



GeoCLIM Manual



By Diego Pedreros and Tamuka Magadzire

FEWS NET

Edited by Claudia J Young and Libby White

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.

US Geological Survey
Famine Early Warning System
Network

UC Santa Barbara
Climate Hazards Group

Table of Contents

Introduction..... 5

 Using the Manual 5

Chapter 1: Overview..... 7

 1.1. Import GeoCLIM Climate Archives  7

 1.2. Download Climate Data  8

 1.3. Define Output Options  8

 1.4. View Available Data  9

 1.5. GeoCLIM Settings.  10

 1.6. Climatological Data Analysis (mean, trend, percentile, etc.)  11

 1.7. Rainfall Summaries  12

 1.8. Make Contours  13

 1.9. Calculate Long-Term Changes in Average  13

 1.10. Batch Assistant for Easily Developing Automation Scripts  14

 1.11. Batch Editor for Editing Automation Scripts 15

 1.12. Spatial Data Viewer  16

 1.13. Extract Statistics from Raster Datasets  16

Chapter 2: Settings 17

 2.1. GeoCLIM Settings  17

 2.2. Making data available on the GeoCLIM 18

 2.2.1. Define climate data filename 18

2.2.2. Define a new dataset in the GeoCLIM	19
2.3. Check availability of the data and compatibility with the selected region.....	22
2.4. Review the GeoCLIM Directory Structure	22
2.5. Change workspace.....	23
Chapter 3: Data Types in the GeoCLIM	28
3.1. Characteristics of the Raster dataset.....	28
3.2. Vector data: shapefiles (*.shp)	29
3.3. Tables (*.csv)	30
Chapter 4: Spatial Data Viewer.....	31
4.1. Displaying a Raster dataset using the GeoCLIM	31
Chapter 5: Climatological Analysis.....	33
5.1. Climatological Analysis	33
5.2. Updating GeoCLIM averages	35
5.3. Analysis Methods	37
5.3.1. Average	37
5.3.2. Median	38
5.3.3. Measuring variability with Standard deviation (SD) and Coefficient of variation (CV)	39
5.3.4. Count.....	40
5.3.5. Trend.....	41
5.3.6. Percentile Ranks Accumulations	43
5.3.7. Frequency.....	44
5.3.8. Standardized Precipitation Index (SPI).....	45
Chapter 6: View and Explore Rainfall Summaries	47
6.1. Requirements.....	47
6.2. Calculate seasonal total and anomalies	47
Chapter 7: Contour Tool.....	49
7.1. Making Contours.....	49
Chapter 8: BASIICS	51
8.1. BASIICS.....	51
8.2. Validate Satellite Rainfall	52

8.3. Improve rasters with stations using the blending algorithm	56
Chapter 9: Extracting Raster Statistics and Time Series.....	62
9.1. Extract Statistics.....	62
9.2. The Results.....	64
Chapter 10: Calculate Long-Term Changes in Average.....	65
10.1 Estimate trends using difference in averages tool 	65
Chapter 11: Creating Archives.....	67
11.1. Create an Archive	67
11.2. Importing archives	68
References.....	70

Acknowledgements: GeoCLIM was developed by Tamuka Magadzire of the Famine Early Warning Systems Network (FEWS NET), and Karthik Vanumamalai, Cheryl Holen and Joshua Sickmeyer of the United States Geological Survey (USGS) 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 created by Diego Pedreros and Tamuka Magadzire with the editing support of 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: [Contour Tool](#) briefly reviews how to visualize spatial rainfall distribution based on contour lines.

Chapter 8: [BASICS](#) is a walkthrough of the process of blending station and raster data.

Chapter 9: [Extract Statistics](#) explains how to create summaries of historical rainfall for a given region.

Chapter 10: [Calculate Difference in Averages](#) shows a different way of estimating trends by comparing changes in averages for two periods.

Chapter 11: [Create archives](#) explains how to create climate data archives that could be shared with other GeoCLIM users.

Chapter 1: Overview

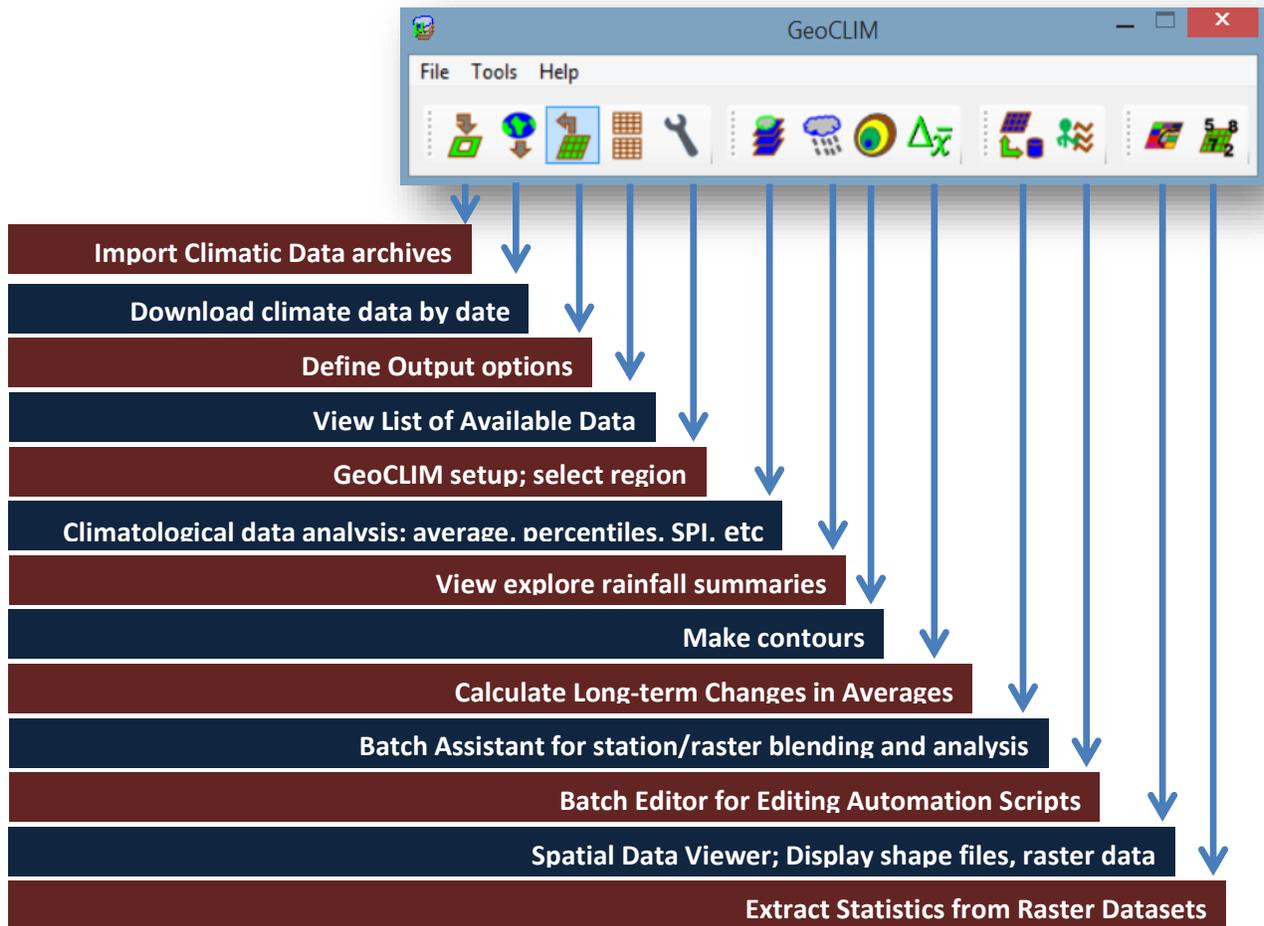


Figure 1.1 GeoCLIM Main Toolbar.

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 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 that can 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 more information on creating data archives, see [chapter 11](#).

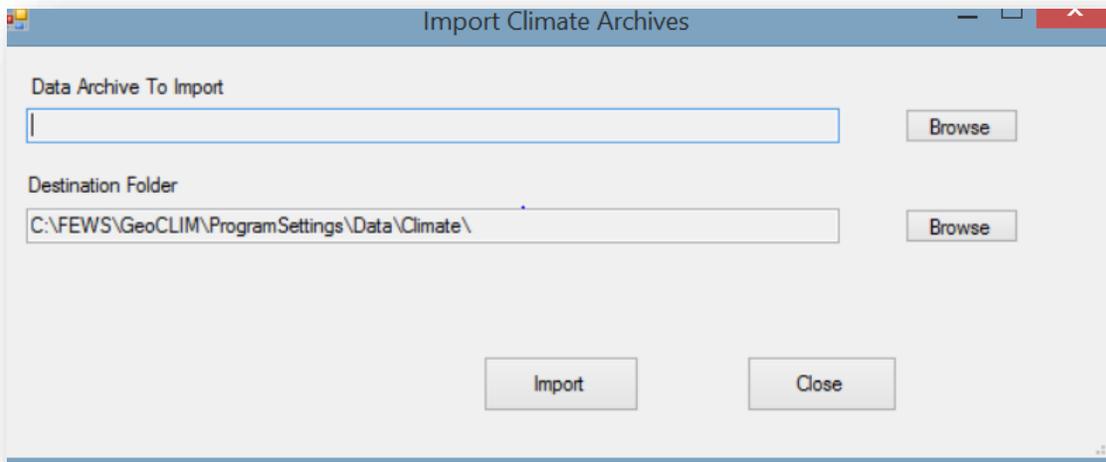


Figure 1.2 Climate data can be made available in the GeoCLIM through downloading archives.

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.). When using GeoCLIM for the first time, the default dataset is the CHIRPS_PPT_AFRICA_MONTHLY. See [chapter 2](#) for more information on selecting datasets, including changing the default dataset.

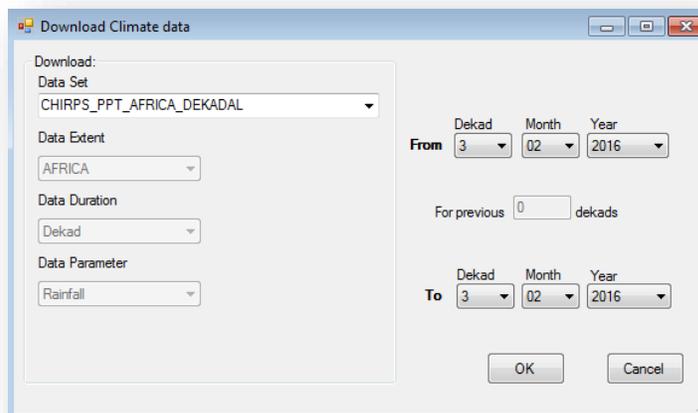


Figure 1.3 Rainfall, Temperature, or Evapotranspiration data can be downloaded directly from an FTP site.

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.

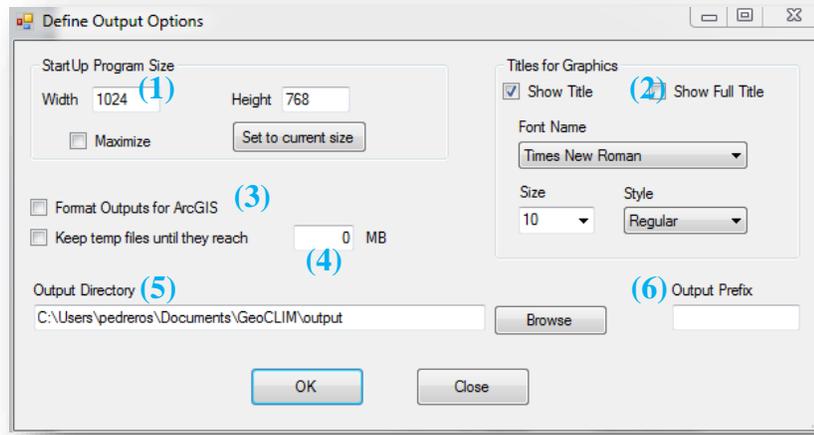


Figure 1. 4 The tool allows the user to define output settings.

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 1981. 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.

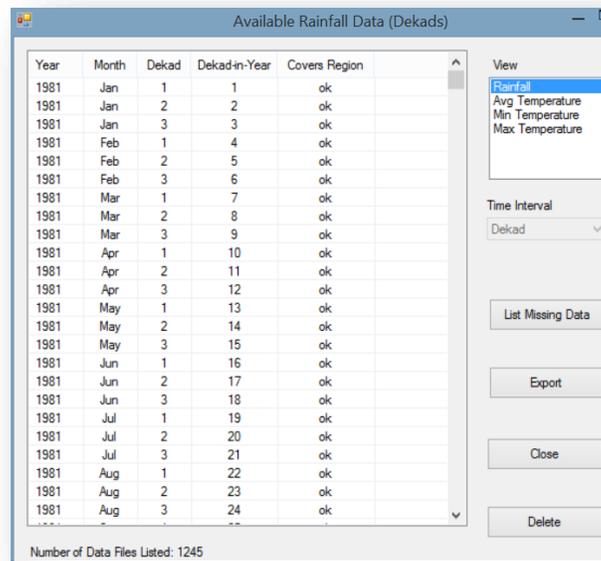


Figure 1.5 The user could list the available data from the different climate variables and their spatial coverage.

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.

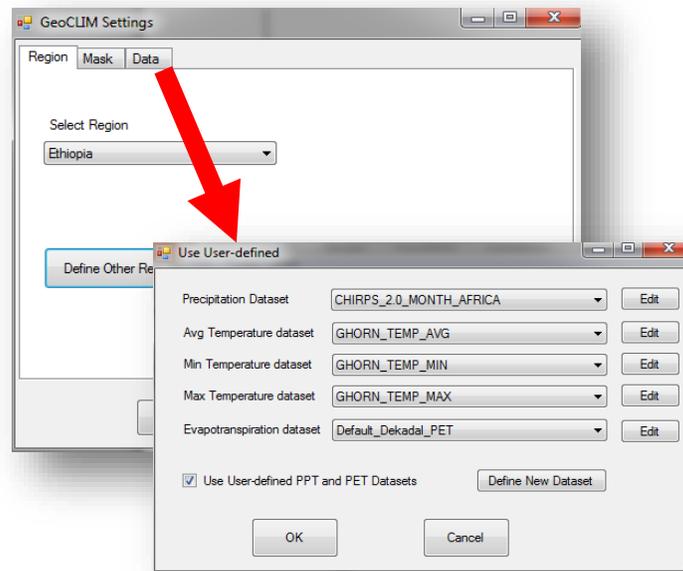


Figure 1.6 The settings tool allows the user to select the climate data and the region for analysis.

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 a long-term 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.

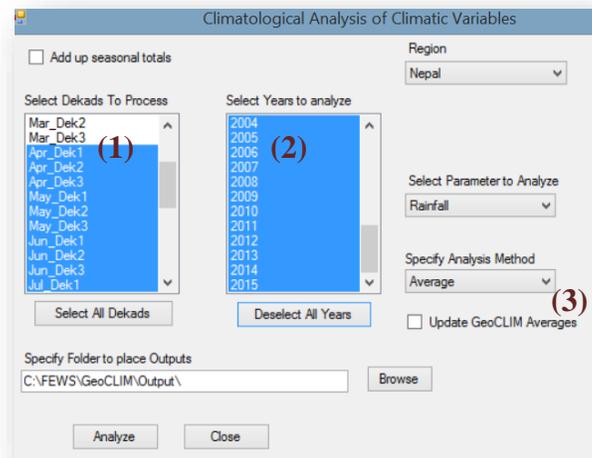


Figure 1.7 The user could calculate statistics, trends, SPI among other functions for specific seasons.

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 for the same period, or a percent of the long-term average. More details on this tool can be found in [chapter 6](#).

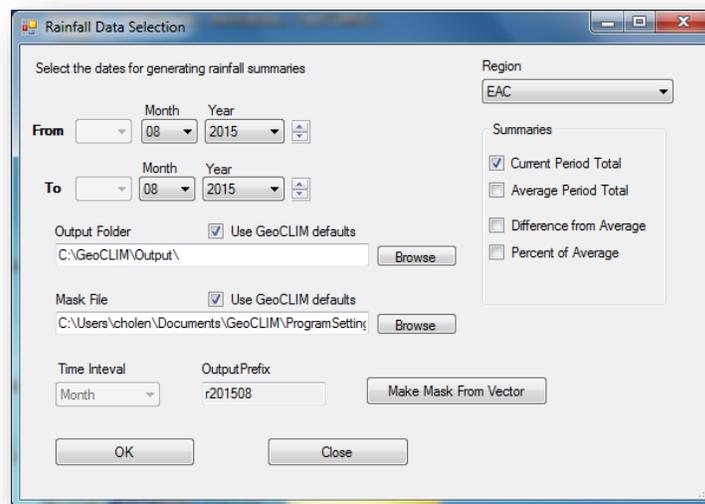


Figure 1.8 The rainfall summaries tool facilitates the calculation of rainfall total and anomalies for a selected period of time.

1.8. Make Contours



The *Make Gridded Contours* tool (Figure 1.9) allows the display of smoothed contours for a specified interval based on a raster BIL file that was generated in GeoCLIM, such as a rainfall accumulation. 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 7](#). The contours displayed in this tool are color-coded for ease of display.

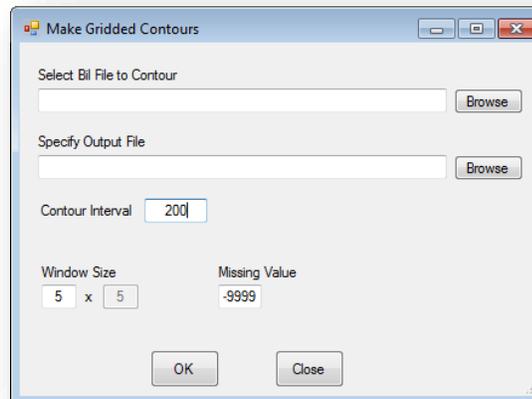


Figure 1.9 Display rainfall data based on contour intervals.

1.9. Calculate Long-Term Changes in Average



The *Climate Trends - Changes in Averages* tool (Figure 1.10) 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 1.10). 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 10](#) for more details.

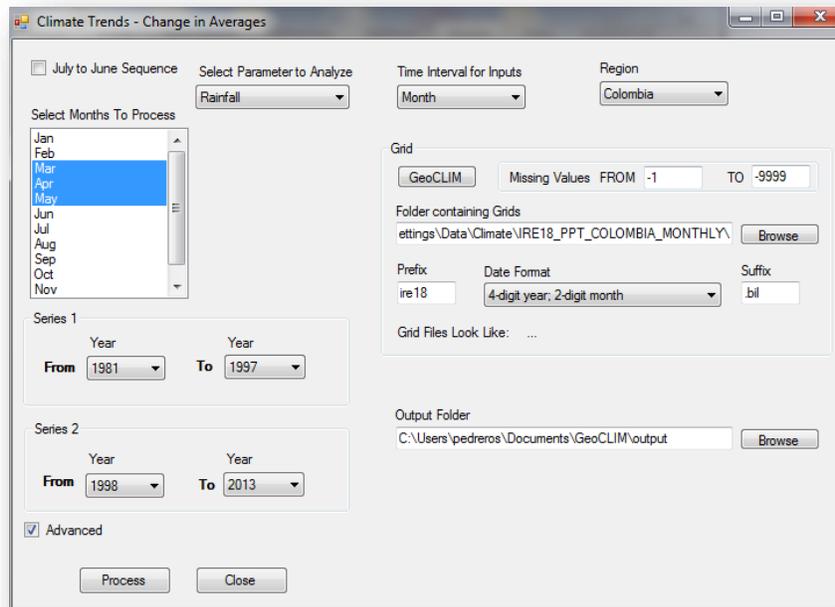


Figure 1.10 The change in average tool compares the average rainfall for two periods of time, identifying trends.

1.10. Batch Assistant for Easily Developing Automation Scripts

The *Batch Assistant* tool 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:

- **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 **Background Assisted Station Interpolation for Improved Climate Surfaces (BASICS)** 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.

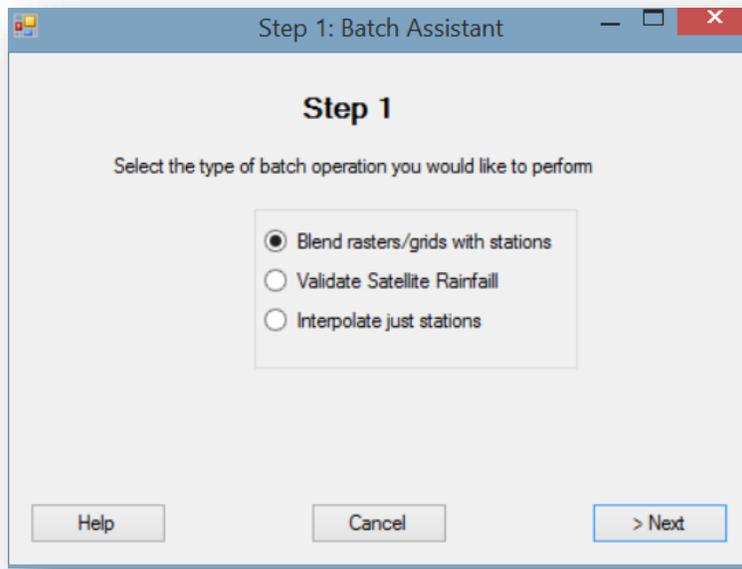


Figure 1.11 The Batch Assistant tool has functions that validate satellite estimated data using station values, blends station data with raster (BASIICS) and interpolates station data.

1.11. Batch Editor for Editing Automation Scripts

```

GeoCLIM Batch Text Editor - ire_Cagbl
File Edit Run Tools Help
BEGIN IRE
MULTISTATIONFILE "D:\Data\precipitation\Ground_stations\americas\Mario_WB_cafe
ca_month2.csv"
BEGIN FILEDETAILS
DELIMITER "comma"
MISSINGVAL -9999
NUMERLINES 1
LONGCOL 2
LATCOL 3
YEARCOL 4
HEADERROW 1
STRTCOL 1
MONTH1TO12COLS 5 16
END FILEDETAILS
GRIDFOLDER "C:\Users\pedreros\Documents\GeoCLIM\ProgramSettings\Data\Climate
\CHIRPS_PPT_GLOBAL_MONTHLY\"
GRIDPREFIX v1p8chirps
GRIDDATEFORMAT YYYYMM
GRIDSUFFIX .b11
MISSINGVALRANGE -9 -9999
DATERACKET FROM 2012/04/1 TO 2013/12/1
OUTPUTSTATFILE "D:\Data\precipitation\GEOClim_archives\ire18_ppt_CA_MONTHLY\ire_ca4.csv"
DIAGNOSTICOUTPUTS True
OUTPUTFOLDER "D:\Data\precipitation\GEOClim_archives\ire18_ppt_CA_MONTHLY\"
OUTPUTPREFIX ire
OUTPUTDATEFORMAT YYYYMM
OUTPUTSUFFIX .b11
WEIGHTPOWER 2
MINSTNS 0
MAXSTNS 10
SEARCHRADIUS 100
FUZZFACTOR 1
MAXRATIO 3
MAXEFFECTIVEDIST 50
FORCEDLONGRANGEVAL 1
INTERPOLATIONALGORITHM 1dw_g
UPPERLEFT_XY -93 20
LOWERRIGHT_XY -76.5 6.5
END IRE

```

Figure 1.12 The BASIICS process could be re-run using the batch assistant tool.

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 function. See Figure 1.12.

1.12. 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 1.13. See the [Spatial Data Viewer](#) section for more details on this tool.

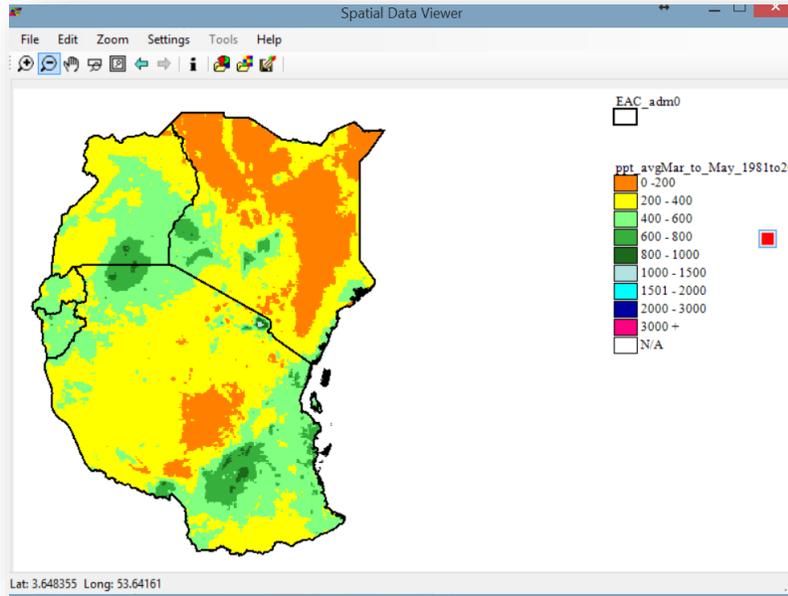


Figure 1.13 The spatial data viewer allows for the display and editing of raster data and shape files.

1.13. Extract Statistics from Raster Datasets

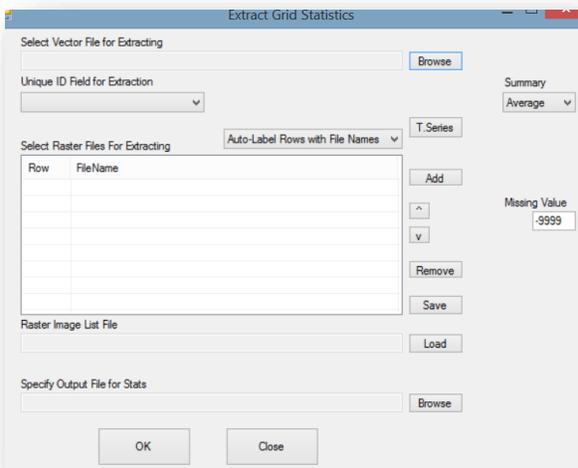


Figure 1.14 The user could extract spatially statistics based on polygons.

The *Extract Grid Statistics* tool produces a table with areal statistics such as averages, totals, and max/min for raster data over a set of polygons. See Figure 1.14. 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 9](#) for more information.

Summary

This chapter focuses on the GeoCLIM settings such as working with 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 the selected region, the mask for the region, or the selected climate datasets by clicking on the  icon from the GeoCLIM toolbar

- 2.1.1. Select the **Region** of interest. If the desired region is not in the predefined list, see [section 2.6](#) to learn about how to create a 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).

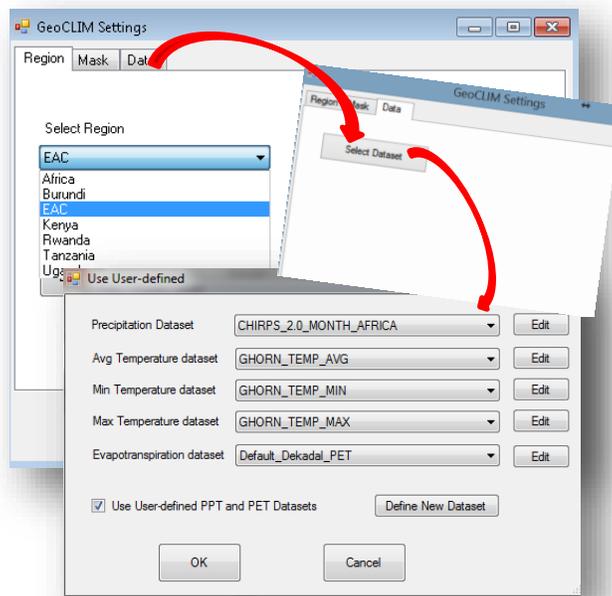


Figure 2.1 The GeoCLIM Settings allow the user to select a new region, a mask and/or climate datasets.

2.2. Making data available on the GeoCLIM

Users can create their own climate datasets to use with the GeoCLIM’s analysis tools. For example, the Kenya Meteorological Department may create a new dekadal rainfall dataset in-house by blending gridded satellite rainfall estimates and rainfall station data 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.

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: yyyyymm means 4 digits year followed by 2 digit month (01, 02, 03...12). More formats are available for pentadal, dekadal, 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 1981 from CHIRPS 2.0, the recommended name of the BIL file could be *v2p0chirps198136.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 (1981) and a 2-digit dekad (36), and the <suffix> is the extension of BIL files (.bil). To learn more about the data formats used on the GeoCLIM, see [Chapter 3](#)

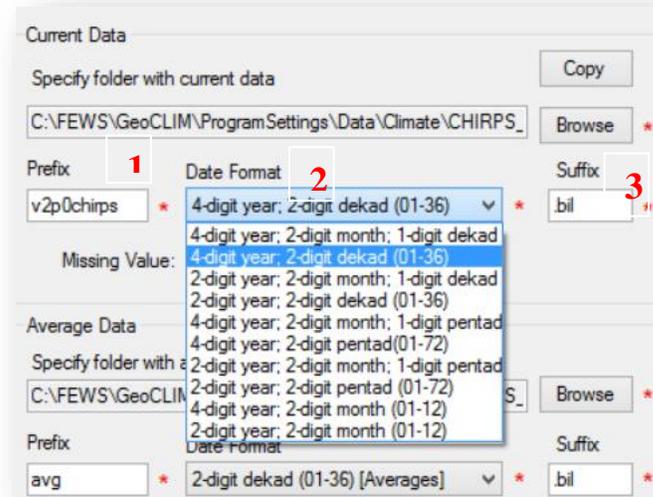


Figure 2.2 The name of the data file must be composed of a prefix, a date (year and period of time) and a suffix.

2.2.2. Define a new dataset in the GeoCLIM

Once the data are ready, the next step is to create a climate dataset definition in the GeoCLIM. The climate dataset definition specifies the folder location of the data, the file-naming convention (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) Type the name of the new climate dataset for the data, for example:
NEW_PPT_CHIRPS_DEKADAL.

NOTE: the name of the climate dataset and its directory must not have spaces.

- 2) Create a new sub-directory which will contain the new data. This new subdirectory should preferably be under the `\GeoCLIM\ProgramSettings\Data\Climate` directory. The new subdirectory **must** have same climate dataset name defined in step 1. In this example, the new location would be
`GeoCLIM\ProgramSettings\Data\Climate\NEW_PPT_CHIRPS_DEKADAL`

- 3) Copy the BIL raster dataset (*.bil and *.hdr files) into the new directory created in step 2

- 5) **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).
- 6) **Select the data Type:** select the type according to the data; Precipitation, Avg Temperature, Min Temperature, Max Temperature, or Evapotranspiration.
- 7) **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*
- 8) **Specify folder with current data:** Browse to the new directory
`GeoCLIM\Programsettings\Data\Climate\NEW_PPT_CHIRPS_DEKADAL`
- 9) **Prefix:** Fill out the prefix as defined above.
- 10) **Select the date-format.** For this example, select *4-digit year; 2-digit dekad (01-36)*
- 11) Fill out the **suffix** *.bil*. NOTE: do not forget the ‘.’
- 12) Fill out the **missing value**, for example for CHIRPS data the missing data value used is - *9999*.
- 13) 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.
- 14) Give a **prefix** to the average files, for example: *avgppt*.
- 15) Fill out the **date-format**. For this example, select *2-digit dekad (01-36) [Averages]*
- 16) The **suffix** is the same *.bil*.
- 17) The **missing value** should be the same as defined for the current data.
- 18) **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** twice to close all the windows. Once a dataset is selected on the settings, all the GeoCLIM tools 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

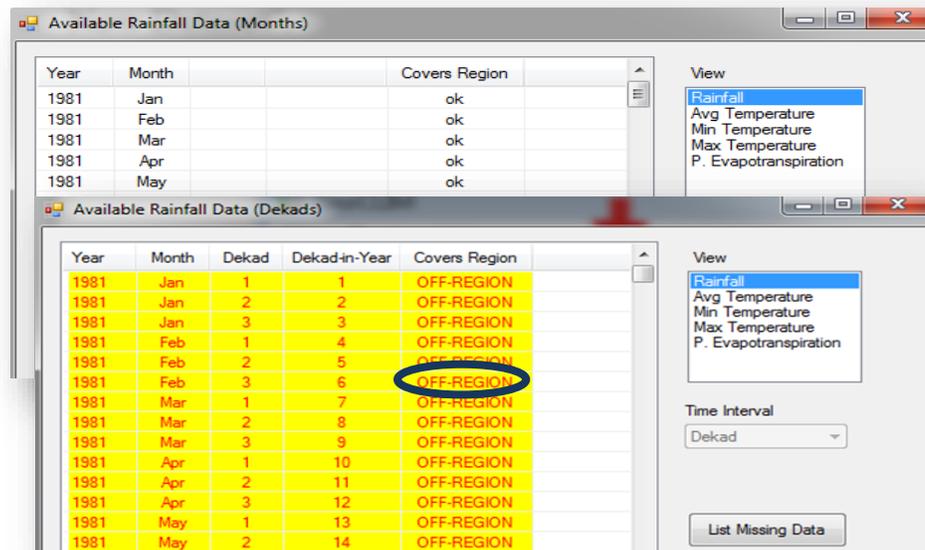


Figure 2.4 The available Data window shows the length of the time series and how well the data extent covers the region of work.

proceed with the analysis tools. 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 8) 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:

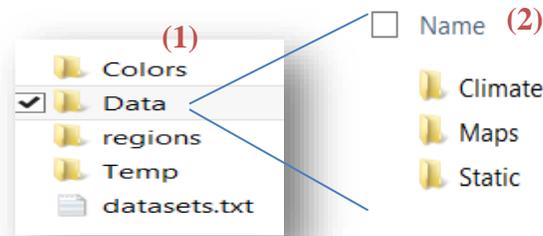


Figure 2.5 The GeoCLIM directory contains two folders; the 'output' where all the results go by default, and the 'ProgramSettings' that contains the 'Data' directory among others.

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

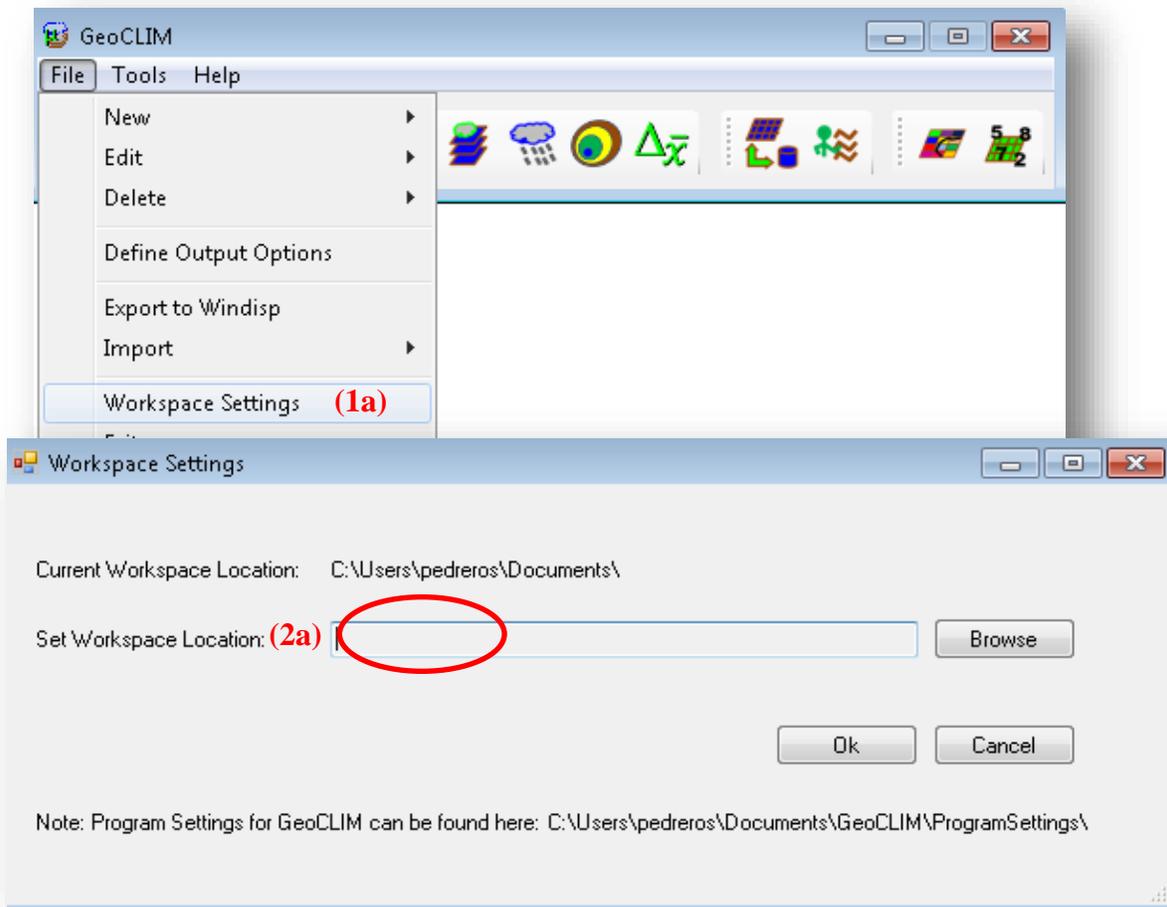


Figure 2.6 The GeoCLIM allows the user to change the workspace directory from the default to a new location.

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

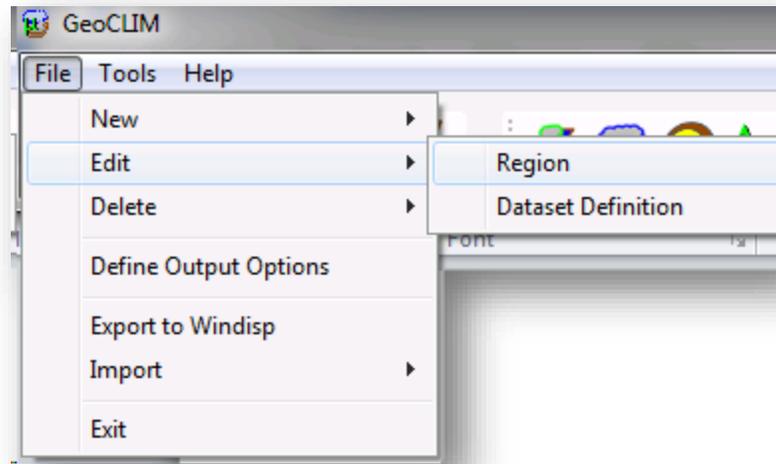


Figure 2.7 To open an existing region, go to File >Edit>Region

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

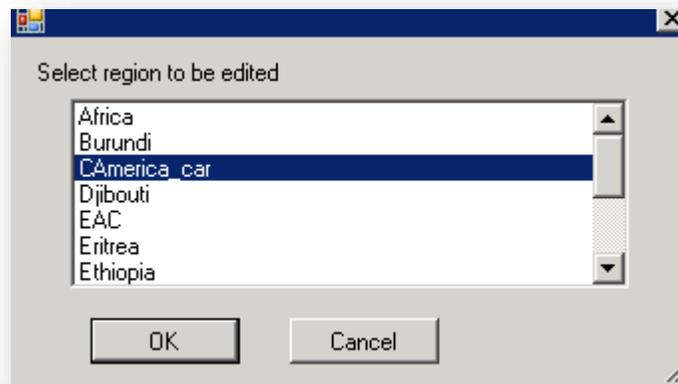


Figure 2.8 The GeoCLIM contains several regions that could be used as example to create new one.

- 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.

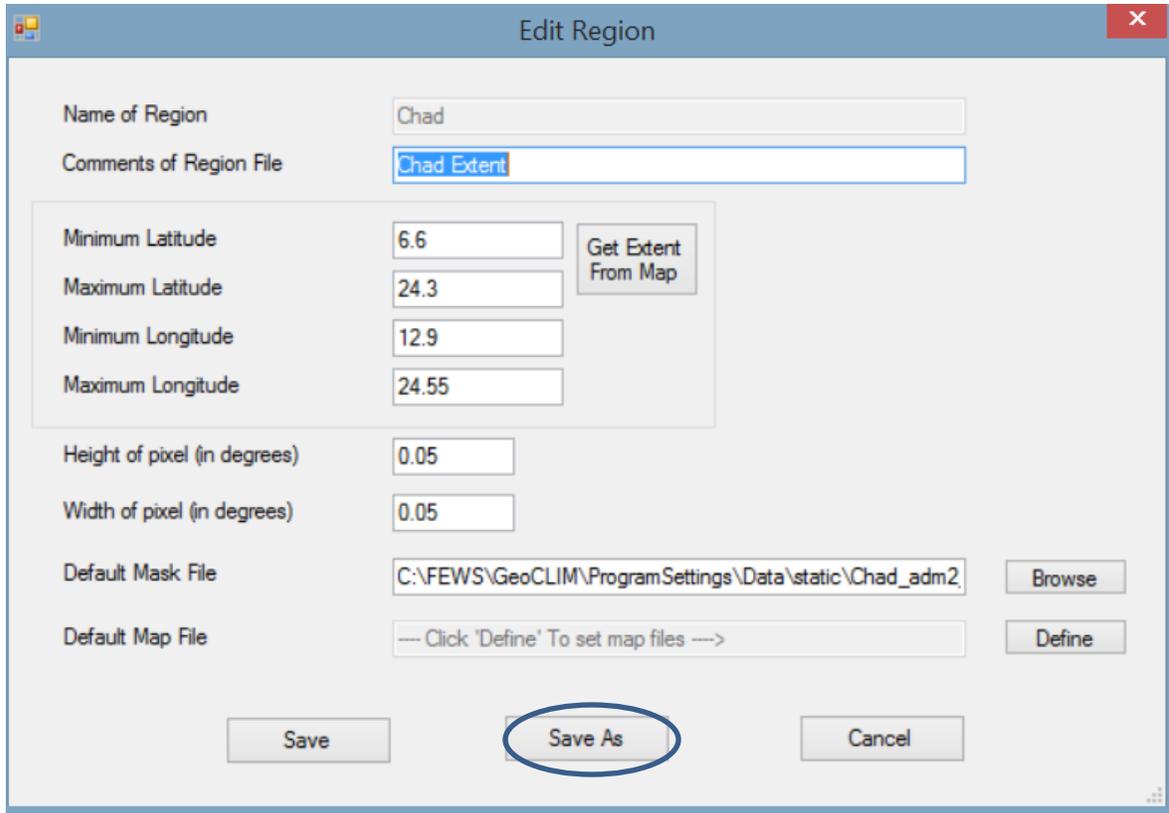


Figure 2.9 The 'Edit Region' form describes the geographic boundaries of the region, the pixel size for the outputs, the mask used and the outline shape file.

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 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.

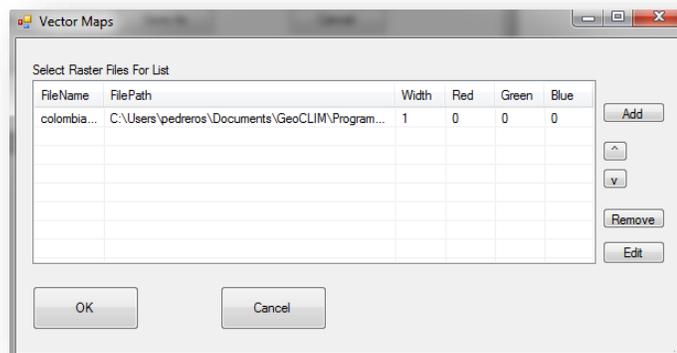


Figure 2.10 A shape file that serves as the outline on the output products.

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.

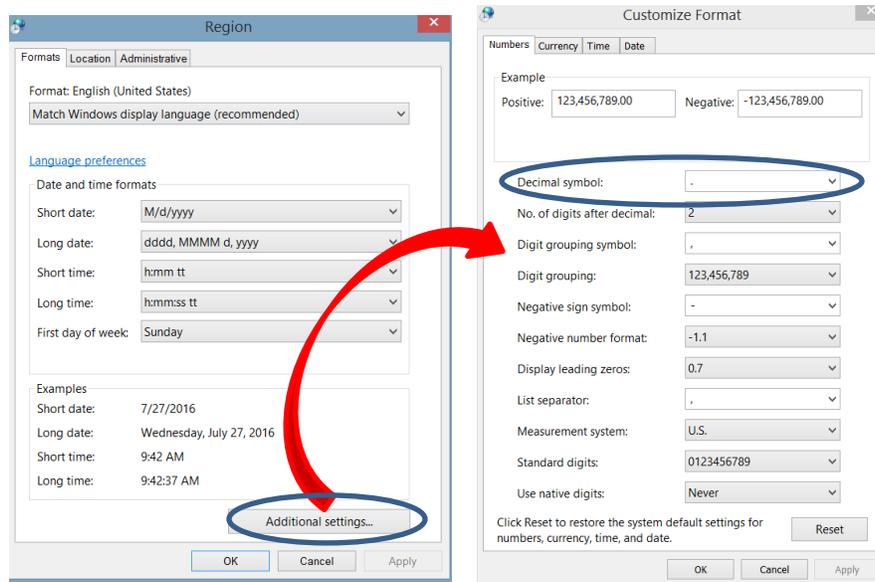


Figure 2.11 One possible error when using the GeoCLIM is the decimal separator. Go to the Additional settings and change the ‘Decimal symbol’

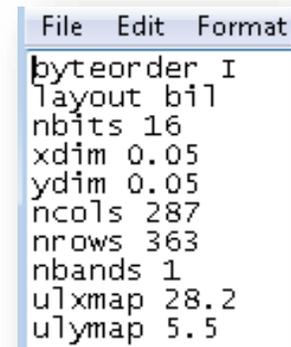
Summary

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.

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 2008](#)). Sometimes it is important to modify the HDR file so that the data is read correctly by the program.



```
File Edit Format
byteorder I
layout bil
nbits 16
xdim 0.05
ydim 0.05
ncols 287
nrows 363
nbands 1
ulxmap 28.2
ulymap 5.5
```

Figure 3.1 Example of a HDR file.

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 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=8` means 8 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 (1.1.2) 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*, *lon*, *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.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	id	lat	lon	year	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
2	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
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
4	566	-3.35	37.33	1900	17.6	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999
5	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
6	569	-5.08	39.07	1901	-9999	-9999	-9999	-9999	-9999	-9999	20.2	20	20	20.9	22.5	24.4
7	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
8	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
9	566	-3.35	37.33	1901	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	17.4
10	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
11	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
12	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
13	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
14	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
15	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
16	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
17	603	-0.35	31.78	1902	-9999	-9999	-9999	-9999	-9999	-9999	15.6	16.5	-9999	15.9	15.7	15.7
18	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
19	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
20	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
21	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
22	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
23	439	9.92	45.25	1903	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	11.7	5	5.6
24	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
25	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
26	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 3.8 Tables of station data values are used in Comma delimited tables (CSV).

Chapter 4: Spatial Data Viewer

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.

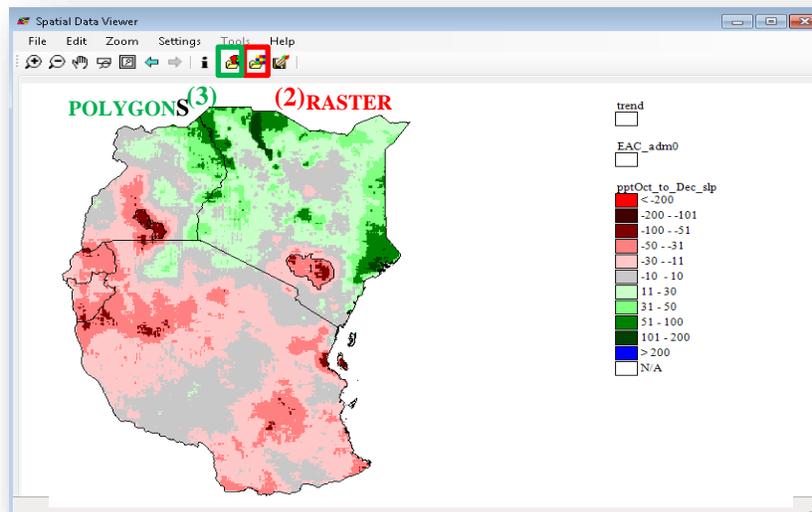


Figure 4.2 The GeoCLIM allows the user to work with both raster (.bil format) and shape files.

- 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 images displays the ranges of values, see Figure 4.3b.

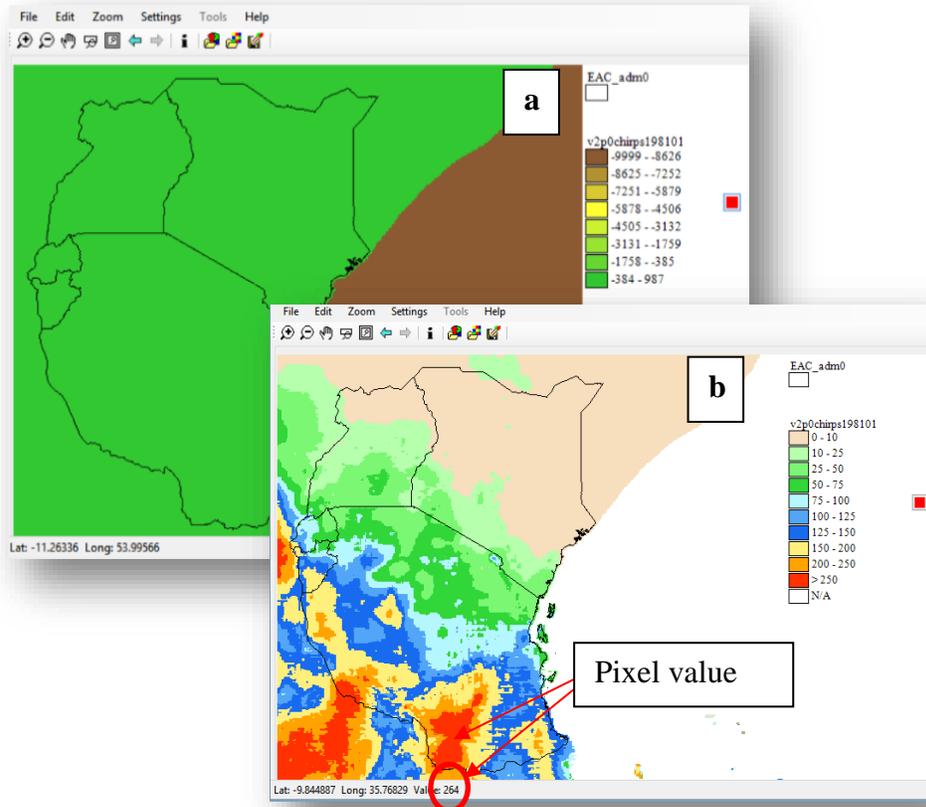


Figure 4.3 The GeoCLIM allows the use of raster data in .bil format. (a) Shows a raster dataset using the default legend (b) Shows the same raster file after assigning a different legend, and also shows how to view actual pixel values.

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 part of it from the selected climate dataset, such as for El Niño years for the March-April-May season

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.

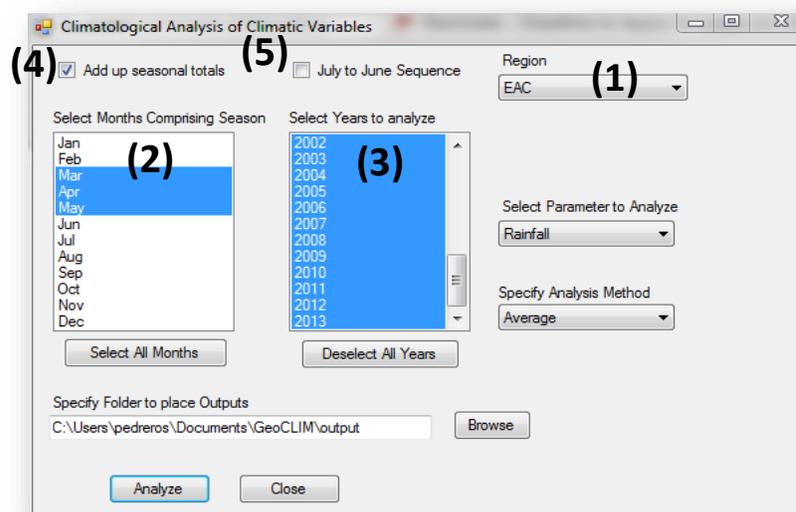


Figure 5.1 The Climatological Analysis tool facilitates the calculation of statistics, trends, SPI among others, using the complete or part of the time series.

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.

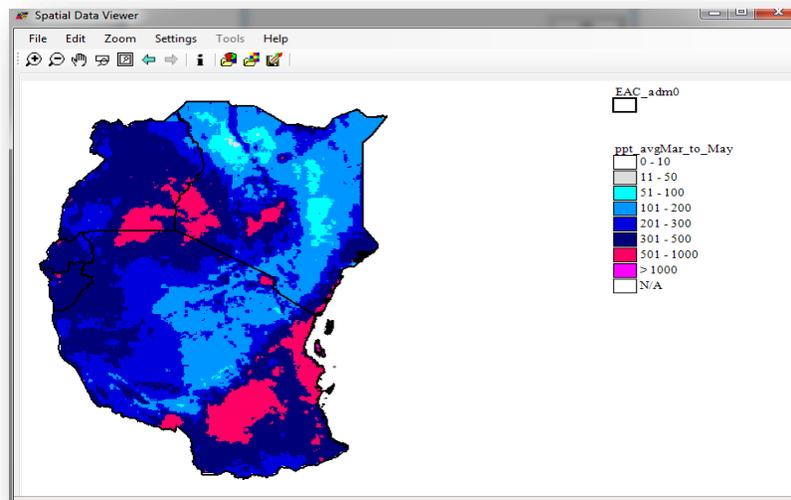


Figure 5.2 Average rainfall for the period March-May 1981-2013

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.

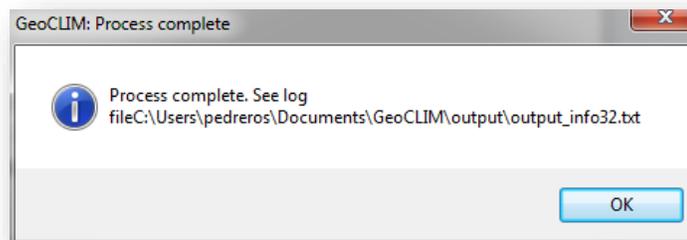


Figure 5.3 A log file is created after each process.

NOTE: If multiple periods are selected 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.

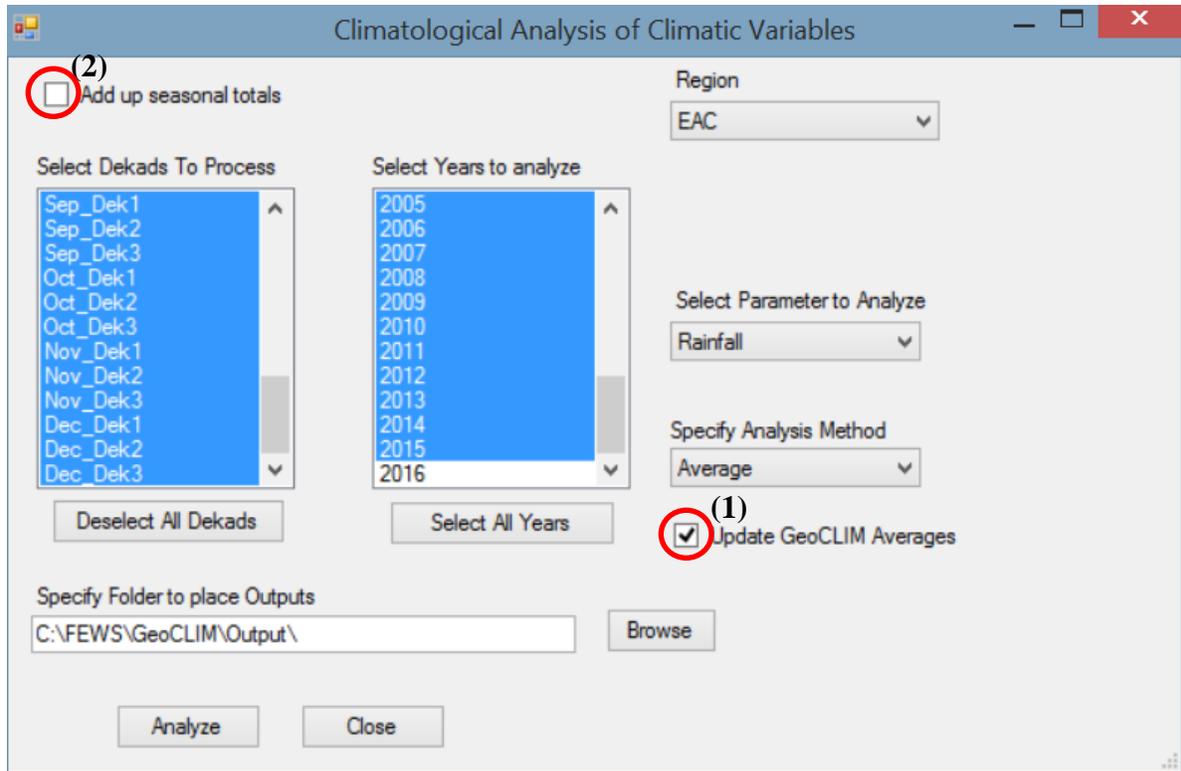


Figure 5.4 Calculate average rainfall for each pentad, dekad or month using the Climatological analysis tool.

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).

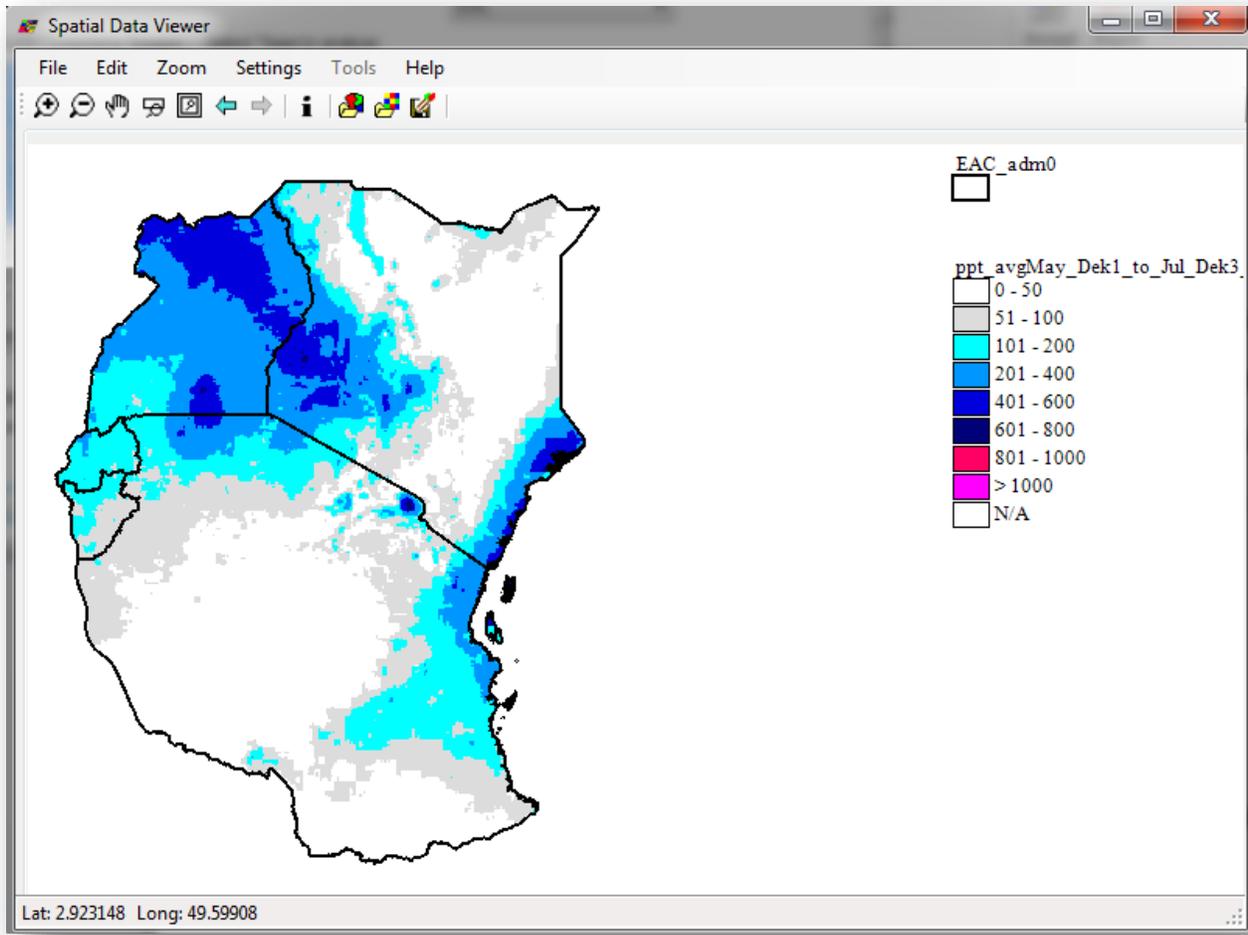


Figure 5.5 Average rainfall for the May-July season for the years 1981-2013.

- Name
-  pptsum1998May_Dek1_to_Jul_Dek3.bil
 -  pptsum1998May_Dek1_to_Jul_Dek3.hdr
 -  pptsum1999May_Dek1_to_Jul_Dek3.bil
 -  pptsum1999May_Dek1_to_Jul_Dek3.hdr
 -  pptsum1996May_Dek1_to_Jul_Dek3.bil
 -  pptsum1996May_Dek1_to_Jul_Dek3.hdr
 -  pptsum1997May_Dek1_to_Jul_Dek3.bil
 -  pptsum1997May_Dek1_to_Jul_Dek3.hdr
 -  pptsum1993May_Dek1_to_Jul_Dek3.bil

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.

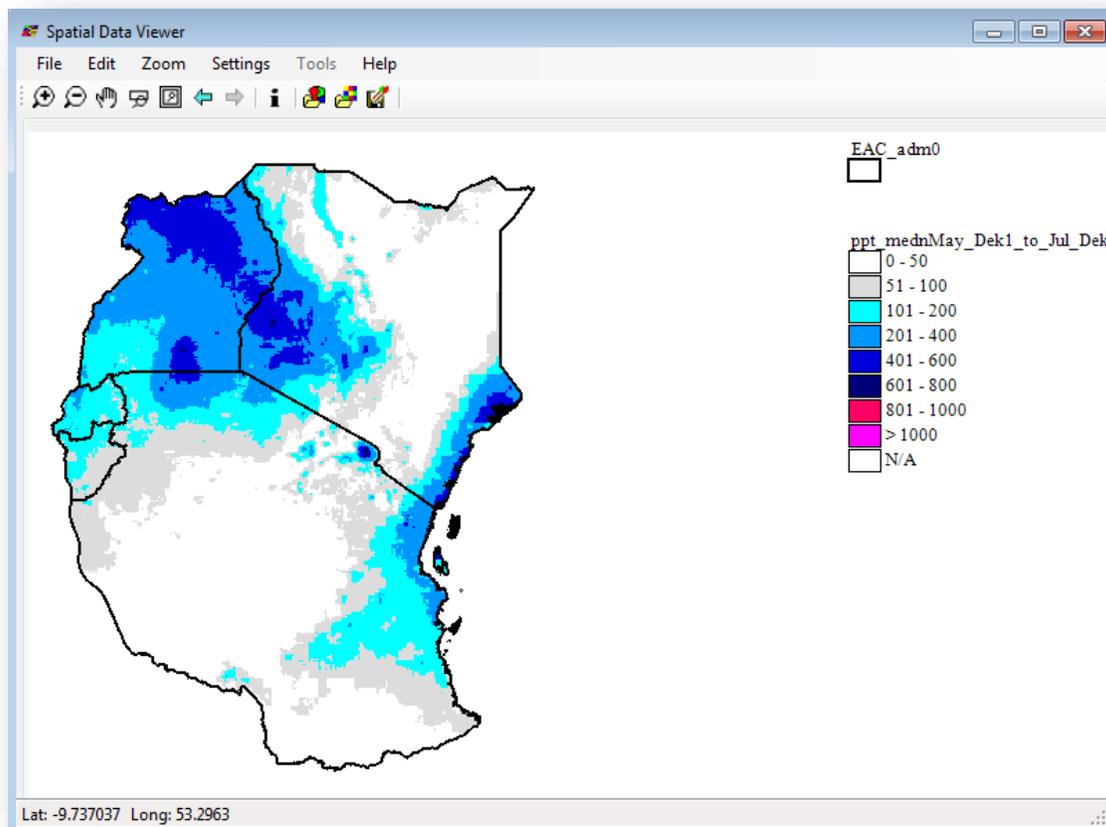


Figure 5.7 Median for the season May-July for the years 1981-2013.

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 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). Note that the actual pixel values generated in the outputs of the SD calculation and the CV calculation are SD multiplied by 100 and CV multiplied by 100 respectively

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 5.8).

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

© 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).

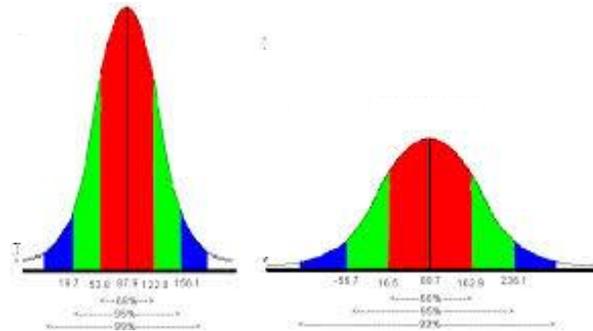


Figure 5.9 The graph on the left shows a low SD, low variability within the data; values are close to the mean. The graph on the right shows the distribution of a more variable data set.

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

$$CV = \text{SD} / \text{average} * 100$$

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).

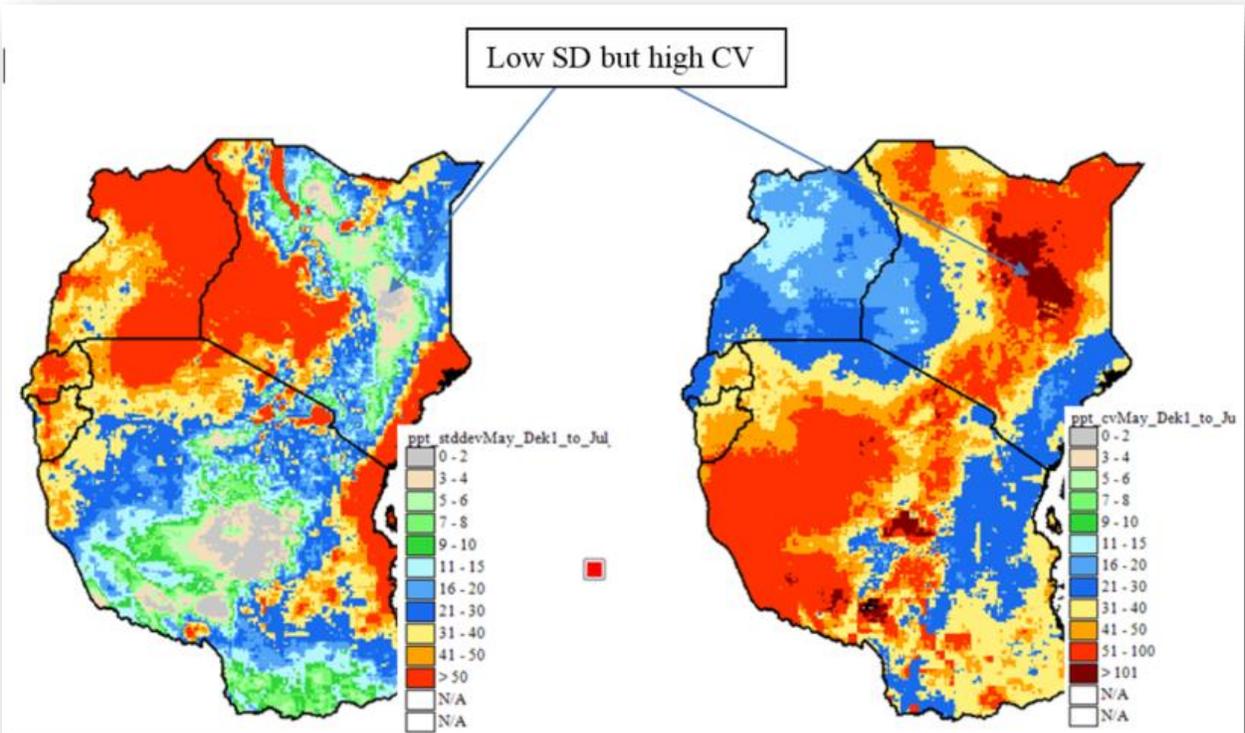


Figure 5.10 Measuring variation; the SD (a) shows the variation of rainfall in mm, while the CV (b) presents the SD as percent of the average allowing the comparison among regions. The SD of the area pointed by arrow shows a low value but the average is so low that it is a very variable region as shown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

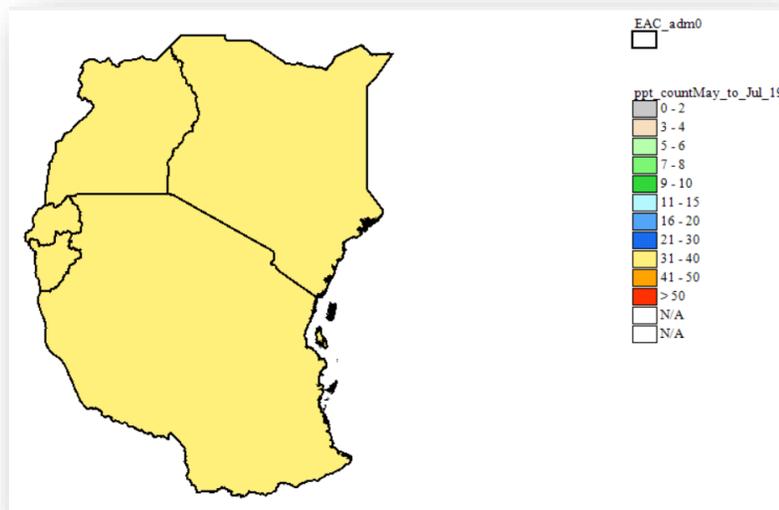
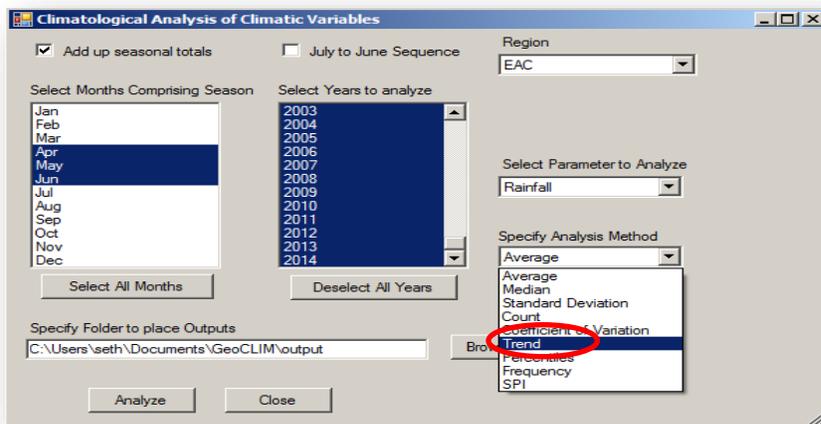


Figure 5.11 The count function looks at the time series for each pixel and counts the number of times there is a valid, non-missing value.

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 totals.



*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.

The trend method produces two maps, Figure 5.13(1) shows the slope of the regression line or mm of rain gained or lost per decade (10 years) on average. The red colors indicate the mm lost per decade and green to blue colors indicate mm gained per 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 consistency of the trend.

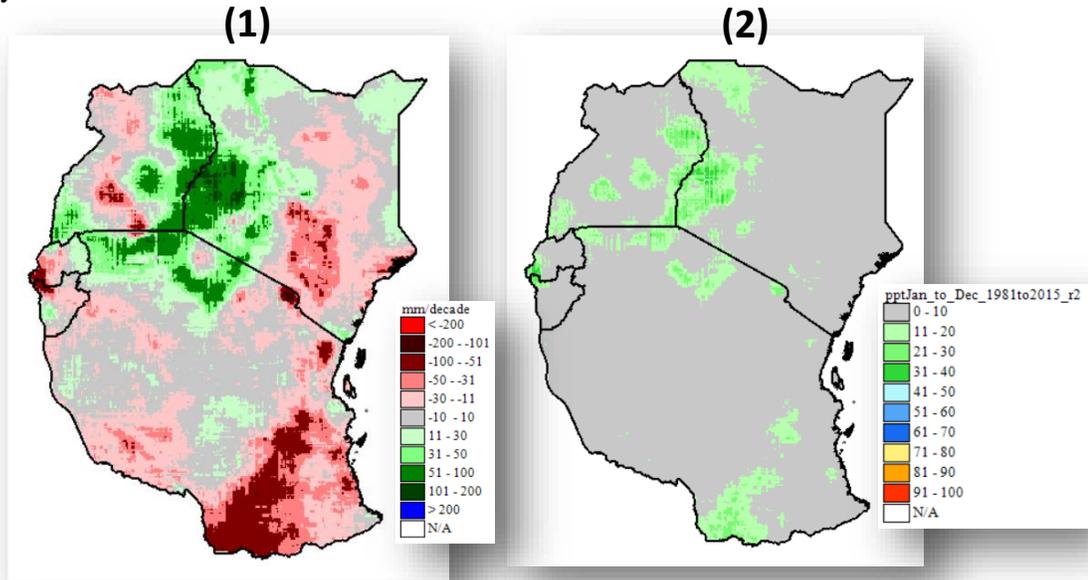


Figure 5.13 The trend function in the GeoCLIM produces two maps. The map on the left (1) shows the slope of the regression in mm lost (- on red) or gained (+ green-blue). The map on the right (2) shows the r^2 of the regression.

5.3.6. Percentile Ranks Accumulations

Percentile rank is the position in a frequency distribution below which a certain percentage of events are equal to or less than the specified percentile rank. For example, a percentile rank of 33% represents the maximum value of the bottom tercile of a distribution, as used in seasonal forecast maps. Because each grid cell in a map has its own unique climatology, each will have its own 33rd percentile value. It is useful to know these values and map them, to appreciate the accumulation of rainfall in millimeters associated with a percentile rank of interest. Figure 5.14 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. On the other hand, 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.

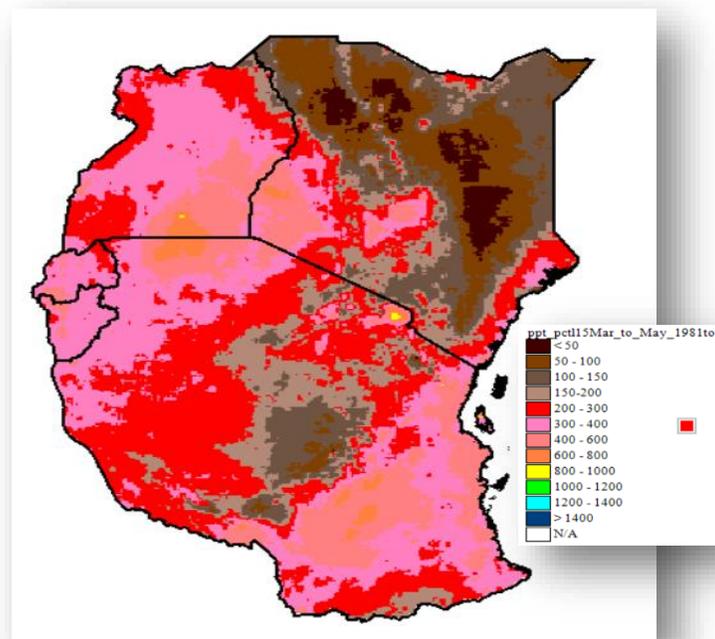


Figure 5.14 The map shows, by pixel, the rainfall value that is lower or equal to 15% of the values in the frequency distribution.

5.3.7. Frequency

The Frequency analysis method in the GeoCLIM Climatological Analysis module (Figure 5.15) calculates the number of times a range of values has occurred in the time series. The legend of the map on Figure 5.16 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.

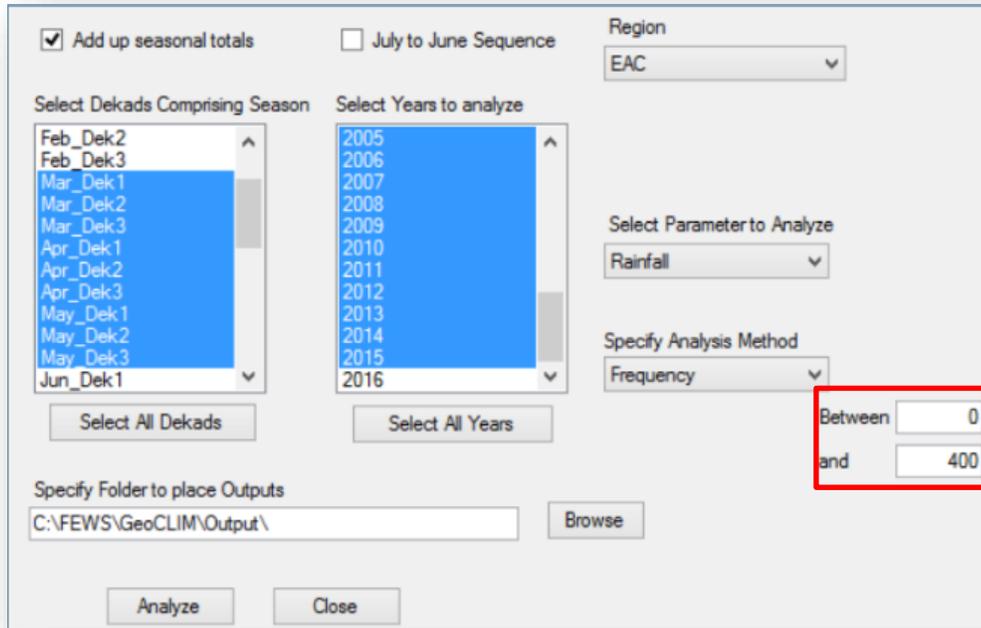


Figure 5.15 Select the season, the number of years and range of rainfall values (red box) to be analyzed.

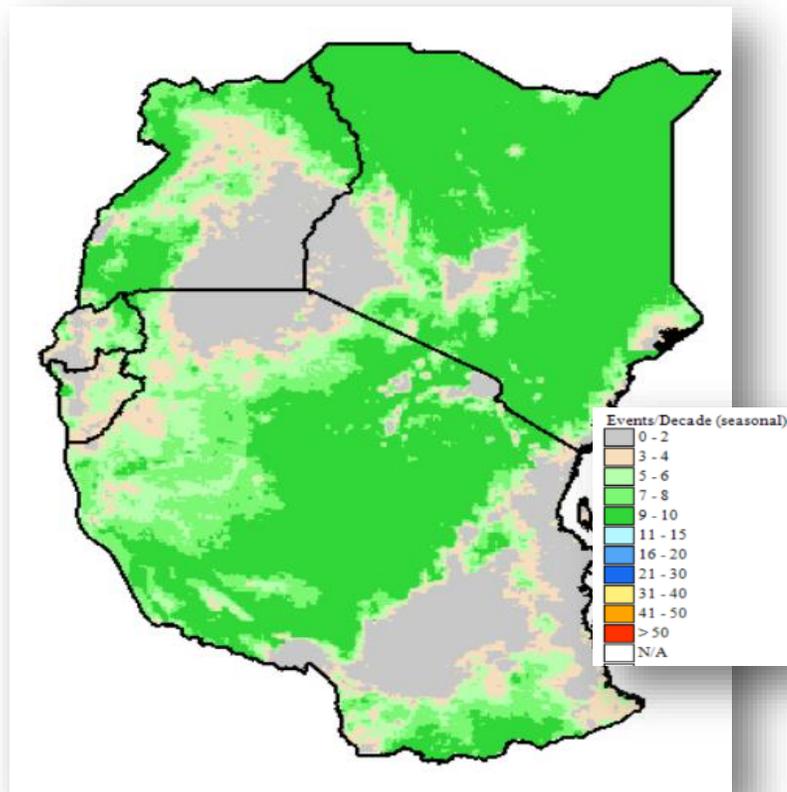


Figure 5.16 The tool calculates the number of times the selected range of values took place during the years selected.

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 5.17). (Early warning USGS-FEWS NET product documentation website)

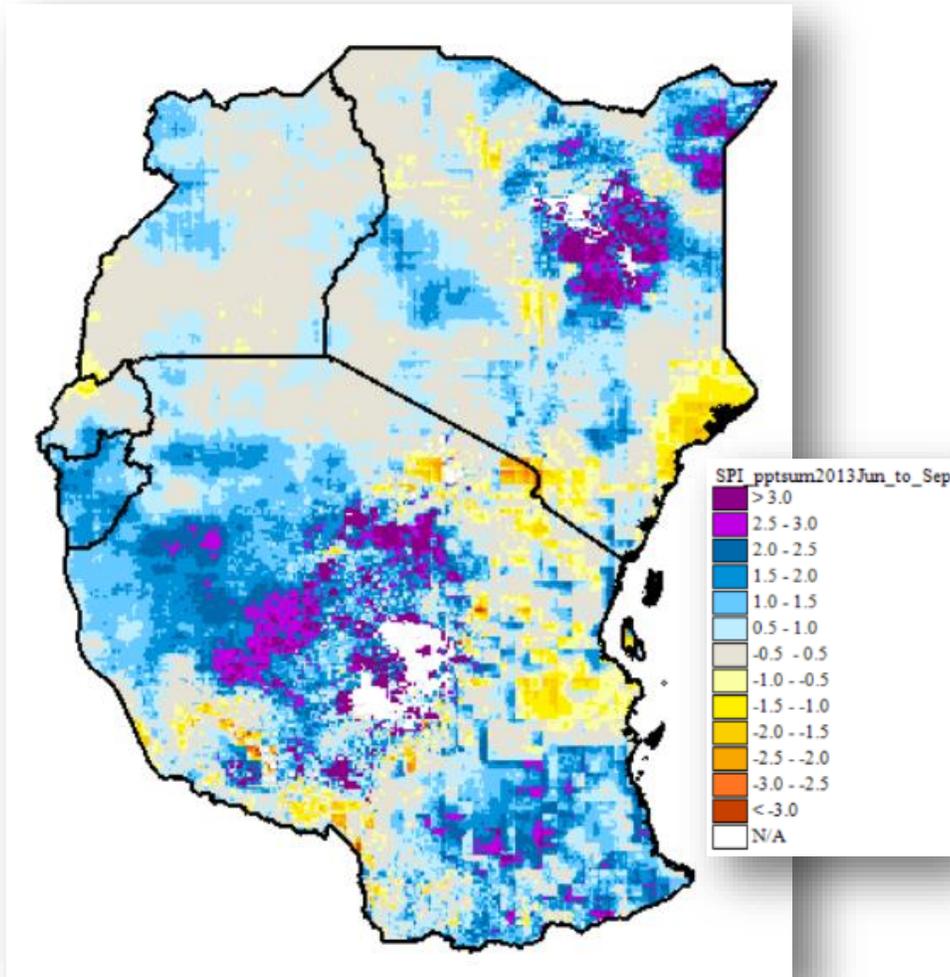


Figure 5.17 The dataset that is produced with the SPI function, in the GeoCLIM, is in units of [SPI multiplied by 100], but the legend shows actual SPI values.

Chapter 6: View and Explore Rainfall Summaries

Summary

The **Rainfall Summaries** tool (Figure 6.1) 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.

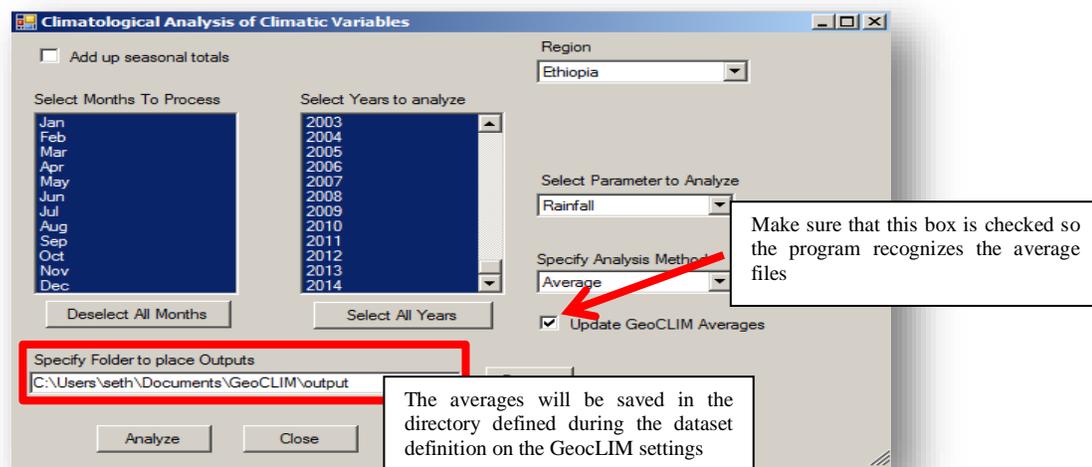


Figure 6.1 In order to run the rainfall summaries tool, it is necessary to calculate the average for each pentad, dekad or month.

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:** $(\text{current} / \text{average}) * 100$

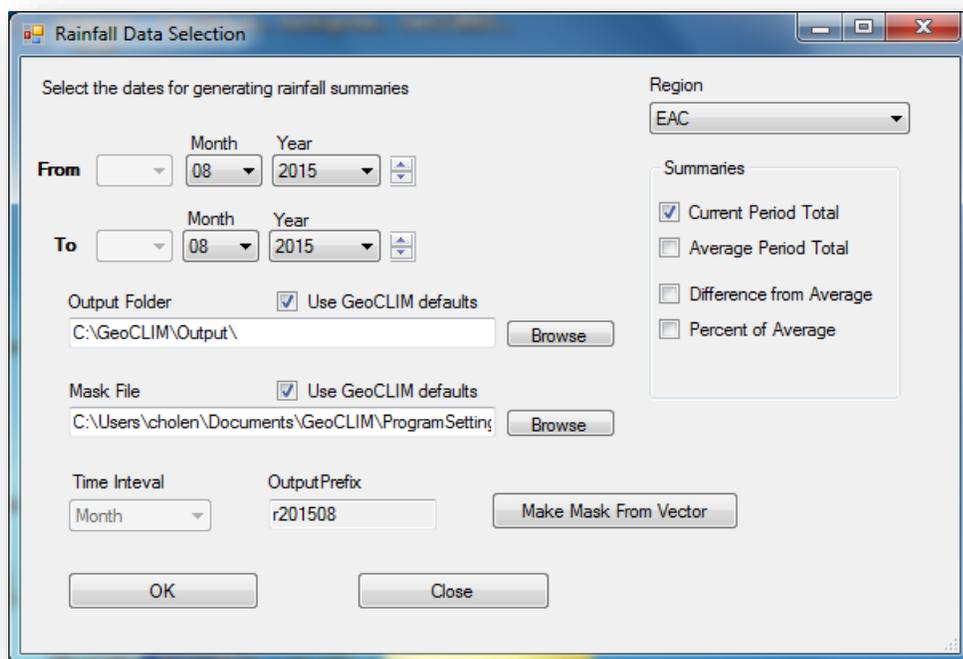


Figure 6.2 The rainfall summaries tool enables the calculation of seasonal total, averages and anomalies for a specific region.

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.

7.1. Making Contours

- 1) Open the **Make Contours** tool from the GeoCLIM main toolbar. See Figure 7.1.
- 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.

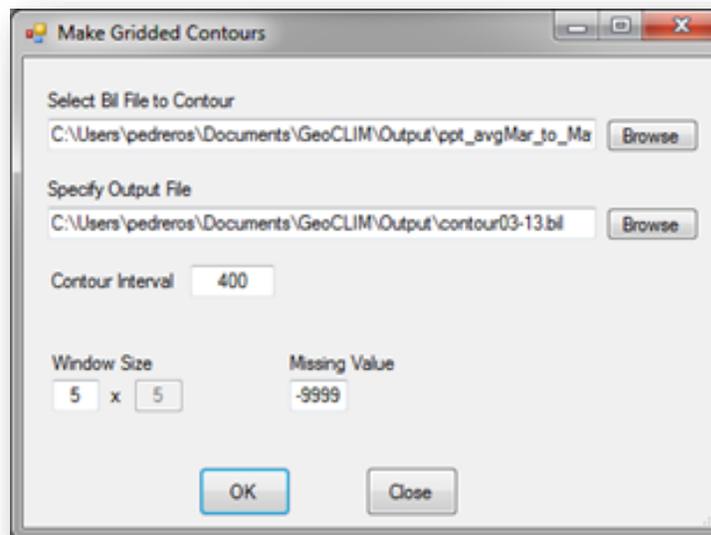


Figure 7.1 Select a rainfall dataset and the contour interval.

The contour tool can be used to analyze changes in average rainfall for different periods of time. For example, Figure 7.2 shows the changes in average for the years 1981-1992, 1993-2004, and 2005-2016 for the East Africa Countries (EAC) for the March-April-May season.

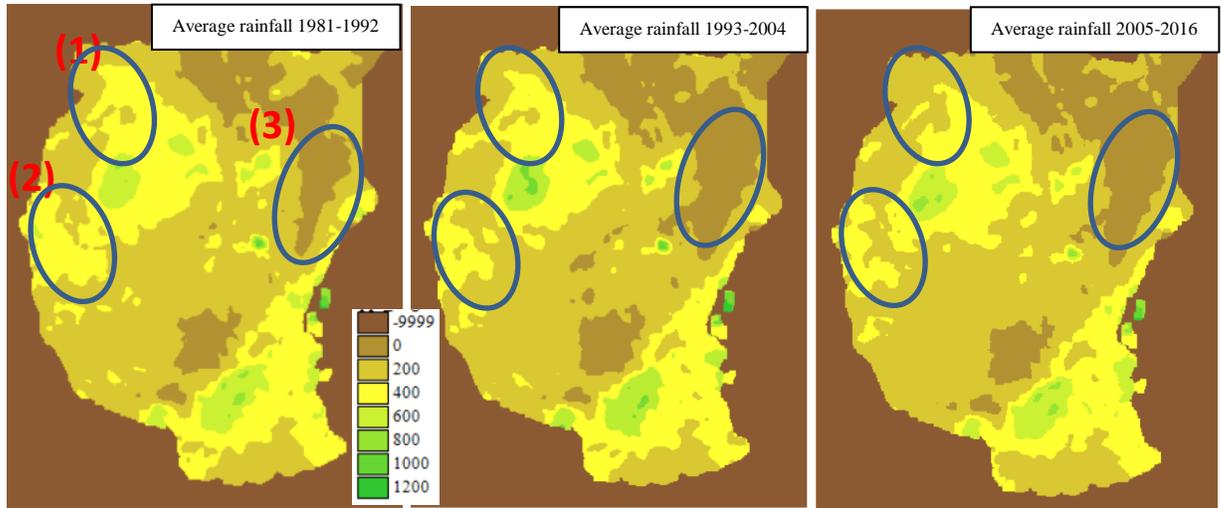


Figure 7.2 The 200 mm interval for the average rainfall of the season March-April-May for the years 1981-1992, 1993-2004, and 2005-2016 show that areas of 400mm in the western part of the region (polygons 1 and 2) are changing into 200mm. Also, areas of 200mm in Kenya are changing into the zero interval (polygon 3). These polygons show that rainfall is decreasing in these regions.

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.

8.1. BASIICS



The **Background-** **Assisted Station Interpolation for Improved Climate Surfaces (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 estimate by blending them with stations values.
4. Use batch files to run the algorithm for subsequent periods.

8.2. Validate Satellite Rainfall

One of the batch operations of the GeoCLIM is the **Validate Satellite Rainfall** operation, see Figure 8.1, which allows to comparison of satellite rainfall with station data.

To use this operation, follow the three steps below:

Step 1

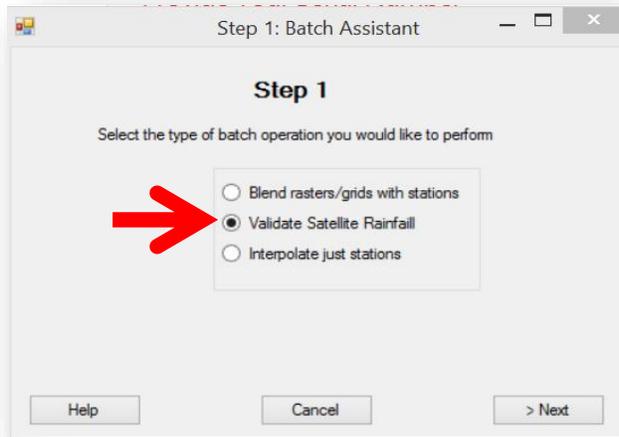


Figure 8.1 There are three functions; Blend stations and raster data, Validate Satellite Rainfall and Interpolate Just Stations.

- Click on the **BASIICS** icon from the GeoCLIM toolbar to open the Batch Assistant dialog box (**Step 1**), see Figure 8.1.
- 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.

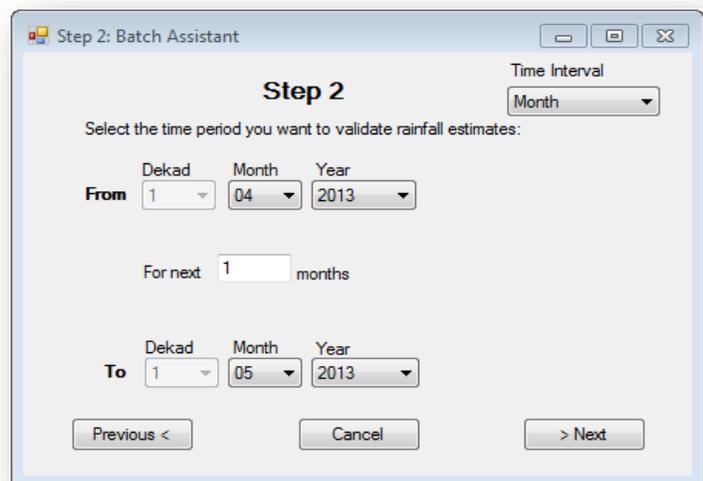


Figure 8.2 Step 2 requires the dates 'from and to' of the data to be analyzed.

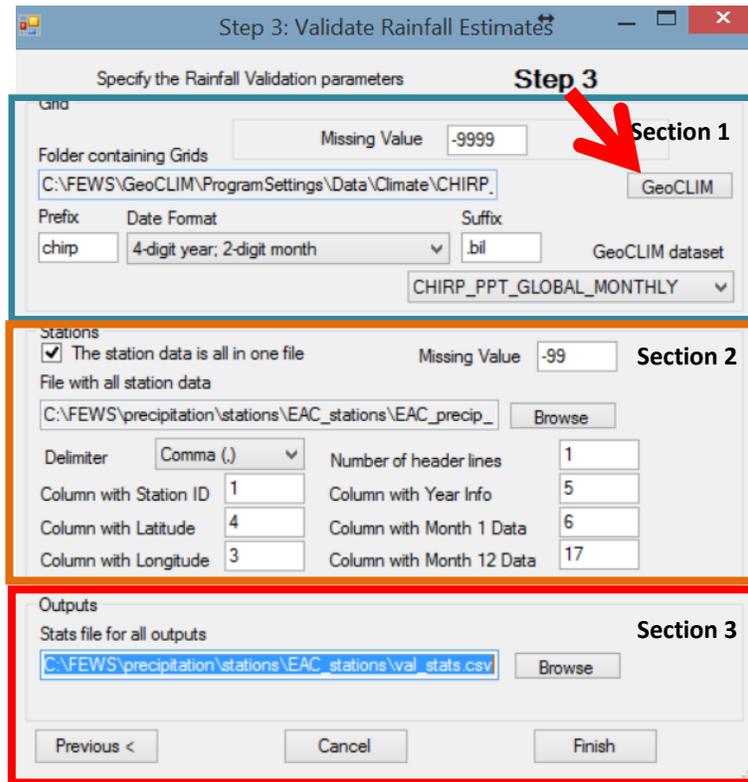


Figure 8.3 To validate raster data using station values, it is required to have the path to the raster data and a table with the station values.

Step 3: Step with three sections.

Section 1 (input grids) in Figure 8.3: 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. Otherwise, fill all the information manually.

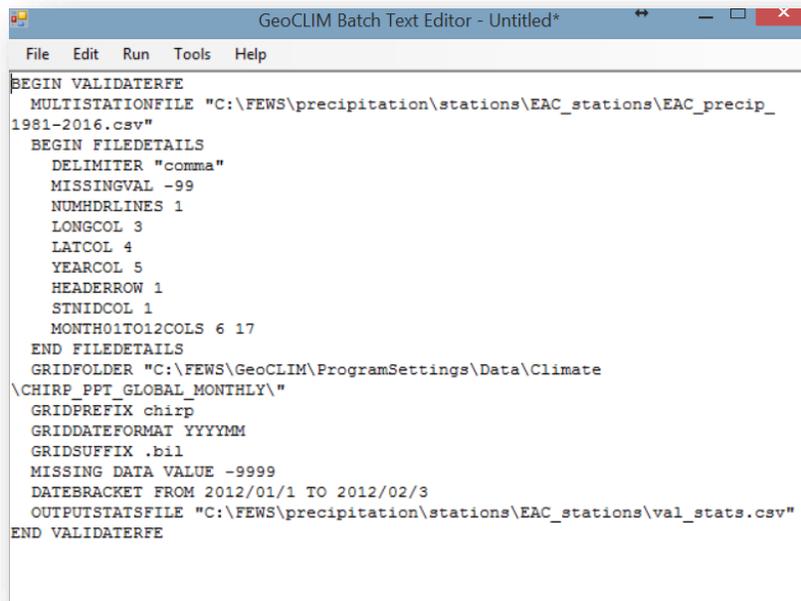
Section 2 (Stations) in Figure 8.3: This sections relates to the station input parameters.

- 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 the 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. Click on any of the boxes to specify any given column.

Section 3 (Outputs) see Figure 8.3:

- 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 8.4. This batch file can be saved for future reference.
- 3) Go to **Run** and select **Run Batch File** (Figure 8.4)



```
GeoCLIM Batch Text Editor - Untitled*
File Edit Run Tools Help
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 8.4 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 8.5), (2) a scatterplot showing the rainfall field values against the stations (Figure 8.6), (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 8.7). 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.

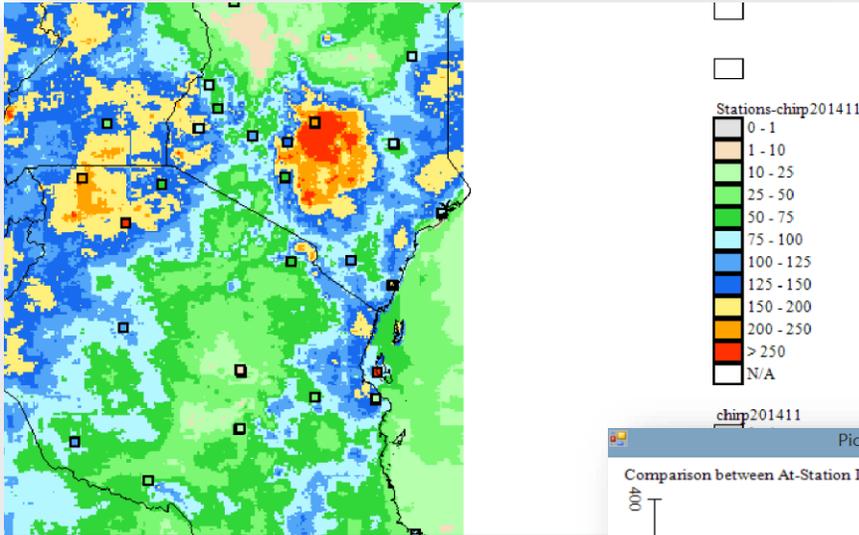


Figure 8.5 a map of the stations that took part in the process of validating CHIRP.

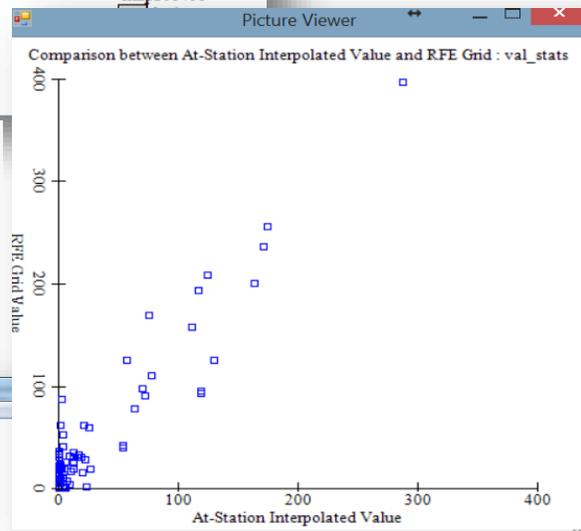


Figure 8.6 scatterplot of station value on X and raster (CHIRP) value on Y.

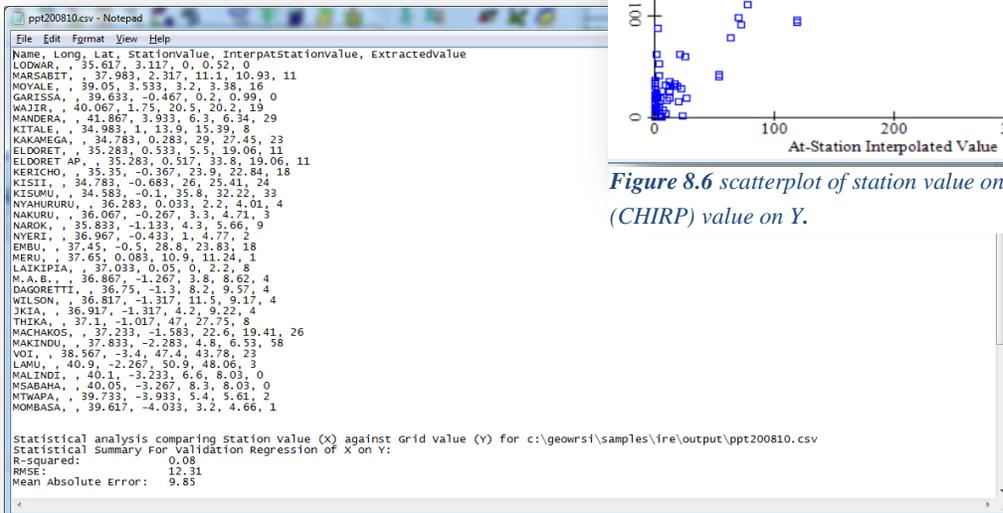


Figure 8.7 text file that includes a list of the station value and its corresponding raster value together with statistics for the two groups.

8.3. Improve rasters with stations using the blending algorithm

To create improved rainfall estimates, follow the steps below:

Step 1

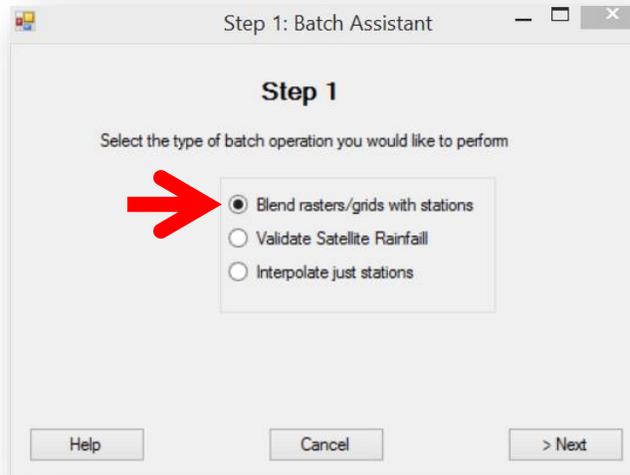


Figure 8.8 Select the Blend raster/grids with stations option.

1) Click on the **BASIICS** button on the toolbar (Figure 8.8).

2) This will open the **Step 1** window, select the **Blend rasters/grids with stations** batch operation, and click on the **> Next** button.

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.

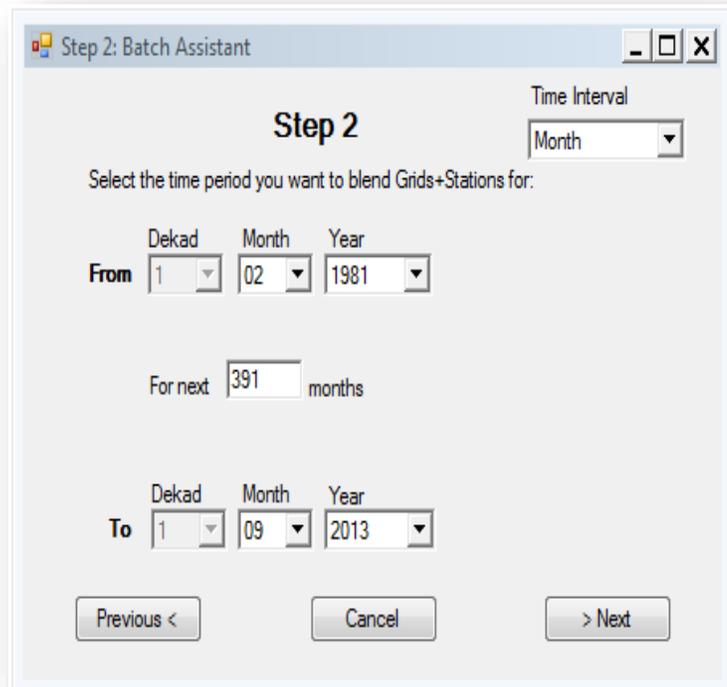


Figure 8.9 Select the starting and ending dates that would be part of the process.

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 8.10).

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 8.10).** 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.

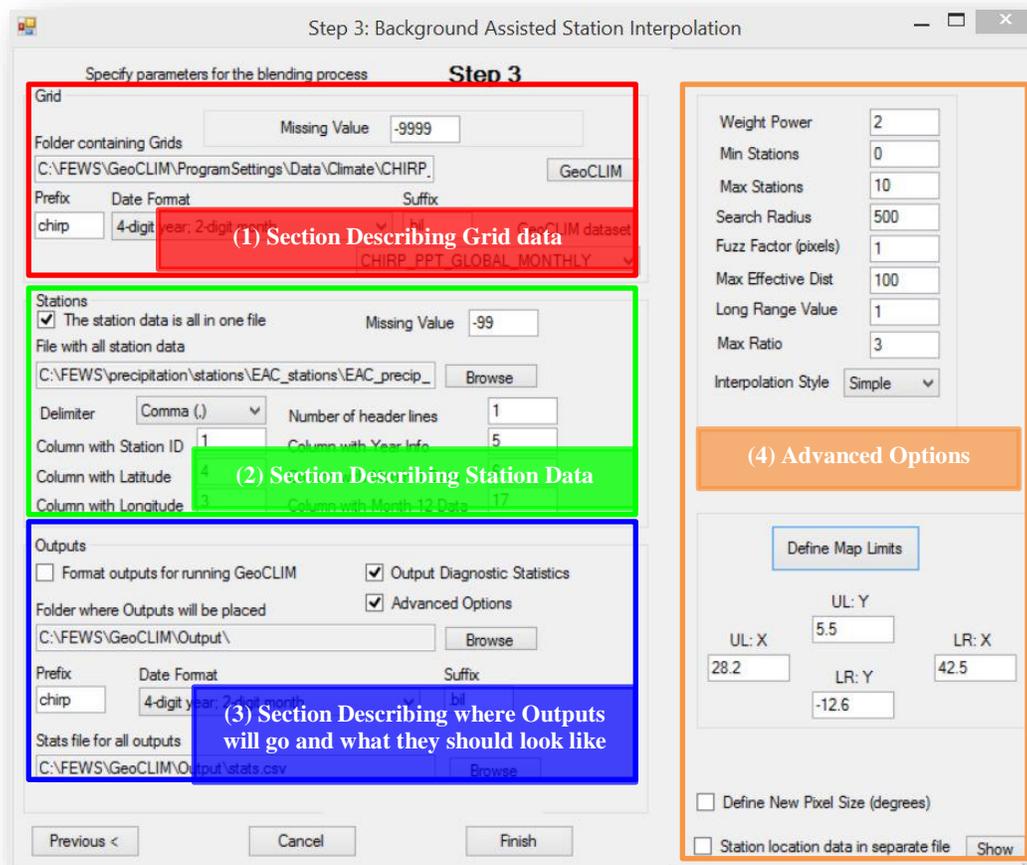


Figure 8.10 Step 3 requires information about the raster data to be improved, the stations with station metadata, and rainfall values.

The station data should be a CSV file and be in the format described in Figure 8.11. 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 8.11 the CSV table with station data must contain a station ID, lon, lat, year and a column for each pentad, dekad or month.

(3) Describing the outputs, (Figure 8.10)

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.

(4) Advanced Options

When the **Advanced Options** checkbox is checked, this will open a set of options to the right of the window. The upper section describes the interpolation parameters, see Figure 8.12 blue box. These parameter are:

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.

Min Stations (MINSTNS): The minimum number of stations to be used in the interpolation. If there are fewer stations available than specified in this parameter at a point, that pixel will not be interpolated, and a missing value will be defined in the output grid at that pixel. Use a value of 0 to produce a value everywhere in the output.

Max Stations (MAXSTNS): The maximum number of stations to be used in the interpolation. For example, if the value defined is 10, and if there are more than 10 stations available at a point for the interpolation, only the nearest 10 stations will be used in the interpolation.

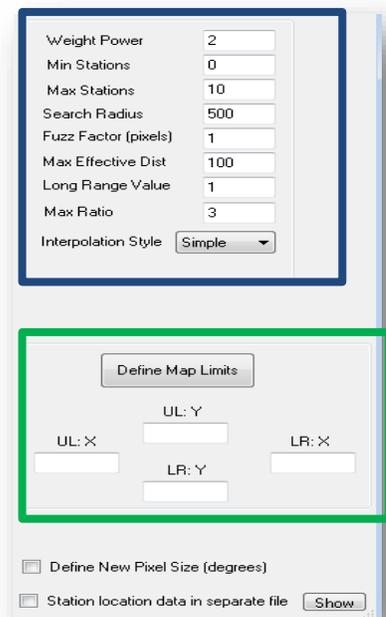


Figure 8.12 The advance option panel allows the user to modify the blending process.

Search Radius (SEARCHRADIUS): The maximum radius (in km) over which to search for stations that will be used in the interpolation of each pixel. If, at a pixel, the distance to a station is larger than this parameter, then that station will not be used in the interpolation at that pixel.

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_o) 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 8.12, 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 8.13) or other spatial data (raster or vector). This option helps to speed up the interpolation process considerably.

For example to run the blending process for the EAC region (Figure 8.13), 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.

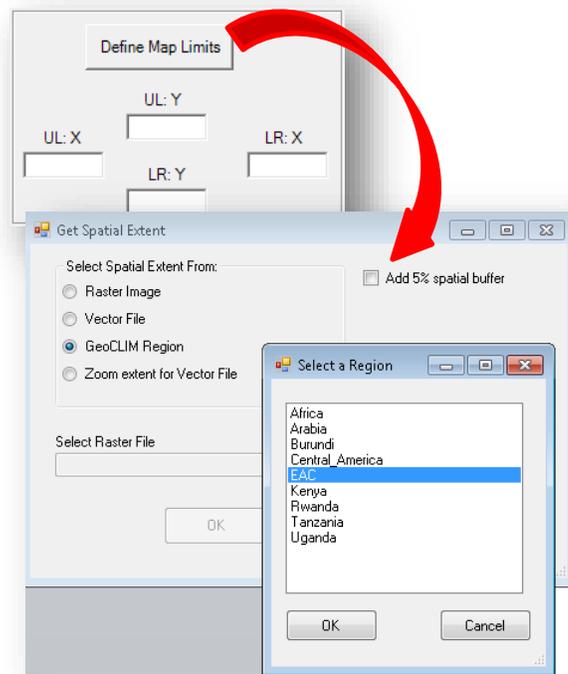


Figure 8.13 Defining the output extend could be done by selecting an existing raster, region or shape file.

5) After all the fields are completed, the Step 3 form should look like Figure 8.10.

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 8.14. This will enable the user to access and edit the saved file later.

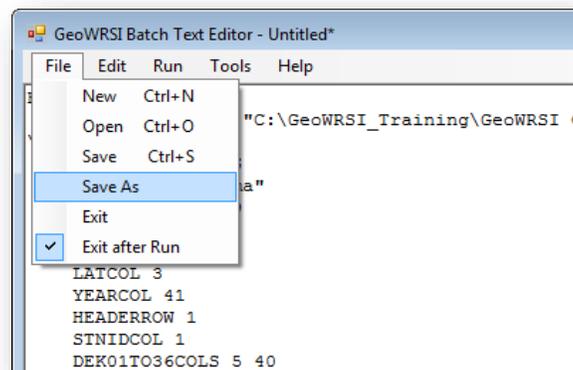


Figure 8.14 The batch file could be saved and re-ran later.

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 8.15.

```
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
STINDCOL 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
UNDETECT_VV 2 2 5 5
```

Figure 8.15 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 9: 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 1981 to 2014 can be calculated.

9.1. Extract Statistics

- 1) Open the **Extract Raster Statistics** tool  from the menu bar (see Figure 9.1).
- 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 9.2.
- 3) Select the type of summary for the pixels within the polygon, see the **Summary** pulldown menu (3 on Figure 9.1).
 - a. Select the raster file(s) to be used. There are two ways to populate the data fields, (see 4 on Figure 9.2): (a) using the **T-series** (Time-series) button and (b) using the

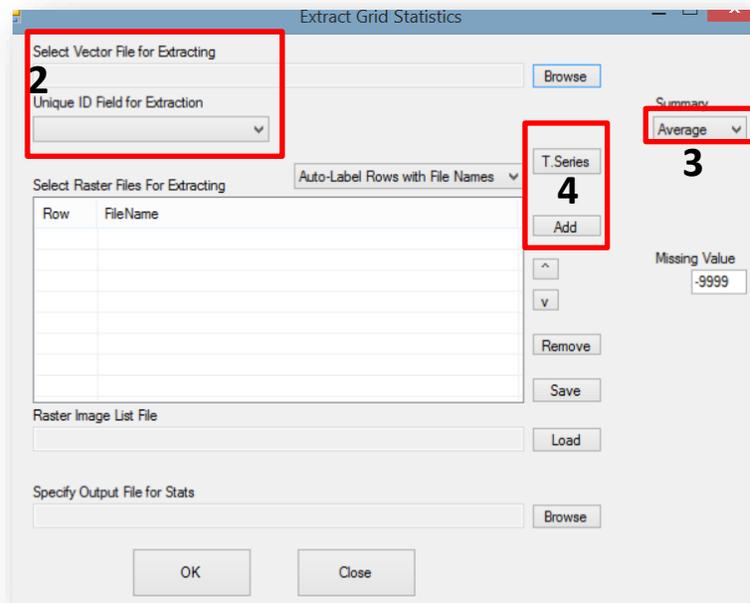


Figure 9.1 The tool calculates the spatial statistics for each raster dataset, based on each polygon on the shp file. The product is a table of a row for every polygon containing a single value (statistics selected) a column based on each raster.

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 9.2).

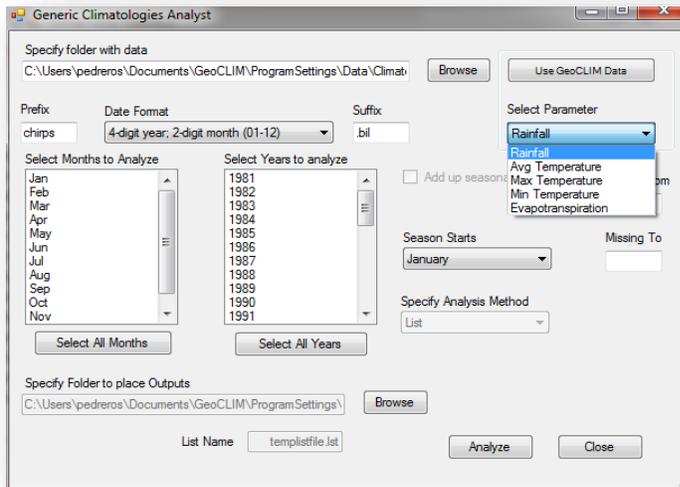


Figure 9.2 Select a T-series from the default data or point the path for a new data source.

- 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.

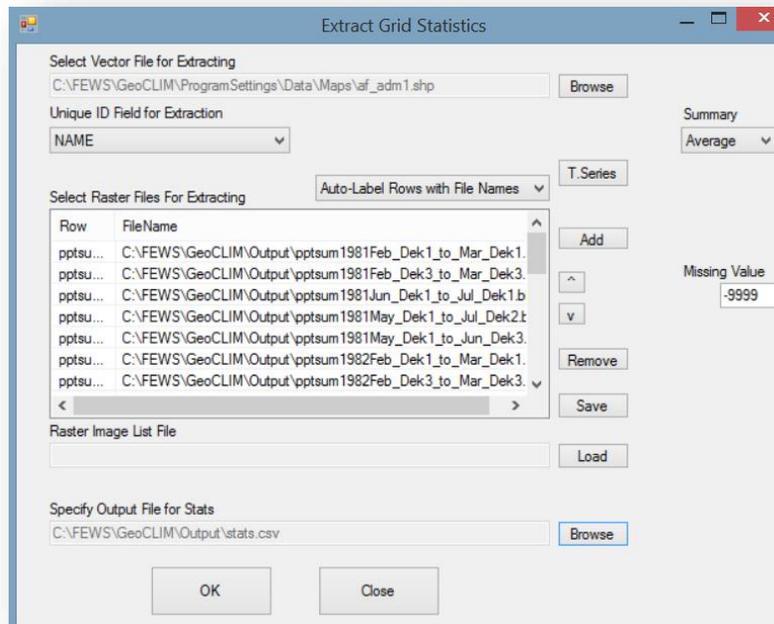


Figure 9.3 Once all the data are selected and the shape files are in place, the form should look like the above.

9.2. 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 selected. Figure 9.4 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.

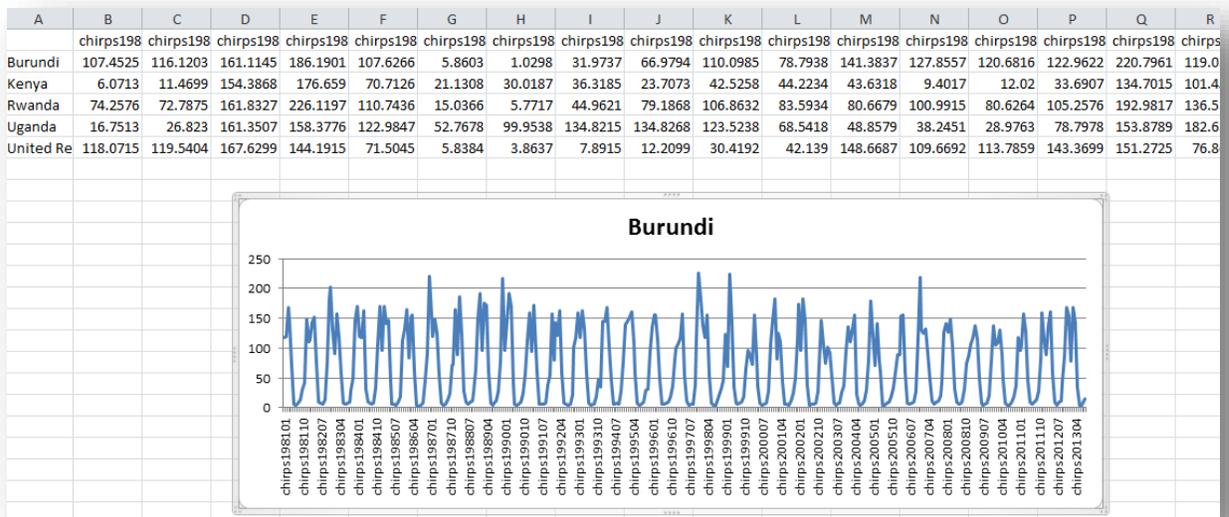


Figure 9.4 The resulting table has a row for every polygon and every column represents the summary value for each raster.

Chapter 10: 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 calculating the difference between the long-term averages of each set of years.

10.1 Estimate trends using difference in averages tool

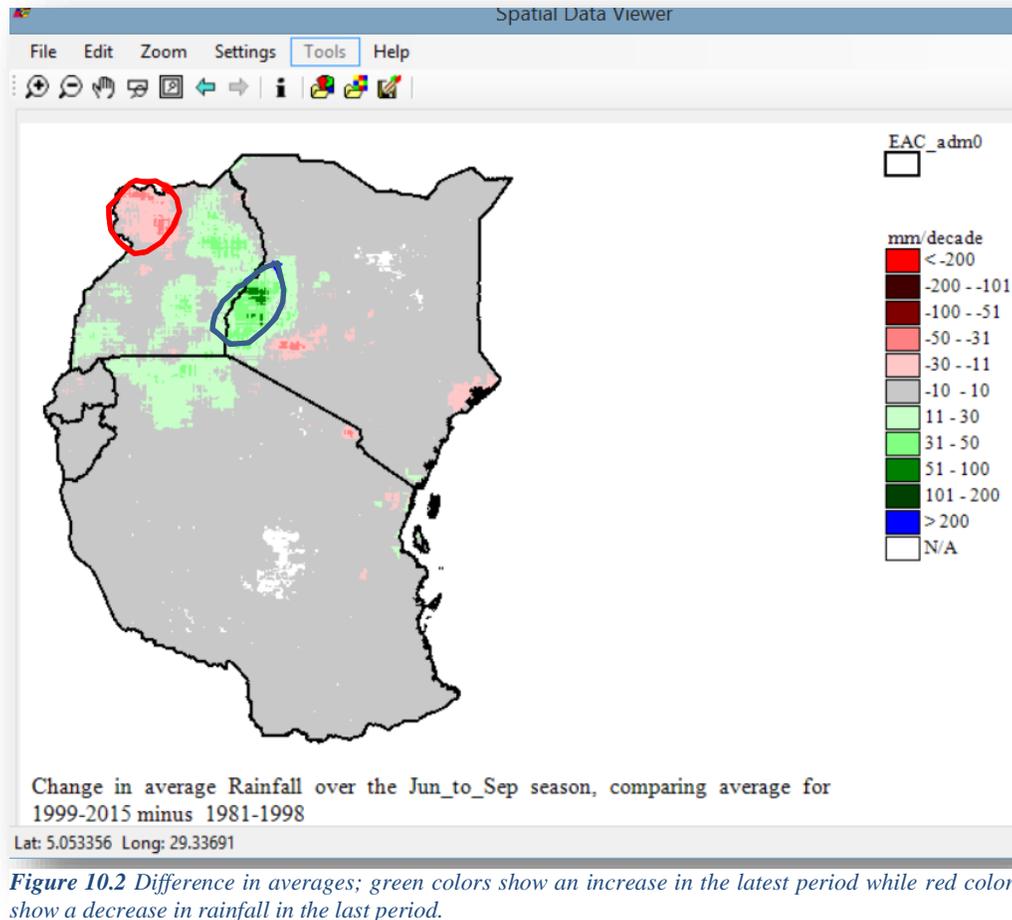


- 1) Open the tool from the main bar menu (see Figure 10.1).
- 2) Select the season to be analyzed.
- 3) On the series 1, select the first period of time.
- 4) On series 2, select the second period.

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

Figure 10.1 The Calculate change in Average tool is another way of estimating trends by comparing changes in averages between two periods.

For example: Figure 10.1 shows the calculation of the change in average from the 1981-1998 period to the 1999-2015 period, for the June-September season. Figure 10.2 shows that there is an increasing trend on the blue polygon while there is a decreasing trend on the red polygon.



Chapter 11: 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](#).

11.1. 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 11.1).
- 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.

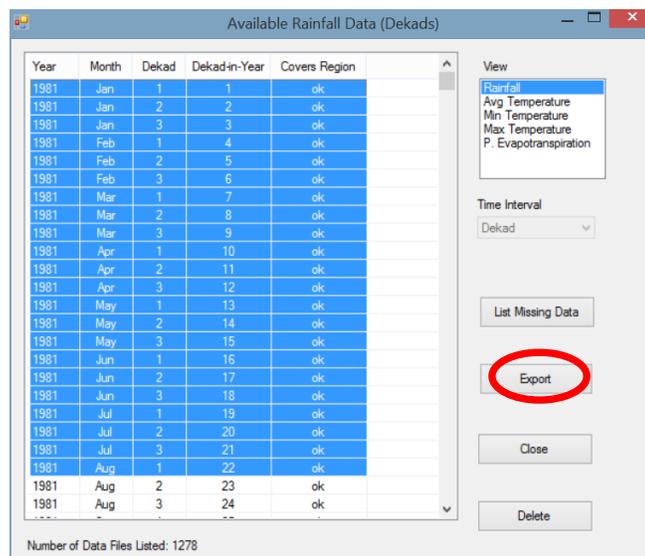


Figure 11.1 Exporting a time series from the Available Rainfall Data tool will create an archive.

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.

11.2. 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

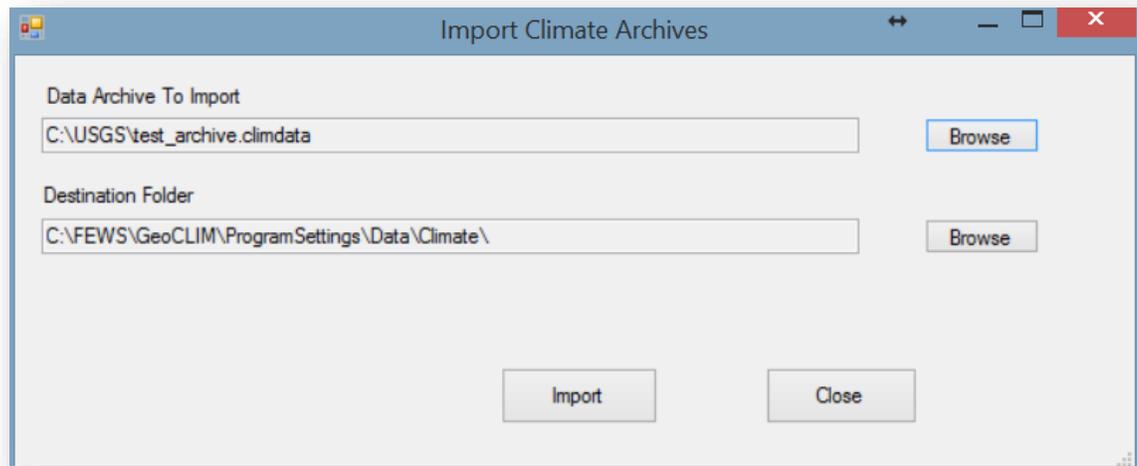


Figure 11.3 The Import Climate Archives tool.

- 3) Close the console once it is done, see Figure 11.3. And make the new dataset the default.

To ensure that the archive imported correctly go to the GeoCLIM **settings> data>select data>** and the new imported dataset should be ready. See Figure 11.4.

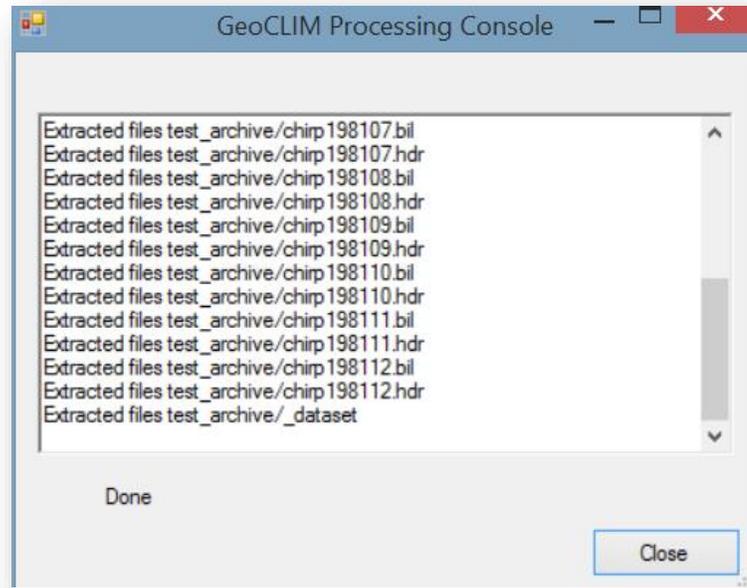


Figure 11.5 The GeoCLIM console shows the progress of the archive import.

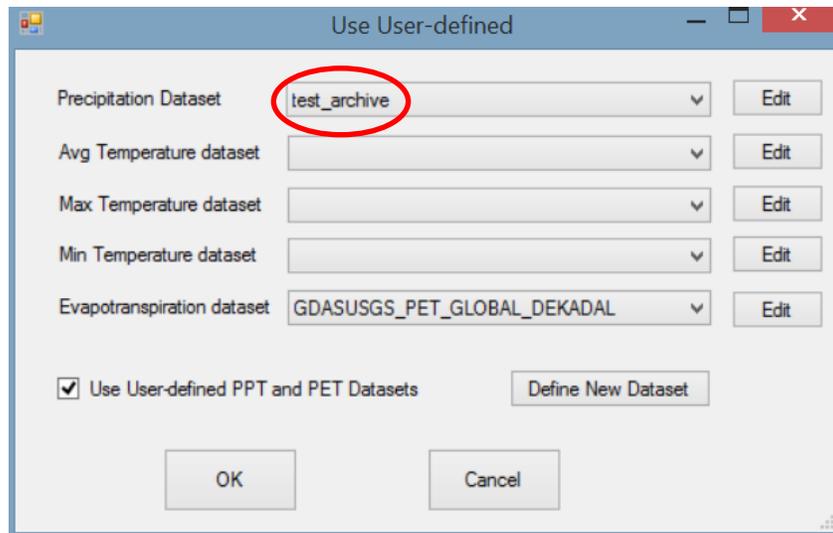


Figure 11.4 Once finished importing the archive, the dataset is ready to be used in the different operation in the GeoCLIM.

References

ESRI support, 2008. 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. 8th 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:10.1038/sdata.2015.66 2015.