLIM console shows the progress of the archive import.*

To ensure that the archive imported correctly go to the GeoCLIM **settings> data>select data>** and the new imported dataset should be ready to use. See Figure 12.6.

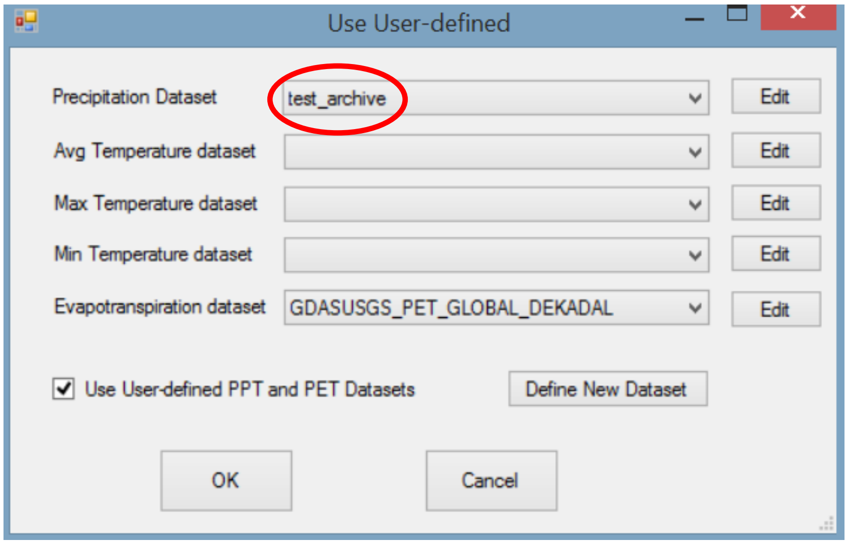


***Figure 12.6** Once finished importing the archive, the dataset is ready to be used in the different operation in the GeoCLIM.*

References

ESRI support, 20010. BIL, BIP, and BSQ raster files. http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=BIL,_BIP,_and_BSQ_raster_files

ESRI support, 2016. What does the Pixel depth mean? http://support.esri.com/technical-article/000006576

Standardized Precipitation Index (SPI), Early warning USGS-FEWS NET product documentation. http://earlywarning.usgs.gov/fews/product/51#documentation

McKee, T. B., N. J. Doesken, et al. (1993). The Relationship of Drought Frequency and Duration to Time Scales. 10th Conference on Applied Climatology.

Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell & Joel Michaelsen. ""The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes". *Scientific Data* 2, 150066. doi:10.10310/sdata.2015.66 2015.