GeoCLIM

Version 1.2.1

Documentation

FEWS NET

April- 2020

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 Climate Hazards Group provide tools to help mitigate or prevent humanitarian crises.

Acknowledgments

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

Note: All Climate Hazards Center FTP URLs can be substituted with HTTPS URLs to the same files.

Table of Contents

Introduction

Summary

The GeoCLIM program is one of several agroclimatic analysis tools developed by the FEWS NET/United States Geological Survey (USGS). It facilitates the analysis of climate data (rainfall, temperature, and evapotranspiration). It also provides an array of accessible analysis functions that allow you to perform the following tasks:

- Analyze large quantities of climate data.
- Create visual representations of climate data.
- Blend station information with satellite data to create improved datasets.
- Calculate seasonal trends.
- Calculate the standardized precipitation index (SPI).
- Compare groups of years within a time series.
- Extract statistics from raster datasets to create a time series for a set of polygons.

Using the Manual

This manual is organized into 12 chapters and presents examples and exercises to help you understand the different applications of the GeoCLIM analysis tools.

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

Chapter 2: Settings provides details on setting up the program and downloading data.

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

Chapter 4: Spatial Data Viewer provides instructions for viewing, editing, and creating shapefiles and rasters using GeoCLIM.

Chapter 5: Climatological Analysis explains how to calculate statistics, trends, and SPI, among other functions, for a set period (e.g., dekad, month, season).

Chapter 6: Rainfall Summaries shows how to calculate totals, averages, and anomalies for a set period (e.g., dekad, month, season).

Chapter 7: Climate Composites describes seasonal analysis among a group or two groups of non-consecutive years within a time series.

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

Chapter 9: Calculate Difference in Averages shows another way of estimating trends by comparing changes in averages between two periods (e.g., dekad, month, season).

Chapter 10: BASIICS explains the process of blending station and raster data.

Chapter 11: Extract Statistics explains how to create spatial summaries of historical data for a given region.

Chapter 12: Working with Climate Data Archives explains how to manage climate data archives.

For GeoCLIM updates and video tutorials, go to chc.ucsb.edu/tools/geoclim

Chapter 1: Overview

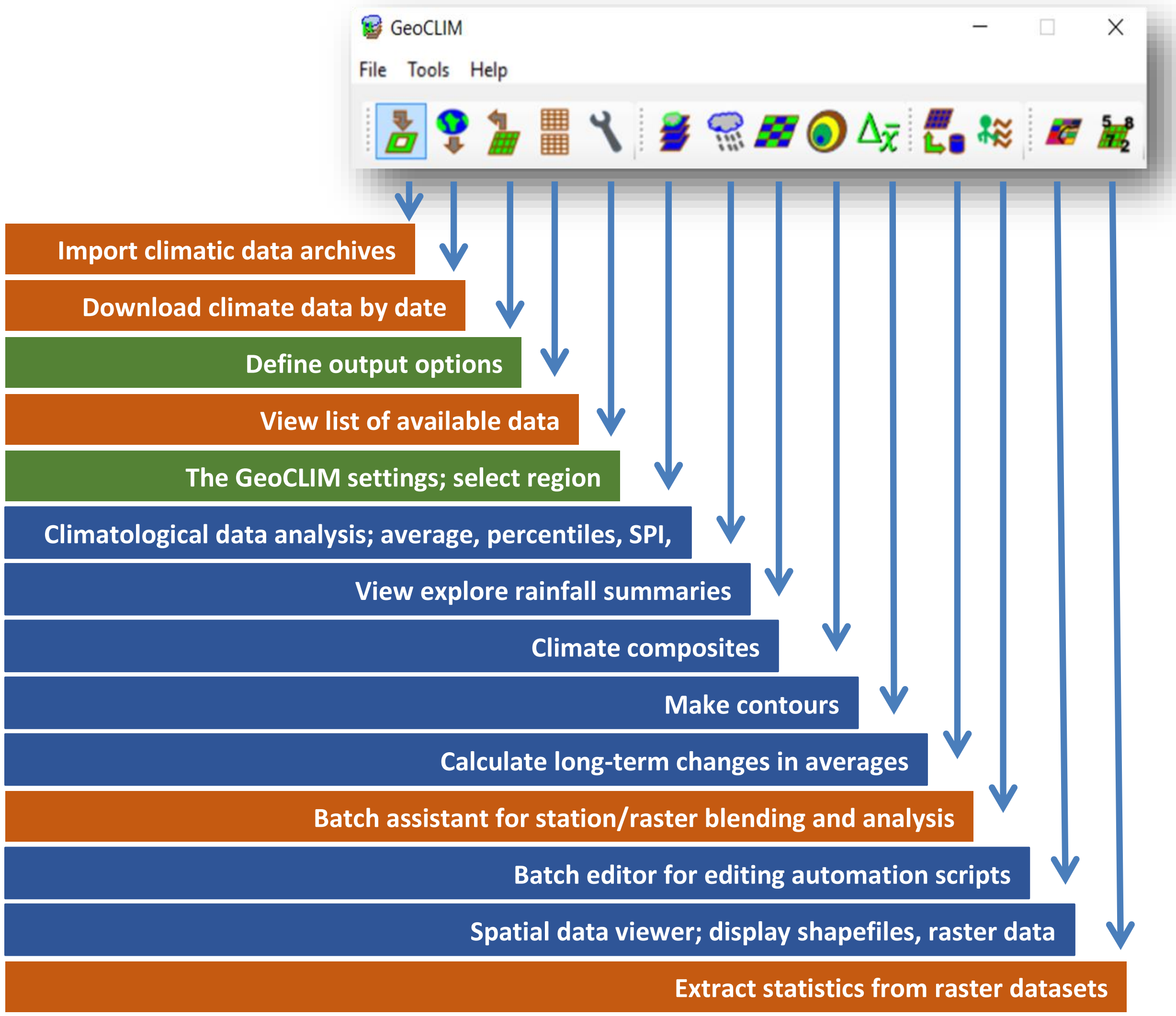
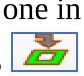


Figure 1.1 *The GeoCLIM main toolbar. Setting tool (green), data management tools (orange) and analysis tools (blue).*

Summary

Figure 1.1 shows the main tools available in the GeoCLIM toolbar. These tools consist of settings, data management, and analysis methods. This chapter briefly describes the main tools in GeoCLIM, and the following chapters look at each one in detail.

1.1. Import GeoCLIM Climate Archives 

A GeoCLIM archive is a compressed file containing data for a given climate variable and specific information to be imported and made ready for use in the program. The ***Import Climate Archives*** tool (Figure 1.2) makes datasets available in GeoCLIM. These archive files are useful for sharing data among GeoCLIM users. For information on creating data archives, see Chapter 12.

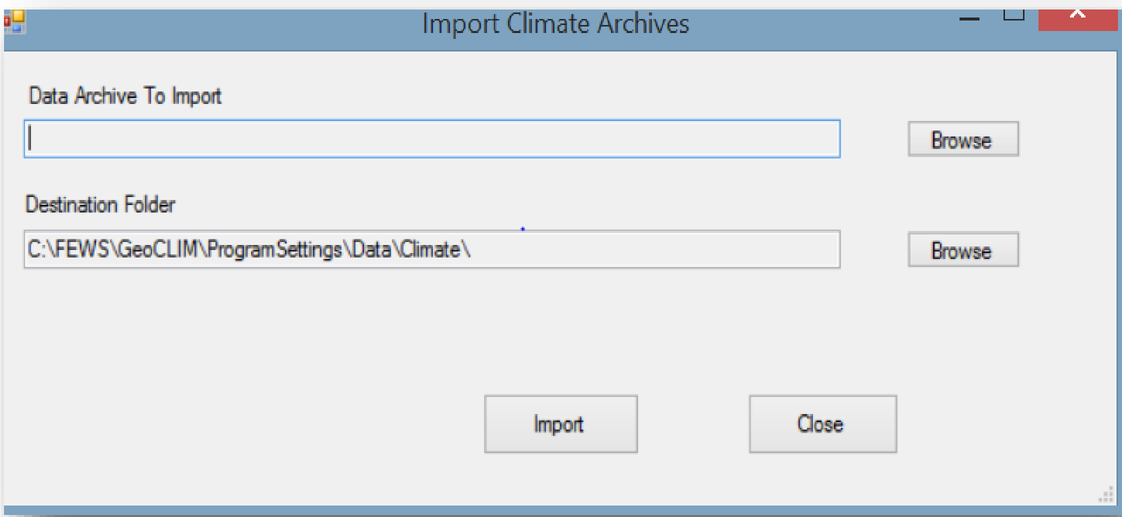


Figure 1.2 *One way of making Climate data available in GeoCLIM is through importing archives.*

1.2. Download Climate Data 

The ***Download Climate Data*** tool (Figure 1.3) facilitates bulk downloads of available climate data via FTP, HTTP, or HTTPS from different sources (e.g., UCSB, USGS, etc.). See chapter 2 for more information on GeoCLIM settings, such as adding new climatological data, changing the workspace, and creating a new area of interest (region) to use with the different analysis tools.

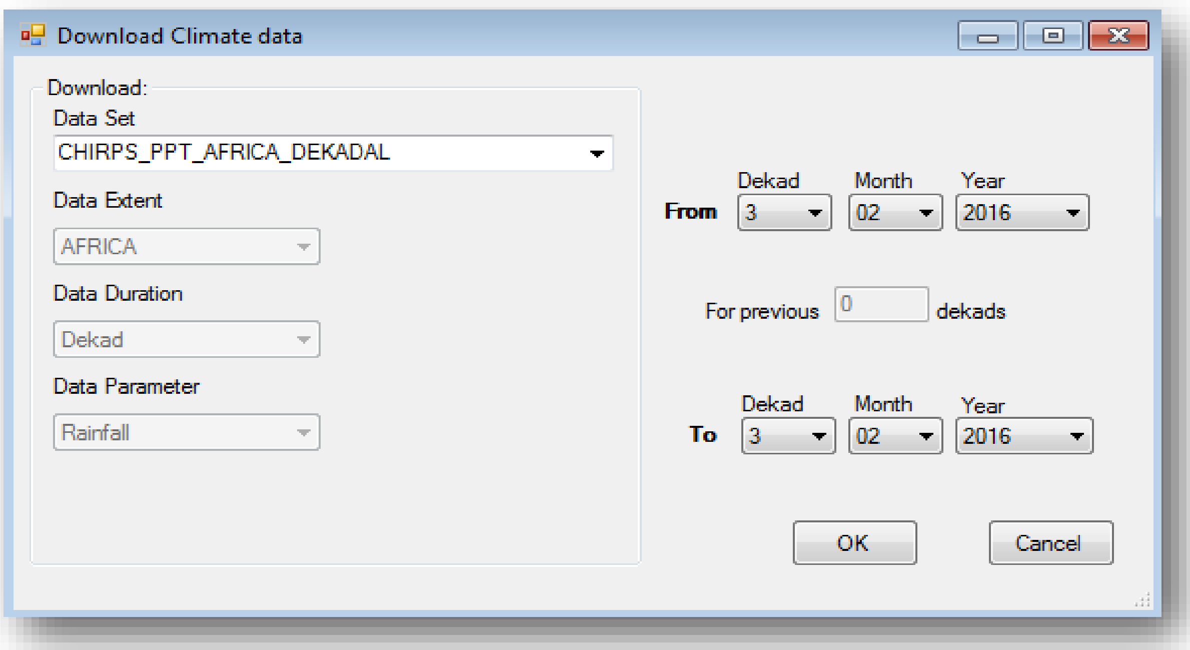


***Figure 1.3** You can download Rainfall, Temperature, or Evapotranspiration data directly from an onlinw location such as an FTP site using the Download Climatic Data Tool.*

1.3. Define Output Options 

The ***Define Output Options*** tool enables you to specify how GeoCLIM outputs should be generated and where they can be saved. These options include (Figure 1.4): **(1)** the size of the GeoCLIM main toolbar window, **(2)** the title fonts for the output graphics, **(3)** whether or not 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.

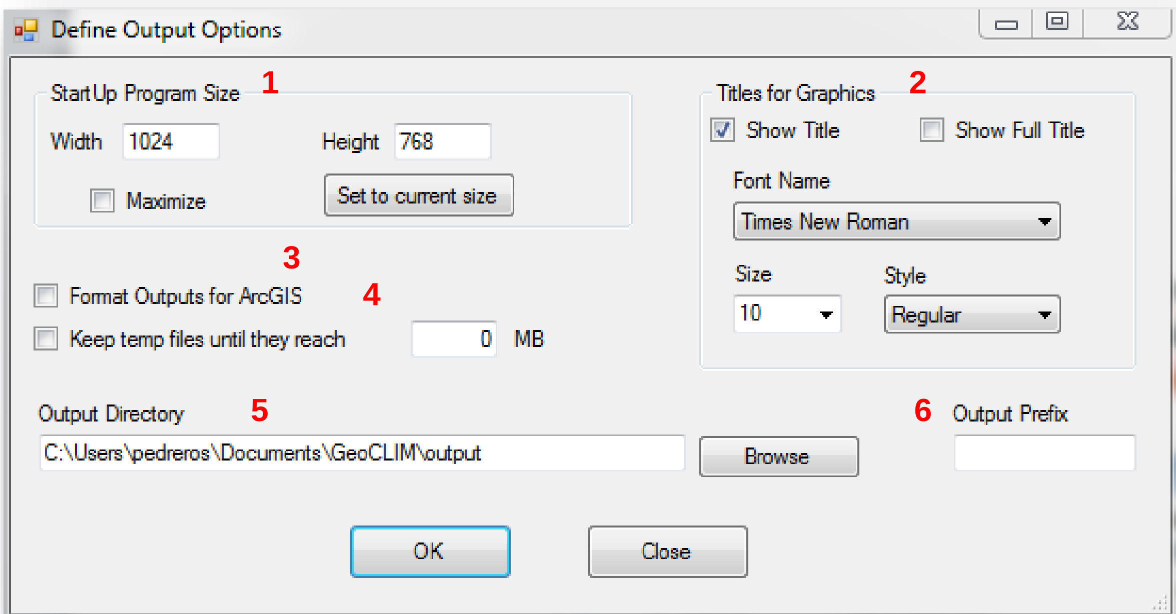
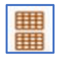


Figure 1.4 *This tool allows you to define output settings.*

1.4. View Available Data 

The ***View Available Data*** tool provides a list of the data available for analysis based on the climate datasets selected (rainfall, mean temperature, minimum temperature, maximum temperature, or potential evapotranspiration). Figure 1.5 shows an example of a list of dekadal (+- 10-day) total rainfall starting on the first dekad of January 1981 (19810101). The List Missing Data button provides a list of any missing dates of the climate dataset selected between the first and the last date in the time series. The Export button is used to export data from the selected climate dataset to different formats (single `BIL` or `NetCDF` files, or as a GeoCLIM archive) for sharing or backup. See chapter 2 for more information.

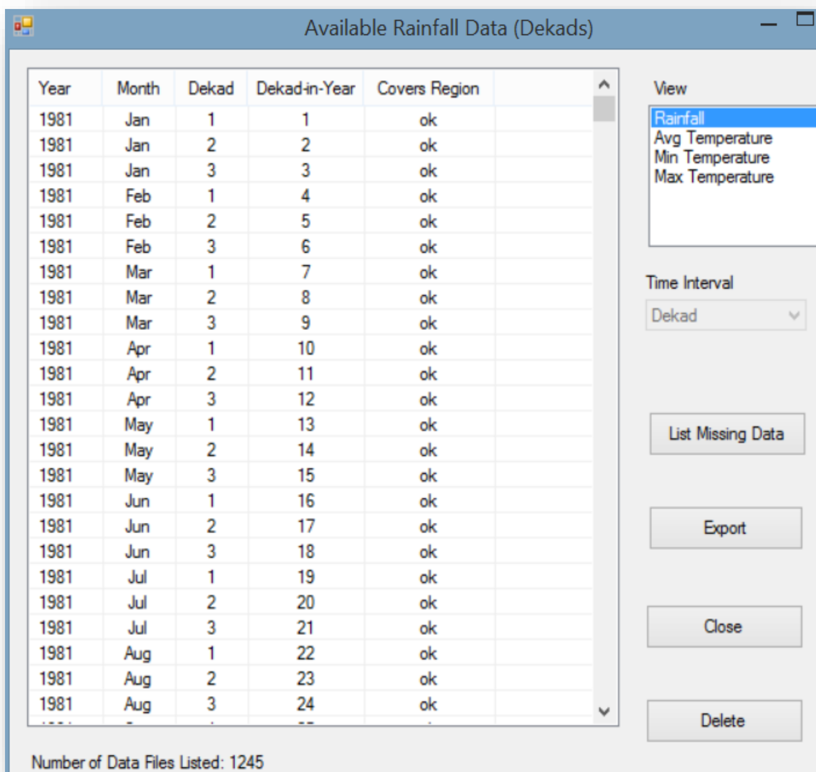


Figure 1.5 *This function shows a list of the available data from the different climate variables for a specific region*

1.5. GeoCLIM Settings

The GeoCLIM ***Settings*** tool allows you to change the region of interest and select/add/edit new datasets. This section contains three main tabs (Figure 1.6):

Region – You can select a pre-defined region available in GeoCLIM or define a new one based on the area of interest (e.g., country, county, pre-selected group of countries, city, or customized region).

Mask – Masks are maps in raster format that are used to define the area of interest (region) and ignore the rest of the data. You can define a new mask or edit an existing one for the selected region; this mask will be used in the analysis tools or for the map viewer (e.g., landmasses, non-desert regions).

Data – You can select a dataset for each of the available climate variables in the program (precipitation, average temperature, minimum temperature, maximum temperature, and evapotranspiration). For more information, see chapter 2.

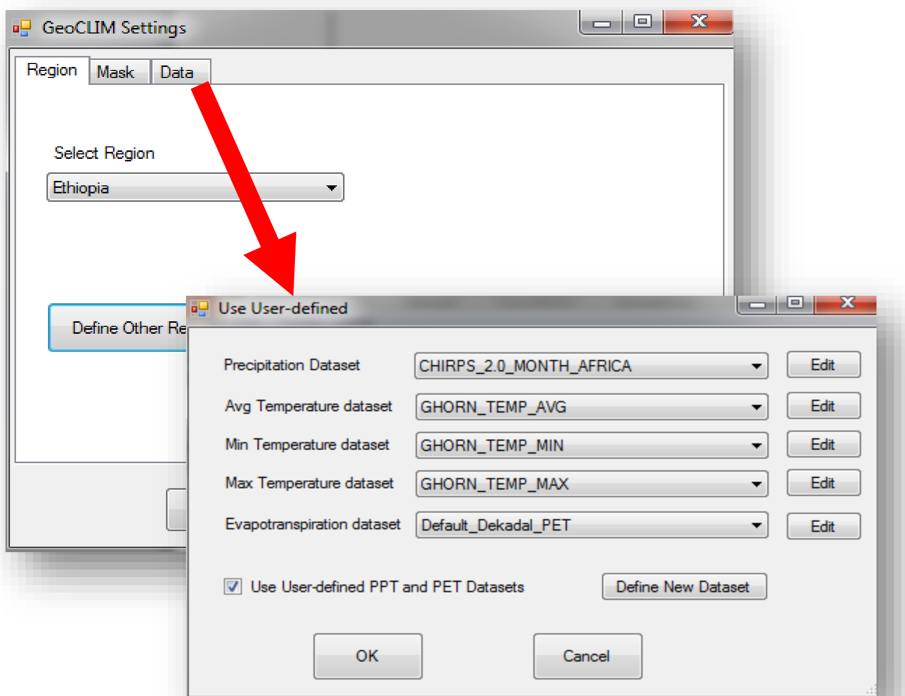


Figure 1.6 *The settings tool allows the user to select the climate data and the region for analysis.*

1.6. Climatological Data Analysis

The ***Climatological Analysis of Climatic Variables*** tool (Figure 1.7) is designed to calculate and display the statistical characteristics of rainfall, evapotranspiration, and temperature data. The tool displays all the years **(1)** and periods **(2)** (months, dekads, or pentads) available for a selected climate dataset (see chapter 5 for a more in-depth discussion of this tool).

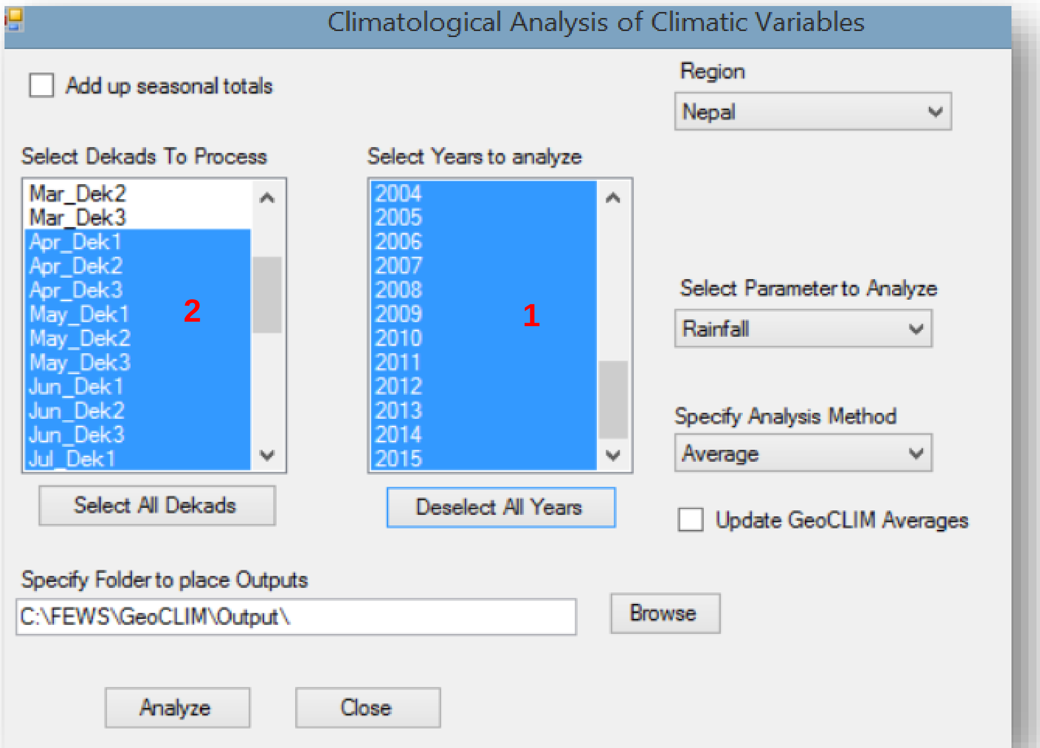


Figure 1.7 *This tool facilitates the calculation of statistics, trends and standardized precipitation index among other functions for specific a data set.*

1.7. Rainfall Summaries 

The ***Rainfall Summaries*** tool (Figure 1.8) calculates the total rainfall, the long-term average, the difference, and the percent of the long-term average for a selected region and range of dates. More details on this tool are available in chapter 6.

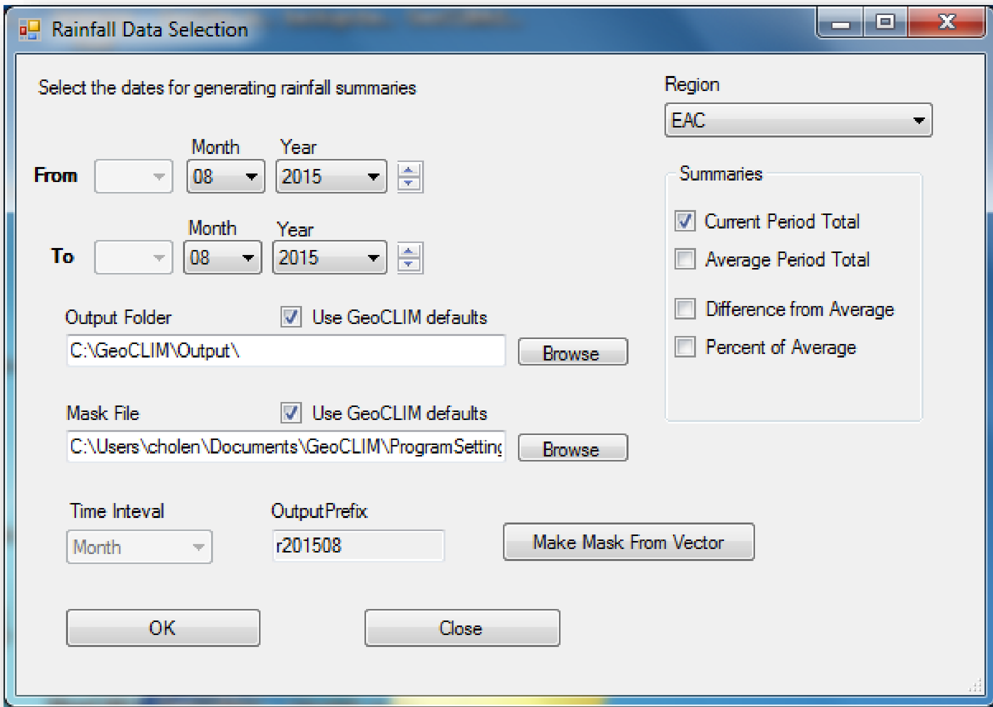


Figure 1.8 *The rainfall summaries tool calculates rainfall total and anomalies for a selected period of time.*

1.8. Climate Composites

The ***Climate Composites*** tool facilitates the seasonal analysis for one or two groups of years that may be non-consecutive. The tool calculates the seasonal average from a group of years, as well as comparing the seasonal rainfall performance between two groups of years, by calculating the percent of average, anomalies, and standardized anomalies (Figure 1.9). See chapter 7 for more details.

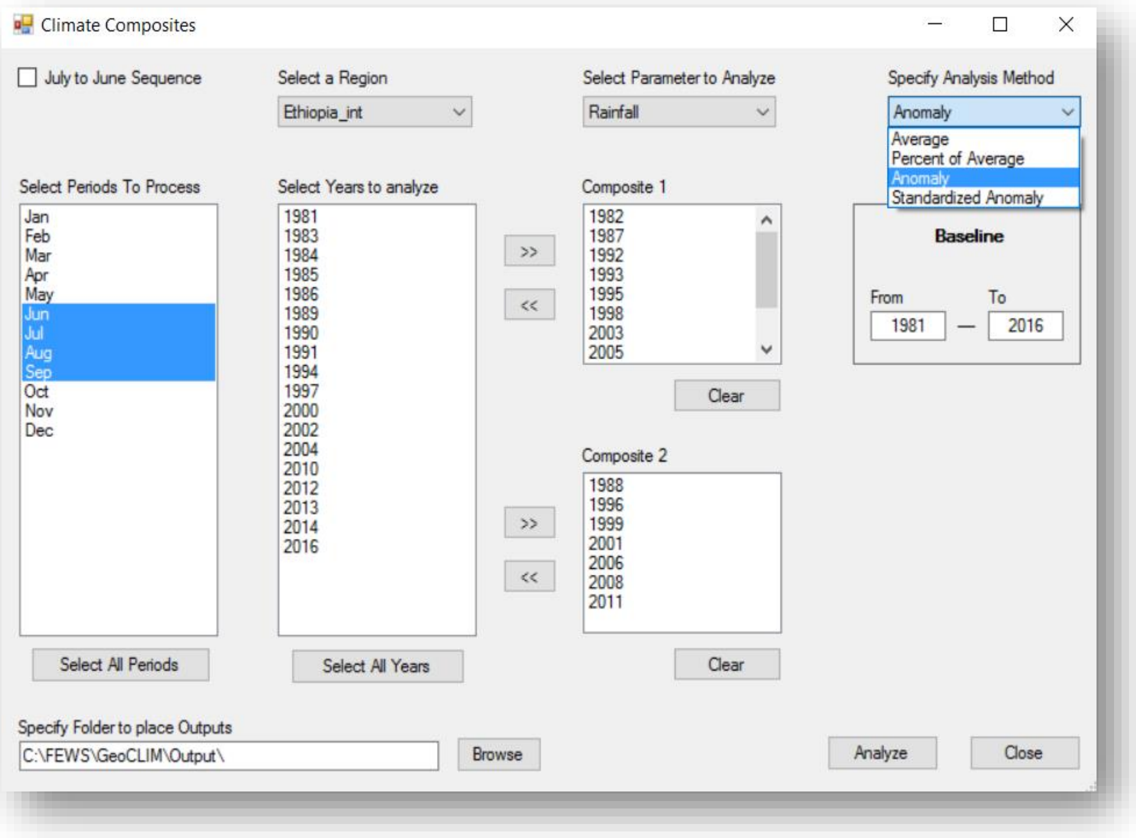


__Figure 1.9__ The Climate Composites tool facilitates the seasonal analysis for one or between two groups of [potentially non-consecutive] years.

1.9. Make Contours

The ***Make Gridded Contours*** tool (Figure 1.10) displays smoothed contours for a specified interval based on a raster (`*.bil`) file. This tool helps identify changes in rainfall patterns within a region of interest. Read more about making contours in chapter 8.

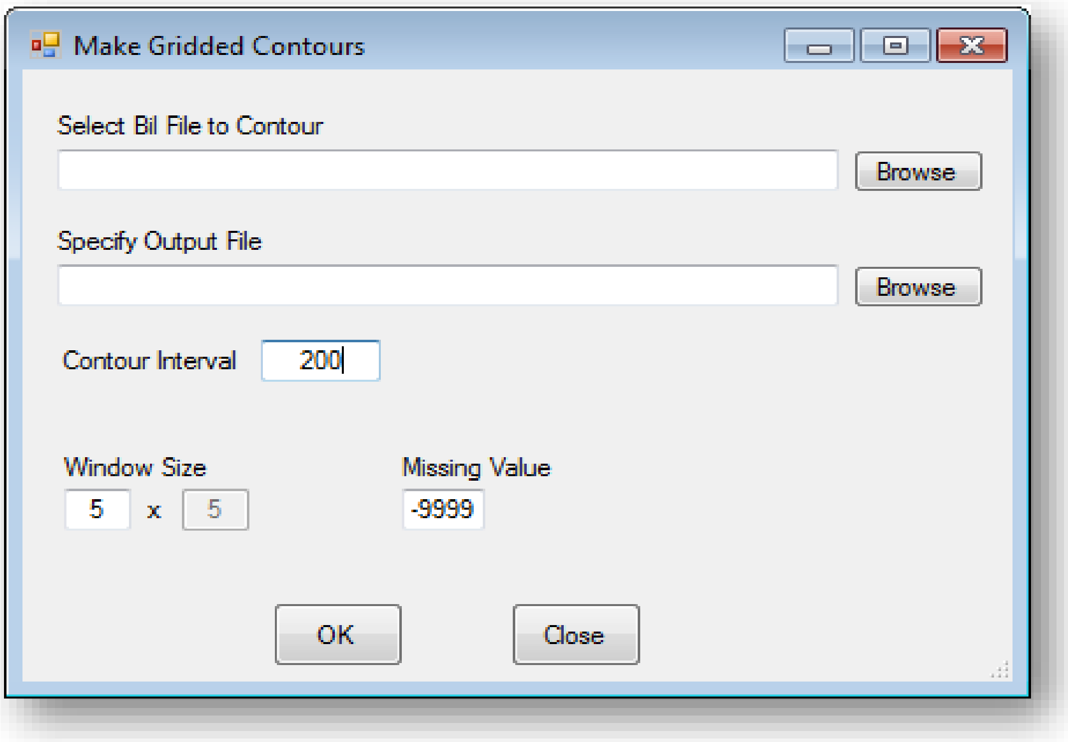


__Figure 1.10__ Display rainfall data based on contour intervals.

1.10. Climate trends - Changes in Average

The ***Climate Trends - Changes in Averages*** tool (Figure 1.11) identifies trends by calculating the difference in average between two periods (denoted as **Series 1** and **Series 2** in Figure 1.11). See chapter 9 for more details.

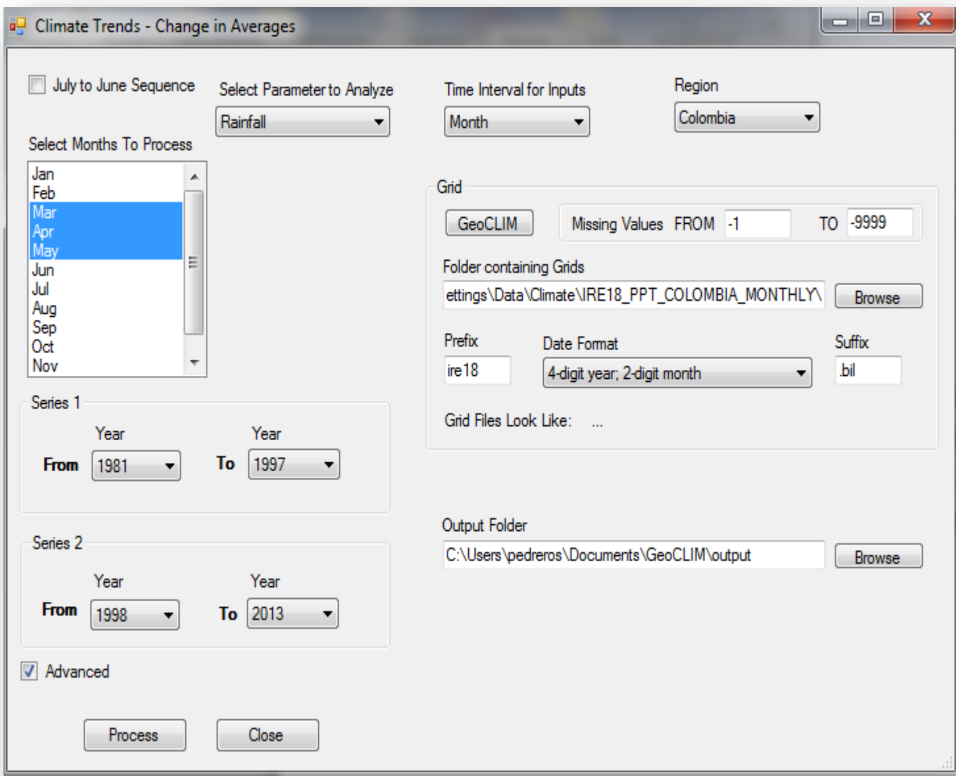


Figure 1.11 *The Climate Trends tool compares the average rainfall for two periods of time, identifying trends.*

1.11. Batch assistant for station/raster blending and analysis

The Batch assistant tool (Figure 1.12) allows you to validate satellite-based data using climatological stations, blend climatological stations with raster data (BASIICS), and interpolate just stations. This section contains the following modules:

1. ***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 and improved climate dataset.
2. ***Validate Satellite Rainfall***: Validates a raster dataset using station data by comparing the point-to-pixel value for each station. The results indicate how different the two datasets are.
3. ***Interpolate just stations***: This function uses a modified inverse distance weighting (IDW) process to interpolate station values. See chapter 10 for more information.

The assistant generates batch operations that are written to a script, allowing the processing of several years of data at once. The script can be used later with the ***Batch Text Editor*** tool to run frequent processes with similar settings.

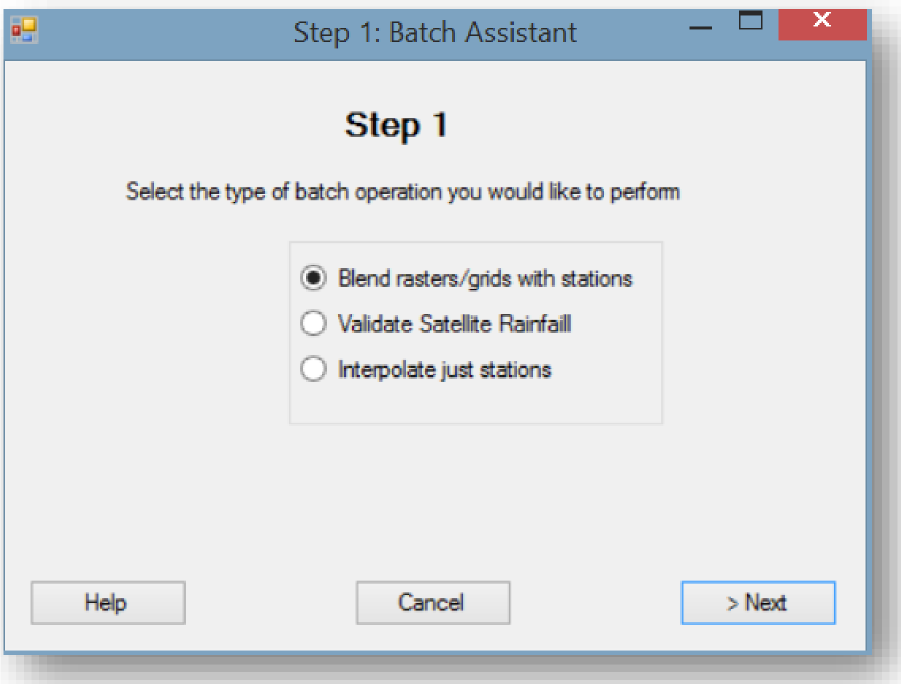
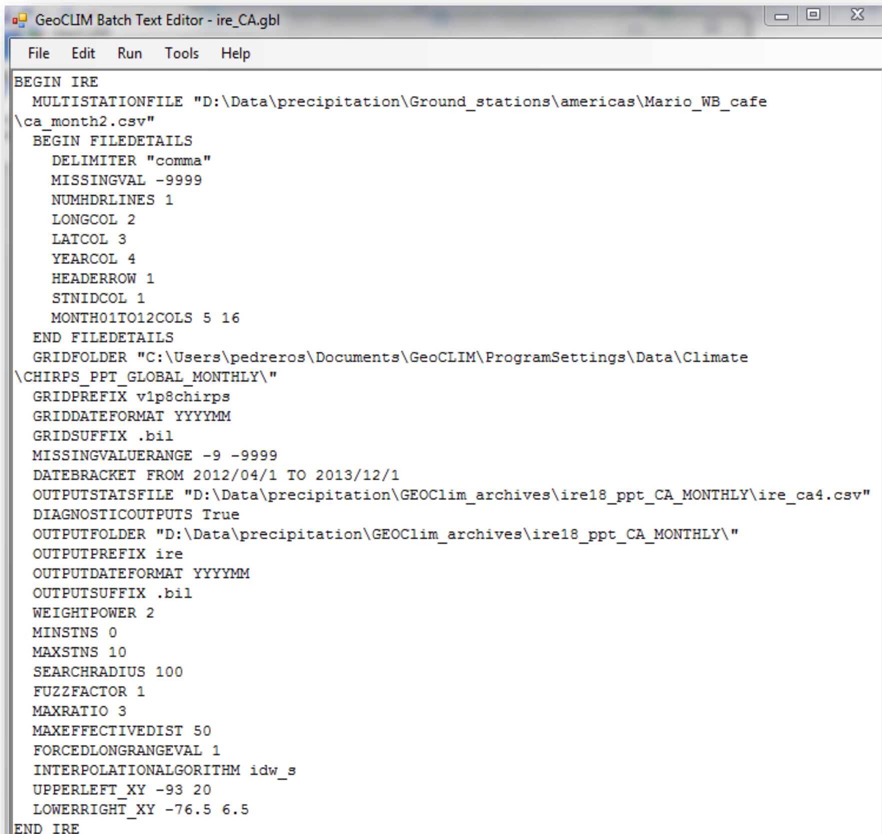


Figure 1.12 *The Batch Assistant tool has functions to: validate satellite estimated data using station values, blend station data with raster (BASIICS) and to interpolate station data.*

1.12. Batch Editor for Editing Automation Scripts

The ***Batch Text Editor*** allows you to change scripts that were previously generated using the Batch Assitant tool. (e.g., a different period in `DATESBRACKET`) (Figure 1.13). See chapter 10 for more information.




```
BEGIN IRE
  MULTISTATIONFILE "D:\Data\precipitation\Ground_stations\americas\Mario_WB_cafe
\ca_month2.csv"
  BEGIN FILEDETAILS
    DELIMITER "comma"
    MISSINGVAL -9999
    NUMHDRLINES 1
    LONGCOL 2
    LATCOL 3
    YEARCOL 4
    HEADERROW 1
    STNIDCOL 1
    MONTH01TO12COLS 5 16
  END FILEDETAILS
  GRIDFOLDER "C:\Users\pedreros\Documents\GeoCLIM\ProgramSettings\Data\Climate
\CHIRPS_PPT_GLOBAL_MONTHLY\"
  GRIDPREFIX v1p8chirps
  GRIDDATEFORMAT YYYYMM
  GRIDSUFFIX .bil
  MISSINGVALUERANGE -9 -9999
  DATEBRACKET FROM 2012/04/1 TO 2013/12/1
  OUTPUTSTATSFILE "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 .bil
  WEIGHTPOWER 2
  MINSTNS 0
  MAXSTNS 10
  SEARCHRADIUS 100
  FUZZFACTOR 1
  MAXRATIO 3
  MAXEFFECTIVEDIST 50
  FORCEDLONGRANGEVAL 1
  INTERPOLATIONALGORITHM idw_s
  UPPERLEFT_XY -93 20
  LOWERRIGHT_XY -76.5 6.5
END IRE
```

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

1.13. Spatial Data Viewer

The ***Spatial Data Viewer*** (Figure 1.14) displays raster and vector data and facilitates the use of basic GIS functionality such as editing legends, and digitizing polygons. See [chapter 4](#) for more details.

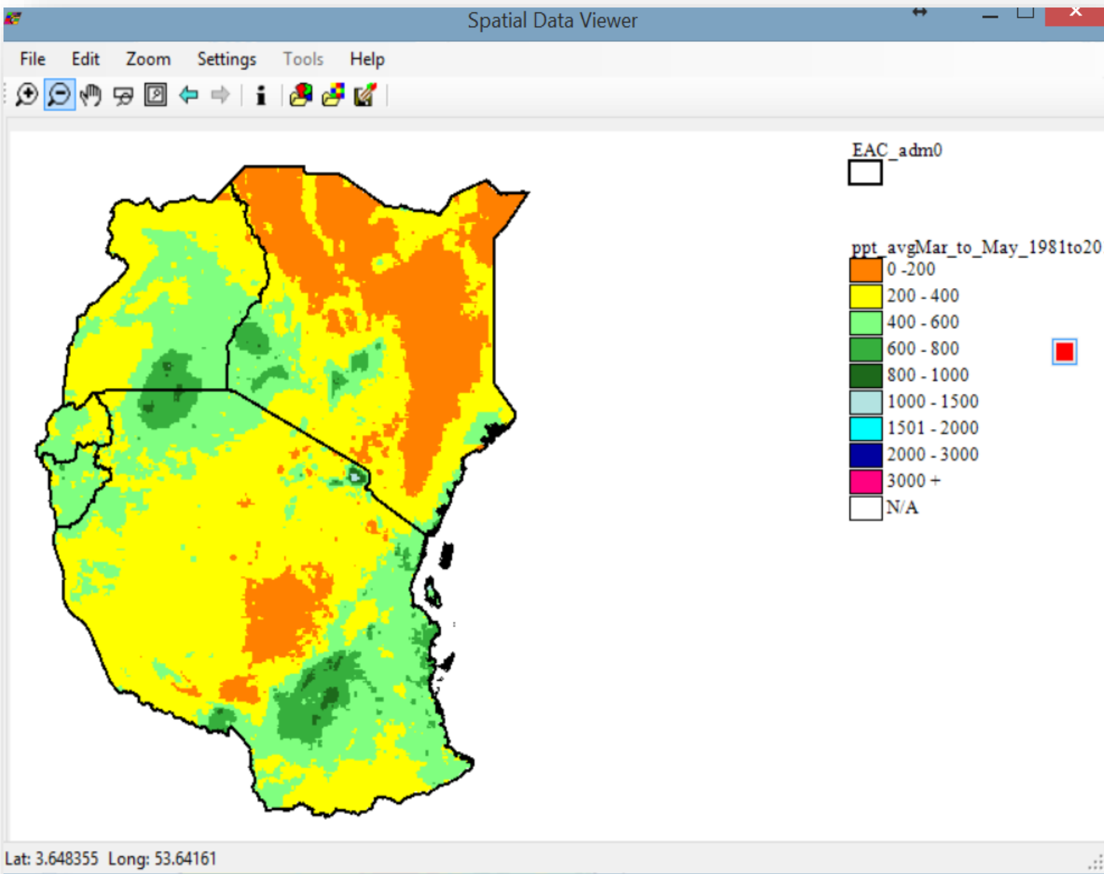


Figure 1.14 *The spatial data viewer; displaying and editing of raster data and shapefiles.*

1.14. Extract Statistics from Raster Datasets

The ***Extract Grid Statistics*** tool summarizes raster information by zones defined in a vector file (polygons such as districts, provinces, etc.). This tool produces a table with summary statistics such as average, sum, maximum, minimum, range, or standard deviation for each polygon. This extraction can be applied to a single climate grid or a set of climate raster files defined in the time series option (Figure 1.15). For more on this tool, see chapter 11.

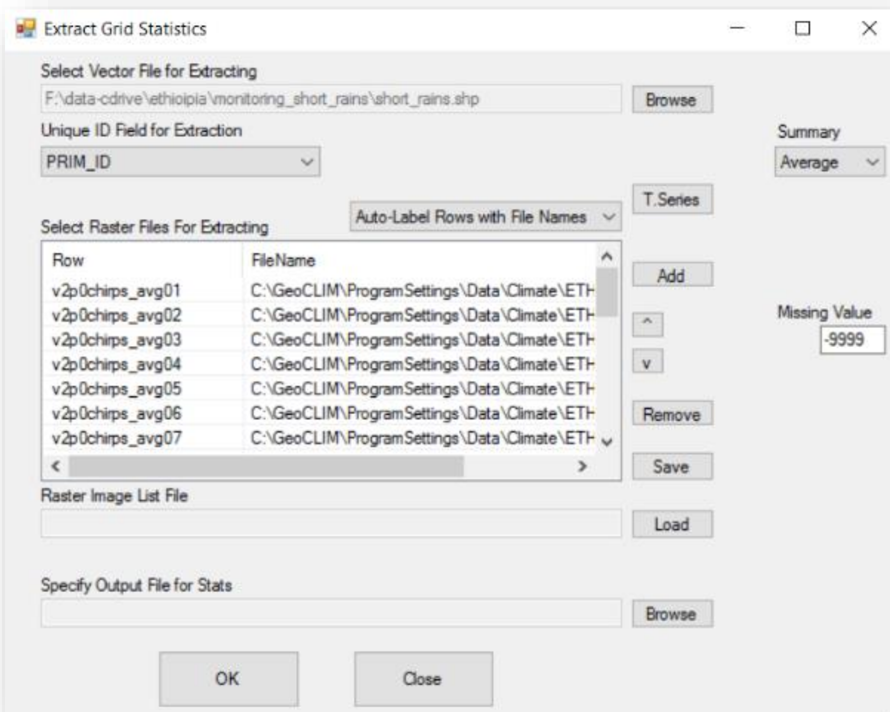


Figure 1.15 *Extract summary statistics based on polygons.*

Chapter 2: Settings

Summary

This chapter describes the GeoCLIM directory structure and settings, including how to add a new climate dataset, how to change the GeoCLIM workspace, how to create a new region of interest, and how to reset GeoCLIM when it breaks, among other items.

2.1. Review of the GeoCLIM Directory Structure

Once the program is installed, the default directory (in Windows Vista, 7, and 10) is:

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

Where `<USER>` is the Windows username. There are two subdirectories in the GeoCLIM folder: `Output` and `ProgramSettings` (Figure 2.1). The `Output` directory holds all the analysis results by default. You can change the default location of the `Output` directory in the ***Output Options*** tool. The `ProgramSettings` directory contains the colors used for the output maps, the data, the region definitions, among other items. See the next section for details. Figure 2. **(a)** shows an outline of the contents of `ProgramSettings`.

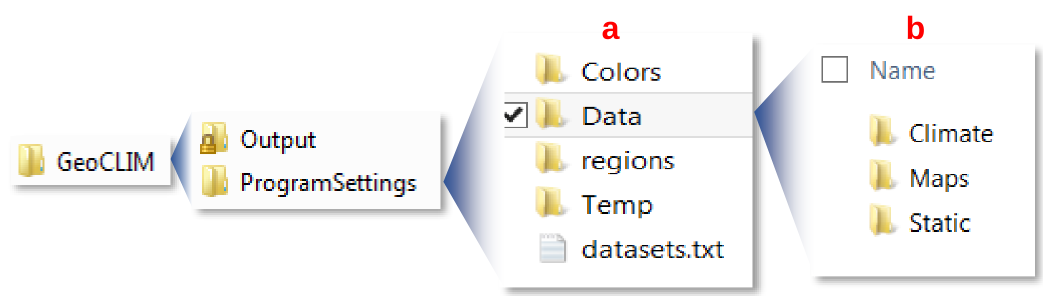


Figure 2.1 *The GeoCLIM directory contains two folders: the 'output' where all the results go by default, and the 'ProgramSettings' that contains the 'Data' directory among others.*

> **NOTE**: The default GeoCLIM directory on `C:\Users\<USER>\Documents\GeoCLIM` contains a set of text files that hold important data. Open each of the text files to get familiar with the information they contain. Don't make any changes or save in these text files unless you are sure of what you are doing as this can break the program.

Contents of the `ProgramSettings` directory:

- **Colors**: Contains color files (look-up tables) for legends and maps produced by GeoCLIM.
- **Data**: (Figure 2.**(b)**) contains the following directories:
  - ***Climate*** – Stores all downloaded and imported data. See section 2.3.3.2 on how to make a dataset available for analysis in GeoCLIM.
  - ***Maps*** – Contains all the shapefiles for the maps of the regions and countries required by the different functions. You can add shapefiles/maps as needed.

  - ***Static*** – Contains the masks for the different regions. Masks are maps in raster format that are used to define the area of interest (region) and ignore the rest of the data. For example, GeoCLIM contains rainfall data for the entire continent of Africa, but the analysis may be needed only for the country of Kenya. The mask would have a value of 1 in the area of interest (e.g., land areas of Kenya) and a value of 0 (zero) outside the area of interest. The results from the different functions will be given only for Kenya.
- **Regions**: Contains the GeoCLIM region files that define the area of analysis/display. The region files specify the min and max longitude of the area of analysis, the pixel size, the mask file, and the shapefile to use when displaying outputs. Learn how to Create a new region in GeoCLIM.
- **Temp**: This directory stores temporary files, such as the downloaded .tar.gz files.
- **Datasets.txt**: This file contains the information for each of the datasets, in the "`Climate`" directory, defined in GeoCLIM. You must know the structure of this file since it sometimes requires editing.

> **NOTE**: The default path to climate data within GeoCLIM workspace is:
> `X:\Users\Documents\<USER>\GeoCLIM\ProgramSettings\Data\Climate.`
> where **X** is the drive letter where the Windows Operating System is installed, and <USER> is the account name of the user that installed GeoCLIM.

2.2. Changing the GeoCLIM workspace

The default workspace is on the `C:\Users\<USER>\Documents\GeoCLIM`, the `C:\` It is, sometimes, too small to hold all the data outputs that GeoCLIM produces. So, it is recommended that users change the workspace to another drive. Once the workspace is changed to a different directory on the computer, the data and data-related files are moved to the new directory while the program files are kept on the original directory. Another benefit to having the workspace at a different location is that, if a new version of the GeoCLIM program is installed, the workspace can be reused. When the workspace remains at the default location, it gets replaced upon re-installation of the GeoCLIM application, and all data downloaded or created with the previous installation is deleted.

> **NOTE**: It is recommended to create a backup before installing a new version of GeoCLIM.

To change the workspace, follow the steps below:

1. From the GeoCLIM menu, go to File > Workspace Settings (**Error! Reference source not found. (1)**).
4. Browse to the new location in the ***Set Workspace Location*** field (**Error! Reference source not found. (2)**).
5. Click OK.

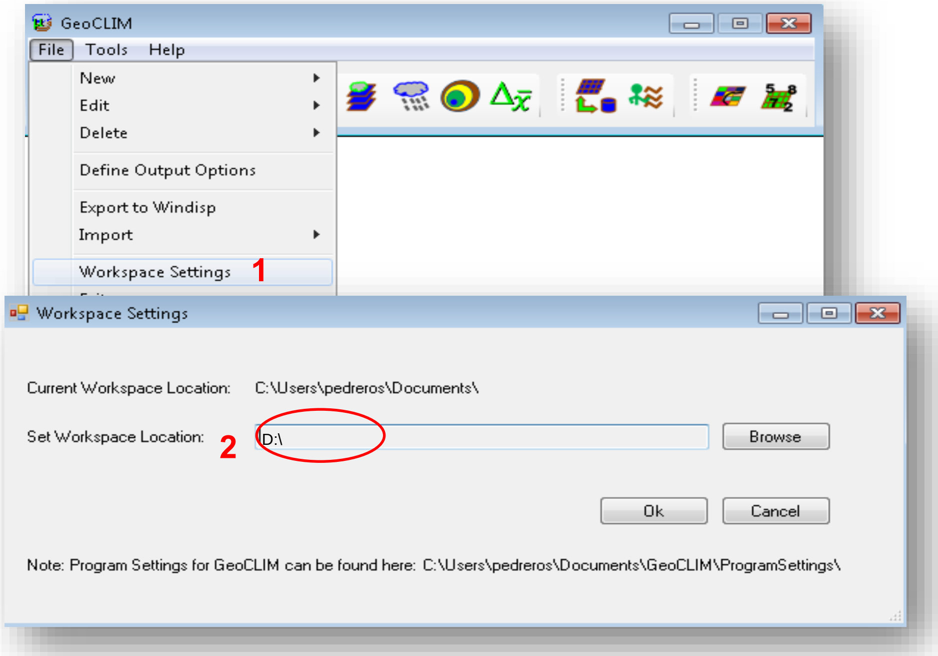


Figure 2.2 The GeoCLIM allows the user to change the workspace directory from the default to a new location.

> **NOTE**: When changing the Workspace, you do not have to create the GeoCLIM directory in the new drive location. GeoCLIM will automatically create it inside the directory you choose. For example, if you want the workspace to be `D:\GeoCLIM`, browse and select `D:\`.

2.3. The GeoCLIM Settings

After installing the GeoCLIM application, you can change some of the program settings, such as the region of interest or the selected climate datasets, by clicking the settings icon on the program toolbar. Once you click on the Settings icon, the ***GeoCLIM Settings*** window opens up (Figure 2.3**(a)**). The window has three tabs ***Region***, ***Mask,*** and ***Data***. See a description of each tab in the following sections:

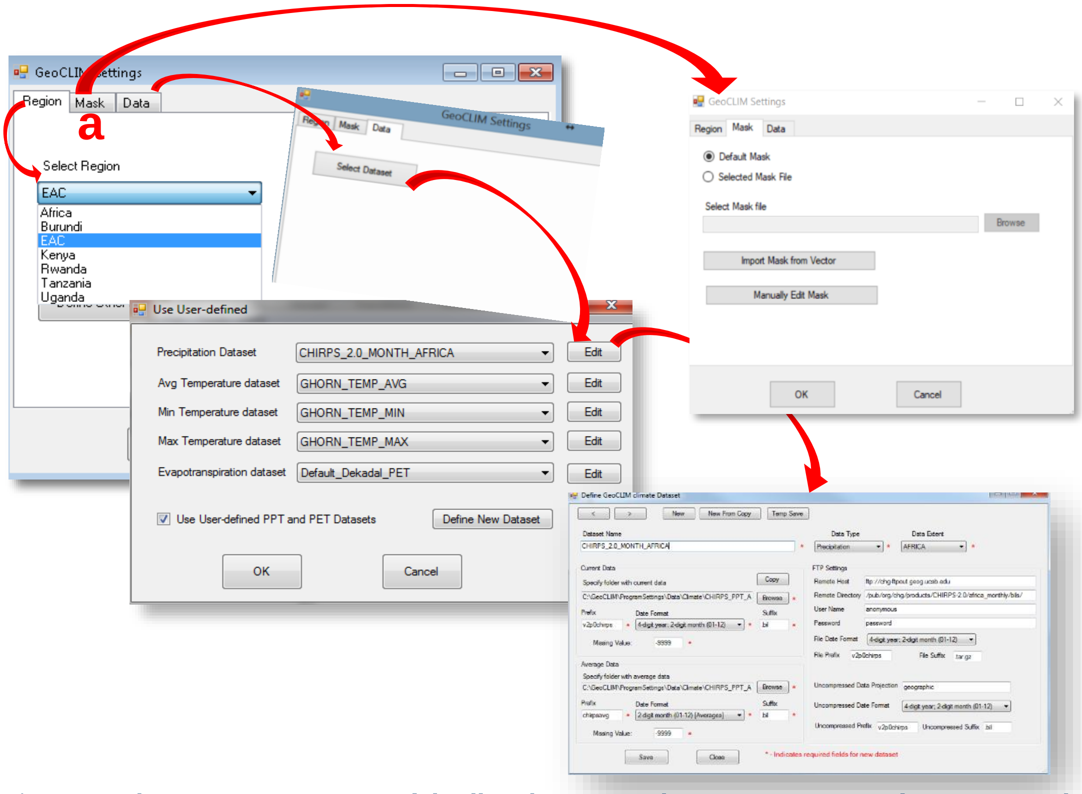


Figure 2.3 *The GeoCLIM Settings module allow the user to select/create a region, edit/create a mask, modify climate datasets, and edit parameters such as ftp information.*

2.3.1. Mask

Masks are raster images that are used to include only the desired area of interest (region) in the analysis and ignore the surrounding areas. A mask is a raster dataset with pixel-value of "1" for the area of interest and "0" for outside of the region. The mask facilitates the execution of the algorithms on the areas where the pixel value=1 while excluding areas where pixel values=0. In the Mask tab, you can select a new mask, create a new mask, or edit an existing one. Once you click on the ***Mask*** tab, a new window opens up (Figure 2.4**(a)**). Here, you can conduct the following tasks:

2.3.1.1. *Edit/create a Mask*

You can use the functions in the ***Mask*** tab if you have to make changes to the default mask for your region of interest or if you have to create a new mask (the default mask is defined in the region settings form, see create a region). For example, if the default mask does not cover the entire area of interest (see the circled area in Figure 2.4**(3)**), you could edit it, or you could use a polygon shapefile to create a new mask.

To edit/create a mask, follow these steps:

1) Open the ***Settings***.
2) Select the region of interest.
3) Click on ***Mask***.

4) To edit the default mask, continue with the following steps:
   a. Click on the Manually Edit Mask button (Figure 2.4**(1)**).
   b. Click on Edit (Figure 2.4**(2)**).
   c. Edit the image as described in, see section 4.1. Use a value of 0 for areas that should be excluded from the analysis and a value of 1 for areas to be included.
   d. Save the new mask and replace it on the ***Region*** form.

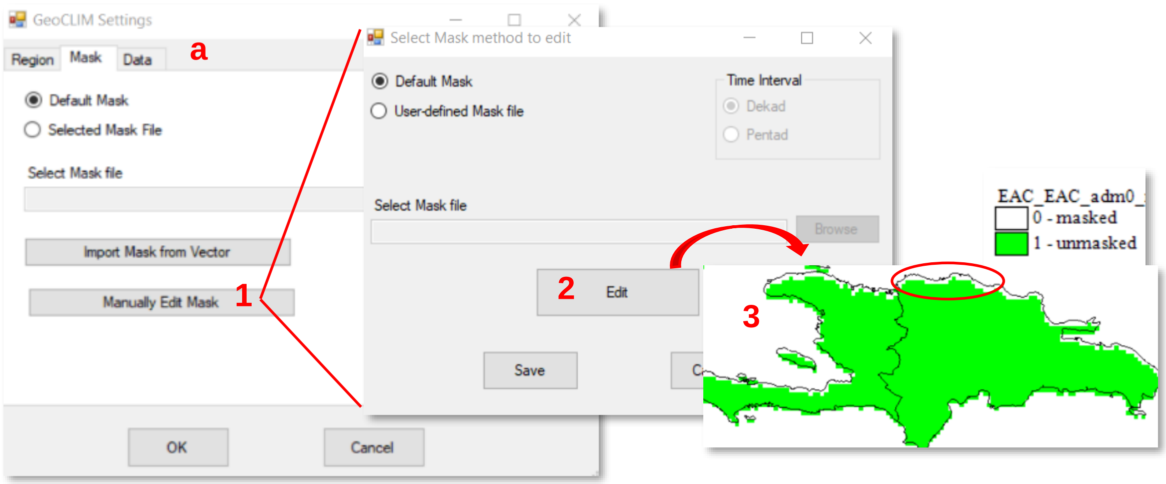


Figure 2.4 The GeoCLIM Settings allow you to make changes to masks.

5) To create a new mask using a Shapefile, continue with the following steps:
   a. Click on Import Mask from Vector button (Figure 2.5**(1)**).
   b. Select the Shapefile to use (Figure 2.5**(2)**). The output directory and name are selected by default.
   c. Click OK.
   d. Once the conversion is completed, open the new file using the Spatial Data Viewer (Figure 2.5**(3)**).

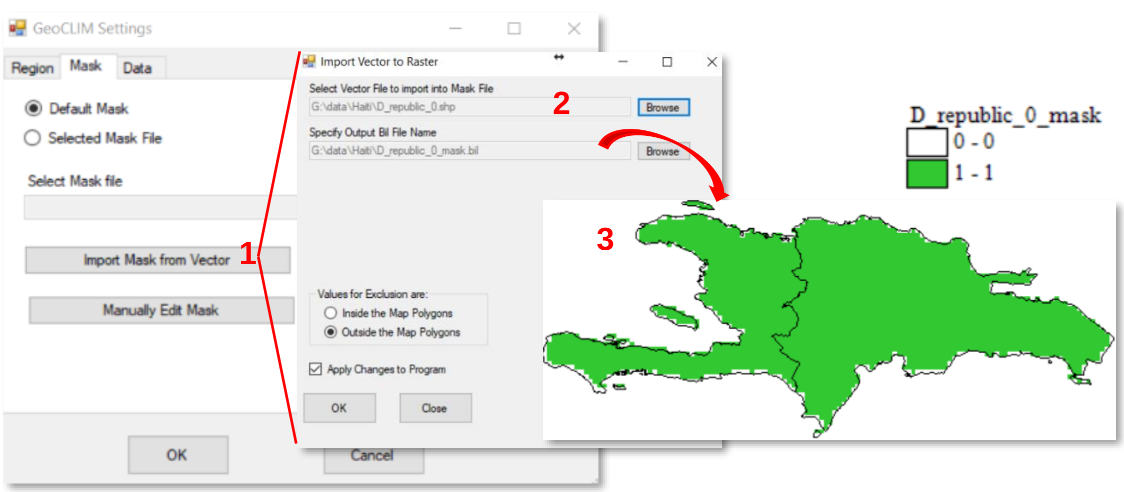


Figure 2.5 The GeoCLIM Settings allows you to create a new mask for the region of interest.

2.3.2. Region

GeoCLIM works on specific areas of interest called Regions. The definition of a Region includes a box of latitude and longitude, a default mask (see the previous section for the explanation of a mask), and a set of vector maps that outline the Region. There is a set of predefined regions in GeoCLIM. If the desired region is not in the predefined list, you could create a new region. See the next section to learn how to create a Region.

2.3.2.1. *Create a new region in GeoCLIM*

There are two options to create a user-defined region in GeoCLIM. The first option is to go to File > New > Region and fill out all the fields. The second option is to modify an existing region. See an example of the second option below:

1. Open an existing region by clicking the File > Edit > Region (Figure 2.6).

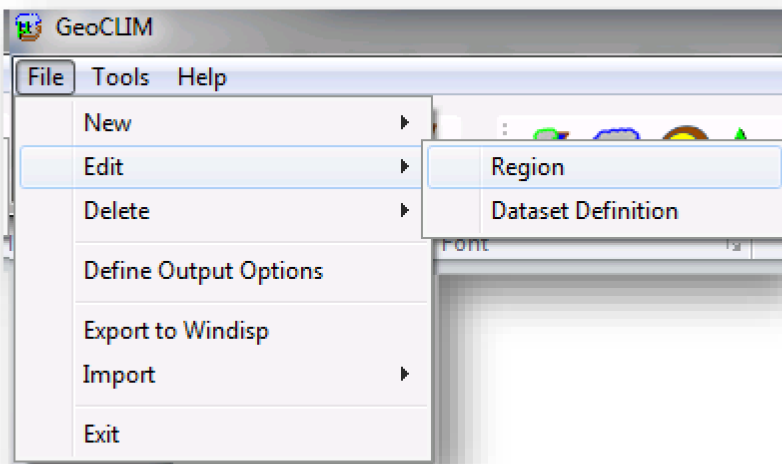


Figure 2.6 *To open an existing region, go to File >Edit>Region*

2. Select an existing region and click OK (Figure 2.7).
3. Change the Comments of Region File field of the existing region to align with the information of the new one and click Save As to save it as a new region (Figure 2.8).

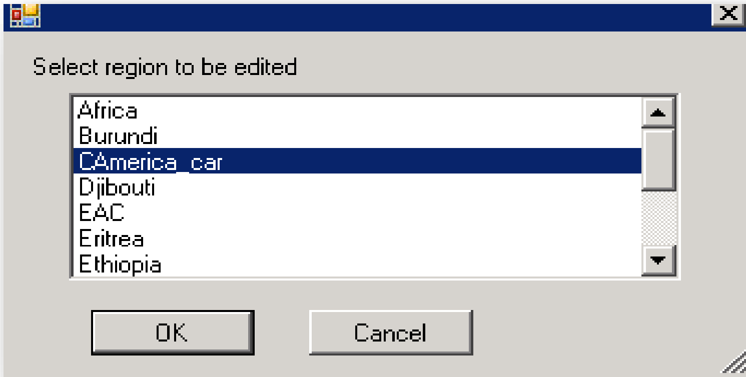


Figure 2.7 *The GeoCLIM contains several regions that you could use as an example to create a new one.*

4. Next, enter the minimum and maximum latitude and longitude for the new region (Figure 2.8 **(1)**). This step could be done in two ways:
   a) Entering the values by hand.
   b) Extracting the coordinates from an existing map as follows: Click on the Get Extent from Map and select one of the GeoCLIM geographic options. The tool will retrieve the coordinates from the selected input.
5. The fields Height of pixel and Width of pixel refer to the pixel size in degrees for the output products using this region (Figure 2.8 **(2)**).
6. For the Default Mask File, enter a BIL file to be used as a mask for the new region. The example in Figure 2. uses a global mask, but the geographic boundaries of the region (min/max latitude and longitude) limit the output area (Figure 2.8 **(3)**).
7. Finally, for Default Map File, specify a vector layer in shapefile format depicting polygons related to the region of interest, such as political boundaries, watersheds, etc. (Figure 2.8 **(4)**).
8. Click on the Define button in Figure 2.8 **(5)**.
   a) Click on the Add button, as shown in Figure 2.9
   b) Select a vector file that outlines the region; this shapefile is used when displaying the results. More than one vector file can be used (e.g., districts and provinces), with different widths or colors to distinguish between them.

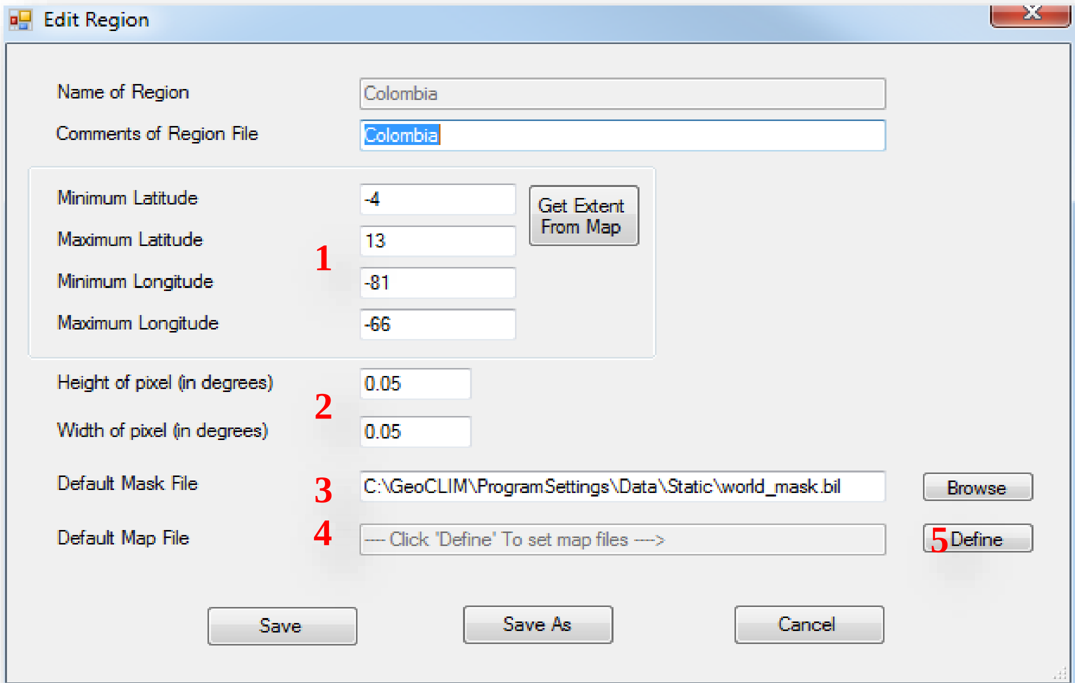


Figure 2.8 *The Edit Region form describes the geographic boundaries* **(1)** *of the region, the pixel size* **(2)** *for the outputs, the mask* **(3)** *used, and the outline shapefile in the Default Map File box* **(4).**

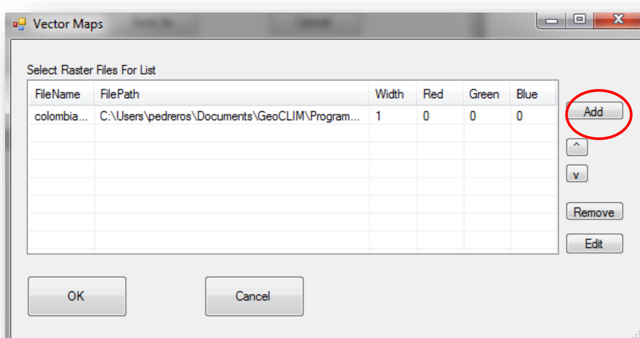


Figure 2.9 *Add a shapefile that serves as the outline on the output products.*

NOTE: The size of the region **must** be smaller or equal to the size of the climate dataset. Otherwise, the error message OFF-REGION will display on the ***View Available Data*** tool, the GeoCLIM tools will display an error message, and they won't run.

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 “,” instead of “.” to separate decimals. To fix this problem, do the following: from Windows, go to `Control panel` > `Clock, Language, and Regions`> `change date, time, or number Formats` click Additional Settings. Make sure that the decimal symbol is “.” (Figure 2.).

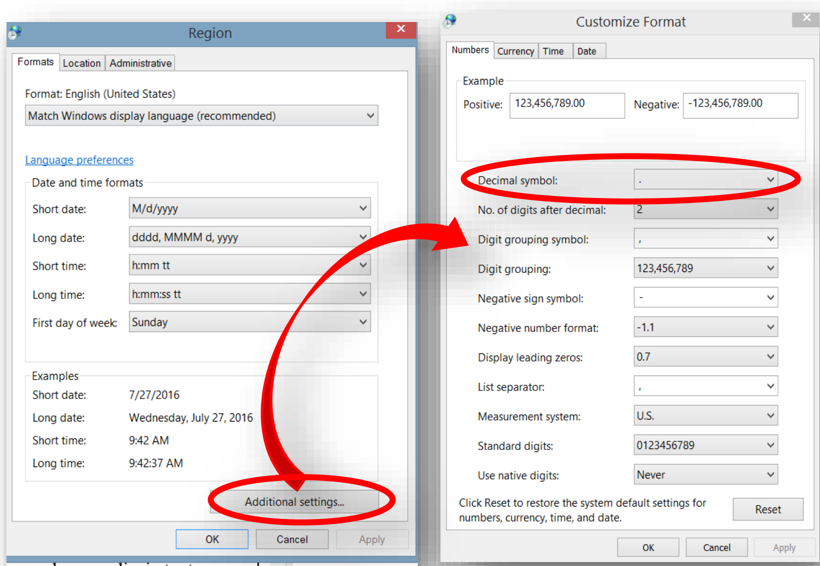


Figure 2.10 *One possible error when using GeoCLIM is the decimal separator. Go to the Additional settings and change the ‘Decimal symbol’*

2.3.3. Data

The ***Data*** tab, as shown in **Error! Reference source not found.**, facilitates the selection of a vailable datasets (rainfall, temperature, and evapotranspiration) for analysis. The ***Data*** tab also allows you to add new climate datasets or edit existing ones and add/edit FTP information to facilitate the update of a dataset online.

DATA for OTHER REGIONS: GeoCLIM provides predefined settings to download final CHIRPS for Africa and Central America windows. To obtain data for other regions, you must download the global dataset. To download dekadal
global preliminary CHIRPS
https://data.chc.ucsb.edu/products/CHIRPS-2.0/prelim/
To download dekadal global final CHIRPS
https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_dekad/

2.3.3.1. *Making new data available for GeoCLIM*

GeoCLIM requires climate datasets in raster (`*.bil`) format. You can use any of the pre-defined datasets (e.g., `CHIRPS_PPT_AFRICA_MONTHLY` or `GDASUSGS_PET_GLOBAL_DEKADAL`) or create your own. For example, you can create a new rainfall dataset by blending gridded satellite rainfall estimates and rainfall station values using the BASIICS function in GeoCLIM. For this new dataset to be available in GeoCLIM, you will need to define it in the ***Settings*** so that the software can read the data. To do this, the file names and the data must have a specific format (see the next two sections to learn how to define a new dataset).

2.3.3.1.1. Define climate data filename

The file name in a climate dataset uses the following format:
`<prefix> <date-format> <suffix>`
where:
`<prefix>` is a set of characters before the date that could be associated with the dataset name, descriptor, or source; (Figure 2. **(1)**).
`<date-format>` The date is composed of the `<year>` and `<period of time` (pentad, dekad, or month)>. The GeoCLIM program has a variety of pre-defined formats for the date; for example, `YYYYMM` corresponds to the 4-digit year followed by the 2-digit month (i.e., `01`, `02`, `03…12`). More formats are available for pentadal, dekadal, and monthly datasets under Date Format ˅, as seen in Figure 2. **(2)**.
`<suffix>` The suffix corresponds to any character after the date, including the extension of the file (e.g., `.bil`) (Figure 2.11 **(3)**).
For example, to name the rainfall total for the 36th dekad of 1991 from CHIRPS 2.0, the name of the `BIL` file is "`v2p0chirps199136.bil.`" In this case, the `<prefix>` is v2p0chirps, to indicate that it is CHIRPS 2.0 data, the `<date-format>` is comprised of a 4-digit year (1991) and a 2-digit dekad (36), and the `<suffix>` is the extension for a `BIL` file including the "." (`.bil`).
To learn more about the data formats used in GeoCLIM, see chapter 3.

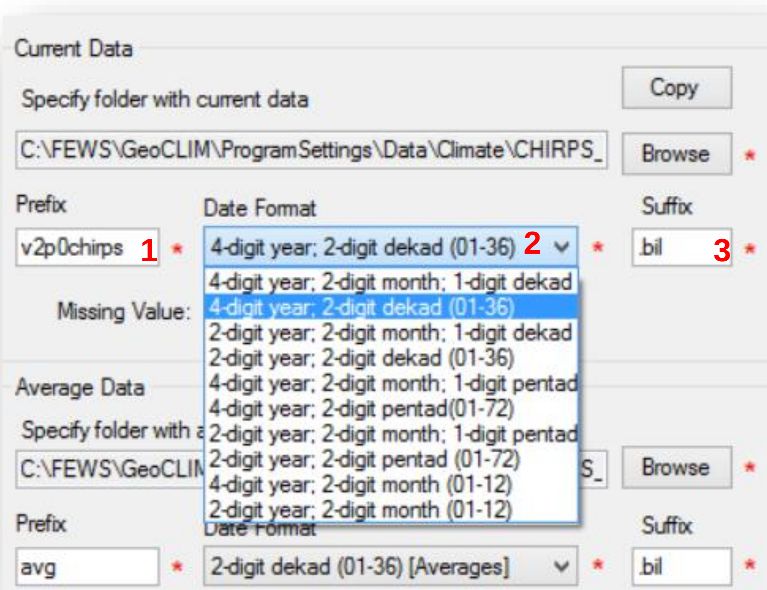


__Figure 2.11__ The name of the data file is composed of a prefix, a date (year and period of time), and a suffix.

2.3.3.1.2. Define a new dataset in The GeoCLIM

The definition of a climate dataset includes the location and name of the folder containing the data, the name of the raster files, the missing data value, and, where applicable, the download information for the dataset updates.

To define a climate dataset follow these steps:

1. Create a new sub-directory in the GeoCLIM data repository **X**:\\GeoCLIM\ProgramSettings\Data\Climate\ and copy the BIL raster data (*.bil and *.hdr files). Make sure that the raster file names follow the format described in the previous section. **X** is the drive where the GeoCLIM workspace is located, see changing the GeoCLIM workspace. In this example, the new subdirectory will be called NEW_PPT_CHIRPS_DEKADAL. Once the data files are in the new directory, the next step is to define it in GeoCLIM, so that the tool can read the data.
2. Open the dataset definition form. There are three ways to get to the dataset definition form:
   a) **Option 2a**: From the GeoCLIM menu, go to File > New > Dataset definition to create a new climate dataset definition (Figure 2.12 **(2a)**).
   b) **Option 2b**: On the GeoCLIM toolbar, click the Settings > Data > Select Dataset > Define new Dataset (Figure 2.12 **(2b)**).
   c) **Option 2c**: Open the definition of an existing dataset by clicking on the Edit button (Figure 2.12 **(2c)**) and make the changes to reflect the new dataset. Save the form with a new name.

Figure 2.12 shows a completed data definition form. The left side of the form, in the blue box, defines the local aspects of the dataset: the name, the path to the data directory, the file-naming convention (prefix, date format, and suffix), the missing (NoData) value, and the location and format of the dataset averages. The GeoCLIM creates the average data for the period that you define, see Updating the GeoCLIM averages for details on how to create averages using GeoCLIM.

NOTE: In case you want to edit an existing dataset, open the dataset definition as describe on 2.c above and edit the form as described in the following sections.

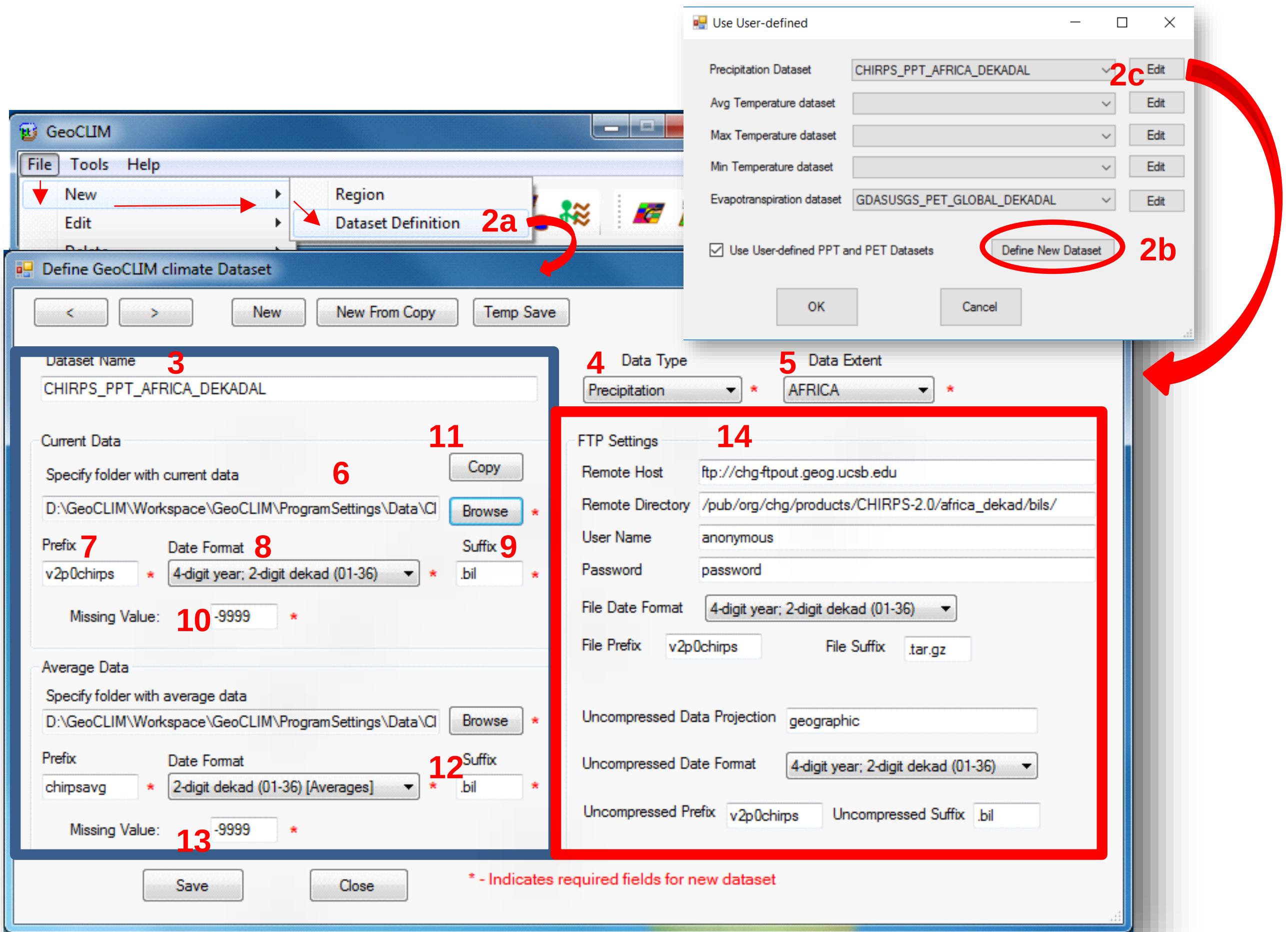


Figure 2.12 There are different ways to get to the Data definition form in GeoCLIM: 2a) File>New> Data Definition, or 2b) Settings>Data>Select Dataset>Define New Dataset.

3. Dataset Name: Type the name of the dataset; it should be the same as the name of the new directory, created in step 1, containing the data, `NEW_PPT_CHIRPS_DEKADAL`, as described in 1 (Figure 2.12 **(3)**). The dataset name should have no spaces.
4. Select the Data Type ∨: Select the type according to the data: Precipitation, Avg Temperature, Min Temperature, Max Temperature, or Evapotranspiration (Figure 2.12 **(4)**).
5. Select the Data Extent ∨: There are only three data extents on the current version of GeoCLIM: Africa, Central America, and Global. If the dataset is not for a location in Africa or Central America, select Global (Figure 2.12 **(5)**).
6. Specify folder with current data: Browse to the new directory that contains the data `\GeoCLIM\Programsettings\Data\Climate\NEW_PPT_CHIRPS_DEKADAL` (Figure 2.12 **(6)**).
7. Prefix: Fill out the prefix as defined above (Figure 2.12 **(7)**). See [Define climate dataset filename](#) for more details.

8. Select the Date Format ∨. For this example, select 4-digit year; 2-digit dekad (01-36) (Figure 2.12 **(8)**).
9. Fill out the suffix "`.bil`" (Figure 2.12 **(9)**).

NOTE: Do not forget to add the '.' before "`bil`"

10. Fill out the missing value; for example, the missing value in CHIRPS is -9999 (Figure 2.12 **(10)**).
11. Click on the Copy button below the Dataset name (Figure 2.12, **(11)**); this copies the data directory path onto the **Average Data** section to ensure that the long-term averages are saved into the same directory as the time series data files. See Updating the GeoCLIM averages for details on how to create averages using GeoCLIM.
12. Fill out the prefix, date-format ∨, and suffix ("`.bil`") for the average files (Figure 2.12 **(12)**).
13. The missing value should be the same as defined for the current data (Figure 2.12 **(13)**).
14. The **FTP Settings** section contains the necessary information to download data updates. If the data does not have FTP information, this section can be empty (Figure 2.12 **(14)**).

Once you complete the form, click Save and then YES on the following screen to create a new dataset or NO to update an existing one. Then click YES to confirm.
To start using the new dataset, go to the GeoCLIM settings > Data > Select Dataset and select the new dataset from the **Precipitation Dataset** list. Then, click OK to close all the windows. Once you select a dataset on the settings, all the GeoCLIM functions will use it as default.

2.3.3.2. *Clipping raster data to your region of interest*

If you are using CHIRPS, but your region of interest is outside of the Africa or Central America windows, you would have to download the global data and clip it to your region of interest. GeoCLIM offers two ways to clip the data; one is by blending the raster CHIRPS with local stations. During this process, you can define the size of the output region; see Blend rasters/grids with stations. Another way is by clipping the CHIRPS data to your region of interest using the Tools > Window Grids function in GeoCLIM. Follow the steps below:

1. Open the Window Grids function from the Tools ∨ pull-down menu (Figure 2.13).
2. Click on Select Files to open the **Raster File List Builder (**Figure 2.14).
3. Click on Add to select the raster files.
4. Click OK to go back to the Window Raster Files.

5. Specify the output folder and the region extent and click OK.

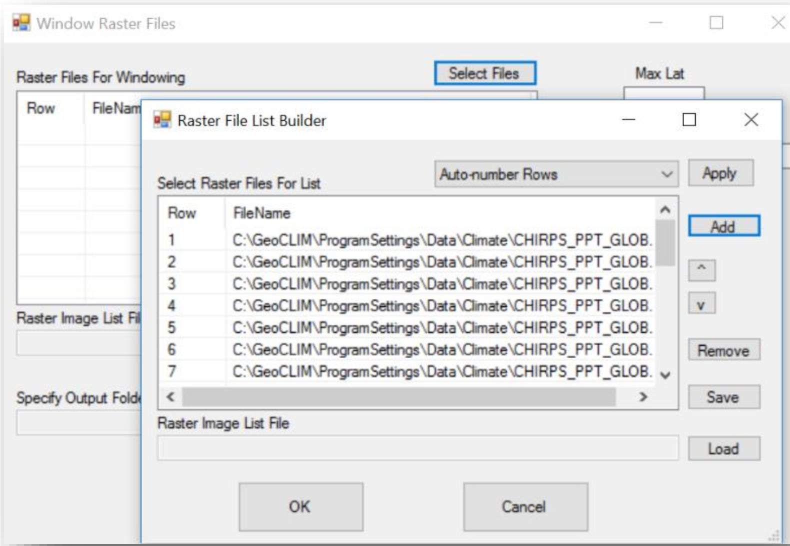


Figure 2.14 Select the raster files to be clipped and enter the lat/lon coordinates for the region of interest.

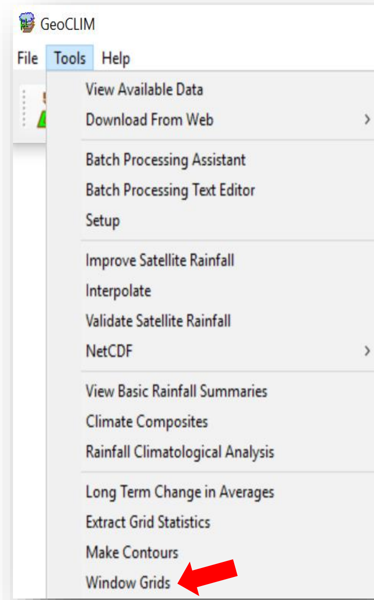


Figure 2.13 To clip the global datasets to your region of interest, open the Window Grids function from the Tools pulldown menu.

2.3.3.3. *NetCDF to bil format*

In many cases, datasets are available in different formats; GeoCLIM allows you to import data in NetCDF format by converting them into "`.bil`" format. To convert NetCDF files to `.bil` do the following:

1. Make sure that the file name in the NetCDF files contains the date in GeoCLIM-readable format, see Define climate data filename for details on the date format.

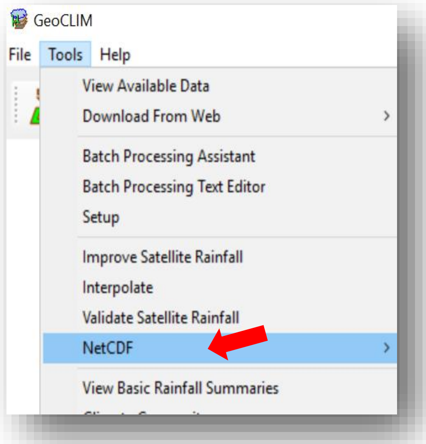


***Figure 2.15** To convert NetCDF files into ".bil" go to the Tools/NetCDF menu and select import NetCDF.*

2. Open the NetCDF converter form by going to the Tools>NetCDF pull-down menu (Figure 2.).
3. Fill out the form to define the input and output directories and name format (Figure 2.).

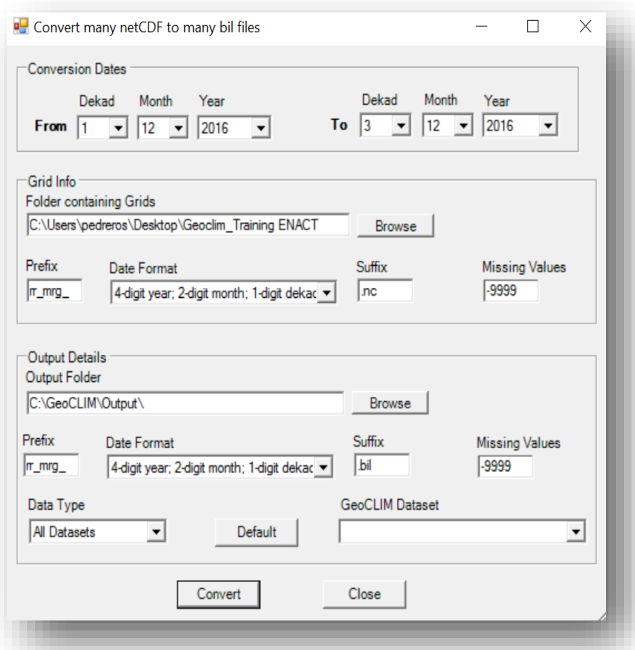


***Figure 2.16** Define the input and output directories and name formats.*

NOTE: The GeoCLIM module es built to handle ENACTS netCDF files; please use other GIS packages for other netCDF types.

2.3.3.4. *Data availability/export data*

After selecting a dataset, click on the ***View Available Data*** icon to see the list of files by date available. Here you can also make sure that the data covers the region selected (Figure 2.). If the ***Covers Region*** column shows **OK**, you can proceed with the analysis functions. If the column shows **OFF-REGION**, the geographic extent of the data does not cover the region. At this point, you must fix the region file.

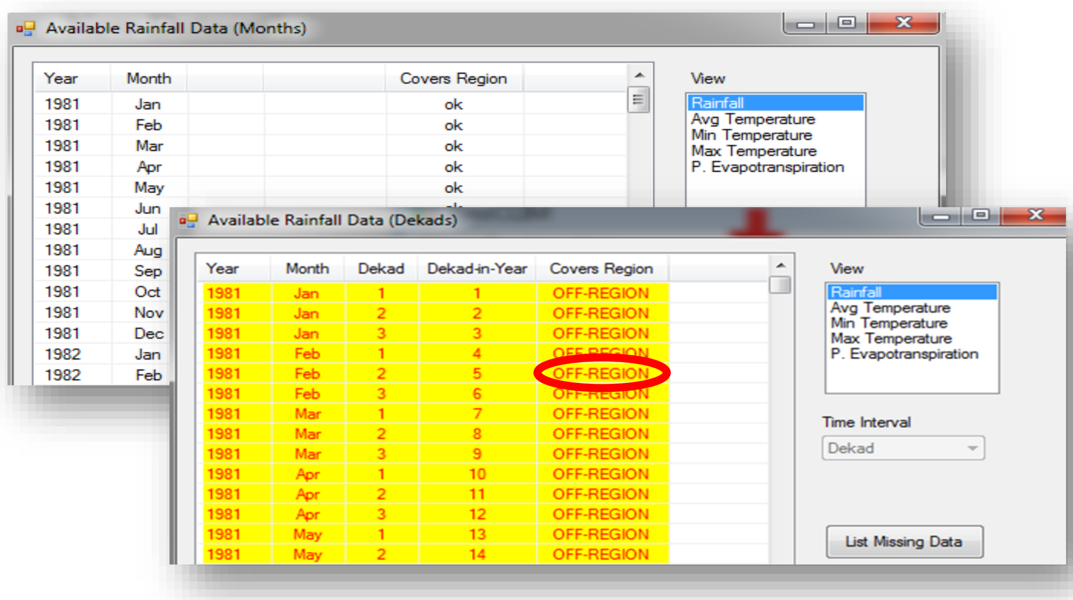


Figure 2.17 *The* ***Available Rainfall Data*** *window allows you to identify the length of the time series, whether the data extent covers the region of interest, missing data files and export files into an archive .*

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 Create a new region in GeoCLIM to learn how to create/edit a region).

The ***Available Rainfall Data*** tool also allows you to identify the missing data in the time series and to export a file or the complete dataset into a GeoCLIM archive or NetCDF. Click on List Missing Data to obtain a list of missing files in the time series. Use the Export button to save selected files into a data archive that could be shared with other GeoCLIM users (see chapter 12 for more information in data archives) or save the files to NetCDF.

2.4. Resetting the GeoCLIM program

GeoCLIM may stop working after displaying an error message. If, after restarting, the program still does not work, it might be that some of the setting files are corrupted. A solution to this problem is to reset the program following the steps below:

1. Close GeoCLIM.
2. Open the Windows Task Manager and check that the GeoCLIM process is not running in the background.
3. Go to the default directory, `C:\Users\<username>\Documents\` and rename the GeoCLIM directory (e.g. `GeoCLIM2`).

4. Run the GeoCLIM installer file as if you were to install the program. There is no need to have administrator permissions.
5. Follow the instructions from the installer and in ***Program Maintenance***, select ***Repair,*** and follow the instructions.
6. Start the GeoCLIM application and follow the same instructions that you initially used.

> **NOTE**: In case you did not change the workspace default installation directory to a new location, after resetting the program, copy all your data files from GeoCLIM2 to the corresponding directory in the new GeoCLIM directory.

Chapter 3: Data Types in The GeoCLIM

Summary

This chapter examines the different types of data formats used in the GeoCLIM program. 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`).

3.1. Characteristics of the raster dataset

A band interleaved by line (`BIL`) dataset contains two files: a (`*.bil`) file and a header file (`*.hdr`). The `.bil` file is a binary file that contains the pixel values (e.g., rainfall, temperature, etc.), while the `HDR` file contains the characteristics of the dataset, such as the geographic location, pixel size, and depth.

The header file is an ASCII text file; it can be generated or edited from a text editor (e.g., Notepad). For example, Figure 3.1 shows that the header file contains information about the number of columns (`ncols`), number of rows (`nrows`), number of bits per pixel (`nbits`), and size of pixel (`xdim` and `ydim`), among others. Figure 3.1 also shows the xdim and ydim values corresponding to the horizontal (x-dimension) and vertical (y-dimension) dimensions of a pixel with a size of 0.05 degrees, which is about 5 kilometers. The `ulxmap` and `ulymap` correspond to the x-axis and the y-axis coordinates of the center of the upper-left pixel of the raster image. There are additional keywords that the header could have (Figure 3.2) (ArcMap 10.3 Help, ESRI). Sometimes, if the header file is incorrect, you may need to modify it so that the data is read correctly by the program.

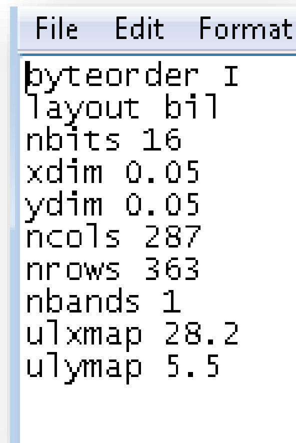


Figure 3.1 *Example of a HDR file.*

> **NOTE**: By default, the BIL dataset pixel type used is unsigned integers, unless the keyword "pixeltype" is used in the HDR file, and its value is "signedint".

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 header file is composed of a series of key words and their respective accepted values. Source: ArcMap 10.3 Help, ESRI*

An important keyword in the header file is the `pixeltype` since it defines the type of value, unsigned (+), or signed (+ or -) a pixel could have. For example, rainfall data could only have unsigned (+) values, while temperature could have signed values (+ or -). In the example in Figure 3.1, the `pixeltype` is missing, so the program assumes that the data is unsigned; in the case of temperature, we would have to add a new line on the header file defining the `pixeltype`. (e.g., `pixeltype unsigned int).` Another keyword to keep in mind is the `nbits` because it indicates the number of bits per pixel or the depth of the raster image (e.g., nbits=16 bit means that a pixel in the raster dataset can have any one of 2^{16} = 65536 unique values). Figure 3.3 (ESRI, Support 2016) shows a list of values a raster dataset could have 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

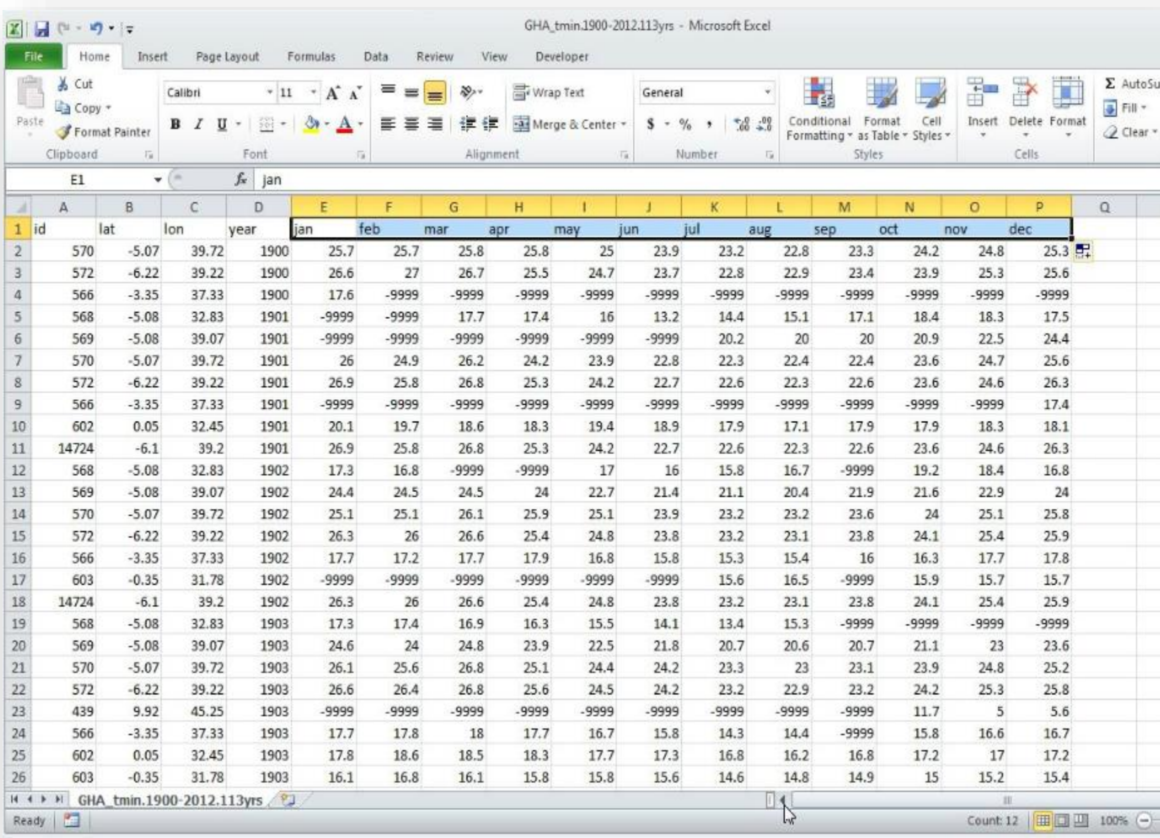
Figure 3.3 *The range of values a dataset could store depends on the nbits*

3.2. Vector data

Another type of data used in GeoCLIM is vector data in shapefile format (`*.shp`). The current GeoCLIM version (1.2.1) only allows editing access to polygons. To get more information about how to open, create, or edit shapefiles, go to chapter 4.

3.3. Tables

The GeoCLIM program uses tables in comma-delimited format (`*.csv`) as input and output data. For example, tables are inputs in the process of blending raster data with station values (***BASIICS***) or validating raster data. For the blending process, the CSV table must have columns for *ID S, latitude (lat), longitude (long), year, and time period* (pentads, dekads, or months), such as the months of January-December in Figure 3.4. The ID, lat, lon, and year columns do not have to be in any particular order, and additional columns are permitted. However, the time-period columns need to be consecutive (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 3.4 *The GeoCLIM accepts tables in comma delimited format (CSV).*

Chapter 4: Spatial Data Viewer

Summary

The ***Spatial Data Viewer*** tool facilitates the visualization and editing of both raster and vector data.

4.1. Working with raster data

4.1.1. Displaying raster data

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

Figure 4.1 The GeoCLIM main tools

2. Next, click on the ***Open Raster Map*** button (Figure 4.2 **(1)**).

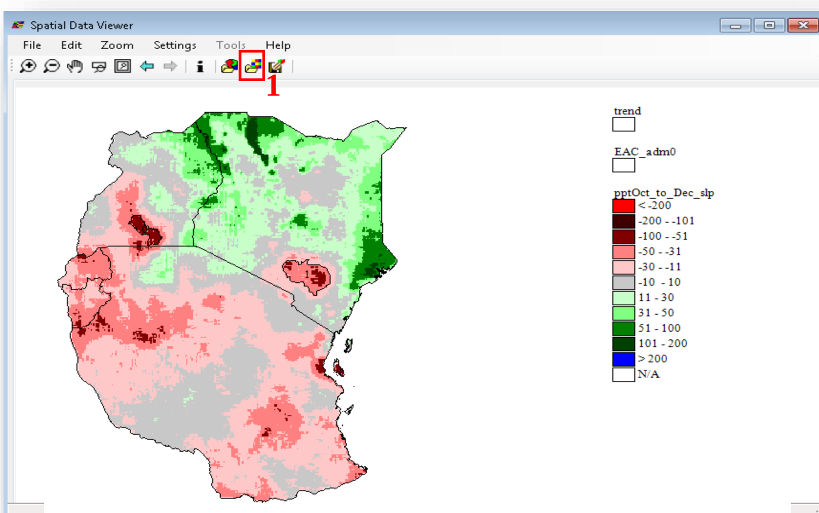


Figure 4.2 GeoCLIM allows you to work with both raster (.bil format) and shapefiles.

3. Once in the ***Select Raster Image window*** (Figure 4.3), click the Browse button **(1)** to select a raster file (`.bil`). Click on the Browse **(2)** to select the legend file, then click OK.

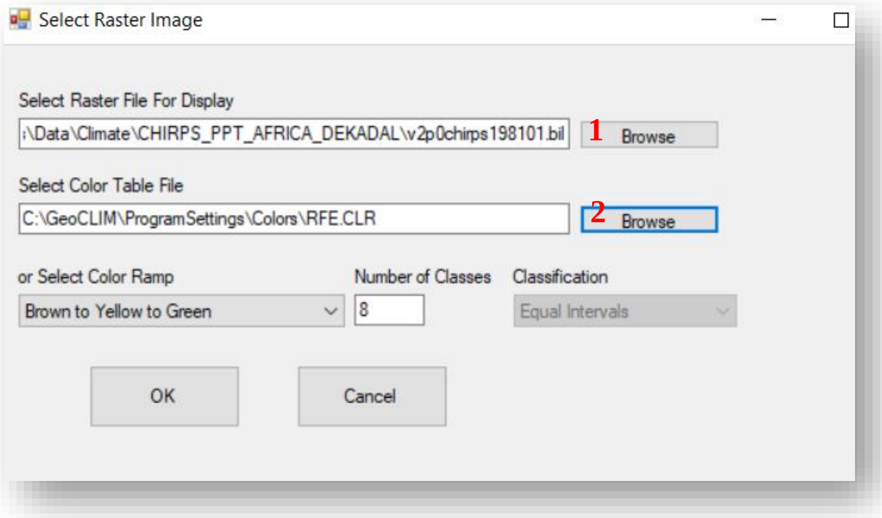


Figure 4.3 *Select the raster file and the color file.*

4. Figure 4.4 **(a)** shows an example of a CHIRPS file. This raster file has a value of `-9999` as the "No Data" value. This data value is being included in the map display and legend. To view the pixel values on the map, click on the legend, a red square appears, then move the mouse over the image to see the pixel value and its coordinates displayed on the lower-left corner of the ***Spatial Data Viewer*** (Figure 4.4 **(b)**).
5. Right-click on the legend and select ***Change Legend***. Browse to `Documents\GeoCLIM\ProgramSettings\Colors` and select `RFE.clr` to display the range of values (Figure 4.4 **(b)**).

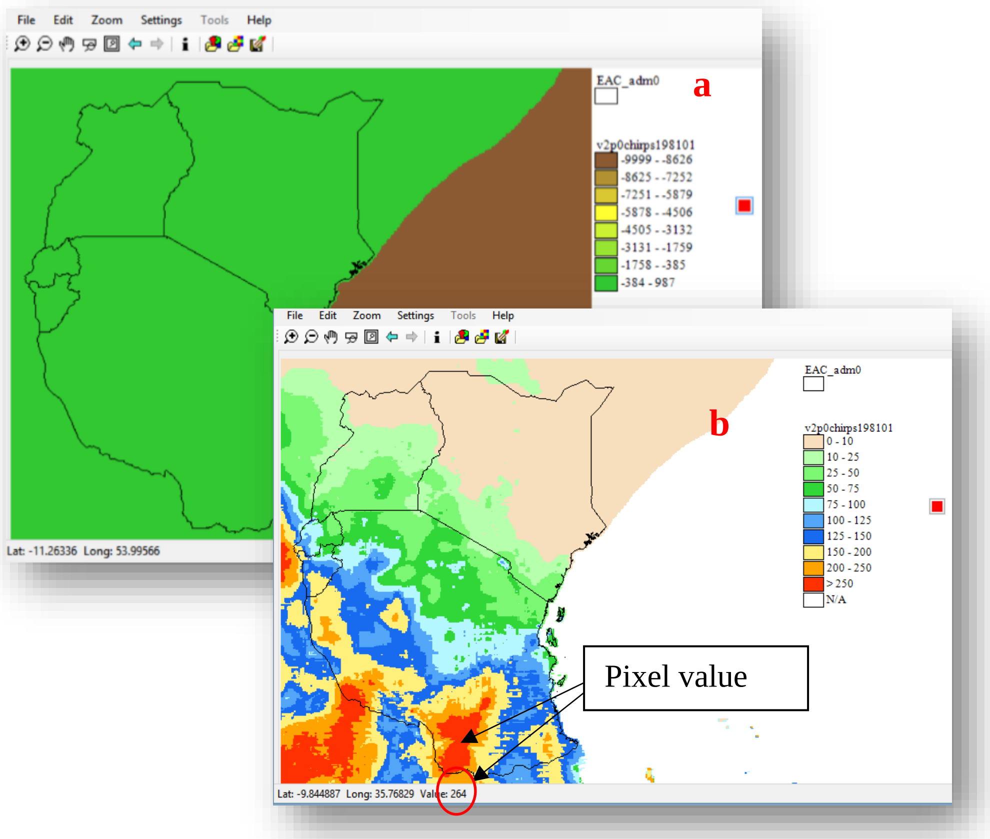


__Figure 4.4__ GeoCLIM allows the use of raster data in .bil format. (a) Shows a raster dataset using the default legend. (b) Shows the raster file after selecting and assigning a legend, and shows how to view actual pixel values.

4.1.2. Changing raster values

The ***Edit Raster*** tool allows the user to change pixel values for a `*.bil` raster file. To change pixel-values, follow these steps:

1. Open the ***Spatial Data Viewer*** tool.
2. Open a raster dataset, as shown in section 4.1.1.
3. Activate the editing tools by clicking the Settings ⋁ dropdown menu (Figure 4.5 **(2)**).
4. Select Show Editing Tools (Figure 4.5 **(3)**).

5. **(4)** Three buttons appear: two menus (the Edit Vector ∨ and the Edit Raster ∨ drop-down menus) and the multicolor icon, which is used to create a new vector.

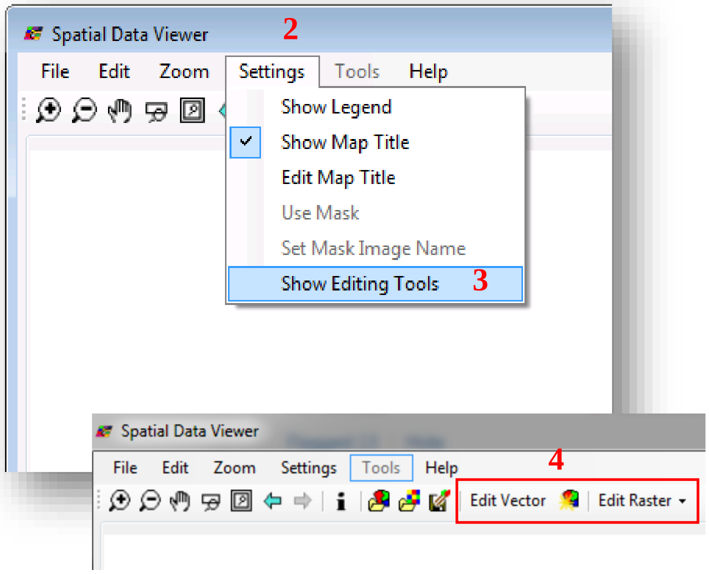


Figure 4.5 activating the editing spatial data tools

6. Select the raster to edit by clicking on the legend. A red square will appear next to the legend, indicating that the raster is ready to be edited (Figure 4.6 **(1)**).

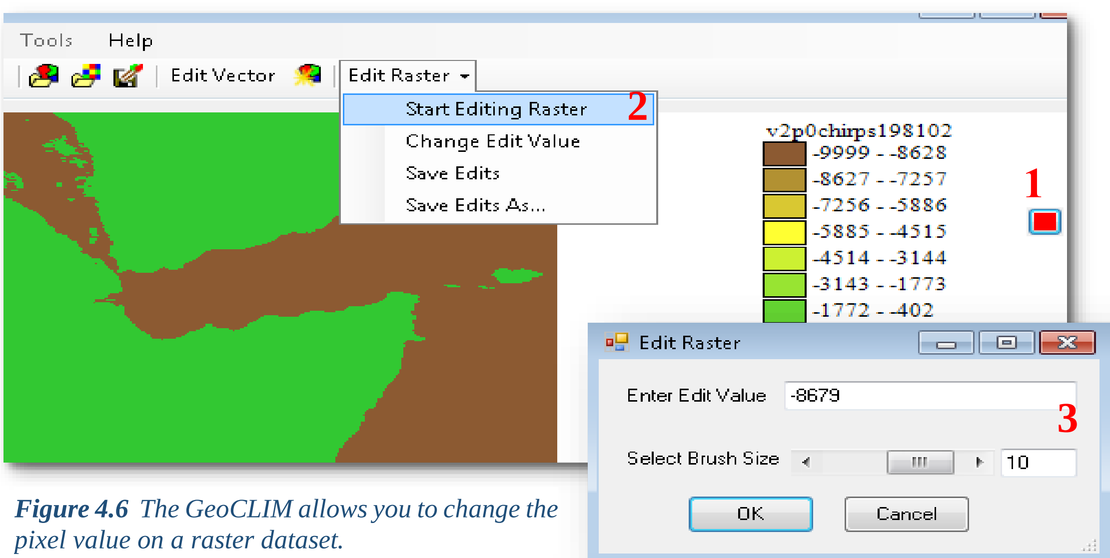


Figure 4.6 The GeoCLIM allows you to change the pixel value on a raster dataset.

7. From the Edit Raster ∨ drop-down menu, select Start Editing Raster (Figure 4.6 **(2)**).
8. Add the new value, then select the size of the brush (1 - 10); this indicates the number of pixels changing values (Figure 4.6 **(3)**).
9. Click OK.
10. Start sliding the mouse over the raster to change the pixel to the new value (Figure 4.7).
11. To finish, click on the Edit Raster ∨ drop-down menu again and select Save Edits to save the changes on the current raster dataset or select Save Edits As to save it as a new raster dataset.

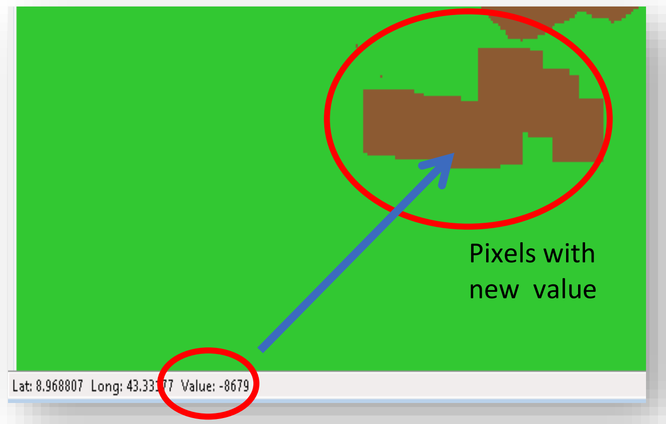


Figure 4.7 *Slide over the raster to change the pixel value.*

4.1.3. Editing legends for raster data

The GeoCLIM default legend file may not represent all the values of a raster dataset on the map viewer. It may be necessary to edit the current legend to create a new one or import an existing legend. To edit the legend for a raster dataset:

1. Open the raster dataset, as shown in section 4.1.1.
2. Right-click on the raster dataset legend and select Change Legend.
3. From the ***Select Raster Image*** window, click on Edit Color Table button.
4. In the ***Color Table Editor*** window, the range of values (***From*** and ***To*** columns), color values in RGB (Red, Green, and Blue columns), label, and color columns from the existing legend are displayed (Figure 4.8 **(1)**).
5. To make changes to a specific line, double click on it. The ***Editor Color Table Item*** window will open, and the values for each field can be changed (Figure 4.8 **(2)**).

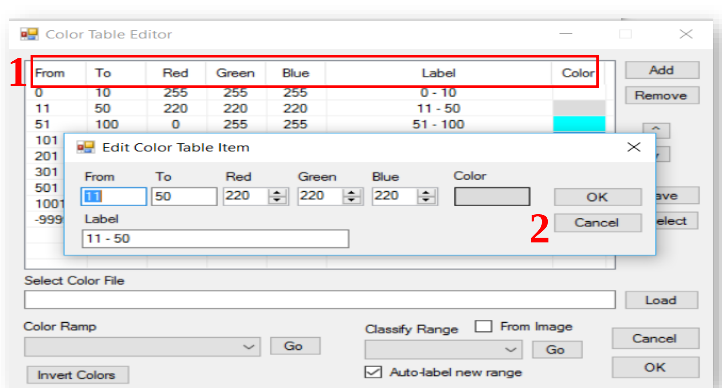


Figure 4.8 *The Color Table Editor allows you change color legends.*

6. If a new range of values is necessary, click on the Add button in the upper-right corner of the ***Color Table Editor*** and enter all the necessary parameters.

7. It is recommended to save the legend as a new CLR file with a different name to the: `\GeoCLIM\ProgramSettings\Colors` directory.

4.2. Working with polygons

Use the ***Spatial Data Viewer*** tool to create and/or edit polygon shapefiles. To display polygons, follow the instructions below (Figure 4.9).

4.2.1. Displaying polygons shapefiles

To display an existing shapefile, follow the steps below:

1. Click on ***Open Vector Map*** icon (Figure 4.9 **(1)**).
2. Browse to the directory and select the polygon shapefile.
3. Click OK to open.

4.2.2. Showing labels for polygons

To display the labels on a polygon shapefile:

1. Open a shapefile as described above.
2. Right-click on the legend and select Show Labels. This option opens the ***Define Labels*** window.
3. Select the field with the values to be displayed as labels and click OK.

4.2.3. Create a new polygon shapefile

1. Make sure that none of the zoom and pan tools are selected (Figure 4.9 **(2)**).
2. Click the Settings ∨ drop-down menu and select Show Editing Tools. Two menus will appear:
   a) Edit Vector button and
   b) Edit Raster ∨ drop-down menu.
3. Click on the Edit Vector button to enable additional tools (Figure 4.9 **(3)**).
   a) ***New Feature*** – add a new polygon.
   b) ***Edit Vertices*** – change the shape of a polygon by moving the vertices.
   c) ***Edit Attribute Data*** – this function is disabled in this version
   d) ***Save As Shapefile*** – saves the new edits.
4. Click on the ***New Vector File*** icon, browse to the directory where the new shapefile will be saved and give it a file a name, leave the other options with default values, and click OK.
5. Right-click on the empty shapefile's legend and select Start Editing. A green box with an "E" indicates that the shapefile is ready to be edited (Figure 4.9 **(4)**).
6. Make sure that the **New Feature** icon is activated (indicated by a blue line around the icon). Next, start clicking on the map display to draw vertices of the new polygon. Every click on the map makes a new vertex. To finish, right-click to close a polygon. Once the polygon is closed, you can start digitizing a new polygon.

7. To save the new polygon(s), click the save button (Save As shapefile). This option asks for the shapefile where you will save the polygons; select the one you initially created in step 4.
8. To stop editing:
   a) Save first and then
   b) Right-click on the shapefile's legend and select Stop Editing. Note that the "E" in the green square disappears.

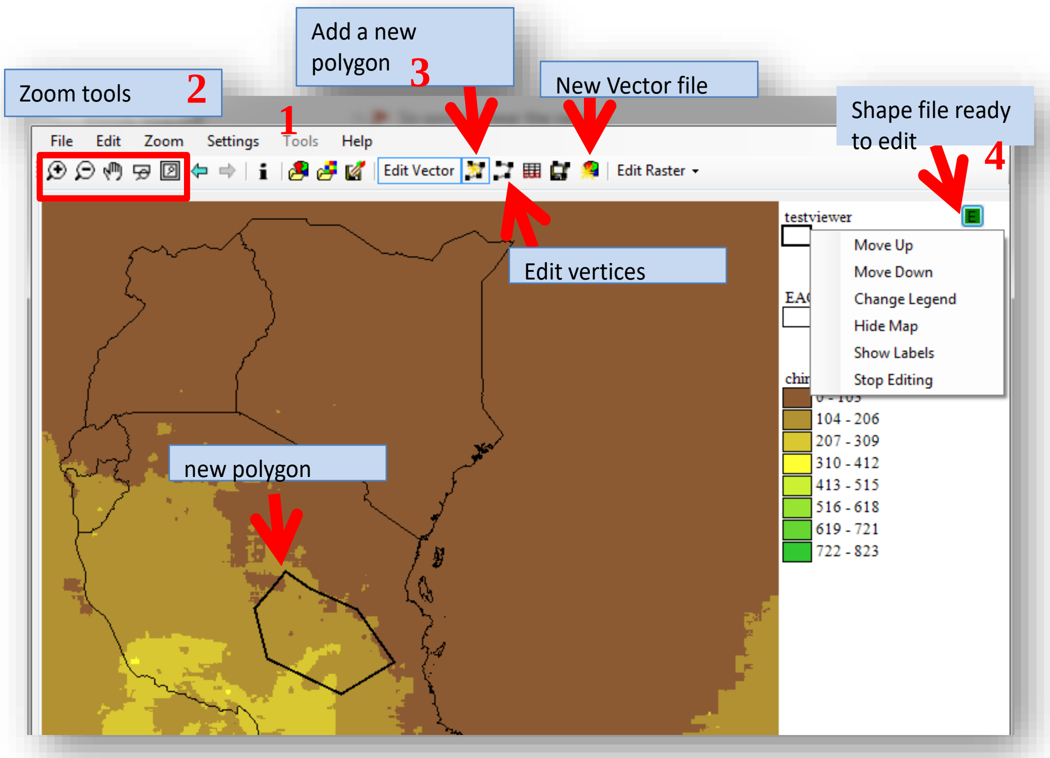


Figure 4.9 You can digitize polygons using the Spatial Data Viewer.

> **NOTE**: GeoCLIM does not have full GIS functionalities. It is missing the ability to 'undo.' Therefore, if you make a mistake, you will have to start over. It is highly recommended to save often.

4.2.4. Editing polygons

To change the shape of a polygon, follow these steps:

1. Open the shapefile.
2. Under the settings menu, make sure the "Show Editing Tools" option is checked.
3. the
4. Click on the ***Edit Vertices*** button (Figure 4.10 **(1)**).
5. Click on the polygon to be edited; all vertices will show up as red dots along the perimeter of the polygon.

6. To adjust the shape of the polygon, click on a vertex dot and move it to the desired location.
7. After all the changes are done, make sure to save the shapefile one final time.
8. Finally, right-click on the shapefile's legend and select Stop Editing.

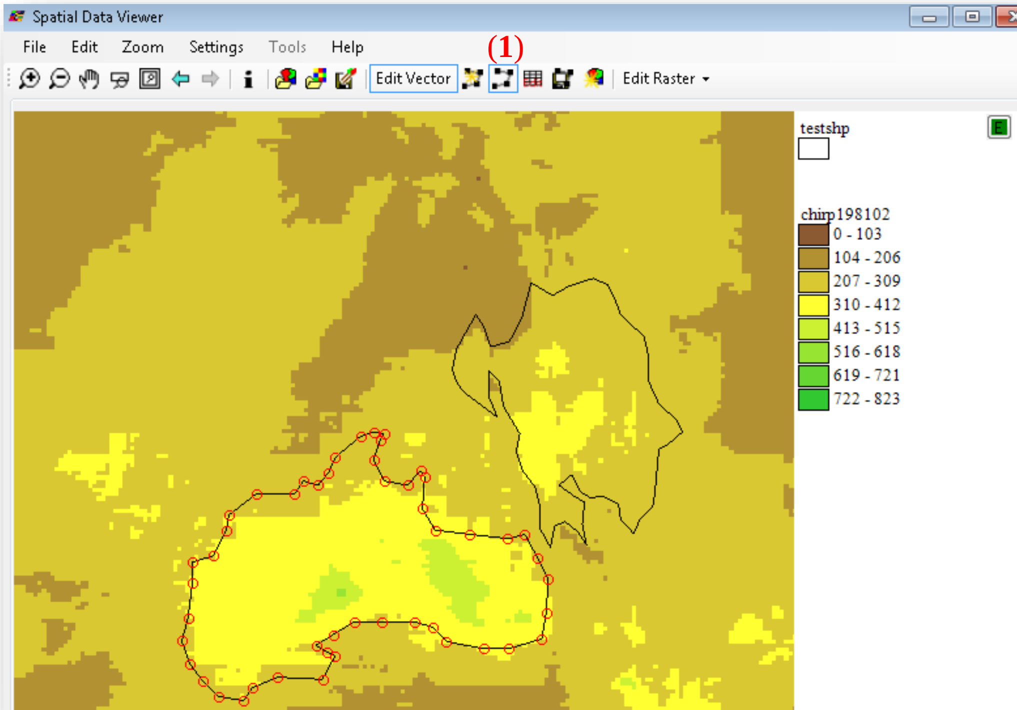


Figure 4.10 *To edit the shape of a polygon move the vertices to the desired position*

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 has already been downloaded or imported into the GeoCLIM data directory (see chapter 2 for how to make data available in GeoCLIM). You can analyze a climate time-series or just a selected subset, such as the March-April-May season for El Niño years; for example, you may select 1982-83, 1986-87, 1987-88, 1991-92, 1997-1998, 2002-03, 2009-10, 2015-16.
The ***Climatological Analyses*** tool includes the following analysis methods:

- **Average**: Calculates the temporal average value for each pixel for a period or group of periods using the years selected.
- **Median**: Calculates the midpoint value of a frequency distribution for the selected climate variable for a group of periods using the selected years.
- **Standard deviation:** Calculates the standard deviation in a frequency distribution for the selected climate variable for a group of periods using the selected years.
- **Count**: Counts the number of valid values by pixel on a time-series.
- **Coefficient of variation**: Calculates the Coefficient of Variation (CV), which is the ratio of the SD to the mean in percent.
- **Trend**: Calculates a linear trend using a regression analysis of the seasonal values and time.
- **Percentiles**: Produces a raster map with the rainfall value for each pixel corresponding to the percentile rank requested.
- **Frequency**: Calculates the number of times a range of values has occurred in the time series.
- **Standardized Precipitation Index (SPI):** Presents a rainfall anomaly as a normalized variable.

5.1. Running climatological Analysis

To open the climatological analysis tool (Figure 5.1), use one of these two options:

1. Click on the ***Rainfall Climatological Analysis*** icon from the GeoCLIM main toolbar, or
2. Click on the Tools dropdown menu from the main GeoCLIM toolbar and select Rainfall Climatological Analysis.

To use the tools, follow the steps below:

1. Make sure to select the region of interest from the Region ∨ drop-down list (Figure 5.1 **(1)**), (see GeoCLIM Settings tool to set up a region).
2. Select the Parameter to Analyze ∨ (Figure 5.1 **(2)**). This field corresponds to the climate

variable to be used in the analysis (Rainfall, Avg Temperature, Min Temperature, Max Temperature, or Evapotranspiration). The selected parameter uses the climate dataset selected in GeoCLIM Settings.

3. To calculate the total of the selected season for each year, check the ■ Add up seasonal totals box in the upper left corner (Figure 5.1 **(3)**).
4. If the season to analyze goes across years, for example, Oct to March, check the ■ July to June Sequence checkbox (Figure 5.1 **(4)**).
5. 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. In this case, the data period is a month (Figure 5.1 **(5)**).
6. Select the years of interest on the right panel (Figure 5.1 **(6)**).

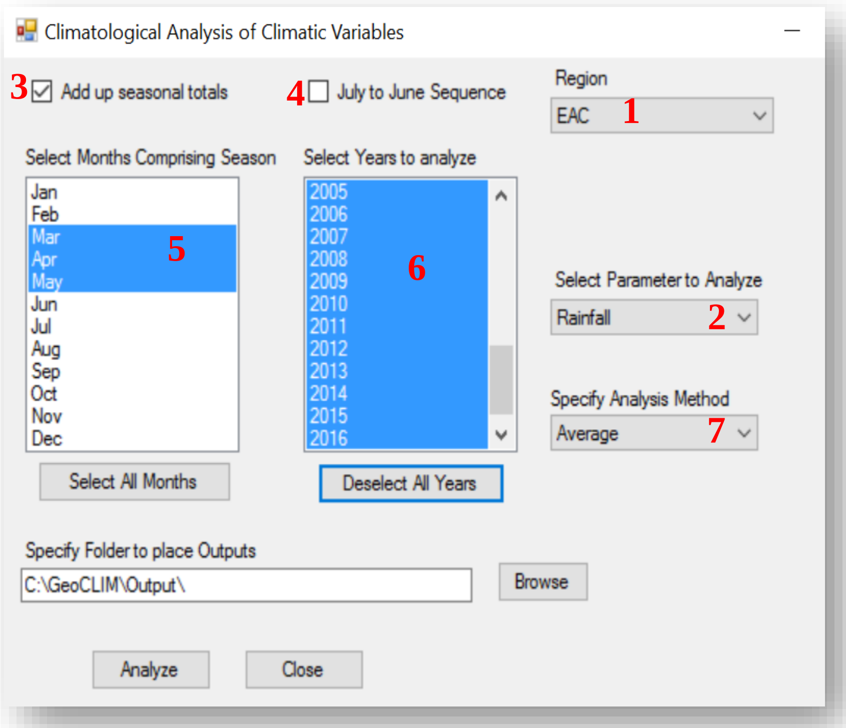


Figure 5.1 *The Climatological Analysis tool facilitates the calculation of statistics, trends, SPI among others, using the complete or part of the time series.*

7. Select the type of analysis from the Specify the Analysis Method ∨ menu (Figure 5.1 **(7)**).
8. Optional, modify the *Specify Folders to place Outputs* field if you want to save outputs in a different location than the default path, `C:\GeoCLIM\Output\`.

The output from this analysis is displayed on the ***Spatial Data Viewer*** (Figure 5.2). This result is also saved in the output folder as a raster dataset in (`*.bil`) format and as an image in PNG (`*.png`) format.

NOTE: Make sure the last year selected contains a complete season; otherwise, there will be a “missing data” error message that would prevent the tool from running.

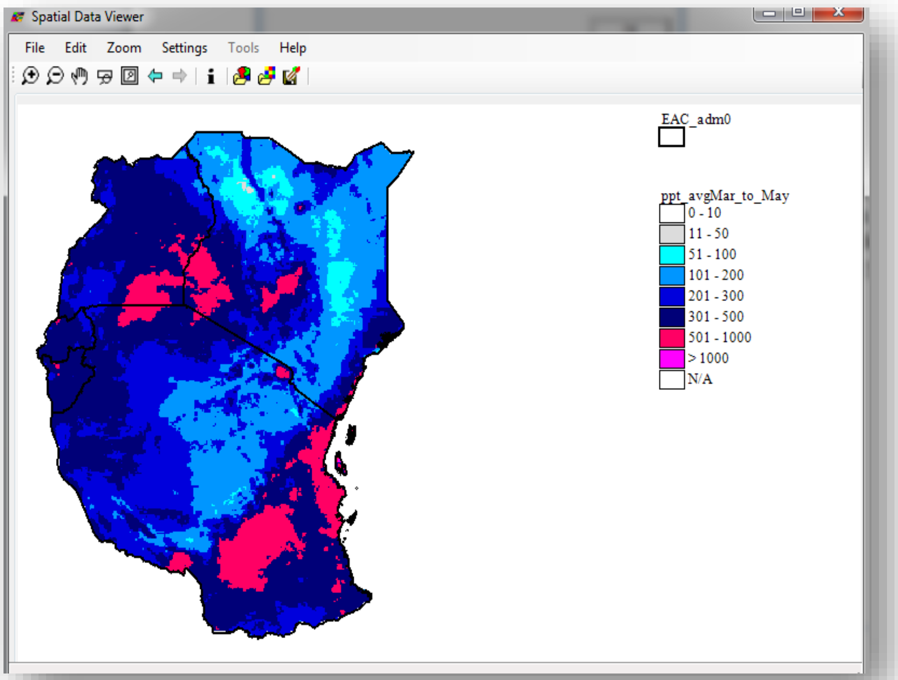


Figure 5.2 *Average rainfall for the period March-May 1981-2013.*

After the viewer window is closed, a message displays information on the location of a log file. This file contains useful information about inputs and outputs from the analysis (Figure 5.3).

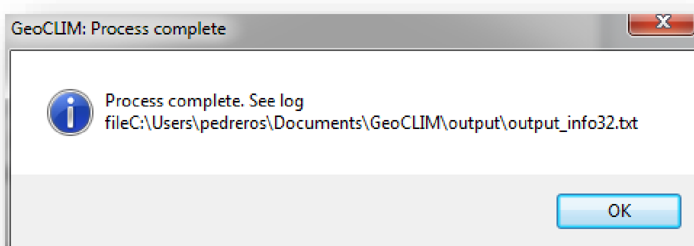


Figure 5.3 *The tool produces a log file, after each process, containing information on the inputs and outputs.*

NOTE: If multiple periods are selected (e.g., March-April-May) and the ■ Add up seasonal totals box is not checked, the process runs for each month, and there is no display because there would be more than one output created.

5.2. Updating the GeoCLIM averages

GeoCLIM uses the average for each period (pentad, dekad, or month) for different calculations, such as seasonal average or to calculate anomalies. The ***Climatological Analysis*** tool calculates the average for every period based on the information saved during the definition of the dataset in the GeoCLIM settings, (see The GeoCLIM Settings, chapter 2). To calculate the average, follow the steps below:

1. Select all the periods and years to be used in calculating the average.
2. Select the Parameter to Analyze, in this case, is rainfall.
3. Check the ■ Update GeoCLIM Averages box (Figure 5.4 **(1)**).
4. Make sure that the □ Add up seasonal totals box is NOT checked (Figure 5.4 **(2)**).
5. Make sure that all the years selected have data for all the periods.
6. Once all the inputs are defined, click on Analyze. The BIL outputs are saved in the directory assigned on the selected climate dataset definition, usually the same directory

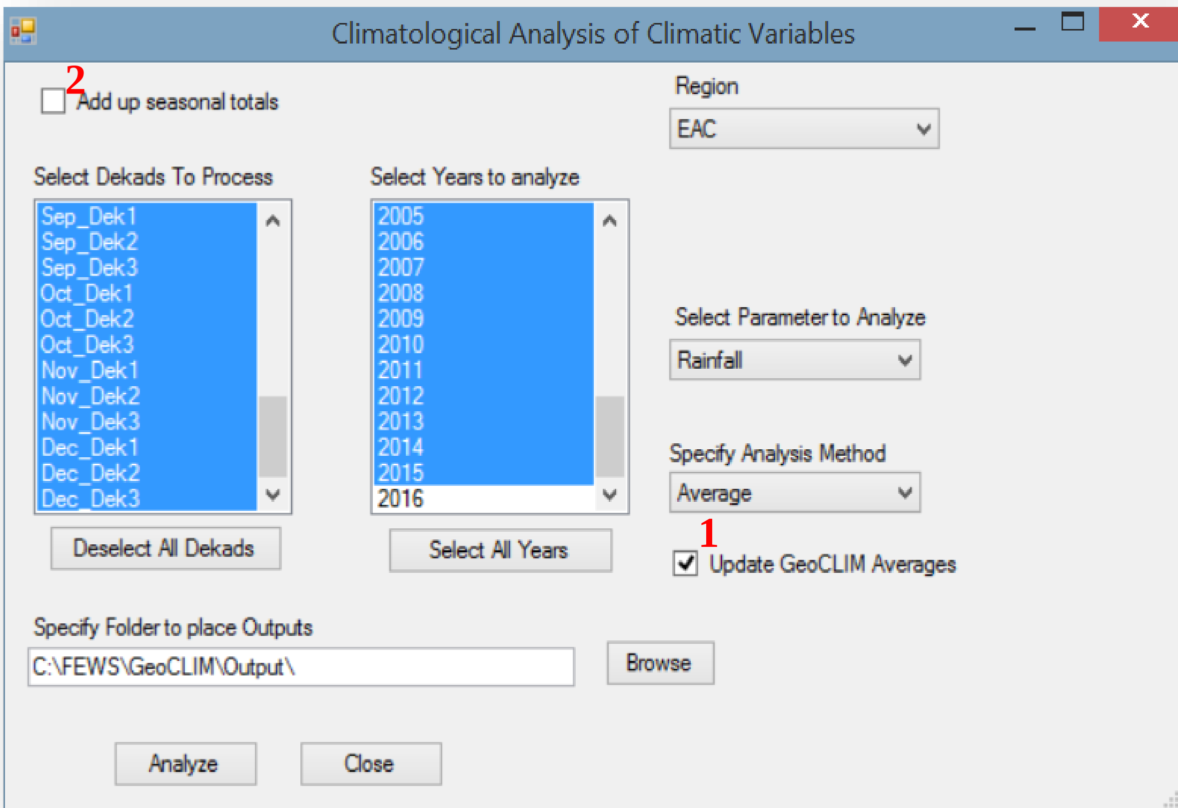


Figure 5.4 The Climatological analysis tool allows us to calculate the average for each period (month, dekad or pentad) for the region of interest.

where the default data files are.

NOTE: The ■ Update GeoCLIM Averages option creates the average for just the extent of the selected region. 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 for the region of interest.

5.3. Analysis Methods

5.3.1. Average

The ***Average*** analysis method calculates the temporal average value for each pixel for a period, for example, the month of May, dekad 3, or a season (May-June-July) using the years and region selected. Figure 5.5 shows the average rainfall for the period May to July 1981-2013, for the selected region (EAC). In other words, the map represents the average of all May-June-July rainfall totals from 1981 to 2013. A by-product of this process is the seasonal total in raster format (.bil) for each year (Figure 5.6).

To calculate the average, follow the steps below:

1. Start the ***Climatological Analysis*** tool, as described in section 5.1.
2. Select the season on the left panel and the years on the right panel.
3. Select the variable or parameter to analyze. For this example, we select Rainfall.
4. Check the ■ Add up seasonal totals option.
5. Select **Average** from the analysis methods list.
6. Click on Analyze to run the analysis.

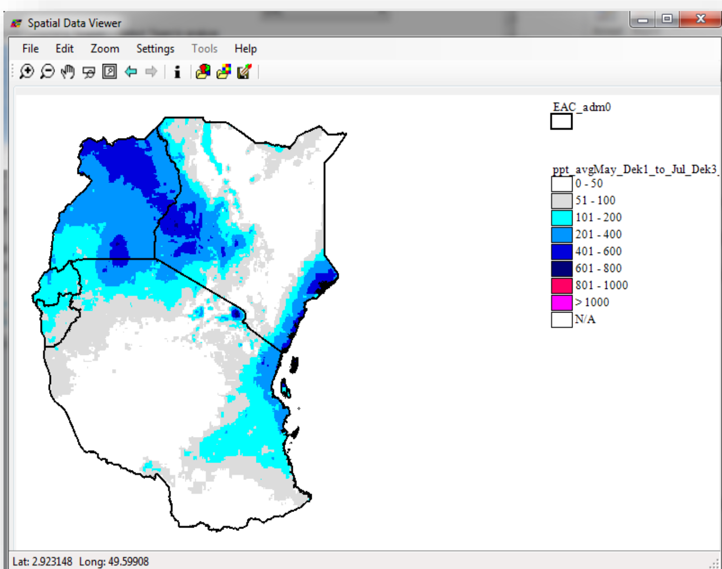


__Figure 5.5__ Average rainfall for the May-July season for the years 1981-2013.

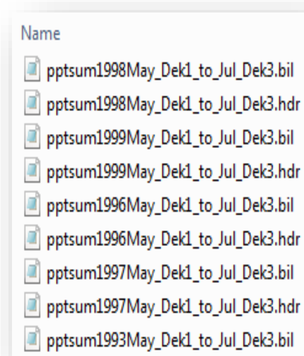


__Figure 5.6__ Seasonal totals for each year for the selected period (May dek1 to July dek3) from 1981 to 2013.

NOTE: When the ■Add up seasonal totals option is not checked, the average is calculated for each period selected (pentad, dekad, or month). In the example above, the module would calculate 1981-2013 average for May dekad 1; 1981-2013 average for May dekad 2; 1981-2013 average for May dekad 3; 1981-2013 average for June dekad 1, etc., until July Dekad 3.

5.3.2. Median

The ***Median*** analysis method calculates the midpoint value of a frequency distribution for the selected climate variable. Figure 5.7 shows an example of median output calculated for May-to-July rainfall totals for the years 1981-2013.
To calculate the median, follow the steps below:

1. Start the ***Climatological Analysis*** tool, as described in section 5.1.
2. Select the season on the left panel and the years on the right panel.
3. Select the variable or parameter to analyze. For this example, we select Rainfall.
4. Check the ■ Add up seasonal totals option.
5. Select **Median** from the analysis methods list.
6. Click on Analyze to run the analysis.

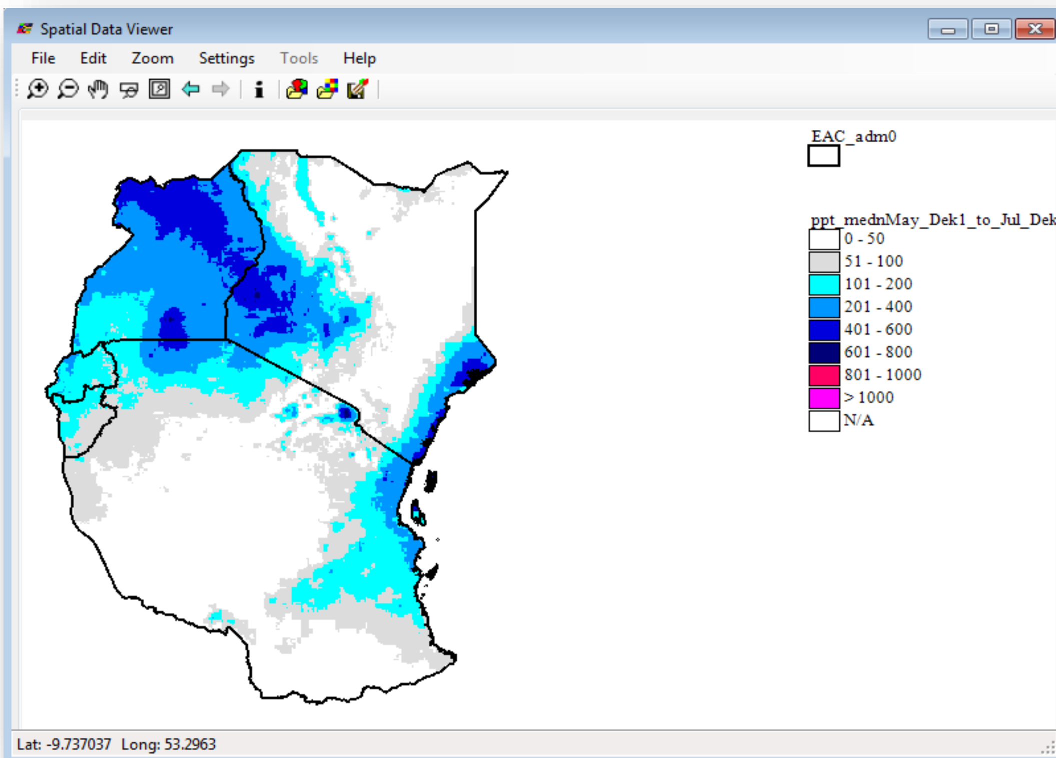


Figure 5.7 Median for the season May-July for the years 1981-2013.

5.3.3. Measuring variability with standard deviation and coefficient of variation

GeoCLIM provides two different methods of estimating variability. The standard deviation (SD) shows the variability within the time series over the selected years (Figures 5.8 and 5.9) for each pixel, while the coefficient of variation (CV) shows the SD as a percent of average, facilitating the comparison of variability among regions.

5.3.3.1. *Standard deviation*

The standard deviation (SD) is a measure of variation or how spread out the data are from the mean. An increase in the SD indicates that the data is more variable (Figure 5.8). See Figure 5.9 **(a)** for an example of an SD product using GeoCLIM.

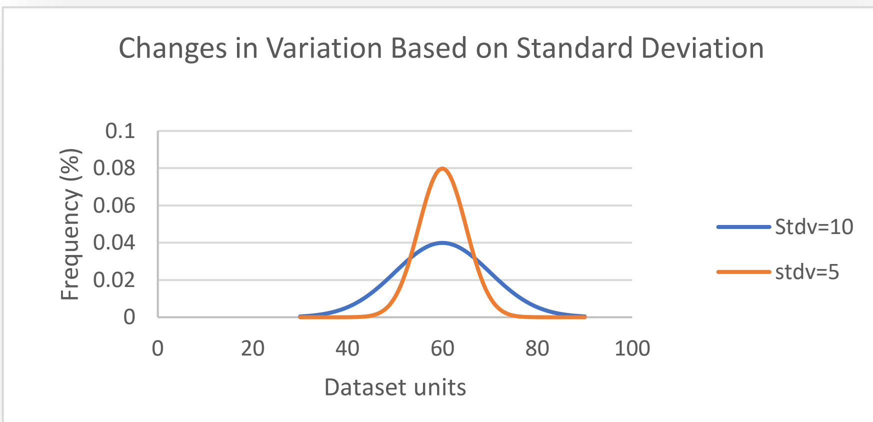


Figure 5.8 *The distribution of two datasets with same mean and different SD. The red line shows a low SD (stdv=5) indicating low variability within the data; values are closer to the mean. The blue line shows the distribution of a more variable data set (stdv=10)*

5.3.3.2. *Coefficient of ariation.*

The Coefficient of Variation (CV) is the ratio of the SD to the mean $CV = \left(\frac{SD}{average}\right) * 100$. Table 5.1.

SD	Mean	CV
171mm	721mm	24%

Table 5.1 *The CV is the ratio of the SD over the mean.*

Figure 5.9 **(a.1)** and **(a.2)** show an example of low and high SD, respectively. But, this information alone does not allow us to determine which area is more variable. The CV allows us to compare among different magnitudes of variation or between regions with different means. Figure 5.9 **(b)** shows that even though regions **1** and **2** have low/high SD when compared to the average amount of rainfall, area **1** is more variable.

To calculate standard deviation or coefficient of variation, follow the steps below:

1. Start the ***Climatological Analysis*** tool, as described in section 5.1.
2. Select the season on the left panel and the years on the right panel.

3. Select the variable or parameter to analyze. For this example, we select **Rainfall**.
4. Check the ■ Add up seasonal totals option.
5. Select **Standard Deviation** or **Coefficient of Variation** from the analysis methods list.
6. Click on Analyze to run the analysis.

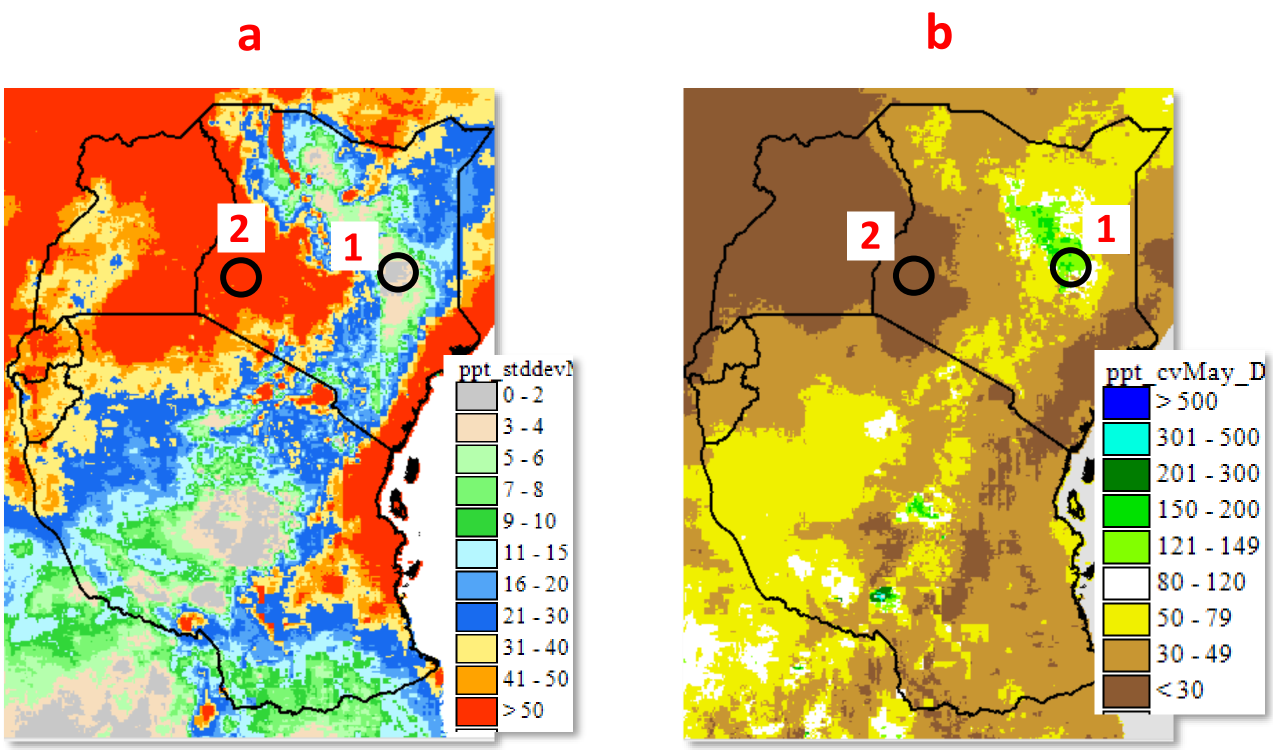


Figure 5.9 **(a)** *shows the SD of rainfall (mm),* **(b)** *presents the CV (SD as a percent of the mean) allowing the comparison among areas. The SD of areas* ***1*** *and* ***2*** *shown as low/high value but* ***1*** *is highly variable compared to area* ***2****, as shown by the CV.*

5.3.4. Count

The count analysis method on the ***Climatological Rainfall Analysis*** tool shows the number of pixels in the selected years, with valid values (any values which are not missing value). The example in Figure 5.11 shows the count as 37 (1981-2017) for all pixels, which means there are no missing values in the time series used in the analysis.

To calculate count, follow the steps below:

1. Start the ***Climatological Analysis*** tool, as described in section 5.1.
2. Select the season on the left panel and the years on the right panel.
3. Select the variable or parameter to analyze. For this example, we select **Rainfall**.
4. Check the ■ Add up seasonal totals option.
5. Select **Count** from the analysis methods list.
7. Click on Analyze to run the analysis.
8. Move the mouse over the map to get the pixel value on the lower-right corner.

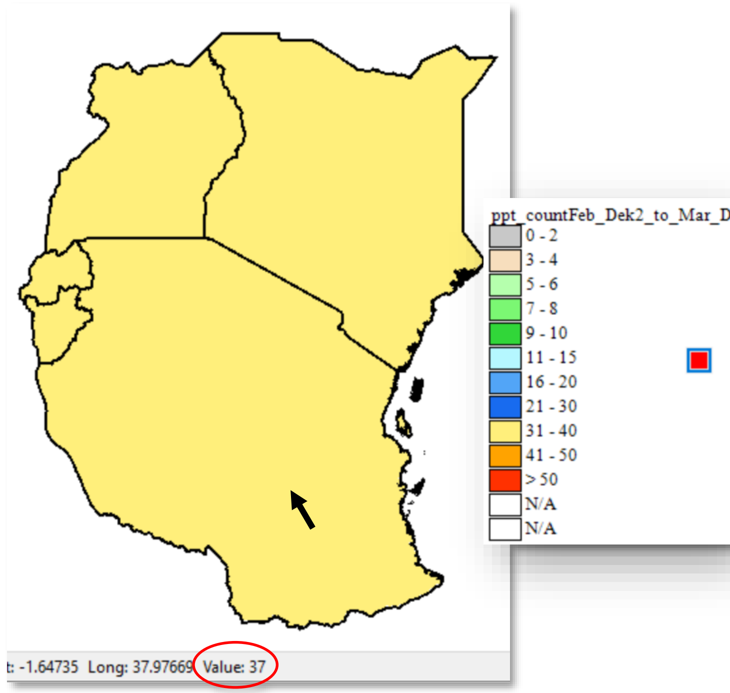


Figure 5.10 The function counts the number of valid values in the time series. In the example, there is no missing data and there are 37 values.

5.3.5. Trend

The trend is an analysis technique that helps us identify a change in the expected value of a variable that occurs over a long period of time. The Trend analysis method in GeoCLIM first calculates the total seasonal rainfall for each selected year and then calculates a linear trend using a regression analysis of the seasonal values and time

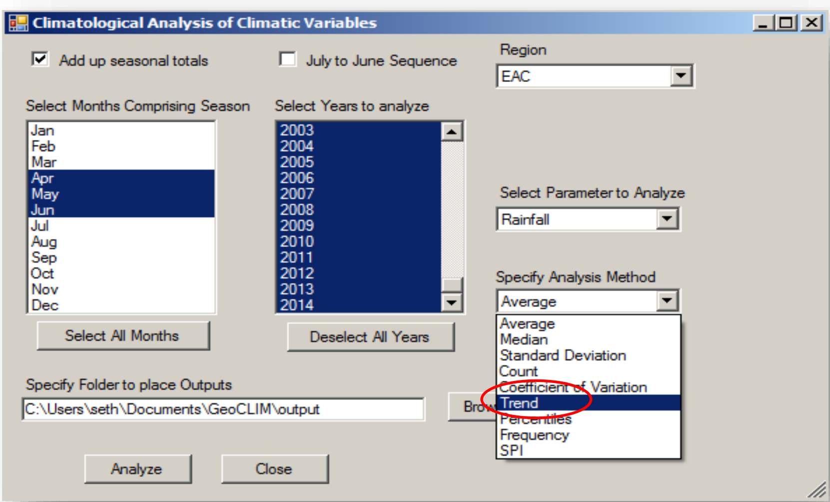


Figure 5.11 To calculate the trend for a climate variable, select the season, make sure that the Add up seasonal totals *option is checked, and select the*

(Figure 5.11). This function in GeoCLIM produces two maps; one is the coefficient of determination (r-squared, or r2), and the other is the slope of the regression. The slope represents the trend.

To calculate the trend, follow the steps below:

1. Start the ***Climatological Analysis*** tool, as described in section 5.1.
2. Select the season on the left panel and the years on the right panel.
3. Select the variable or parameter to analyze. For this example, we select **Rainfall**.
4. Check the ■ Add up seasonal totals option.
5. Select **Trend** from the analysis methods list.
6. Click on Analyze to run the analysis.

Figure 5.12 shows the results of the **Trend** analysis method in GeoCLIM for the annual rainfall total in Ethiopia for the period 1981 – 2016. Figure 5.12 **(1)** shows the slope of the regression line, or the trend for each pixel, in mm/decade of increasing (green-blue) or decreasing (pink-red) rainfall, the legend shows these results per decade (10 years). Figure 5.12 **(2)** shows the coefficient of determination (r-squared, or r2) (multiplied by 100) of the linear regression between the variable and time as an indication of the reliability of the trend. It is important to use both maps to develop a conclusion about trends in an area. For example, points a, b, and c show three sites with strong trends and different r2.

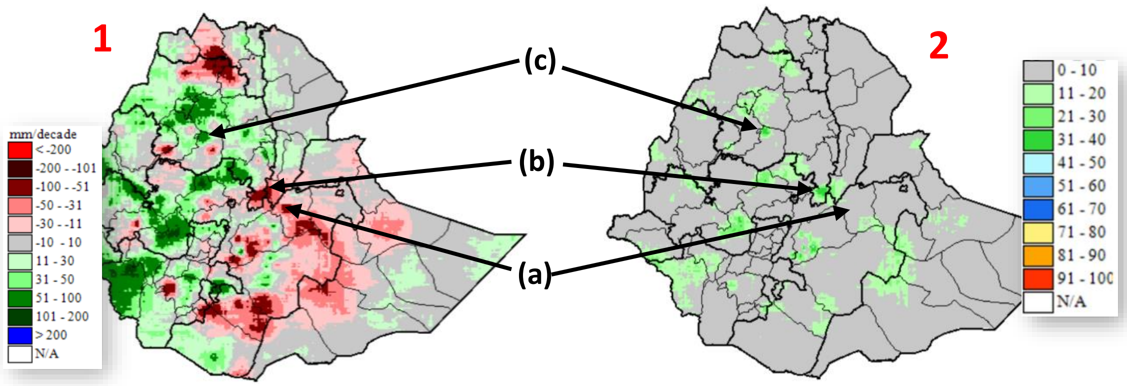


Figure 5.12 *The trend analysis method in GeoCLIM produces two outputs. (1) Shows the slope of the regression in mm per decade decrease (pink-red) or increase (+ green-blue) and (2) shows the r2 of the regression.*

Based on Figure 5.12 **(1)** and **(2)**, site (a) has a 71mm decrease per decade (dark red) with r2= 9% (grey), site (b) shows a 75mm decrease per decade (dark red) with r2 = 36% (dark green), while site (c) shows 89mm increase per decade (dark green) with r2 = 35% (dark green). Sites (a) and (b) have similar trends, but the r2 values show that site (b) has the strongest correlation. Also, sites (b) and (c) have similar r2 shown as green color. Figure 5.13 shows the regression plots of total annual rainfall against time for sites a, b, and c. The annual total for the period 1981–2016 was extracted using the Extract Statistics function in GeoCLIM, for each site, and plotted using Excel. The plots in Figure 5.13 corroborate the difference in r2 by showing how close the points are to the regression line. Site (a) shows the points scattered, while sites (b) and (c) show the points closer to the regression line.

NOTE: It is important to use both maps to develop a conclusion about trends in an area, since the trend map shows how much change there has been in the time period we are analyzing, and the r2 map shows the reliability of the trend. The trend with a larger r2 value suggests a more robust trend, while the weaker r2 indicates that this trend may be by chance.

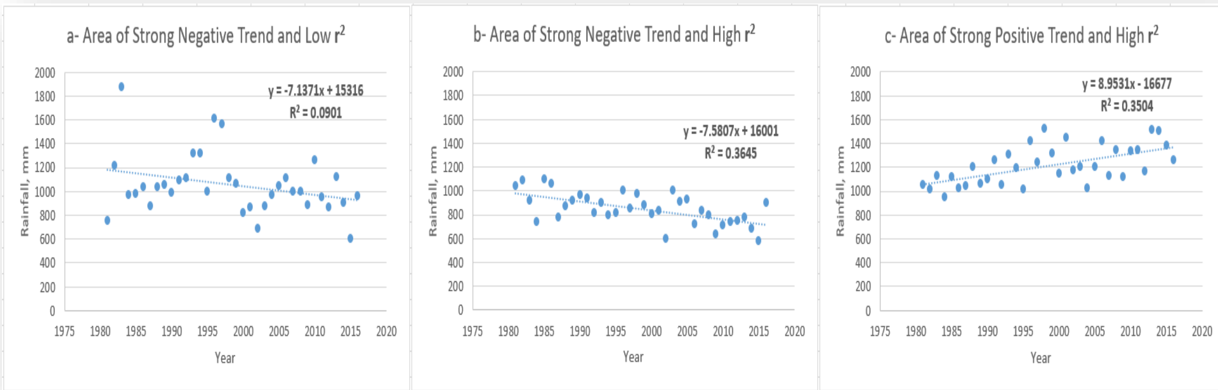


Figure 5.13 It is important to evaluate the strength of the relationship (r2) before making conclusions about the trend. Plots show three regions that present strong trends on Figure 5.12(1) with different r2.

5.3.6. Percentiles

A percentile is a statistic that specifies the value below which a certain percent of observations in a ranked dataset will fall. Percentiles are calculated at breakpoints ranging from 0 to 100. The 0th percentile corresponds to the lowest value. The 100th percentile is the highest. The 50th percentile is the median value. To calculate a percentile value, we first have to rank the time-series, and then identify the value associated with the nth percentile position.

For example, if the 20th percentile is 80 mm of rainfall, then we would expect that 20% of the time, rainfall would be less than or equal to 80 mm. One way of using percentiles is to answer questions like: "if we have the time series for the total FMAM season from 1981-2017 (table 5.2), what would we expect a 1-in-5 year dry event to look like?" To explore this question, we could calculate the 20th percentile. Statistically, we would expect rainfall of this amount or lower once every five years.

Another use of percentiles is when we have a value, let's say the rainfall total for the FMAM for 2017=216 mm, and we would like to know what percentile that value represents, or how frequent a value like this happens. Using the data in table 5.2 (see Note below on how the data was obtained) and the PERCENTRANK function in Excel, we find that 216 mm is the 71st percentile or greater than 71% of the values in the dataset. The ***Percentiles*** function in GeoCLIM produces a raster map with the rainfall value for each pixel corresponding to the percentile rank requested. To calculate a given percentile for your region of interest, follow the steps below:

1. Start the ***Climatological Analysis*** tool, as described in 5.1.
2. Select the season on the left panel and the years on the right panel.
3. Select the variable or parameter to analyze. For this example, we select **Rainfall**.

4. Check the ■ Add up seasonal totals option.
5. Select ***Percentile*** from the analysis methods list.
6. Enter the percentile rank desired (Figure 5.14 **(1)**).
7. Click on Analyze to run the analysis.

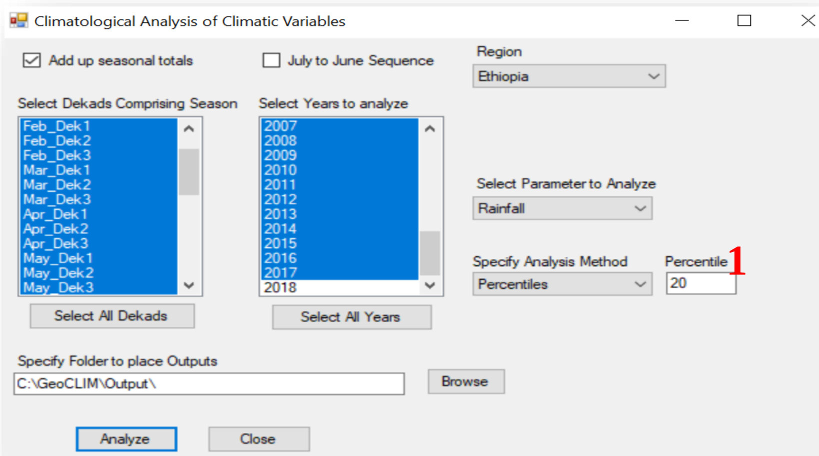


Figure 5.14 *The Percentiles method in GeoCLIM produces a raster map with the rainfall value for each pixel corresponding to the percentile rank*

This function in GeoCLIM helps answer questions such as, what are the low/high values (e.g., 15th/90th percentiles) in the distribution for the time-series? (Figure 5.15). Table 5.2 shows the time-series for the FMAM seasonal total for the period 1981-2017 for point **(A)** in Figure 5.15. The result of the PERCENTILE.EXC function in Excel shows that the 20th percentile is = 105.

	Feature	prec_FMAM	
1	2009	35	
2	2008	44	
3	1984	61	
4	1999	66	
5	2011	79	
6	2015	94	
7	2000	103	105->20th percentile
8	1994	107	
9	2013	120	
10	1992	121	
11	1998	121	
12	2007	123	
13	1997	133	
14	1982	134	
15	1988	146	
16	2001	153	
17	1991	154	
18	2003	162	
19	2004	163	
20	2012	163	->50th percentile
21	1990	171	
22	2014	175	
23	2010	180	
24	2005	181	
25	2006	197	
26	1995	207	
27	2017	216	->71th percentile
28	1983	217	
29	1989	225	
30	1996	227	
31	1993	228	
32	1986	232	
33	2002	239	
34	1981	244	
35	2016	277	
36	1985	291	
37	1987	298	

Table 5.2 *Seasonal total FMAM 1981-2017 for point A in Figure 5.15*

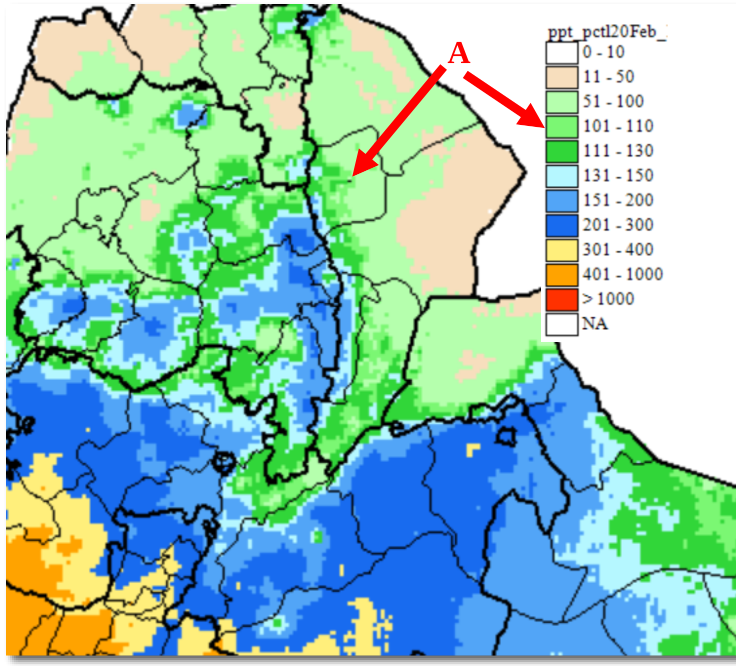


Figure 5.15 *An example of rainfall accumulations (mm) corresponding to the 20th percentile rank for the FMAM season. This percentile rank defines a set of low frequency dry events. The default legend was modified to represent the data. See chapter 4 on how to modify the legend.*

NOTE: Table 5.2 was created using the ***Extract Statistics*** tool for a small polygon shown in Figure 5.15(**A**).

5.3.7. Frequency

The Frequency analysis method in the GeoCLIM Climatological Analysis module (Figure 5.16) gives us the number of times a range of values has occurred in the time series. The Frequency method helps to answer questions such as, "How many times has the total seasonal rainfall been less than 400 mm?" Answering these types of questions can help users to decide whether an area is suitable for a particular climatically-driven activity (such as farming certain crops or livestock). The legend in Figure 5.17 represents the number of events per decade (ten years).
To calculate the frequency of a range of values, follow the steps below:

1. Start the ***Climatological Analysis*** tool as described on 5.1.
2. Select the season on the left panel and the years on the right panel.
3. Select the variable or parameter to analyze. For this example, we select **Rainfall**.
4. Check the ■ Add up seasonal totals option.
5. Select ***Frequency*** from the analysis methods list.
6. Fill in the values Between and And to define the range of values.
7. Click on Analyze to run the analysis.

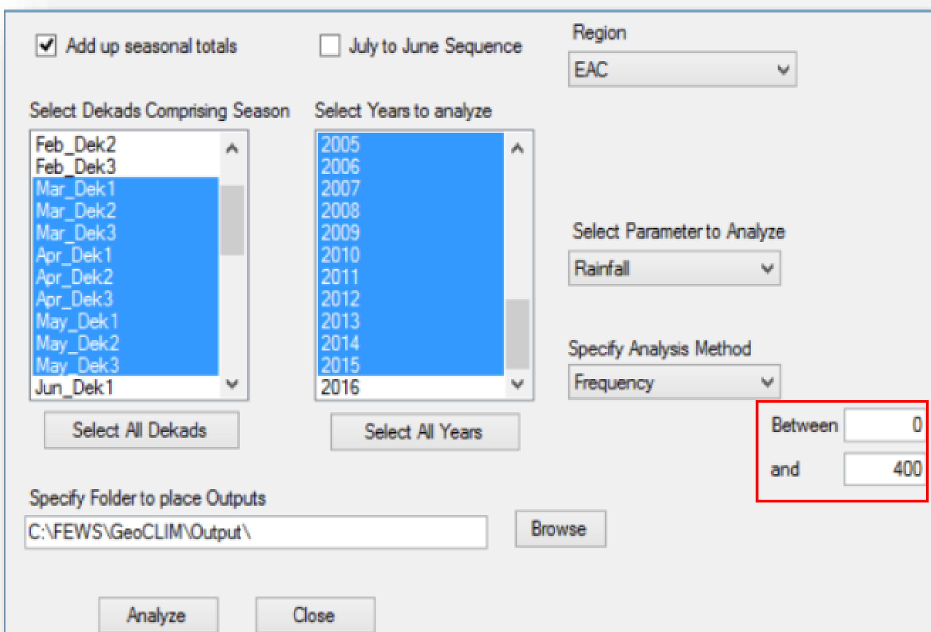


***Figure 5.16** Frequency function allows for the selection of a range of values (red box) and identifies the number of times this range has occurred in the time series.*

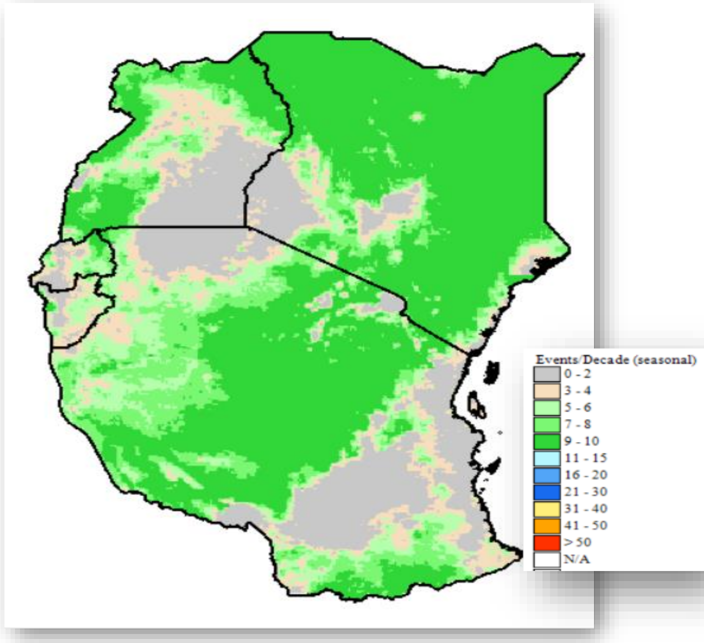


Figure 5.17 The tool calculates the number of times the selected range of values took place during the time series selected. The legend is in events per 10 years.

5.3.8. Standardized Precipitation Index (SPI)

The ***Standardized Precipitation Index*** (SPI) 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 either anomaly (which can be relatively large, but not very significant in high rainfall areas) or percent of average (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-plus time-series (starting in 1981, operationally updated) 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 indicates drier-than-median conditions. For drought analysis, an SPI less than -1.0 indicates that the observation is roughly a one-in-six dry event and is termed "moderate." An SPI less than -1.5 indicates a one-in-fifteen dry event and is termed "extreme." Values less than -2.0 are typically referred to as "exceptional," indicating that it is in the driest 2% of all events (Figure 5.19). (https://earlywarning.usgs.gov/fews/product/51#documentation).

To calculate the SPI for a year or multiple years, follow the steps below:

1. Start the ***Climatological Analysis*** tool as described on 5.1.
2. Select the season on the left panel and the years on the right panel.

3. Make sure that the variable or parameter to analyze is Rainfall.
4. Check the ■ Add up seasonal totals option.
5. Select ***SPI*** from the analysis methods list.
6. Select the year or the group of years for which you want to calculate the SPI (Figure 5.18 **(1)**).
7. Click on Analyze to run the analysis.

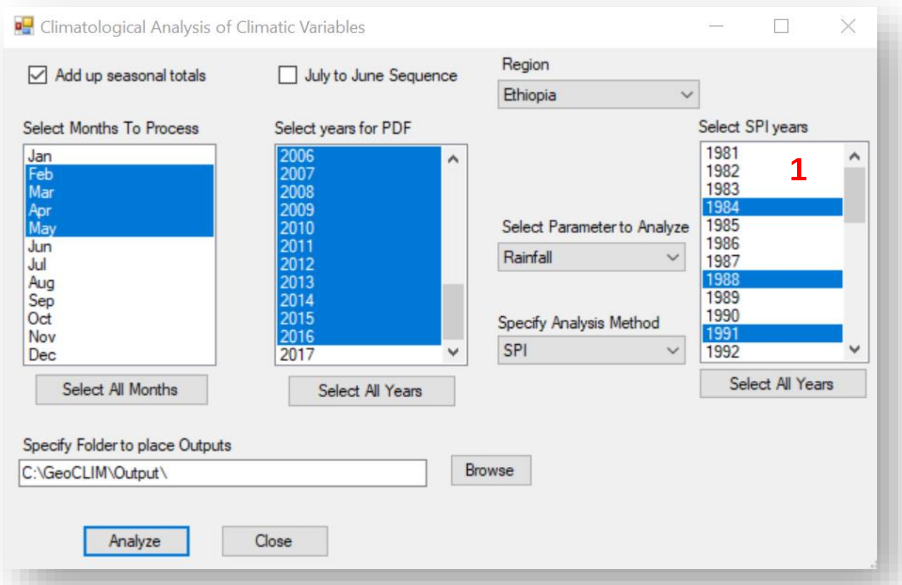


Figure 5.18 *The GeoCLIM allows you to calculate the SPI for a single year or multiple years.*

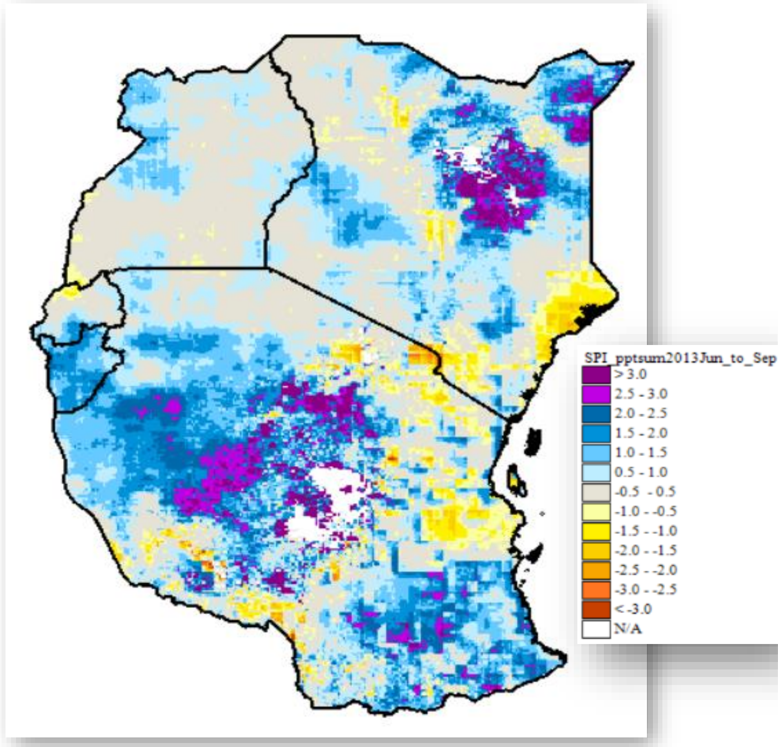


Figure 5.19 *The GeoCLIM SPI raster output is in units of [SPI * 100], but the legend shows actual SPI values.*

Chapter 6: View and Explore Rainfall Summaries

Summary

When analyzing a rainy season, it is important to estimate how different it is from average. The ***Rainfall Summaries*** tool (Figure 6.1) facilitates the calculation of:

1. **Current Period Total**: total rainfall for the selected period.
2. **Average Period total**: long-term average for the selected period.
3. **Difference from Average**: `(current period total – average period total)`
4. **Percent of Average**: `(current period total / average period total) *100`

The outputs provide answers to questions like, “What is the average for the period June dek1 to Augt dek3” also “How different was June dek1- Aug dek3 2012 from average?”

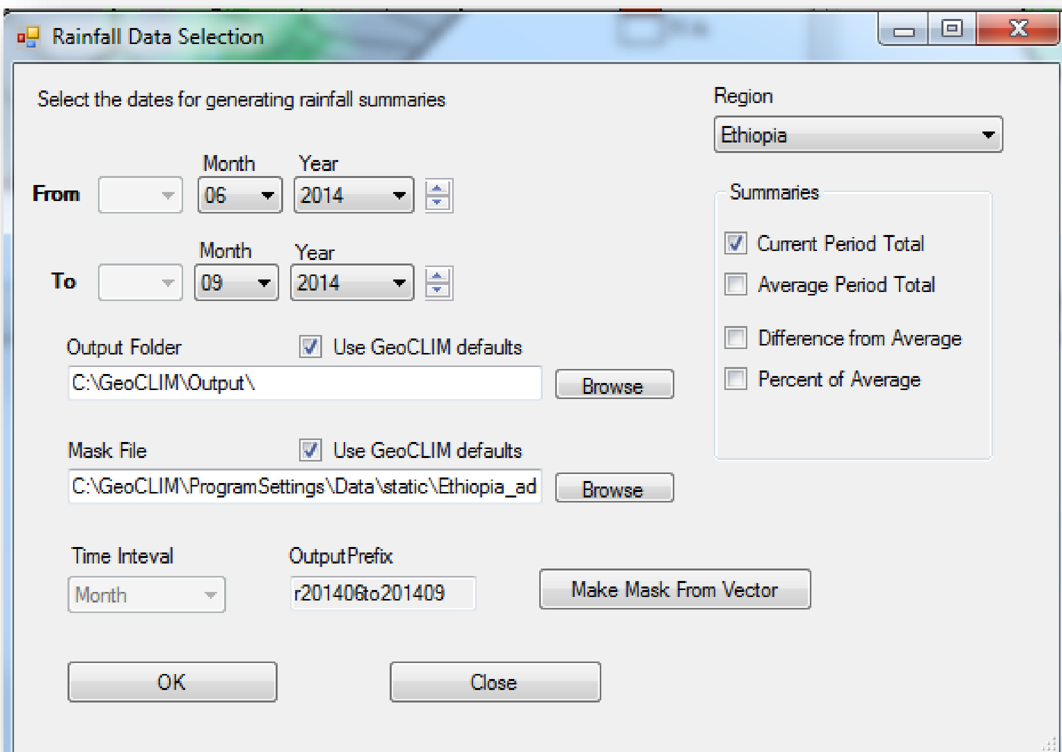


Figure 6.1 The rainfall summaries tool calculates seasonal total, averages and anomalies for a specific time period.

6.1. Requirements

To use the ***Rainfall Summaries*** tool, the selected climate dataset must have GeoCLIM averages available for the selected region. If the averages do not exist, a window message shows up, and the ***Climatological Analysis of Climatic Variables*** tool will open to calculate the averages (Figure 6.2).

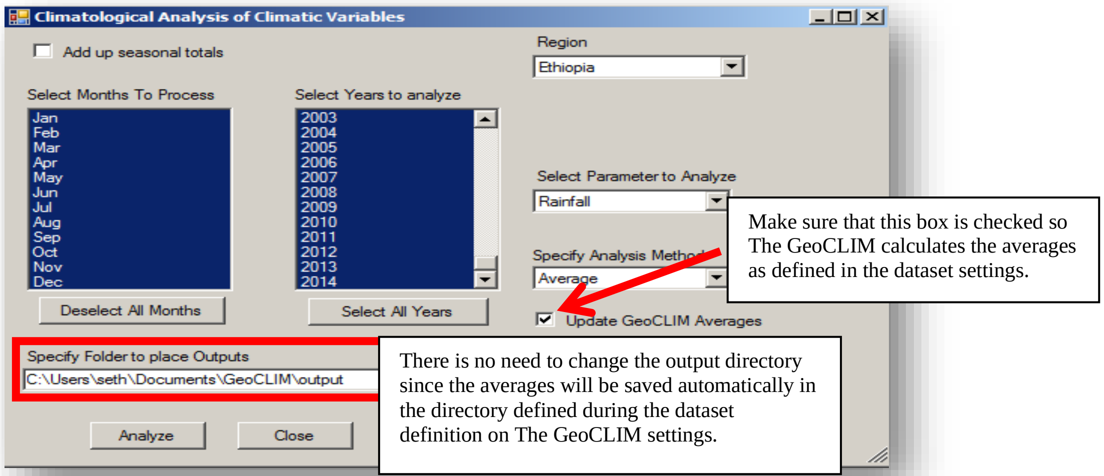


Figure 6.2 *The Climatological analysis tool opens from the rainfall summaries tool to calculate the average for each period on the selected climate dataset.*

6.2. Calculate seasonal total and anomalies

After the averages are calculated, close the Climatological Analysis tool to go back to the ***Rainfall Summaries*** tool.

1. Select the period of analysis (defined by the From and To date).
2. Select the type of summary.
   a) ***Current Period Total***
   b) ***Average Period Total***
   c) ***Difference from Average***
   d) ***Percent of Average***
3. Click OK to run the tool.
4. You might have to edit the legend to represent your data better. See chapter 4.

NOTE: To save outputs in a different directory, uncheck the box □ Use GeoCLIM defaults and browse to the new directory.

NOTE: Even though the tool calculates all four products, it only displays the one you select.

Chapter 7: Climate Composites

Summary

The ***Climate Composites*** tool facilitates the analysis of a season among a group of non-consecutive years or compares the seasonal rainfall performance among two groups of years. For example: comparing the difference in rainfall condition of the May–July (MJJ) season during El Niño and La Niña years in Central America. The ***Climate Composites*** tool calculates the seasonal average for a single group of years, the percent of average, as well as anomalies or standardized anomalies for one or two groups of years using an average baseline period.

7.1. Average

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

To calculate the seasonal average for a group of years, follow the steps below:

1. Select the region of interest (Figure 7.1 **(1)**).
2. Select the season to be analyzed (Figure 7.1 **(2)**).
3. Select the years to be included for composite 1 (Figure 7.1 **(3)**).
4. Move the selected years to the ***composite 1*** box by clicking the >> button (Figure 7.1 **(4)**).
5. Select the method to run: in this case, select Average (Figure 7.1 **(5)**).
6. Click on Analyze

Figure 7.2 shows the results; in this case, the legend was modified to display the range of values better. See chapter 4 for how to change legends.

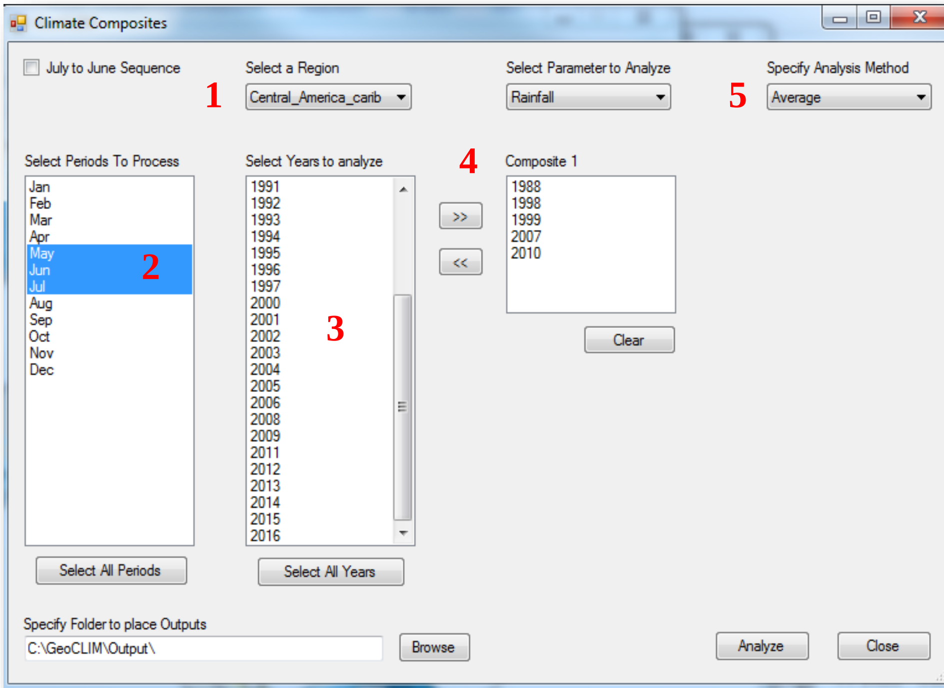


Figure 7.1 *The Composites tool calculates the seasonal average from a group of years and displays the output on the Spatial Data Viewer.*

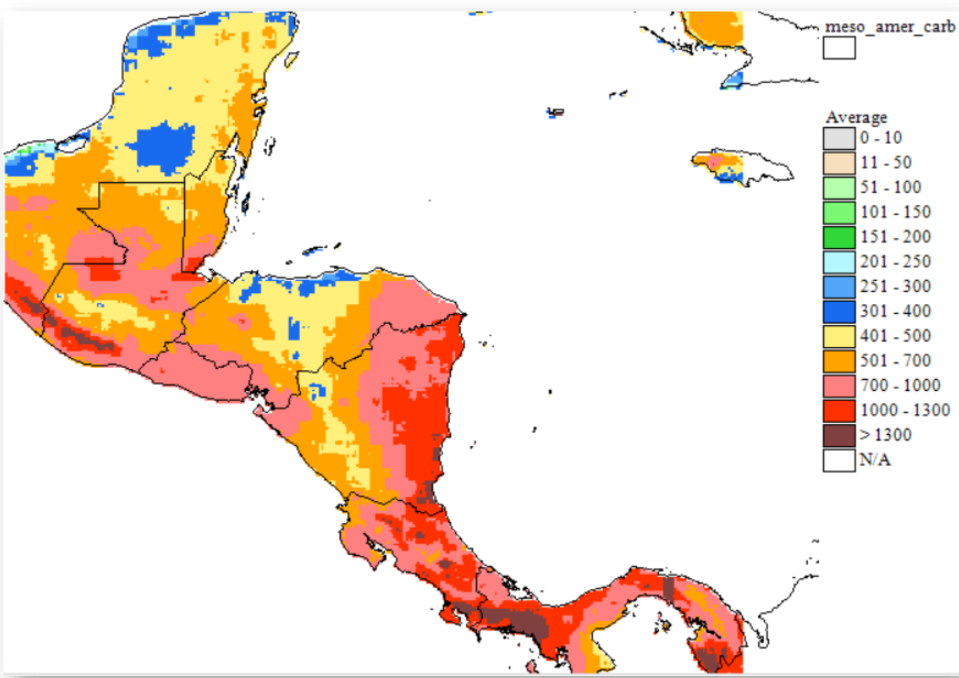


Figure 7.2 *Average rainfall for La Niña years in Central America for the May - July season, composite1. In this case the default legend was edited to represent large values.*

7.2. Percent of Average: (Applies to composite 1 and composite 2)

The ***Percent of Average*** allows for the analysis of a single group of years or the comparison between two groups of years. Figure 7.3 shows the input parameters: **(1)** La Niña years (composite 1), **(2)** El Niño years (composite 2), and **(3)** the Baseline, which indicates the period to be used to calculate the average.

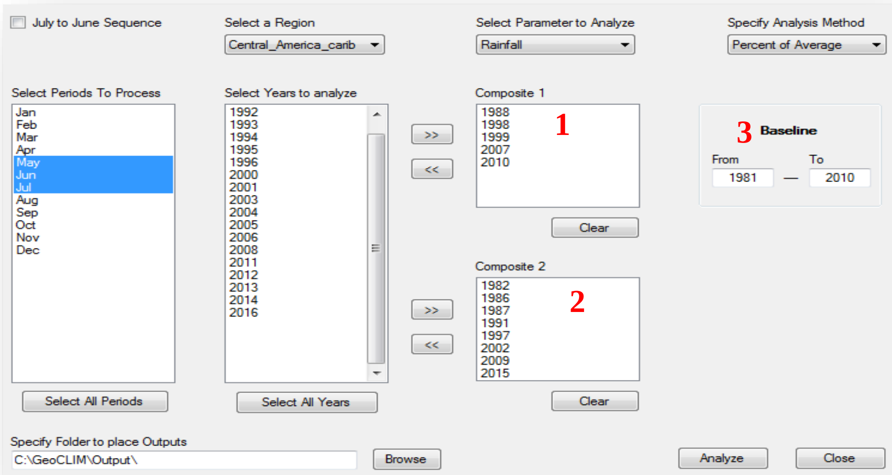


Figure 7.3 The composites function calculates the percent of average for a single group of years or compares two groups of years.

To calculate the percent of average for composite 1 (Eq. 7.1):

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

1. If composite 2 is empty, the program saves pct_comp1 as the final output and displays it on the ***Spatial Data Viewer***.
2. If composite 2 is not empty, the program calculates the difference between the two composites as shown by Eq. 7.2:

$$pct_{comp} = \left(\frac{average_{comp1} - average_{comp2}}{average_{baseline} + 0.1}\right) * 100 \qquad (7.2)$$

If (average_comp1 - average_comp2) = 0 or average_baseline = 0 then ((average_comp1 - average_comp2) / average_baseline) = 0, otherwise the results will come from Eq. 7.2. The tool displays the results on the ***Spatial Data Viewer*** (Figure 7.4). In this case, the results show that rainfall is higher during La Niña years in most of the Pacific coast.

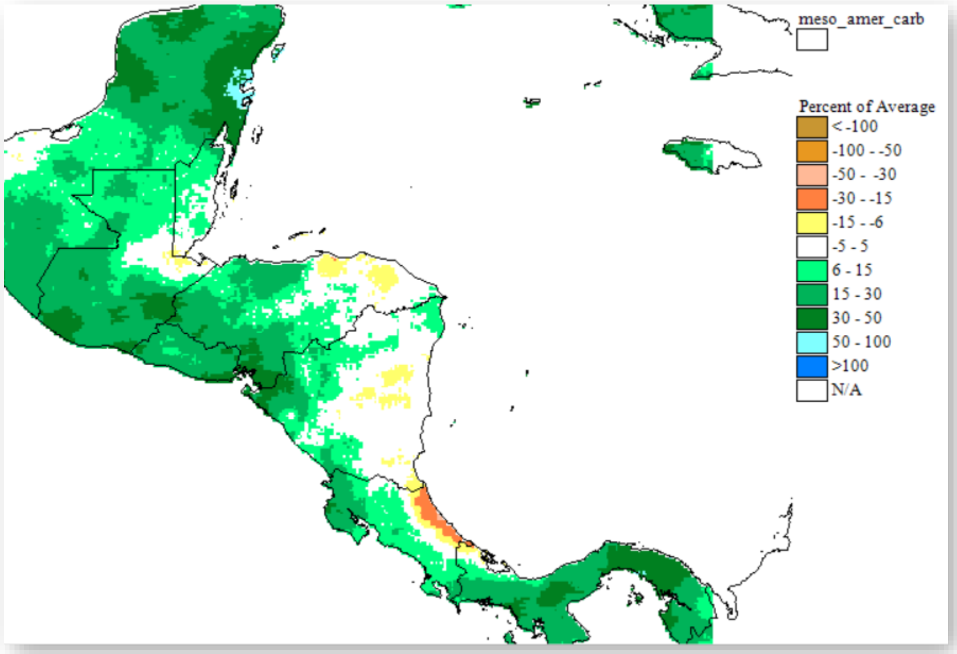


Figure 7.4 *Percent of average for composites 1 (La Niña) and 2 (El Niño). In this example, the positive values indicate that precipitation during La Niña years is higher, on average, than during El Niño years. The default legend was edited for display. See chapter 4 to see how to edit the legend.*

7.3. Anomaly: (Applies to composite 1 and composite 2)

This analysis method calculates the average for each composite and the baseline; then, it calculates the anomaly for each composite (Eq. **Error! Reference source not found.**).

$$anom_{compN} = average_{compN} - average_{baseline} \quad (7.3)$$

1. If composite 2 is empty, the $anom_{comp1}$ is saved as the final output and displayed on the ***Spatial Data Viewer***.
2. If composite 2 is **not** empty, the program calculates the difference between the anomalies of the two composites (Eq. 7.4).

$$anom_{comp} = anom_{comp1} - anom_{comp2} \quad (7.4)$$

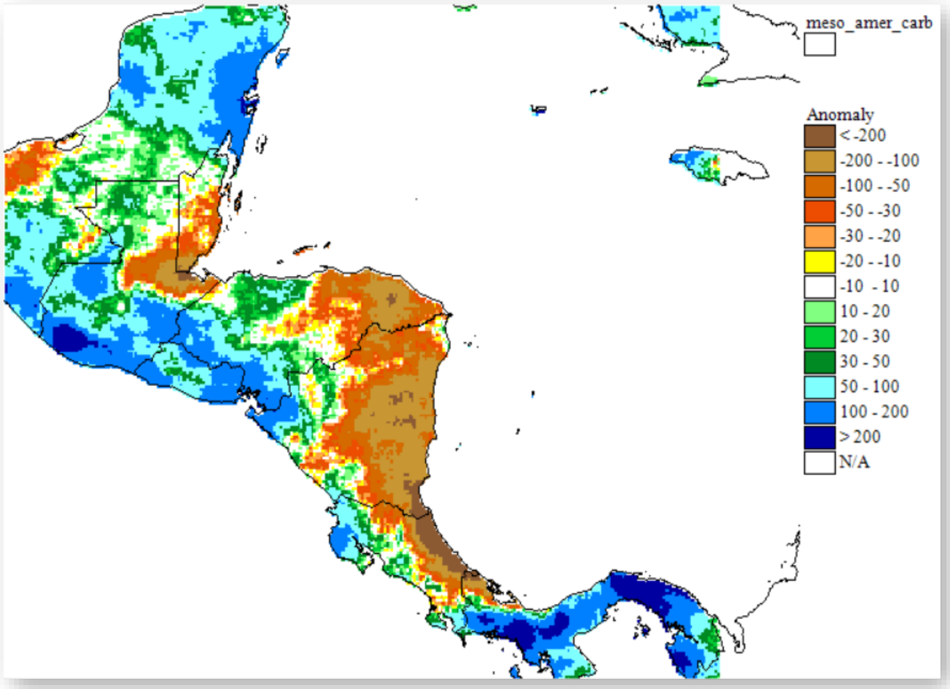


***Figure 7.5** The positive anomalies show areas were, on average, La Niña years have higher values than El Niño. The results are shown in mm. The default legend was modified based on the range of values.*

7.4. Standardized Anomaly: (Applies to composite 1 and composite 2)

This analysis method calculates the difference anomaly, for the average seasonal precipitation, for each group of years and expresses it as a percent of the standard deviation. The function then subtracts the results of compsite 2 from composite 1 and expresses it in terms of standard deviations away from the mean.

The method:

1. Validates if data exist for the selected years for composites 1 and 2, and baseline.
2. Calculates the standard deviation, including all the available years.
3. Calculates the average for the composites and baseline.
4. Calculates anomaly for each composite.
5. Calculates the standardized anomaly for each composite (Eq. 7.5).

$$stdanom_{compN} = \left(\frac{(average_{compN} - average_{baseline}) + 0.1}{stdev_{available\ years} + 0.1}\right) * 100 \qquad (7.5)$$

Where $stdev_{available\ years}$ is the standard deviation for all the years in the climate dataset for the selected period (e.g., period: May-July, composite1: El Niño years, baseline: 1981-2010, climate dataset: 1981-2017).

1. If composite 2 is empty, the function saves $stdanom_{comp1}$ as the final output and displays it in the ***Spatial Data Viewer***.
2. If composite 2 is not empty, the function calculates the difference between the two composites as follows:

$$stdadnom_{comp} = stdanom_{comp1} - stdanom_{comp2} \qquad (7.6)$$

The results in Figure 7.6 show the difference between composite 1 and 2 in terms of the standard deviation of the complete climate dataset. The areas in blue/purple show how much wetter on average El Niño years are than La Niña years, with the difference expressed in terms of standard deviations from the mean.

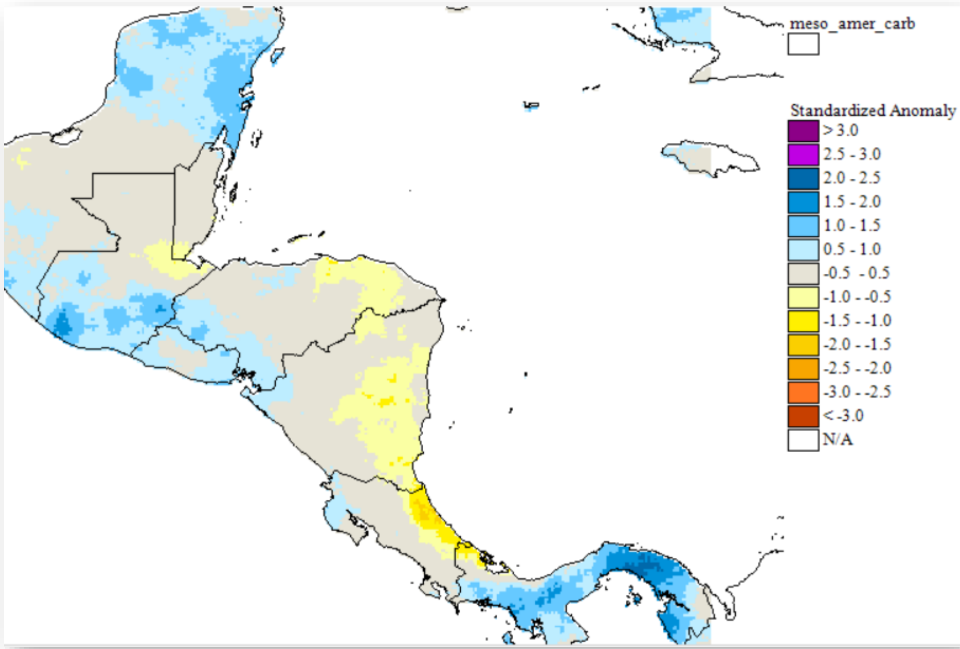


Figure 7.6 *This function calculates standardized anomaly which is the difference anomaly of the average precipitation for a group of years (composite 1) expressed as a percent of the standard deviation. If composite 2 exists, the function calculates the difference between the two standardized anomalies.*

NOTE: The raster values on the map shown in Figure 7.6, are numbers with a scale factor of 100 since GeoCLIM does not work with decimals values. However, the legend shows the number of standard deviations from the mean.

Chapter 8: Contour Tool

Summary

The ***Make Contours*** tool delineates areas within a defined interval of rainfall. Analyzing contours from different periods of time helps to identify changes in a given variable within a region.

To run the tool, follow the steps below:

1. Open the ***Make Contours*** tool from the GeoCLIM main toolbar.
2. Specify the BIL input file; in this example, we are using the average March-May rainfall season for the period 1981-1992 (Figure 8.1**(1)**).
3. The tool automatically specifies the output file (Figure 8.1**(2)**).
4. Select a contour interval value. In this case, 400 for an interval of rainfall of 400 mm (Figure 8.1**(3)**).
5. Change the missing value if necessary.
6. Keep the window size as default 5 X 5.
7. Click OK to run the tool.

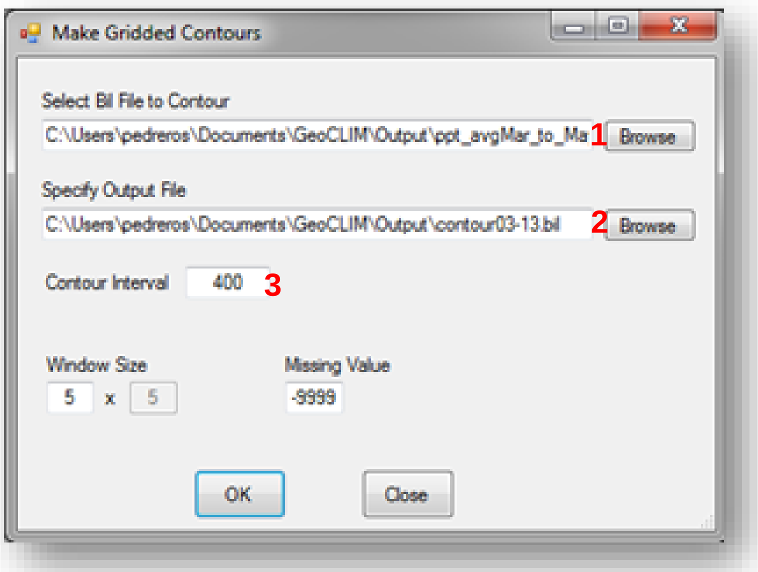


Figure 8.1 *Select a rainfall dataset and the contour interval.*

You can use the contour tool to analyze changes in average rainfall for different periods of time. For example, Figure 8.2 shows the changes in average for the periods 1981-1992, 1993-2004, and 2005-2016 for the East Africa Countries (EAC) for the March-May season. In this example, we use a 200mm interval. The results show a decrease in rainfall in polygons 1, 2, and 3 in Figure 8.2; areas of 400mm in the western part of the region (polygons 1 and 2) are decreasing toward 200mm. Also, areas of 200mm in Kenya are changing into the zero interval (polygon 3).

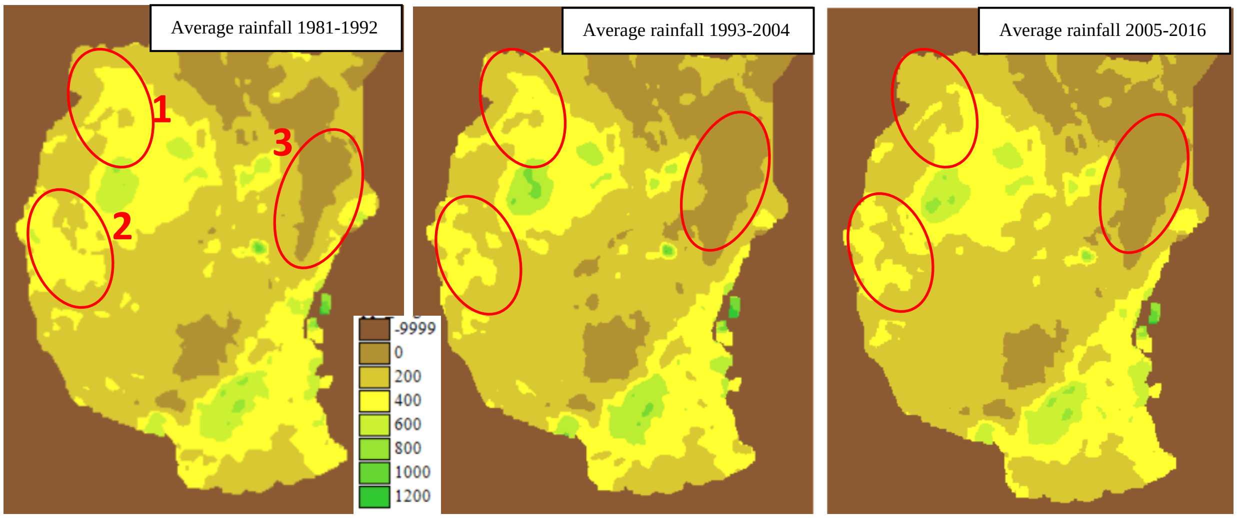


Figure 8.2 *200 mm interval of average rainfall for the March-May season, during the periods 1981-1992, 1993-2004, and 2005-2016*

Figure 8.3 shows another example of the use of the ***Contour tool***. In this case, we are comparing the seasonal average for the period 1981-1998 (in red) to that of 1999-2018 (in green) for the Feb-June season in Ethiopia at the 300mm interval. The image shows that the area that receives more than 300mm is shrinking.

The map in Fig 8.3 was produced as follows:

1. Use GeoCLIM to calculate average 1981-1998 and 1999-2018.
2. Use the Contours tool at a given interval (300mm for this example).
3. Bring the resulting layer to a GIS package and run a raster to vector function.
4. Overlay the two resulting polygon shapefiles.

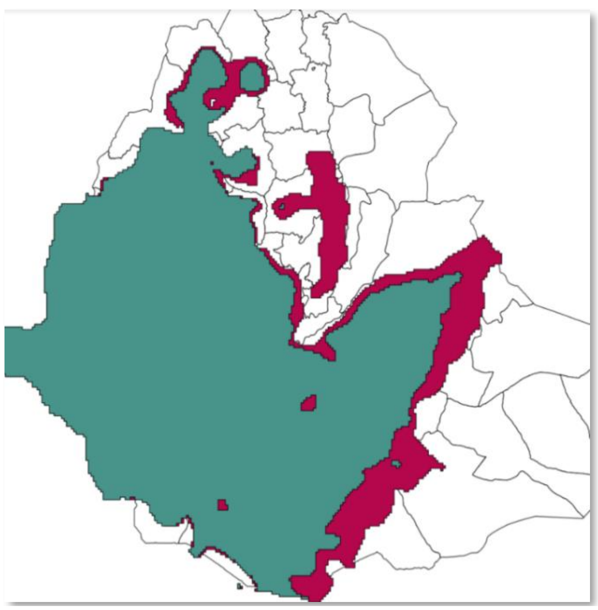


Figure 8.3 *Changes in rainfall patterns during the Belg season in Ethiopia. The image shows the areas with average rainfall above 300 mm (FMAMJ) for the period 1981-1998 (red) and 1999-2018 (green).*

Chapter 9: Calculate Long-Term Changes in Averages

Summary

Another way to estimate trends is by comparing the change in averages between two periods within a time-series. ***The Climate Trends - Change in Averages*** tool allows you to estimate the trend by dividing the time-series into two groups of years and calculating the difference in average between the two groups.
To run the tool, follow the steps below:

1. Open the ***Climate Trends*** tool from the GeoCLIM main toolbar, see Figure 9.1.
2. Select the season to be analyzed, see Figure 9.1 **(1).**
3. In series 1, select the first period of time, see Figure 9.1 **(2)**.
4. In series 2, select the second period, see Figure 9.1 **(3)**.

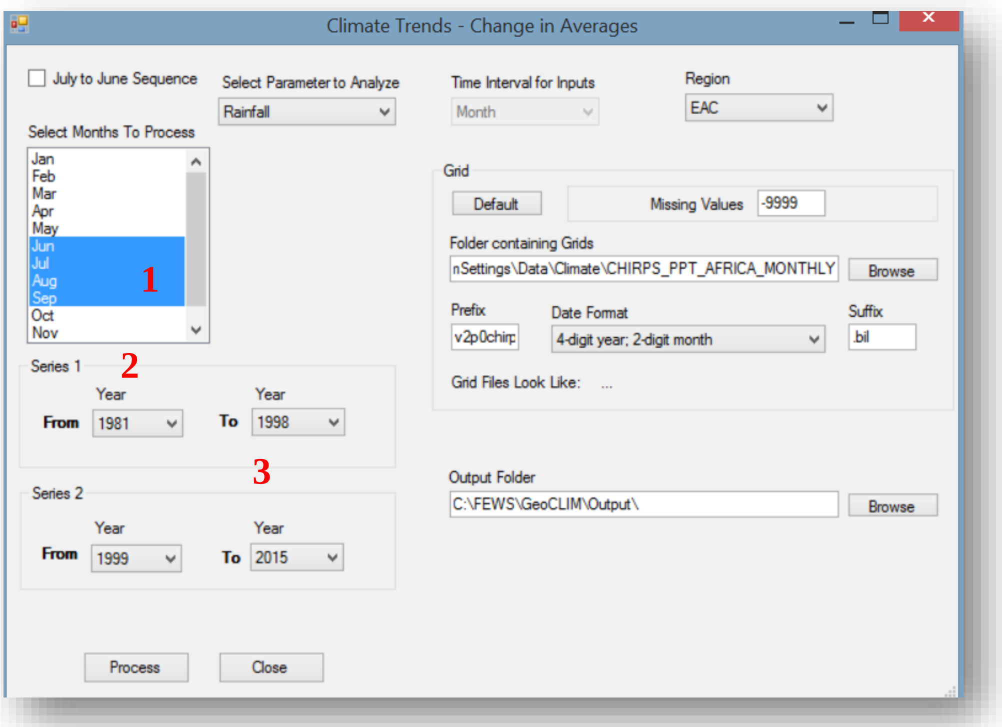


Figure 9.1 *The Climate Trends - Calculate Change in Averages tool is another way of estimating trends by comparing changes in averages between two periods.*

The tool fills the right side of the form automatically with the information from the selected rainfall dataset. Finally, click Process to finish.

Figure 9.2 shows the output map result of the change in average from the 1981-1998 period (series 1) compared to the 1999-2015 period (series 2), for the June-September season. The result shows that there are areas with increasing (blue polygon) and decreasing trends (red polygon). The legend is given in mm for every ten years.

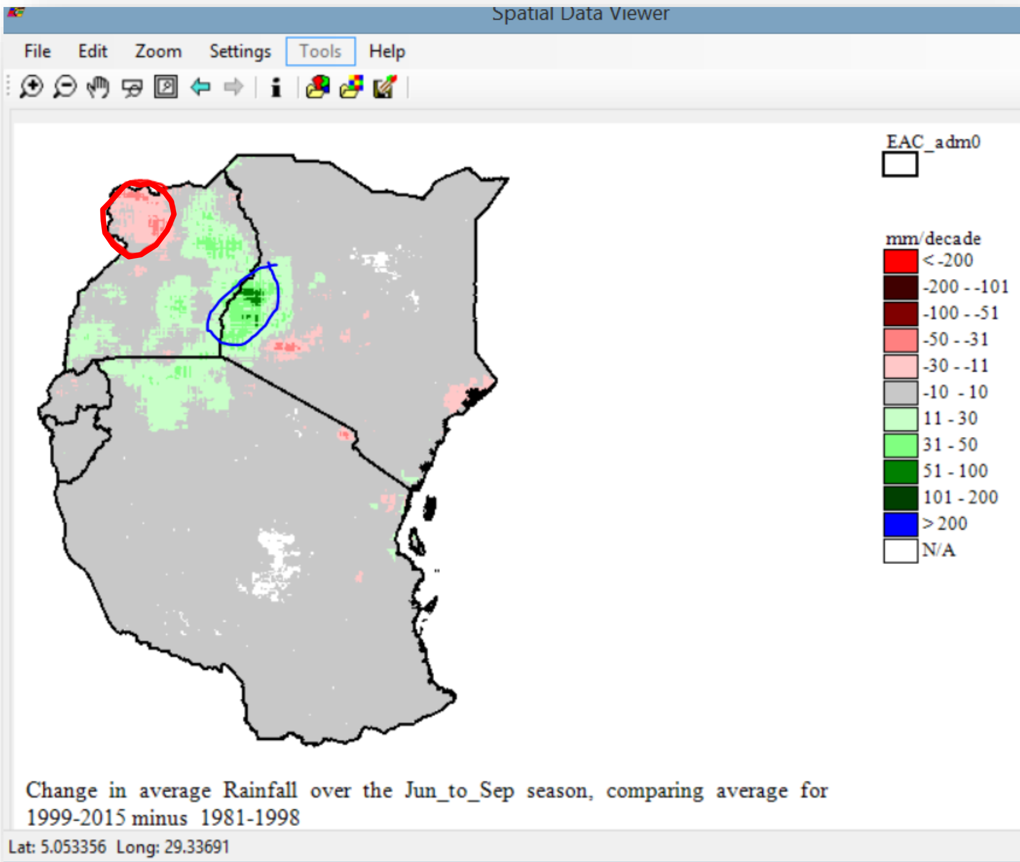


Figure 9.2 *Difference in averages; green-blue colors show an increase in the latter period (series 2) while pink-red colors show a decrease in rainfall in that same period. The legend is given in mm per decade.*

Chapter 10: Background-Assisted Station Interpolation for Improved Climate Surfaces (BASIICS)

Summary

Satellite data provide useful information on climate variables (rainfall, temperature, and evapotranspiration) patterns. However, sometimes, satellite-estimated data contain biases and inaccuracies due to incorrect or limited ground data used during calibration. Some raster data also have a low spatial resolution, meaning the size of the pixel is too large for the area of interest. Such data could be improved by combining them with ground station information using the ***Background-Assisted Station Interpolation for Improved Climate Surfaces***, (BASIICS) algorithm in GeoCLIM. The BASIICS tool comes bundled in the Batch Assistant, which includes the following processes:

- Validate satellite data using ground station values.
- Blend climate raster/grids with stations (BASIICS).
- Interpolate stations only, see section 10.2.3.4 for the description of interpolation using GeoCLIM.

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

1. Use the GeoCLIM to download or import the raster datasets to be improved, see chapter 2.
2. Use the Validate Satellite rainfall to determine if the satellite and station data are correlated.
3. If they are correlated, blend the two datasets to produce improved rainfall estimates. Save the set up as a batch file.
4. Use the batch file to run the blending process to update improved rainfall datasets in the future.

10.1. Validate Satellite Rainfall

The ***Validate Satellite Rainfall*** option allows you to evaluate grid/raster datasets (e.g., satellite-based rainfall estimates) using discrete points in space (e.g., rain gauges). The validation helps to determine if the two datasets are correlated to help in deciding if the blending option can be used with the two datasets to improve the raster using the points. The first step of this process is to extract values from a raster/grid at all locations where the point data have valid values (i.e., non-missing values. Missing values can be specified in the inputs). The result is a dataset of grid values that can be directly compared to the point values. Another result is a diagnostic file with the information on the least-squares regression between the observed/in situ data value at the points being evaluated and the extracted grid values along with an R-squared output value. Once the correlation has been determined, then the raster and station data can be blended into an improved dataset.

To validate raster data, follow the three steps below:

10.1.1. Step 1: Select the BASIICS option.

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

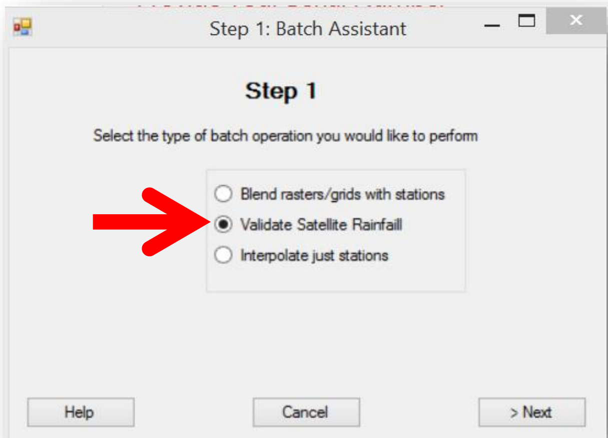


Figure 10.1 *There are three Batch Assistant options available; (1)Blend stations and raster data, (2) Validate Satellite Rainfall and (3) Interpolate Just Stations.*

10.1.2. Step 2: Select the validation time interval and period.

4. Select the time interval (e.g., month, dekad, or pentads) and period (e.g., 012013 to 122018) of the raster data to validate. The time period and time interval are based on the selected climate dataset definition. In this example, we are using monthly data (Figure 10.2). And we are validating from January 2013 to December 2018.
5. Click on the > Next button to proceed to ***Step 3***.

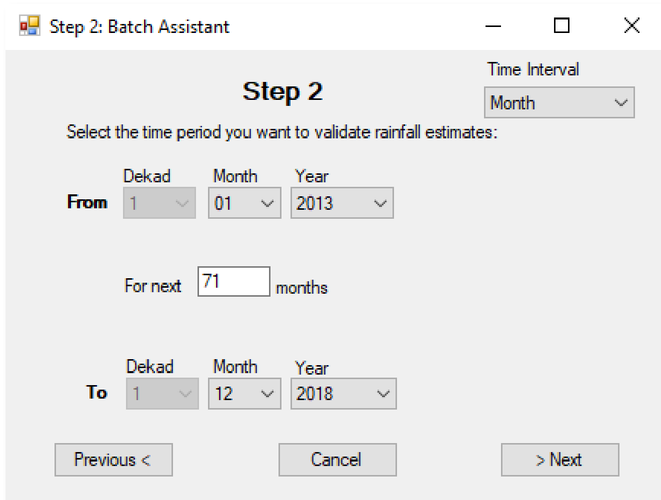


Figure 10.2 Step 2 requires specifying of the period and Time Interval of the data to be analyzed.

10.1.3. Step 3: This step has three sections (Figure 10.3).

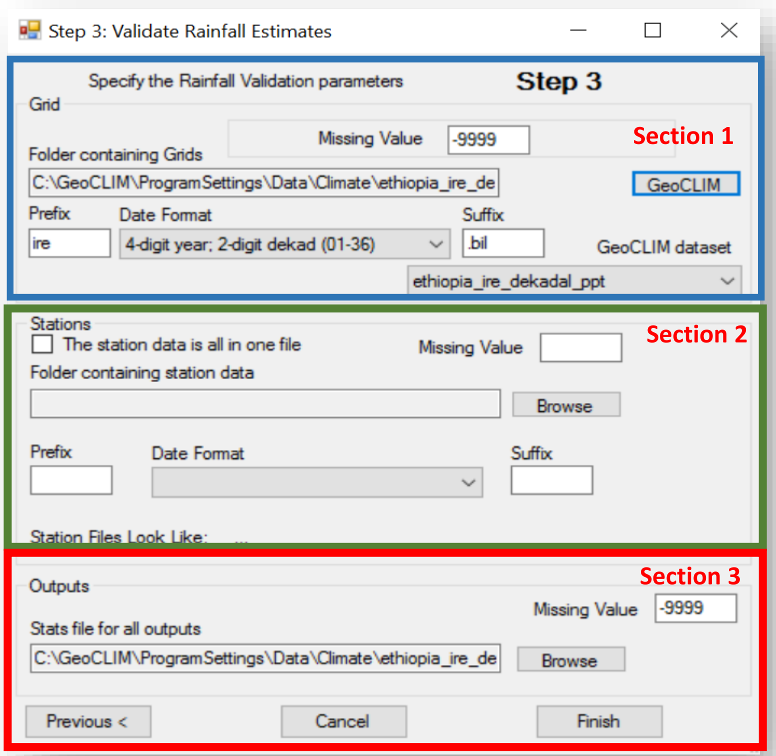


Figure 10.3 To validate raster data using station values, it is required to specify the path to the raster data and a table with the station values.

10.1.3.1. *Section 1: Grid*

This section relates to the raster/grid input parameters. This process allows validation of climate datasets that have already been registered in GeoCLIM. To select the climate dataset to be validated, use the GeoCLIM dataset ∨ pull-down menu, then click on the GeoCLIM button to automatically populate all the fields in this section with the information of the selected dataset.

10.1.3.2. *Section 2: Stations*

This section relates to the station input parameters.

1. Check □The station data is all in one file box, then select the file which contains the station data. The file must be in CSV format, see an example in Figure 10.4. See the Data Types chapter for more information on the format of the table and other file types in GeoCLIM. If the station data are in separate files, leave this box unchecked.

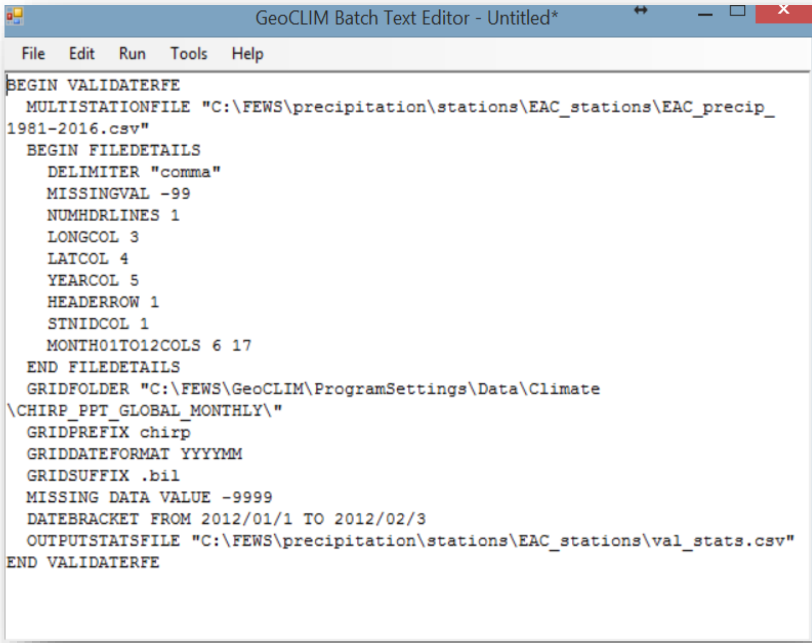
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
ID	FEWSID	lon	lat	year	month01	month02	month03	month04	month05	month06	month07	month08	month09	month10	month11	month12
14741	63932002	33.6	-9.3	1985	115	269	258	915	250	82	46	18	30	68	234	90
14688	63844007	38.4	-5.1	1984	72.8	19.2	172.7	913	306.9	295.4	114	40.1	98.8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

Figure 10.4 *The CSV table with station data must contain a statin ID, lon, lat, year and a column for each pentad, dekad or month.*

2. After selecting the station file(s), the ***Define Delimited Data Text File*** dialog box will open showing the format of the station file: 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**, the first and last period (the period could be pentad, dekad, or month), and the missing value have all been specified.

10.1.3.3. *Section 3: Outputs*

1. Specify the file location where the statistical outputs will be written.
2. Click on Finish.
3. A batch file is generated and displayed on the GeoCLIM ***Batch Text Editor*** (Figure 10.5). The information displayed comes from the inputs entered on the previous steps. This batch file can be saved for future reference or edited to run a new process.
4. Go to the Run pull-down menu and select the Run Batch File (Figure 10.5) to start the validation. Alternatively, press the F5 key on the keyboard.




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

Figure 10.5 *The batch file is a text file with a list of all the inputs from the three steps of the BASIICS validation form.*

Once the validation process is completed, it creates the following four outputs:

1. A graphic map in PNG format showing the stations used for validation overlaid over the time-corresponding rainfall field (Figure 10.6).
2. A scatterplot showing the satellite rainfall field values against the station values (Figure 10.7).
3. A CSV file with columns containing the station values, the raster values for the points where the stations fall, and the at-station interpolated observed data. These at-station interpolated values are produced to improve comparability between the gridded/raster data and the station data. The CSV file includes some statistics showing the correlation of the rainfall field and station data (Figure 10.8).
4. A shapefile containing all the stations that were used in the process.

These outputs provide the basis to decide if it is appropriate to blend the stations and the raster datasets.

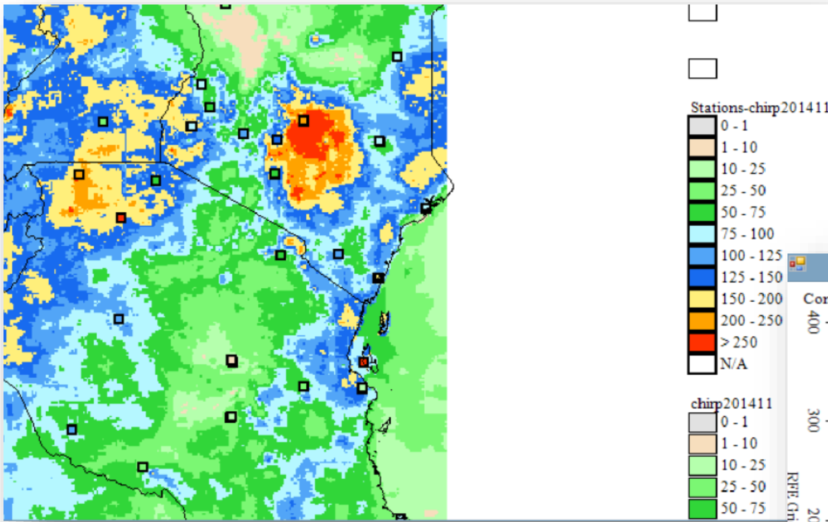


Figure 10.6 A map of the stations that were used in the process of validating CHIRP.

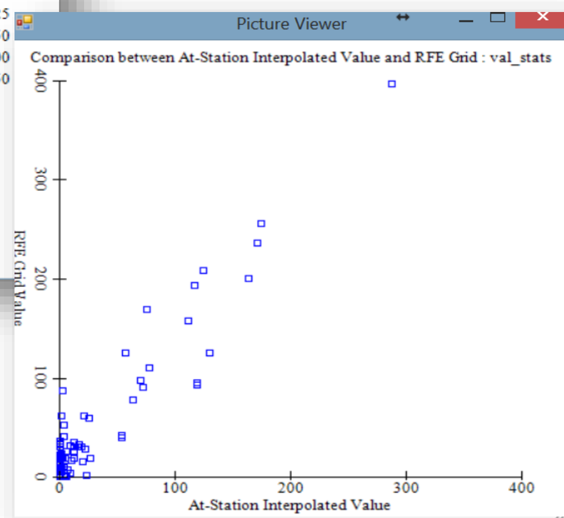


Figure 10.7 Scatterplot of station value on X and raster (CHIRP) value on Y.

```
File  Edit  Format  View  Help
Name, Filename, Long, Lat, StationValue, InterpAtStationValue, ExtractedValue
id1, v2p0chirps200001, -78.374, 0.295, 115, 116.52, 84
id2, v2p0chirps200001, -78.01, 0.023, 130, 135.93, 98
id3, v2p0chirps200001, -77.88, 0.001, 162, 155.02, 116
id1, v2p0chirps200002, -78.374, 0.295, 113, 115.53, 116
id2, v2p0chirps200002, -78.01, 0.023, 146, 151.7, 129
id3, v2p0chirps200002, -77.88, 0.001, 179, 171.52, 182
id1, v2p0chirps200003, -78.374, 0.295, 122, 125.66, 73
id2, v2p0chirps200003, -78.01, 0.023, 128, 157.94, 95
```

Figure 10.8 Text file that includes a list of the station value and its corresponding raster value for each date together with statistics describing their relationship.

10.2. Blend rasters/grids with stations (BASIICS)

The blending algorithm is a methodology designed to combine stations such as rain gauges with raster/grid data, such as satellite-based estimates, to produce a more accurate gridded dataset. The algorithm combines the spatially discrete point data with spatially continuous grid data by interpolating ratios between the point and the grid value, where these two data are collocated, then multiplying the interpolated ratios by the original grid. The blending is done using a modified Inverse Distance Weighting (IDW) method, which uses some concepts from the kriging method of interpolation, particularly simple and ordinary kriging. See a complete description of the blending process in section 10.2.2.4.
To create improved rainfall estimates, follow the steps below:

10.2.1. Step 1: Select BASIICS option

1. Click on the ***BASIICS*** button from the GeoCLIM main toolbar.

2. This will open the ***Step 1*** window, with the ■ Blend rasters/grids with stations option already selected. Click on the > Next button (Figure 10.9).

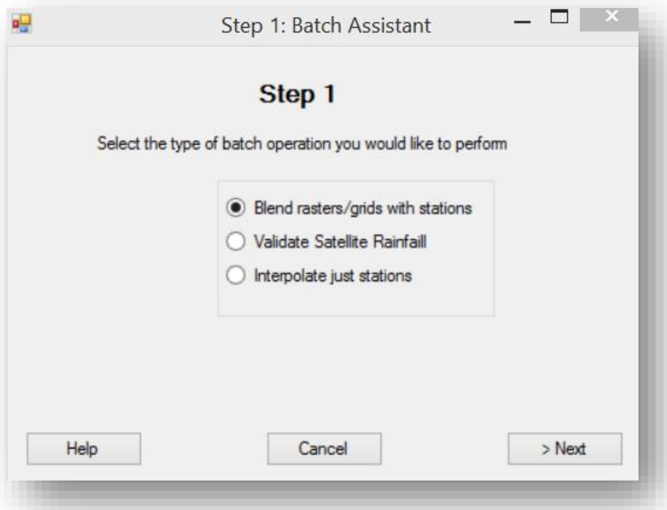


Figure 10.9 *Select the Blend raster/grids with stations option.*

10.2.2. Step 2: Select blending time intervals and periods

Select the Time Interval V (Figure 10.10(1)) (e.g., Month, Dekad, or Pentad) and start and end periods to be improved (Figure 10.10(2)) (e.g., month-year). Make sure that there are stations available for the same time interval and time period to be improved. Click on > Next to continue.

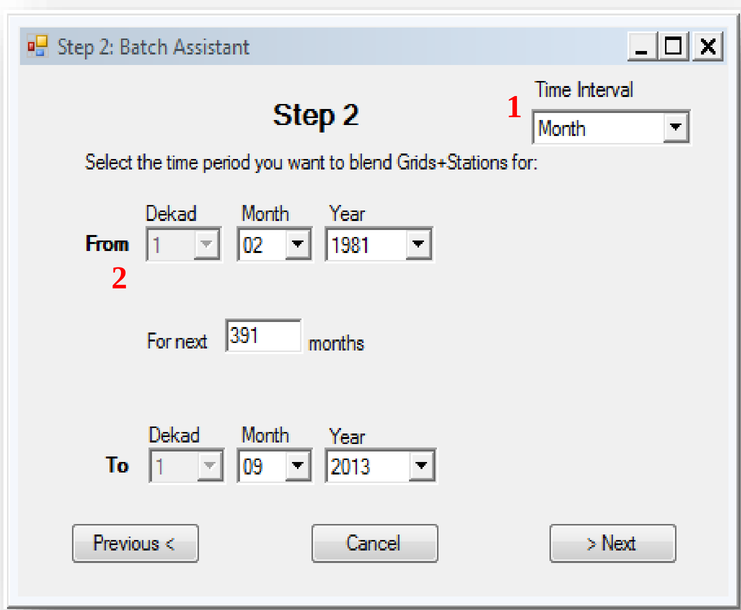


Figure 10.10 *Select the starting and ending dates that would be part of the process and click* Next.

10.2.3. Step 3: Parameter for the blending process

This step contains four sections:

10.2.3.1. *Section 1 - Grid*

This section relates to the raster/grid input parameters. The BASIICS data improvement routine is implemented only on climate datasets that have already been registered in GeoCLIM. Select the climate data to be improved using the GeoCLIM dataset ∨ pull-down menu then click on the GeoCLIM button to automatically populate all the fields in this section with the information of the selected dataset (Figure 10.11(1)).

10.2.3.2. *Section 2 – Stations*

This section describes the file(s) containing the station data; this includes the location of the file(s), the missing value, and the column numbers for each of the required inputs (Station ID, latitude, longitude, etc.). The tool retrieves this information automatically, but it is important to verify that it is correct (Figure 10.11(2)).

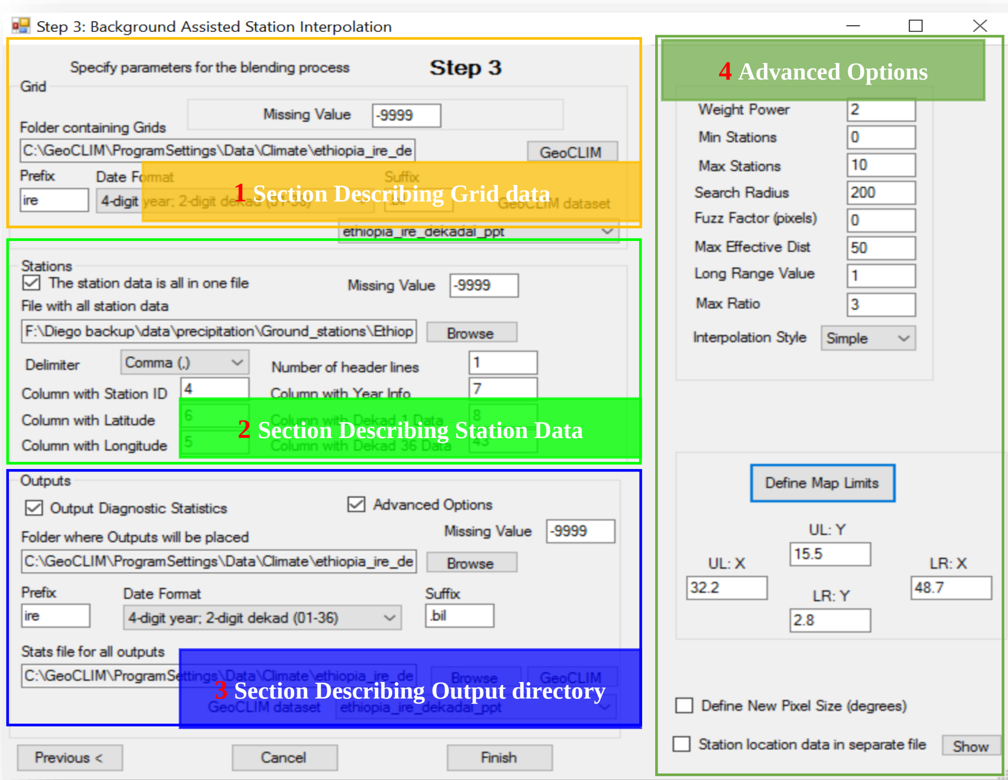


Figure 10.41 Step 3 of the blending process requires information about the raster data to be improved, the stations and the output location.

The station data should be in a CSV file(s) in the recommended format described below and shown in Figure 10.4. The order of the columns is not important, but must include:

- A unique station identifier ID in a single column.
- A longitude column.
- A latitude column.

- 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). Any missing data should be completed with the Missing Value.

10.2.3.3. *Section 3 – Outputs*

In the third section, you can specify the output directory where to save the blended products. At this point, you have two options: (1) create a new dataset or (2) update an existing one.

1. **Create a new dataset:** This first option allows you to create a new dataset; for example, you are blending, for the first time, your stations with the historical data of CHIRPS and want to create a new dataset from the results. To do this:
   a) Browse to the GeoCLIM data repository, Figure 10.12 **(1),** and create a new directory with the name of the new dataset. For example: `C:\GeoCLIM\ProgramSettings\Data\Climate\new_dataset`
   b) Provide a prefix and select the file name format according to the period of time, Figure 10.12 **(2)**.
   c) Make sure you complete the Advanced Options (see Section 4 below) before continuing.
   d) Click Finish.
   e) A dialog box shows up with a question, click Yes to create a new dataset.
   f) A dialog box opens up, enter the name of the dataset as defined in (a) above. Select the Data Type ⩒, in this case, Precipitation. Select Africa for Data Extent ⩒.
   g) A batch file is generated and displayed in the GeoCLIM ***Batch Text Editor***. The information displayed comes from the inputs entered on the previous steps (Figure 10.5). This batch file can be saved for future reference or edited to run a new process.
   h) Go to the Run ⩒ pull-down menu and select Run Batch File. Alternatively, hit the F5 key on the keyboard.
2. **Update an existing dataset:** The second option is to add the latest record to an existing dataset. For example, you are blending the latest CHIRPS dekad with the stations and updating the time series you created previously.
   a) Browse to the existing directory where you want to save the new data (Figure 10.12 **(1)**).

b) Find the existing dataset on the GeoCLIM dataset ∨ pull-down menu (10.12 **(4)**). Once you select the dataset, click the GeoCLIM button (Figure 10.12 **(3)**) to identify this dataset as an existing one. You will see that the prefix changes according to the name of the files on the existing directory (Figure 10.12 **(2)**).

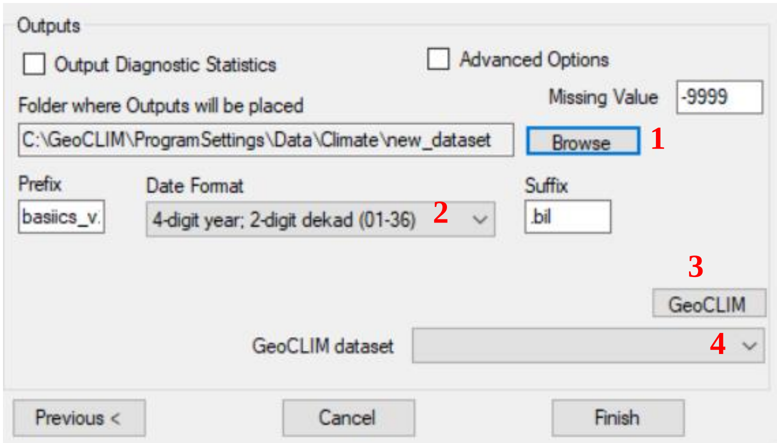


Figure 10.5 *The output section allows you to create a new dataset or update an existing one.*

c) Click Finish.

d) A window message shows up, make sure that the name of the dataset is correct and click Yes to confirm.

e) Run the process.

10.2.3.4. *Section 4 - Advanced Options- the Blending Process*

To enable this section, check the ■ Advanced Options box in section 3. By checking this box, the program opens a set of options to adjust the parameters of the interpolation (Figure 10.13 **(1)**). And define the geographic limits of the output grids (Figure 10.13 **(2)**).

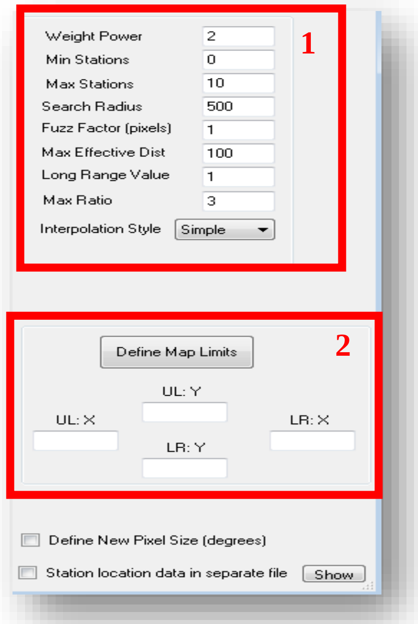


***Figure 10.6** The advance option panel allows you to modify the blending process.*

The Blending Process:
(1) A point dataset with values at discrete locations in space (example: rain gauges)
(2) A grid dataset with values varying continuously over space (for example, a satellite-based rainfall estimate grid or a climatic average). For the algorithm to be used effectively, the two datasets need to be correlated.
Step1. 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 at the point locations that can be directly compared to the point values, see section 10.2 for validation.
Step2. A ratio is calculated between stations and grid value; these ratios are interpolated using a modified IDW method, giving a maximum effective distance to each station. Once the maximum effective distance is reached, the interpolated layer takes the value of 1.
Step3. The original rainfall layer is multiplied by the interpolated ratio layer. The pixels within a maximum effective distance of a station adjust the raster value based on the ratio, the pixels outside the influence of a station that took the value of 1 (step2), take the value of the original raster layer.

The technique is similar in principle to the SEDI technique that originates from the Southern African Development Community (SADC)/FAO Regional Remote Sensing Project, developed by Peter Hoefsloot. The interpolation technique is done using the Inverse Distance Weighting

(IDW) approach, borrowing some concepts from the kriging method, particularly the use of simple and ordinary kriging, as described below.

For example, assume that we are blending a rain gauge dataset with a satellite rainfall estimate. At station point A, the rainfall value is 10 mm, while the grid pixel value is 1 mm. Although the absolute difference between the two estimates is only 9 mm, the station/grid ratio is 10, or 1,000 percent. The ratio from all the points will be interpolated and then multiplied by the original grid. Assume that 50 km away from point A, the grid pixel had a value of 30 mm. This 30 mm will be multiplied by a value close to 10, depending on the surrounding ratios in the interpolation, and the resultant value may be close to 300 mm. This error can be limited by capping the ratio and instructing the program to cut off any ratios than a certain value (MAX RATIO). A cut-off ratio of 3 is used by default in the algorithm, meaning that any ratio greater than 3 is reset to 3 (in the example above, the ratio would be 3 instead of 10). However, this ratio can be set to any value by the user (a very large cut-off can be used; for example, 100,000) if you do not want to have the ratios capped.

The following are the different parameters from the ***Advanced Options*** that could be changed to adjust the interpolation. Please make sure you have a full understanding of the parameters before making any changes, otherwise, leave the default values.

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.

Max Ratio **(MAXRATIO):** The maximum value allowed for the station/grid ratio. The program calculates the ratio between the station and grid values at each point location. The MaxRatio value limits this ratio to avoid "run-away" values in the process.

Search Radius, Min Station, and Max Stations:

- **SEARCHRADIUS –** the radius within which to search for points to be interpolated.
- **MINSTNS** – minimum number of stations used in the interpolation.
- **MAXSTNS** – maximum number of stations used in the interpolation.

The interpolation algorithm needs input values from the Min Stations **(MINSTNS)**, the Max Stations **(MAXSTNS**), and the Search Radius **(SEARCHRADIUS**) fields for the pixel value estimation. At every pixel, the algorithm will search for the nearest stations within the **SEARCHRADIUS** from that pixel location and use the **MINSTNS** and the **MAXSTNS** to limit the number of stations to use during the interpolation.

For example, assume you defined the number of stations between 2 (**MINSTNS**) and 10 (**MAXSTNS**) to be used within a search radius of 200 km (**SEARCHRADIUS**). For this case, the algorithm will search for the nearest 10 stations within a radius of 500 km. If the number of stations found is less than 10 stations, for example, 7, then those 7 stations will be used. However, if the number of stations found is less than 2, then that location will have a missing value. Hence, for BASIICS, it is recommended to use an input value of 0 for **MINSTNS**, to produce a value everywhere in the output and avoid missing values.

Fuzz Factor (pixels) **(FUZZFACTOR):** The fuzz factor hides the location of the station by the number of pixels indicated in this field. A Fuzz Factor = 0 makes the value of the pixel near the station as close as possible to the station value.
Max Effective Distance **(MAXEFFECTIVEDIST):** This parameter is the maximum distance for which a station has influence over. This parameter only works with the Simple interpolation Style (idw_s, see the interpolation style section below). It is very important to take into account the local characteristics of the region to choose a proper value for this parameter. We recommend you test different values for the Max Effective distance to avoid localized (bulls eye) effect around the station's location.
Interpolation Style ∨ **(INTERPOLATIONALGORITHM)**: The program provides two interpolation algorithms, Simple (idw_s) and Ordinary (idw_o) inverse distance weighting (IDW). In the ordinary IDW, the interpolation weights are dependent only on the surrounding stations. The Simple IDW method uses a background field to complete the interpolation. The background grid also contributes as a weight to the interpolation routine, and the relative weight of the background grid increases with increasing distance to surrounding stations.
Define Map Limits option: Allows you to define the interpolation area (Figure 10.144). Make sure that the area is smaller or equal to the gridded dataset. This area can be defined by using the extent of an existing GeoCLIM Region or other spatial data (raster or vector). This option helps to speed up the interpolation process.
To run the blending process for the EAC region, as shown in Figure 10.144, follow the steps below:

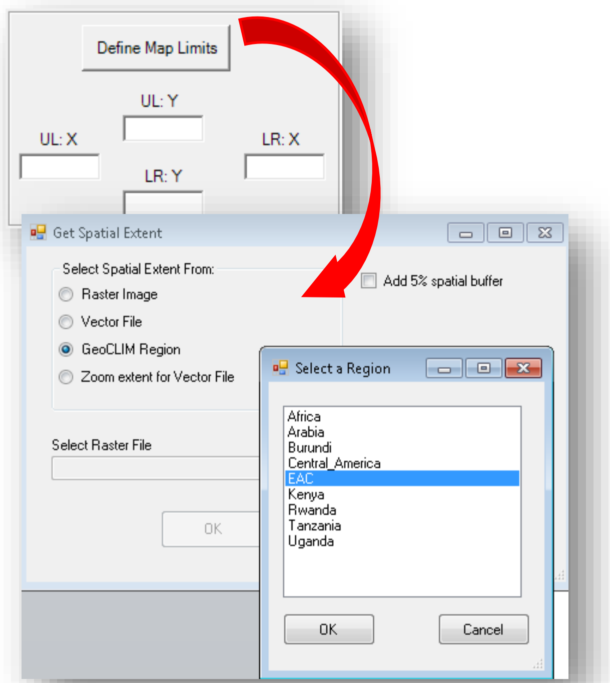
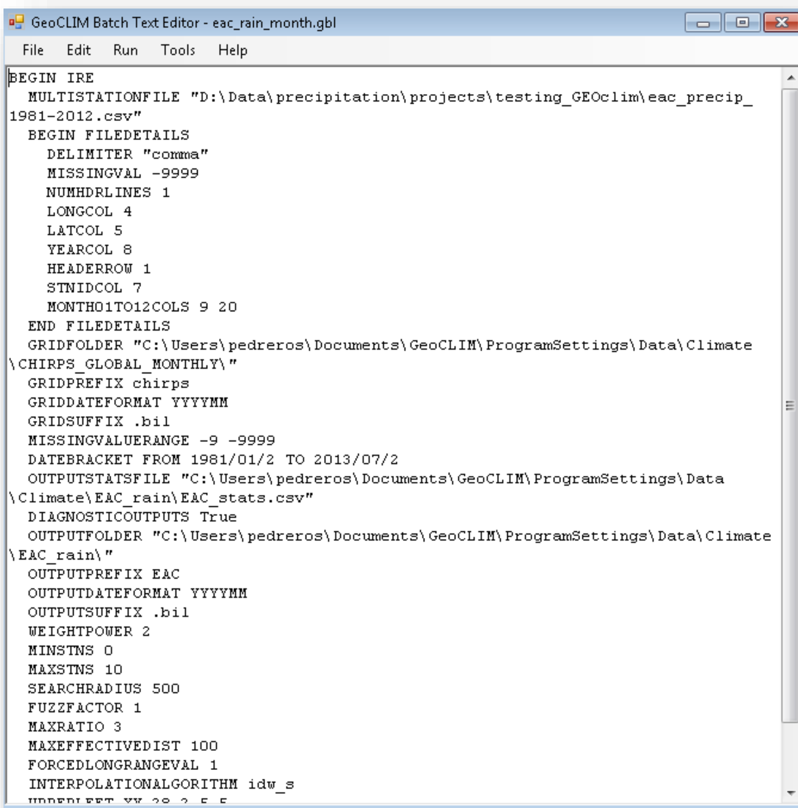


Figure 10.7 *Defining the output extend could be done by selecting an existing raster, region or shapefile.*

1. Click on Define Map Limits button.
2. From the ***Get Spatial Extent*** window, select the ■ GeoCLIM Region option. This will bring up the ***Select a Region*** window.
3. Choose the **EAC** region from the list.
4. Click on OK.
5. Then, on the ***Get Spatial Extent*** window, click OK. The geographic extent of the selected region populates the map limits on the ***Define Map Limits*** window.
6. After you complete all the fields, the Step 3 form should look like Figure 10.111.
7. Next, click on the Finish button at the bottom of the Step 3 window. This will generate a batch file and display it on the ***Batch Text Editor*** with all the inputs in command form Figure 10.15). Review the file to make sure that all the inputs are correct.
8. Save the batch file, go to File > Save As from the ***Batch Text Editor*** menu, this will enable you to have access and edit the saved file later.
9. 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 3 of the blending process. The commands are self-descriptive (Figure 10.15).




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

__Figure 10.15__ The batch file describes, in command form, all the fields from step 3. You can edit the batch file and re-run the process.

NOTE: You can save the batch file for later use. You can edit the code by opening the file using the batch editor and rerun it for different periods.

Chapter 11: Extracting Raster Statistics and Time Series

Summary

The ***Extract Grid Statistics*** tool calculates summary statistics for a polygon (or a set of polygons) from a shapefile, using a raster or set of rasters from the selected climate dataset. For example, you can calculate the spatially averaged rainfall for each district for each month from 1981 to 2017. This produces a CSV table that could be analyzed using Excel.

11.1. Extract Statistics

To extract summary statistics for a set of polygons, follow the steps below:

1. Open the ***Extract Statistics from Raster Data using Shapefile*** tool from the GeoCLIM toolbar.
2. Select a shapefile containing the polygons of interest (e.g., districts) and select a unique ID field (a data field in the shapefile that will uniquely identify each polygon, such as district names) Figure 11.1 **(1)**.

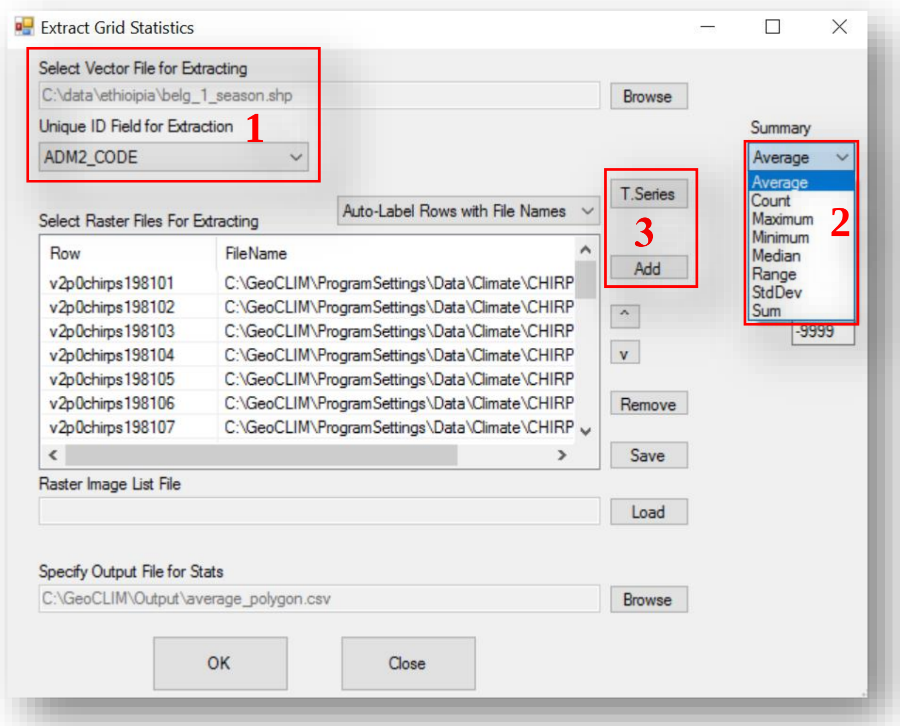


Figure 11.1 *The Extract Statistic tool calculates the spatial statistics for each raster data set, using polygons of the selected shapefile. The output is a table containing a row for each polygon and a single column value (statistics selected) for each raster.*

3. Select the type of summary for the pixels within the polygon, see the Summary∨ pull-down menu, Figure 11.1 **(2)**.
4. Select the raster file(s). There are two options to populate this input, Figure 11.1 **(3)**: (a) using the T-series (Time-series) button or (b) using the Add button to select the raster datasets in BIL format. The T-Series button generates a list of raster files for a time series of raster files (e.g., monthly rainfall for each month of each year from 1981 to 2010) from which to extract a time series, while the Add button allows the selection of individual files.

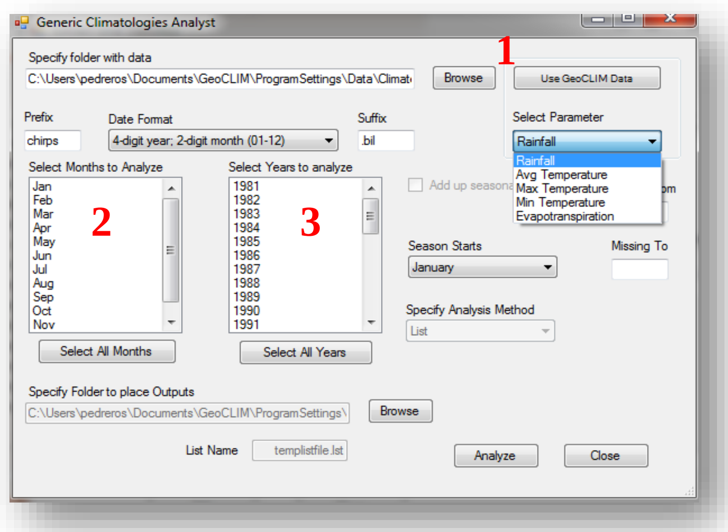


Figure 11.2 Select a T-series from the default data or point to the path for a new data source.

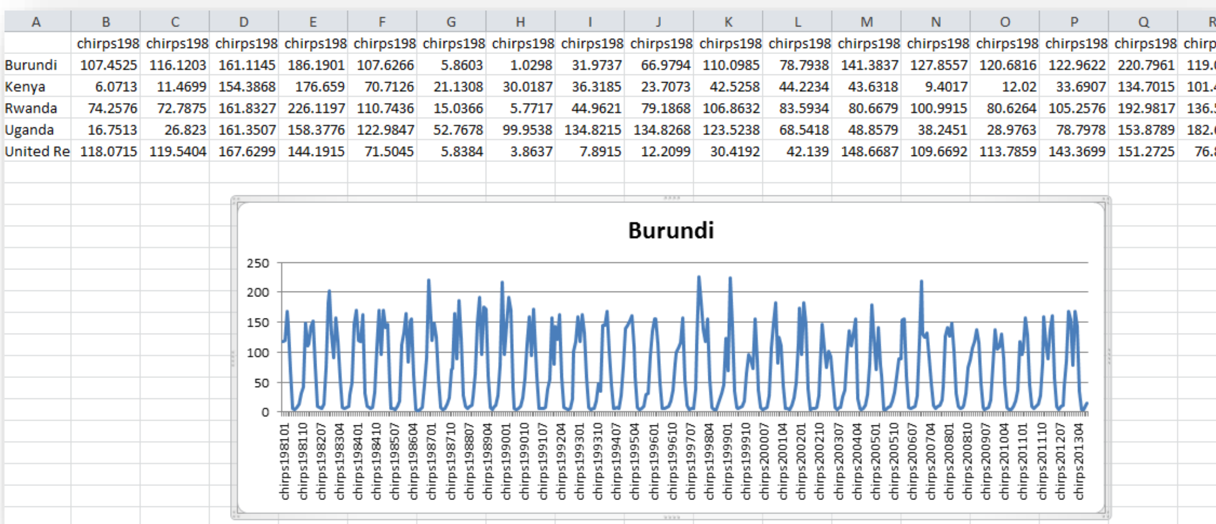
a) **Option 1**: click on the T-series button. This option will open a new window to define the climate dataset and time period to extract from.
   i) Click on the Use GeoCLIM Data button to retrieve all the information from the selected climate dataset or browse to the directory containing the data to be used, Figure 11.2 **(1)**.
   ii) Select the periods, Figure 11.2 **(2)**.
   iii) Select the years, Figure 11.2 **(3)**.
   iv) Click on 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 shown in Figure 11.1. A window message will confirm the creation of a list file.
   v) Click OK to close the window message and then click on the Close button to close the ***T-series*** window.
   vi) Back in the ***Extract Grid Statistics*** tool, the **Select Raster Files for Extracting** field is now populated.
   vii) Specify the output location where the CSV table output file will be saved.
   viii) Click OK to run the process.

b) **Option 2**: Click on the Add button and browse to the directory where the `*.bil` files are located.

i) Select the files to be used in the process.
ii) Click Open.
iii) Back on the ***Extract Grid Statistics*** window, specify the output directory if necessary.
iv) Click OK.

11.2. Results

The ***Extract Grid Statistics*** tool produces a CSV table file with rows corresponding to the polygons from the input shapefile. Columns contain the summary value for each raster file selected. Figure 11.3 shows the output CSV table in Excel for 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 the production of time series graphs, open the CSV file in Microsoft Excel (or other spreadsheet program).



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

Figure 11.3 *The resulting table has a row for every polygon and every column represents the summary value for each raster.*

Chapter 12: Working with Climate Data Archives

Summary

As mentioned in chapter 2, there are different options to make data available in the GeoCLIM program. One option is to download the data directly from an FTP site into the GeoCLIM workspace (`GeoCLIM\ProgramSettings\Data\Climate` directory). Another option is to import a GeoCLIM archive. An archive is a group of raster files compressed into a single file that contains specifications such as name format, (e.g., '`v2p0chirpsyyyymm`' - prefix, 4-digit year, 2-digit month) and time period (pentad, dekad, month) so that GeoCLIM can read the data. Once GeoCLIM imports the archive, it creates a new folder under the `Climate` directory in the workspace, with the name of the archive. This new folder contains all the data from the archive file and makes it available in GeoCLIM Settings > Data > Select Dataset. GeoCLIM archives are a great way to share data among GeoCLIM users. This chapter goes over how to create and import a data archive.
For a review of GeoCLIM data settings, go to chapter 2.

12.1. Create an Archive

To create an archive, follow the steps below:

1. Open the ***View Available Data*** tool.
2. Make sure that the correct climate variable is selected.
3. Select the entire or part of the time series to be added to an archive.
4. Click Export, Figure 12.1.
5. Select ■ GeoCLIM Archive and click OK, a message ensuring the correct name of the new archive will pop up.
6. Click OK and name the new archive.

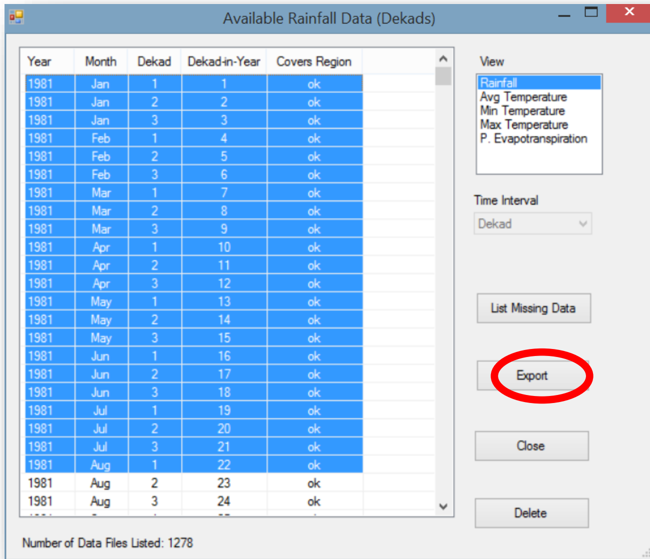


Figure 12.1 Exporting a time series or part of it from the Available Rainfall Data tool will create an archive.

A new file with the CLIMDATA extension (`*.climdata`) is created in the assigned directory. This file contains all the information necessary so that GeoCLIM can read the data.

12.2. Importing archives

To import a GeoCLIM archive, follow the steps below:

1. Open the ***Import Climate Data Archives*** tool from the toolbar.
2. Select the archive to be imported and click Import, Figure 12.2.

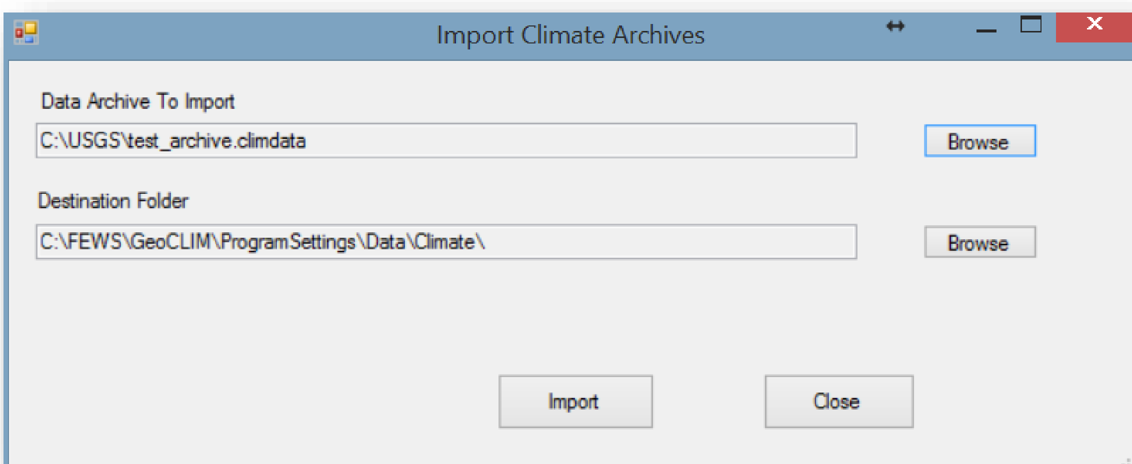


Figure 12.2 The Import Climate Archives tool.

3. Close the console, Figure 12.3, once it is done.

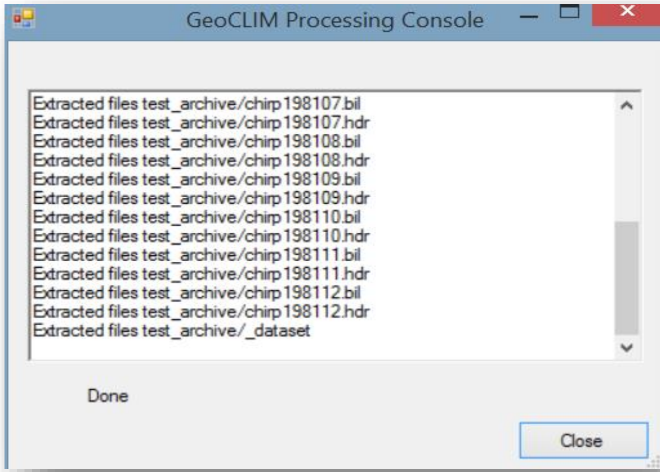


Figure 12.3 *The GeoCLIM console shows the progress of the archive import.*

To ensure that the archive imported correctly, go to the GeoCLIM Settings > Data > Select Data, and the newly imported dataset should be available on the list of datasets, Figure 12.4.

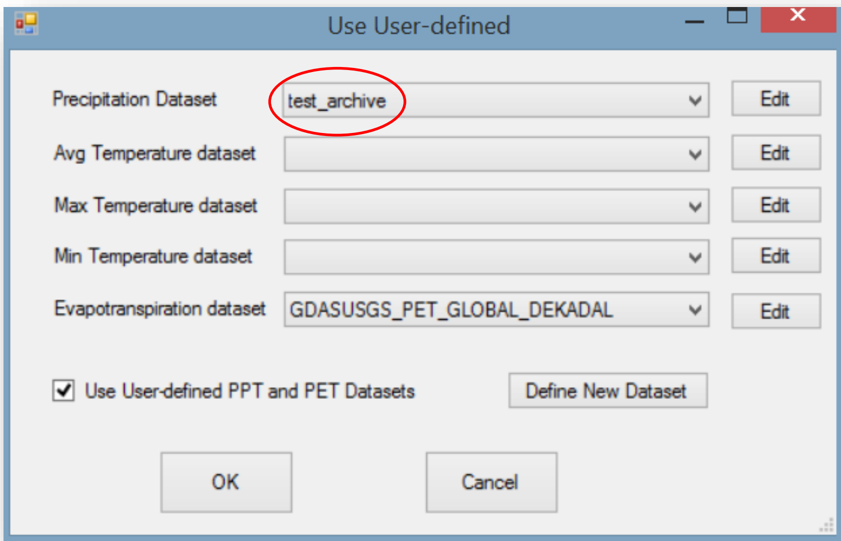


Figure 12.4 *Once finished importing the archive, the dataset is ready to be used in the different operations of the GeoCLIM program.*

References

Environmental Systems Research Institute. (2008). ArcGIS Desktop Help 9.2 - BIL, BIP, and BSQ raster files. Retrieved July 3, 2018, from http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=BIL,_BIP,_and_BSQ_raster_files

Environmental Systems Research Institute. (2016). FAQ: What does the pixel depth mean? Retrieved July 3, 2018, from https://support.esri.com/en/technical-article/000006576

Famine Early Warning Systems Network. Standard Precipitation Index (SPI) Product Documentation. Retrieved July 3, 2018, from https://earlywarning.usgs.gov/fews/product/51#documentation

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific Data, 2, 150066. https://doi.org/10.1038/sdata.2015.66

McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference on Applied Climatology (Vol. 17, pp. 179–183). American Meteorological Society Boston, MA. pdf here