Version 3.1.0 for Qgis

Documentation

FEWS NET

May 2023

FEWS NET provides tools to help mitigate or prevent humanitarian crises. GeoCLIM is a powerful user-friendly tool for climate data analysis. This manual is intended for users of the GeoCLIM plugin for QGIS.

Table of Contents

Introduction

Summary

GeoCLIM is part of a set of agroclimatic analysis tools developed by the FEWS NET/United States Geological Survey (USGS) and the Climate Hazards Center (CHC) of the University of California Santa Barbara (UCSB). GeoCLIM is designed for climatological analysis of rainfall, temperature, and evapotranspiration data. GeoCLIM provides an array of accessible analysis functions for climate analysis of agricultural development. The new version of the GeoCLIM 3.0.19 is a plugin for QGIS 3.22.10 or higher (see Appendix A for basic concepts of QGIS). These user-friendly tools can be used to:

- Analyze large quantities of climate data in raster format.
- Create visual representations of climate-data summaries.
- Calculate seasonal anomalies.
- Blend climate station values with satellite data to create improved datasets.
- Calculate seasonal trends.
- Facilitate drought analysis.
- Compare groups of years within a time series.

Using the Manual

This manual is organized into 10 chapters and presents examples and exercises to help you understand the different applications of the GeoCLIM analysis functions.

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 Management in GeoCLIM provides a review of the different data types used in GeoCLIM, data availability and more.

Chapter 4: Climatological Analysis explains how to calculate statistics, trends, and SPI, among other functions, for a set period (e.g., dekad, month, season) using GeoCLIIM functionality.

Chapter 5: Rainfall Summaries shows how to calculate seasonal totals, averages, and anomalies.

Chapter 6: Climate Composites describes seasonal analysis among a group or two groups of non-consecutive years within a time series.

Chapter 7: Contour Tool explains how to visualize spatial rainfall distribution based on contour lines.

Chapter 8: Calculate Difference in Averages shows a way of estimating trends by comparing changes in averages between two periods within a time series.

Chapter 9: BASIICS explains the process of blending station and raster data to create improved climate fields.

Chapter 10: Extract Statistics explains how to create spatial summaries of historical data for a single or a set of polygons.

Appendix A: QGIS-overview provides links to basic QGIS instructions for viewing, editing, and creating shapefiles and rasters.

Download installation manual from here
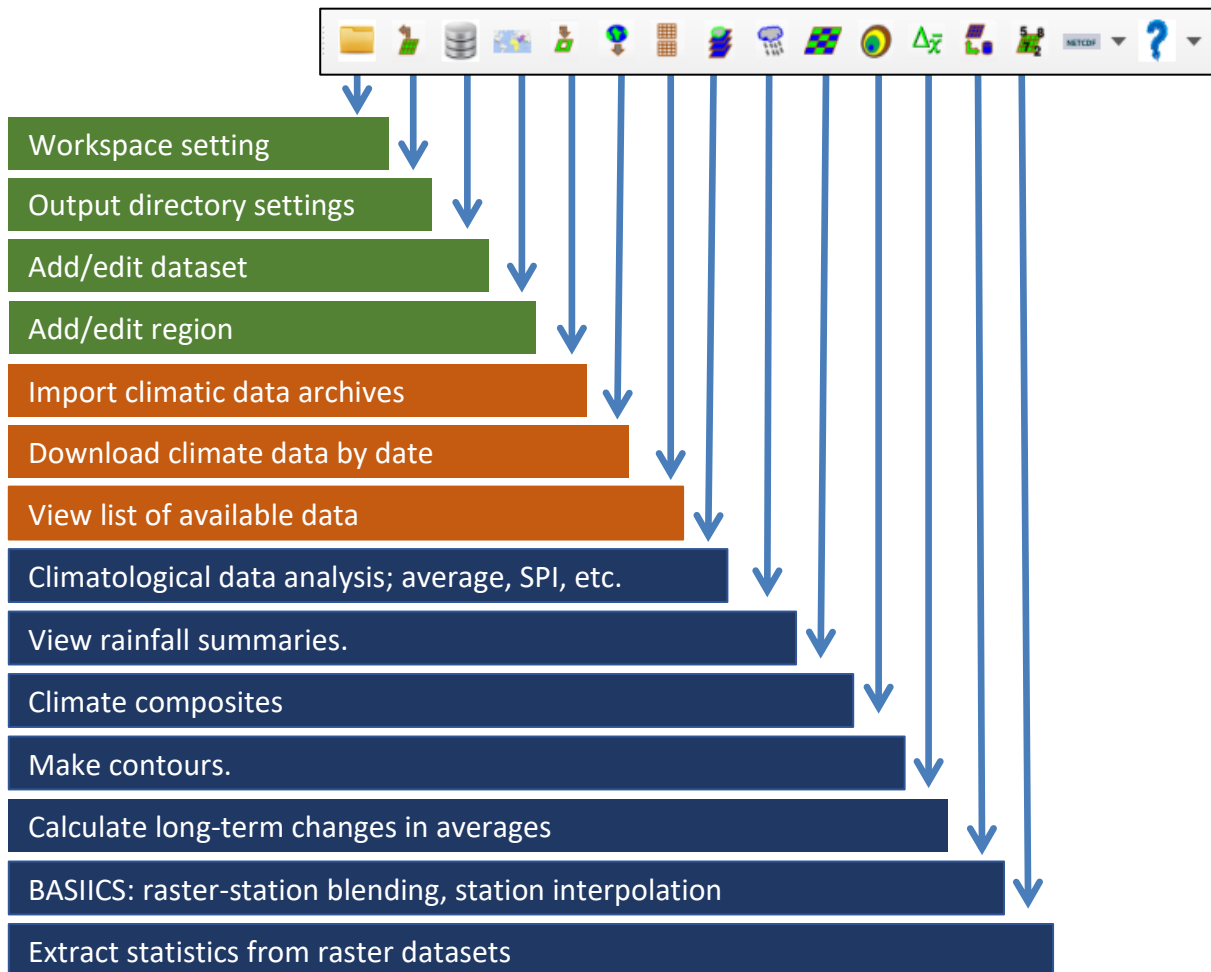
For GeoCLIM updates and video tutorials, go to https://chc.ucsb.edu/tools/geoclim

Glossary

1. **CHIRPS** – Climate Hazards InfraRed Precipitation with Stations
2. **CHIRP** – Climate Hazards InfraRed Precipitation
3. **Dekad** – 10 days total precipitation
4. **IRE** – Improved rainfall Estimates, a blend of station with raster data using GeoCLIM
5. **Pentada** – Five days total precipitation
6. **SPI** – Standardized Precipitation Index

Chapter 1: Overview
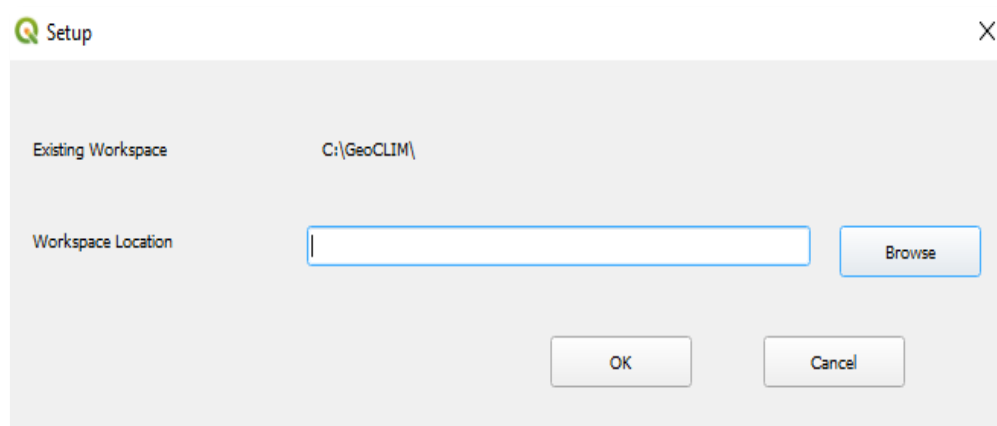
Figure 1-0 *The GeoCLIM main toolbar. Setting functions (green), data management functions (orange) and analysis functions (blue).*

Summary

Figure 1-0 above 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 tool in detail.
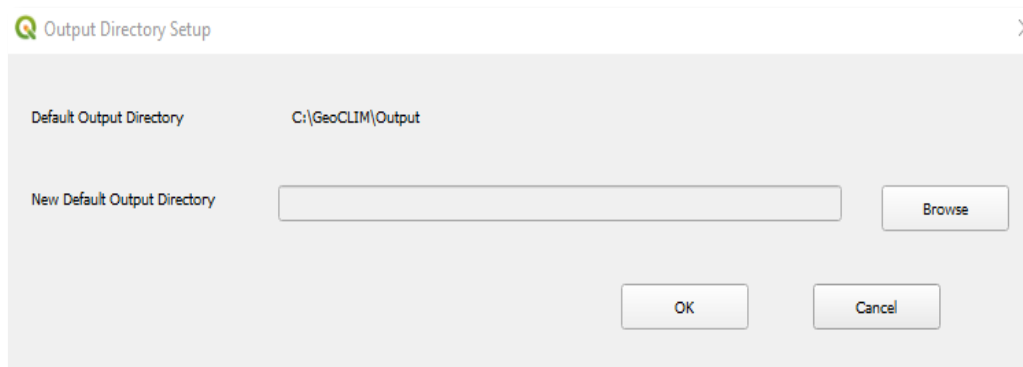
1.1. Setup workspace

Figure 1-1 Setup allows you to change the workspace directory.

The default location of the workspace is `C:\Users\[USER]\Documents\GeoCLIM`. However, you can change the location to a different drive based on memory availability and access, see Figure 1-1. See more about how to change the workspace on chapter 2.

1.2. Output Options

Figure 1-2 You can change the default output directory.

You can change the default output directory where the results from the different functions are deposited. Browse to the new directory and click the OK button, see Figure 1-2.

1.3. Add/Edit Datasets

Figure 1-3 *GeoCLIM requires datasets to be registered before they could be used by the different functions.*

GeoCLIM requires that raster data be registered in the database to be used by the different functions. The registration is done by completing the ***Add/Edit Datasets*** form, Figure 1-3. The tool includes a few datasets by default, but you could add others. See chapter 2 for more information on how to add new datasets.
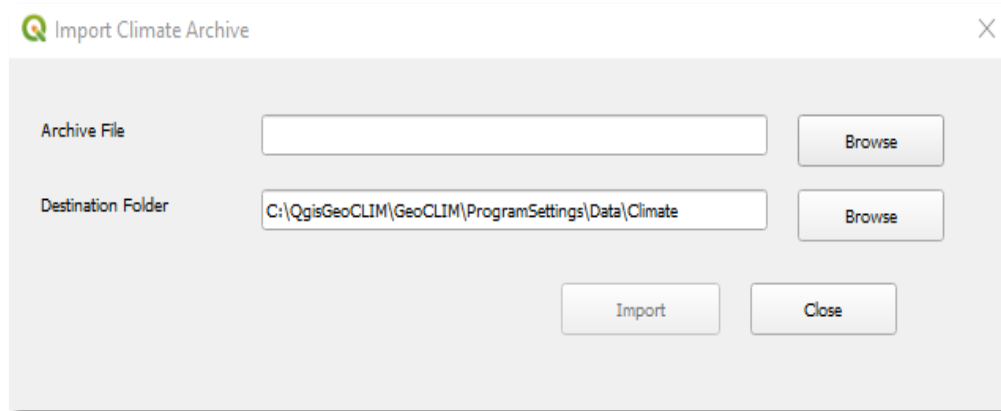
1.4. Add/Edit Regions

Figure 1-4 A region is defined by the lat/lon box and a base mask.

GeoCLIM runs on predefined areas of study or regions. A region is composed of lat/lon box and a mask which is a raster dataset that further defines the region of interest (Figure 1-4). See more about how to create or edit regions on chapter 2.
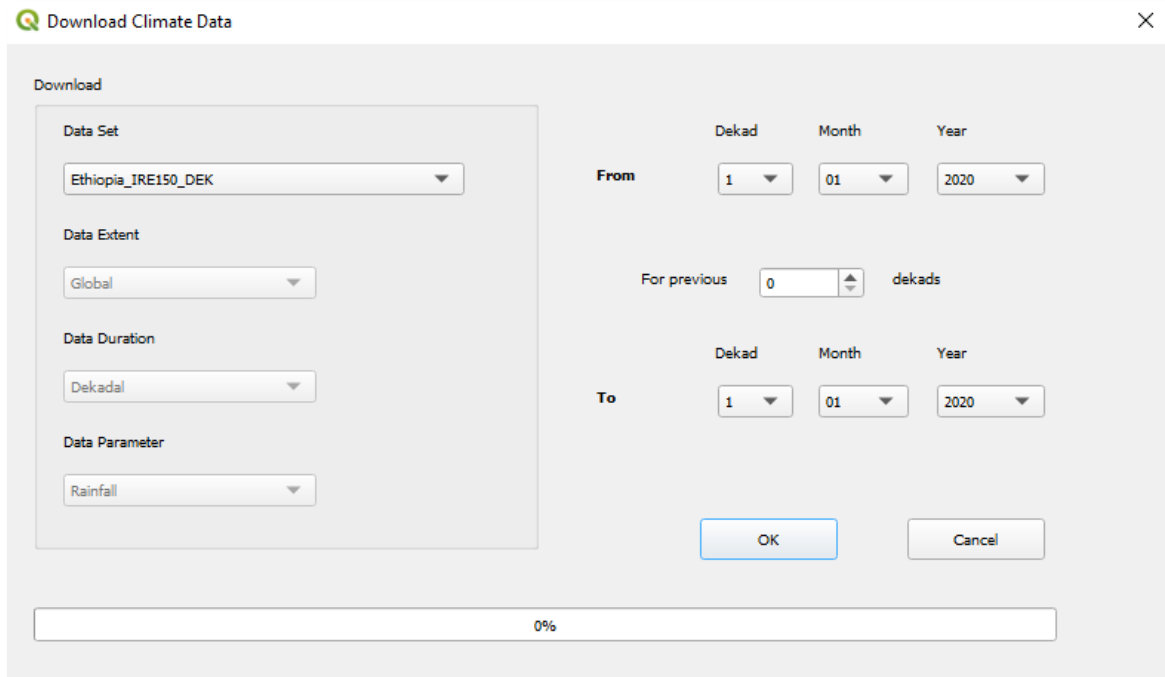
1.5. Import GeoCLIM Climate Archives

Figure 1-5 *One way of making Climate data available in GeoCLIM is through importing archives.*

A GeoCLIM archive is a compressed file containing data for a climate variable including all the settings to make the dataset ready to be used in the program. The ***Import Climate Archives*** tool, Figure 1-5, makes datasets available in GeoCLIM. These archive files are useful for sharing data among GeoCLIM users. For information on creating data archives, see Chapter 3.

1.6. Download Climate Data
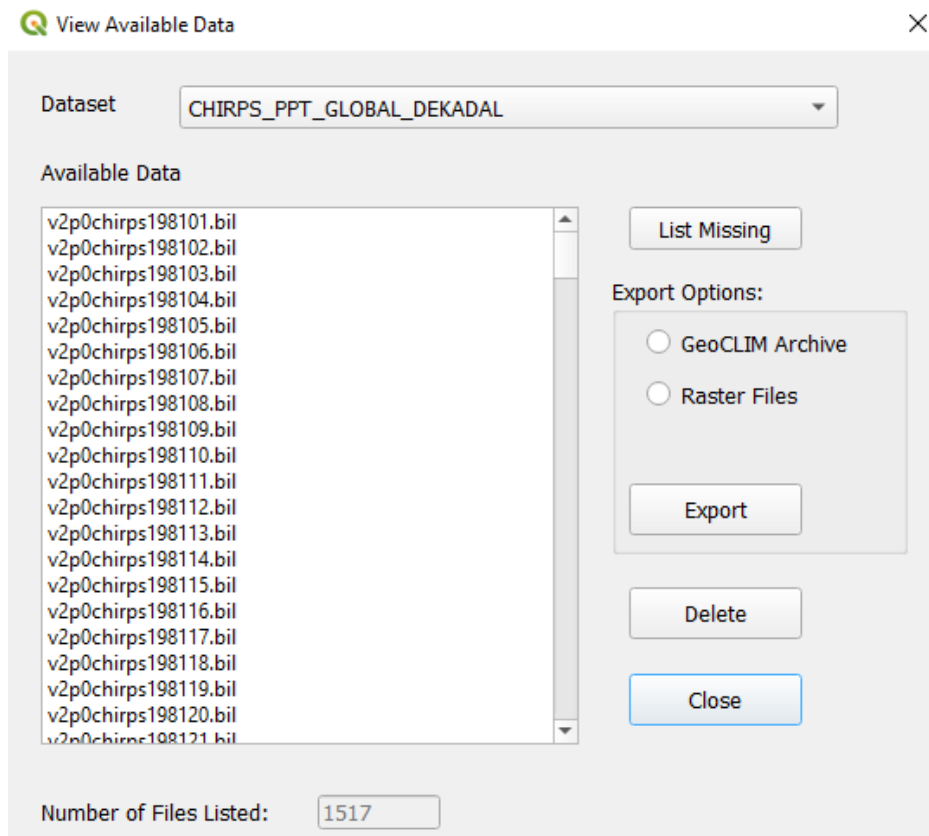
Figure 1-6 *You can download Rainfall, Temperature, or Evapotranspiration data directly from an online location such as an FTP site using the Download Climatic Data Tool.*

The ***Download Climate Data*** tool, Figure 1-6, facilitates bulk downloads of available climate data via FTP, HTTP, or HTTPS from different sources (e.g., UCSB, USGS, etc.). See chapter 3 for more information on GeoCLIM data management.

1.7. View Available Data

The **View Available Data** tool, Figure 1-7, provides a list of the data files available for the climate dataset selected (rainfall, mean temperature, minimum temperature, maximum temperature, or potential evapotranspiration). Figure 1-7 shows an example of a list of dekadal
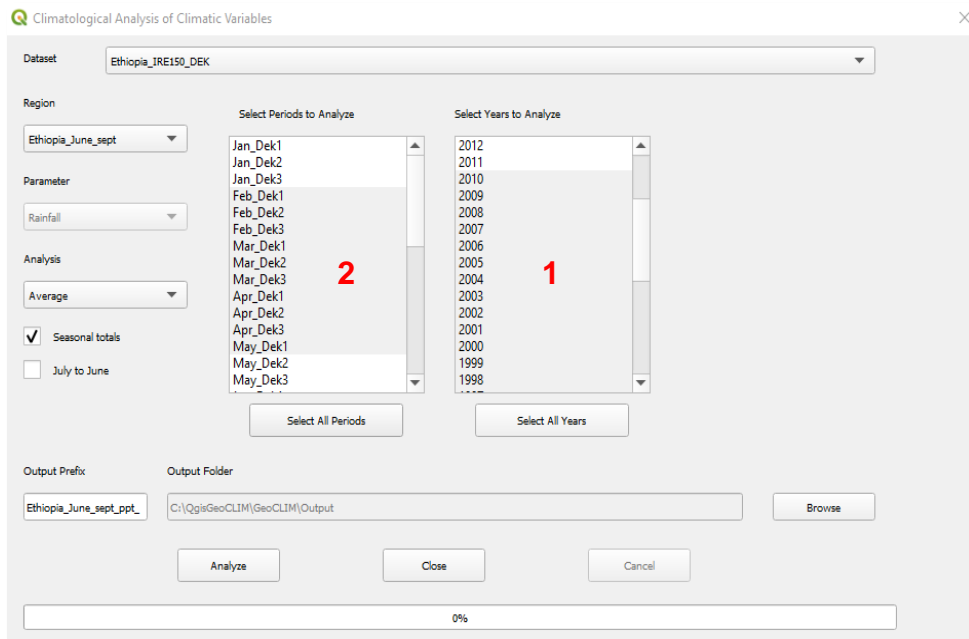
Figure 1-7 *This function allows you to list the available data for each of the climate variables.*

(+- 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. You can 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 3 for more information.
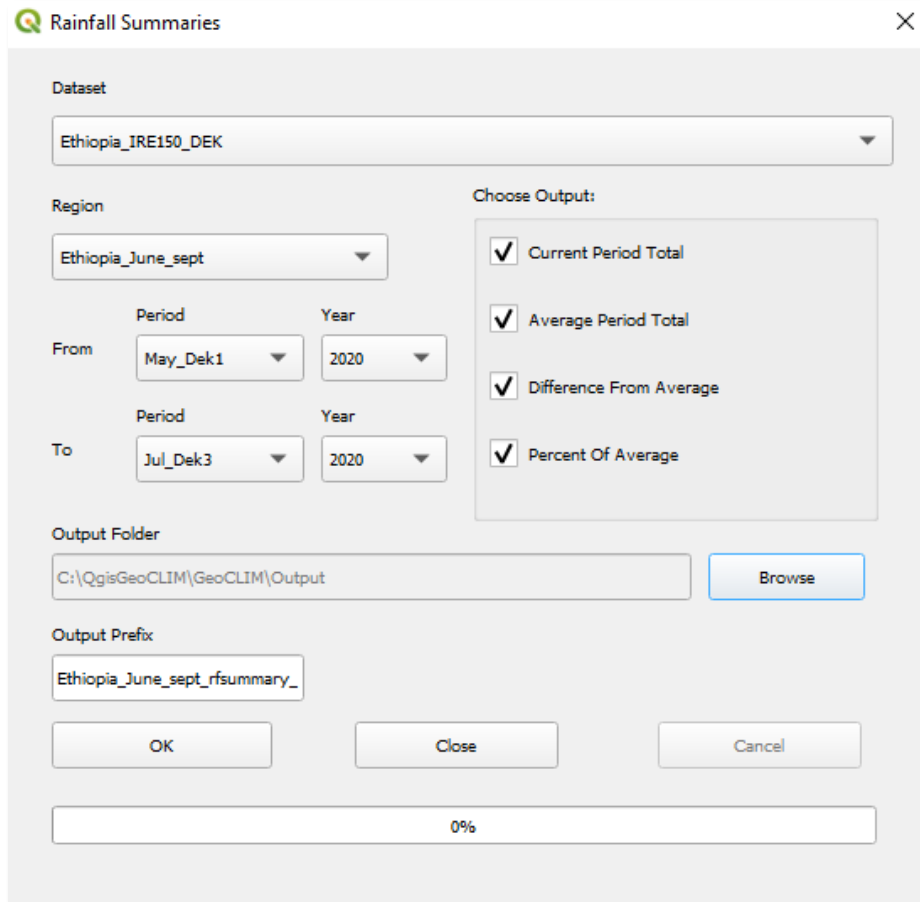
1.8. Rainfall Climatological Analysis

Figure 1-8 This tool facilitates the calculation of statistics, trends and standardized precipitation index among other functions.

The ***Climatological Analysis of Climatic Variables*** tool, Figure 1-8, is designed to calculate and display the averages, trends, SPI, and other 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 4 for a more in-depth discussion of this tool).
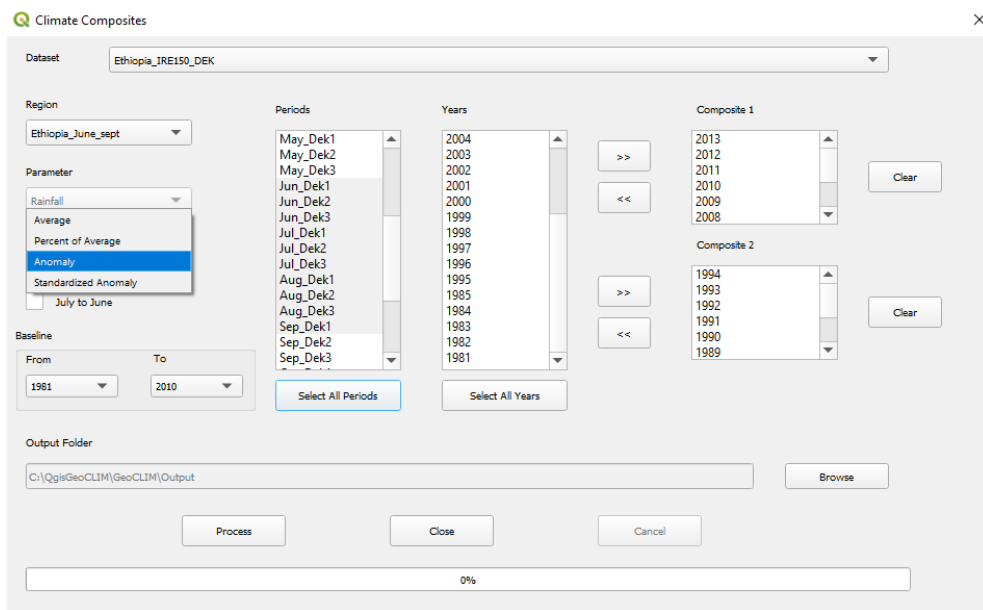
1.9. Rainfall Summaries 🌧️

Figure 1-9 *The rainfall summaries tool calculates rainfall total and anomalies for a selected period.*

The ***Rainfall Summaries*** tool (Figure 1-9) 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 5.
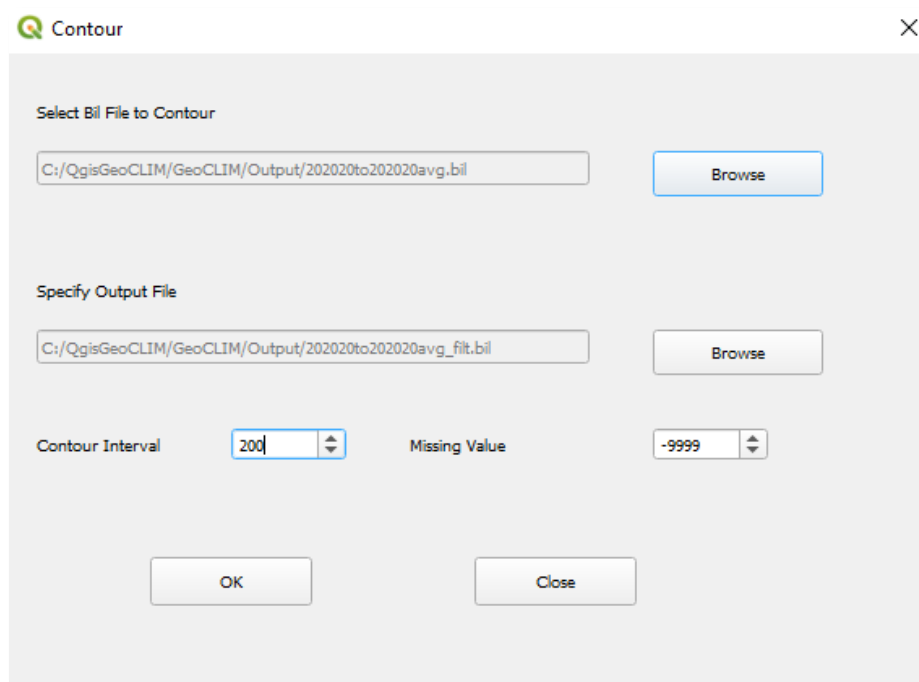
1.10. Climate Composites

Figure 1-20 *The Climate Composites tool facilitates the seasonal analysis for one or between two groups of [could be non-consecutive] years.*

The *Climate Composites* tool facilitates the seasonal analysis for one or two groups of years. 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 1-20). See chapter 6 for more details.
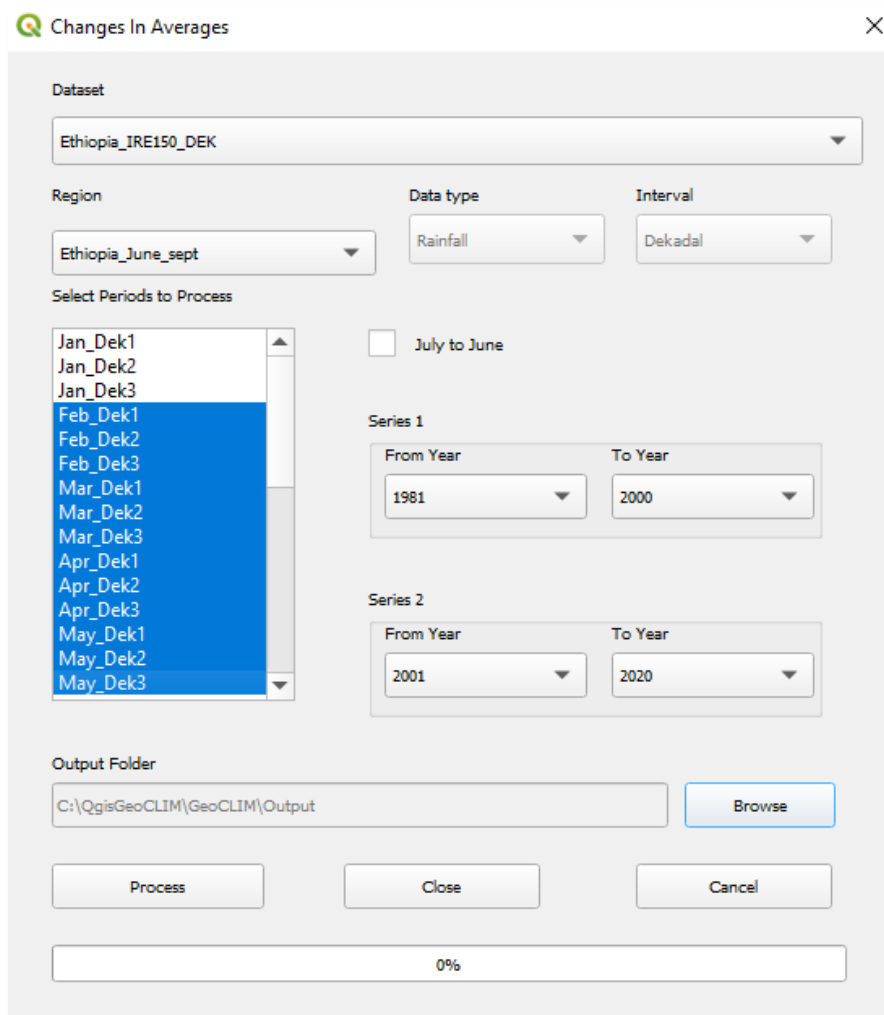
1.11. Make Contours

Figure 1-11 *Display rainfall data based on contour intervals.*

The *Make Gridded Contours* tool (Figure 1-11) displays smoothed contours for a specified interval based on a raster (`*.bil or *.tif, *tiff`) file. This tool helps identify changes in rainfall patterns within a region of interest. Read more about making contours in chapter 7.
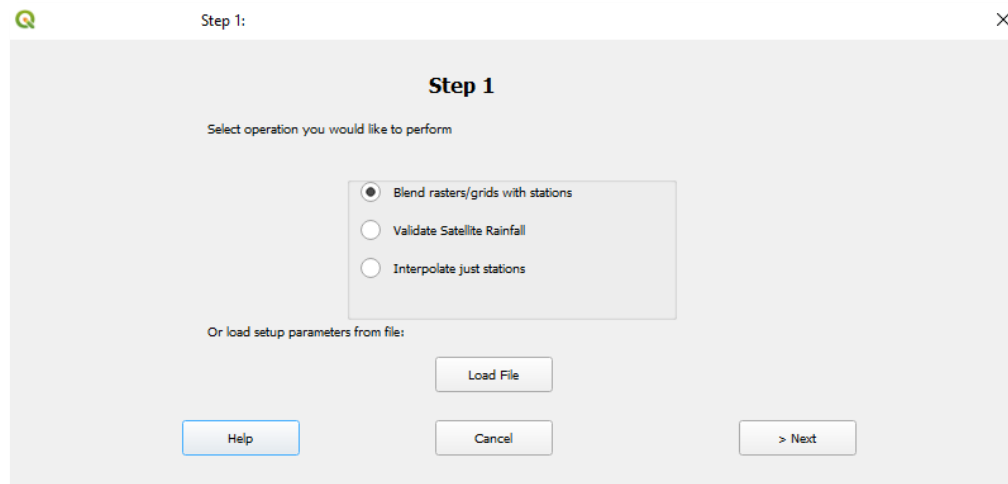
1.12. Climate trends - Changes in Average $\Delta\bar{x}$

Figure 1-12 The Climate Trends tool compares the average rainfall for two periods of time, identifying trends.

The ***Climate Trends - Changes in Averages*** tool (Figure 1-12) identifies trends by calculating the difference in average between two periods (denoted as **Series 1** and **Series 2**). See chapter 8 for more details.

1.13. Batch assistant for station/raster blending and analysis
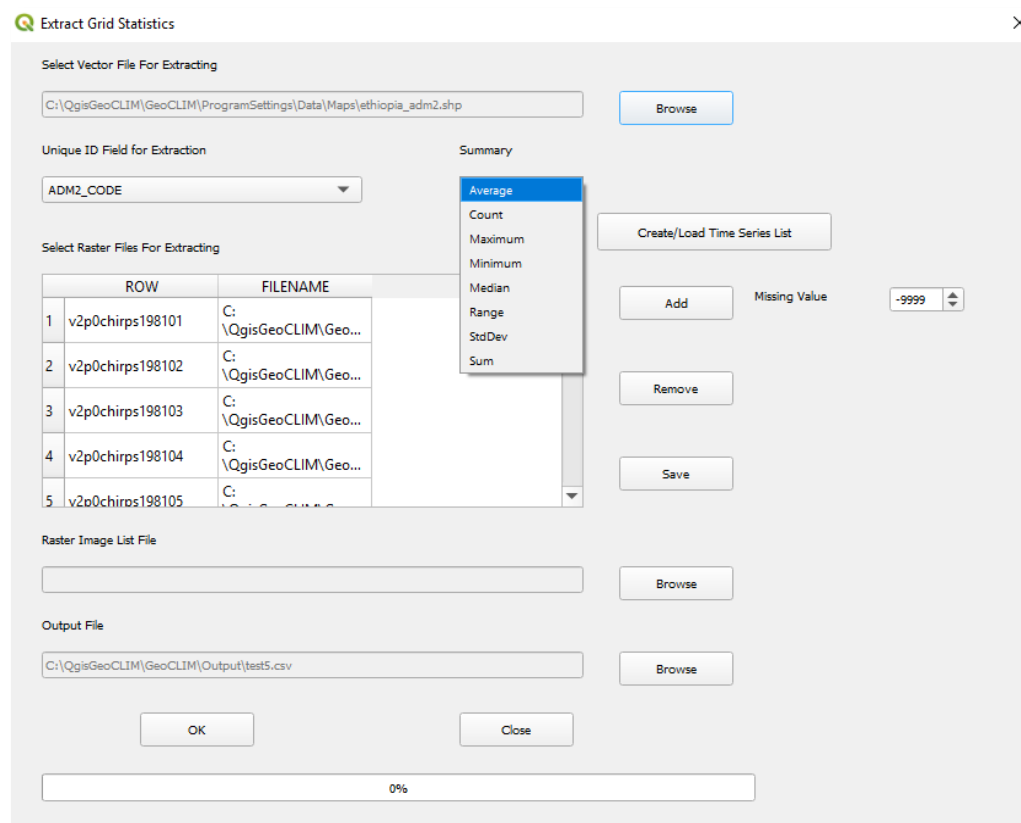
Figure 1-13 *The Batch Assistant tool has functions to: validate satellite estimated data using station values, blend station data with raster (BASIICS) and to interpolate station data.*

The Batch assistant tool (Figure 1-13) 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.
3. ***Interpolate just stations***: This function uses inverse distance weighting (IDW) process to interpolate station values. See chapter 9 for more information.
4. The function allows you to save the settings to run frequent processes, such as updating the time series.

1.14. Extract Statistics from Raster Datasets

Figure 1-14 *Extract summary spatial statistics such as average, count, maxi/min, etc., for each polygon.*

The ***Extract Grid Statistics*** tool summarizes raster information by zones defined in a vector file (polygons such as districts, provinces, watersheds, etc.). This tool produces a table with one of the summary statistics by polygon (sum, maximum, minimum, range, or standard deviation). This extraction can be applied to a single or a set of climate raster files defined in the time series option (Figure 1-14). For more on this tool, see chapter 10.

Chapter 2: Settings
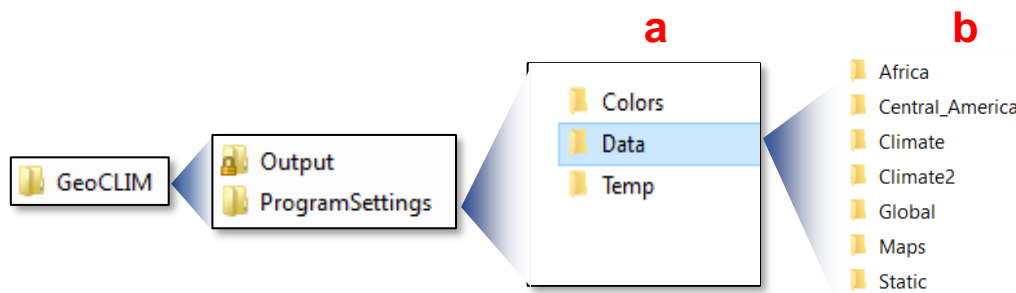
Summary

There are a few settings you must do before you start working with GeoCLIM. You must decide where you want to setup the workspace, this is a directory that will hold all the map related files, and all the raster data that you download or create. Also, you must set the area of work and the dataset that you are going to use to conduct your analysis. This chapter provides detailed description on how to setup GeoCLIM including the following topics:

- GeoCLIM directory structure.
- How to change the workspace.
- How to change the default output directory.
- How to add/edit a new climate dataset.
- How to add/edit a new region.

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 results by default. The `ProgramSettings` directory contains the colors used for the output maps, the Data directory that contains climate data, shapefiles and masks, and a Temp directory where temporary files are saved. Figure 2-1 shows an outline of the contents of `ProgramSettings`.

Figure 2-1 *The GeoCLIM directory contains two folders: the 'output' where all the results are saved by default, and the ProgramSettings that contains the 'Data' directory among others.*

> **NOTE**: The default path to climate data within GeoCLIM workspace is:
> `C:\Users\<USER>\Documents\GeoCLIM\ProgramSettings\Data\Climate.`

Contents of the `ProgramSettings` directory:

- Colors: Contains color files for map legends produced by GeoCLIM.
- Data: (Figure 2-1**(b)**) contains the following directories:
  - *Climate* – Stores all downloaded and imported data. See <u>section 2.4</u> 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.
- Temp: This directory stores temporary files.

> **NOTE:** Another important path to keep in mind is that of the `fews_tools` directory that contains all the program files. Make sure you back up the geoclim.sqlite file that is inside the `fews_tools` directory.
> `c:\Users\[users]\AppData\Roaming\QGIS\QGIS3\profiles\default\python\plugins\fews_tools\`

2.2. Changing the location of the GeoCLIM workspace

The default workspace is on the `C:\Users\<USER>\Documents\GeoCLIM`, the `C:\` drive is, sometimes, too small to hold all the data outputs that GeoCLIM produces. Also, if you are planning to work a lot with GeoCLIM, it is easier if the workspace is on the root directory. So, it is recommended that you change the workspace to another drive. Once the workspace is changed to a different directory, the data and data-related files are moved to the new directory while the program files are kept on the original `c:\users\[users]\AppData\Roaming\QGIS\QGIS3\profiles\default\python\plugins\fews_tools\` directory, mentioned above.

To change the workspace, follow the steps below:

1. From the GeoCLIM toolbar, select the ***Initial Setup/Workspace Settings*** icon, see red box in Figure 2-2.

2. Browse to the new location in the ***Workspace Location*** field.
3. Click OK.

Figure 2-2 *To change the workspace, select the workspace setting icon, red box, and define the new path.*

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

2.3. Changing the default output directory

Sometimes it is necessary to separate the results depending on the project you are working on. For this GeoCLIM allows you to select the default output directory. Click on the icon in the red box in Figure 2-3 and browse to the new directory.

Figure 2-3 *GeoCLIM allows you to select the default output directory.*

2.4. Making new data available for GeoCLIM

GeoCLIM requires climate datasets in raster (*.bil or *.tif, *.tiff) format. There are four ways of adding climate data to GeoCLIM:

1) By downloading existing data such as CHIRPS, CHIRP or any other datasets that comply with the format that GeoCLIM requires.

2) By creating your own data by blending stations values with any raster datasets.

3) By interpolating climate stations. See chapter 9 for instructions on how to blend stations values with a raster dataset or how interpolate stations values only.

4) By importing a data archive, see chapter 3 on how to import archives.

Raster datasets in GeoCLIM are managed by the ***Add/Edit Dataset*** function, see red box in Figure 2-4.

Figure 2-4 *To define a new dataset, complete the Add/Eddit Dataset form.*

If you select methods 1), 2) or 3), mentioned above, to add data to GeoCLIM, the dataset must first be defined in the GeoCLIM in order to work with the different functions. Method 4) adds the data definition automatically see Import archives.

2.4.1. Define a new dataset in The GeoCLIM

The definition of a climate dataset includes the following:

- The location and name of the folder containing the data
- The name-format of the raster files
- The missing-data value
- Where applicable, the FTP information for the dataset updates.

To define a climate dataset, follow these steps:

1. Create a new sub-directory in the GeoCLIM data repository in **X:**\GeoCLIM\ ProgramSettings\Data\Climate\, (where **X:**\ is your drive, i.e., C:\ or D:\) and copy the raster data (tif, tiff or *.bil and *.hdr files) if you have them. Otherwise, if you are planning to download the data, we'll show you how to add the FTP information. Let's add the CHIRPS global dekadal dataset, as an example. To do this, create a new directory in the GeoCLIM data repository, called CHIRPS_PPT_GLOBAL_DEKADAL. The CHIRPS files are in the following ftp site *.bil https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_dekad/bils/. *.tif https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_dekad/tifs/ Visit the site to get familiar with the data format.
2. Open the dataset definition form by clicking on the ***Add/Edit Dataset*** icon, in the red box in Figure 2-4. If this is the first time you open the form, the information you see is for a default dataset. You could either copy the current dataset, by clicking on the Copy Dataset button, and modify the existing information to reflect the dataset that you want to define or click the New button to start entering the information for the new dataset.
3. Complete the Dataset Name, for this example, CHIRPS_PPT_GLOBAL_DEKADAL, (Figure2-5**(1)**).
4. Default Dataset? Select Yes to ensure that all the functions will have this dataset selected by default, (Figure2-5**(2)**).
5. Data type, in this case is Rainfall. (Figure2-5**(3)**).
6. Dataset Extent, select global, (Figure2-5**(4)**).

7. Periodicity, select Dekadal, (Figure2-5**(5)**).
8. In the ***Current Data*** section (blue box in Figure 2.5), browse to the ***Data*** folder that you created on step 1, (Figure2-5**(6)**).
9. Define climate data file name as follows:

   The file name in a climate dataset must have 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; (Figure2-5**(7)**).

   Date format V   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`). The date formats available depend on the periodicity selected on (7) above. Try changing the periodicity and see how the date formats also change. See Figure2-5**(8)**.

   Suffix V   The suffix corresponds to any character after the date, including the extension of the file (e.g., `.bil`) (Figure 2.5 **(9)**).

   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 V 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`).
10. Fill out the Missing Value; for example, the missing value in CHIRPS is -9999 (Figure2-5**(10)**).
11. In the ***Average Data*** section, orange box in Figure 2.5, Click on the Match Data Folder button (Figure2-5**(11)**); this copies the path from the **Data folder** onto the **Average Data folder** to ensure that the long-term averages are saved into the same directory as the dataset files. This step becomes handy when Updating the GeoCLIM averages.
12. Fill out the prefix, date-format V, and suffix (`.bi, .tif, or .tiff`) for the average files (Figure2-5**(12)**).
13. The missing value should be the same as defined for the current data, Figure2-5**(10)**.
14. The **FTP Settings** section, yellow box in Figure 2.5, contains the necessary information to download data updates. If the dataset that you are defining does not have FTP information, this section can be empty (Figure2-5**(13)**). To setup the FTP for global CHIRPS dekads https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_dekad/bils/, follow the steps below:
    a) Add the name of the remote host https://data.chc.ucsb.edu/. (Figure2-5**(13a)**).
    b) Add the remote directory where the data are located /products/CHIRPS-2.0/global_dekad/bils/. (Figure2-5**(13b)**).
    c) Add "anonymous" as the Username. (Figure2-5**(13c)**).
    d) Add "password" as the Password. (Figure2-5**(13d)**).

e) Enter the date format for v2p0chirps198101.tar.gz. in this case 4-digits year 2-digits dekad. (Figure2-5(**13e**)).

f) Enter the file prefix, in this case "v2p0chirps", and suffix ".tar.gz" (Figure2-5(**13f**)).

g) The Uncompressed Data Projection is "geographic." (Figure2-5(**13g**)).

h) The scale factor (Figure2-5(**13h**)).

i) The Uncompressed Date format in our example is 4-digits year 2-digits dekad. (Figure2-5(**13i**).

j) Enter the uncompressed prefix and suffix, "v2p0chirps" and ".bil" respectively. (Figure2-5(**13j**).

15. Once the form is completed click on the Save button. See chapter 3 for how to download data.

The complete ***Define GeoCLIM climate Dataset*** form for `*.bil` data should look like Figure 2-5. Figure 2-5-a shows the form for data in `*.tif` or `*.tiff` format.

Figure 2-5 *All datasets must be defined before they could be used in the GeoCLIM.*

Figure 2-5-a *This is an example of the complete form for the definition of data in *.tif format to be used in GeoCLIM.*

> **NOTE**: To learn more about the data formats used in GeoCLIM, see chapter 3.

2.5. Regions

GeoCLIM works on specific areas of interest called Regions. The definition of a Region includes the maximum/minimum latitude and longitude, a default mask, and a set of vector maps that delineate the Region. There is a set of predefined regions in GeoCLIM. If the desired region is not in the list, you can define a new one.
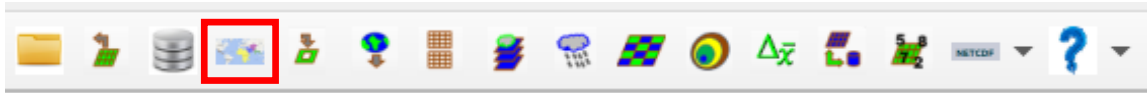
To define your own region, you must have the following information ready:

- Maximum and minimum latitude for your area of interest.
- Maximum and minimum longitude for your area of interest.
- A mask file, a raster file with value of 1 for the area of interest and "0" for the areas outside (see how to create a mask on section 2.5.1).
- A polygon shapefile that delineates the area of work (i.e., watershed boundaries, administrative unit, country, etc.)
- At least one polygon shapefile to overlay the results.

2.5.1. Create a new region in GeoCLIM

To define a new region in GeoCLIM, follow the steps below:

1. Open the ***Define Region*** form by clicking on the ***Add/Edit Region*** button, see red box in Figure 2-6 below.
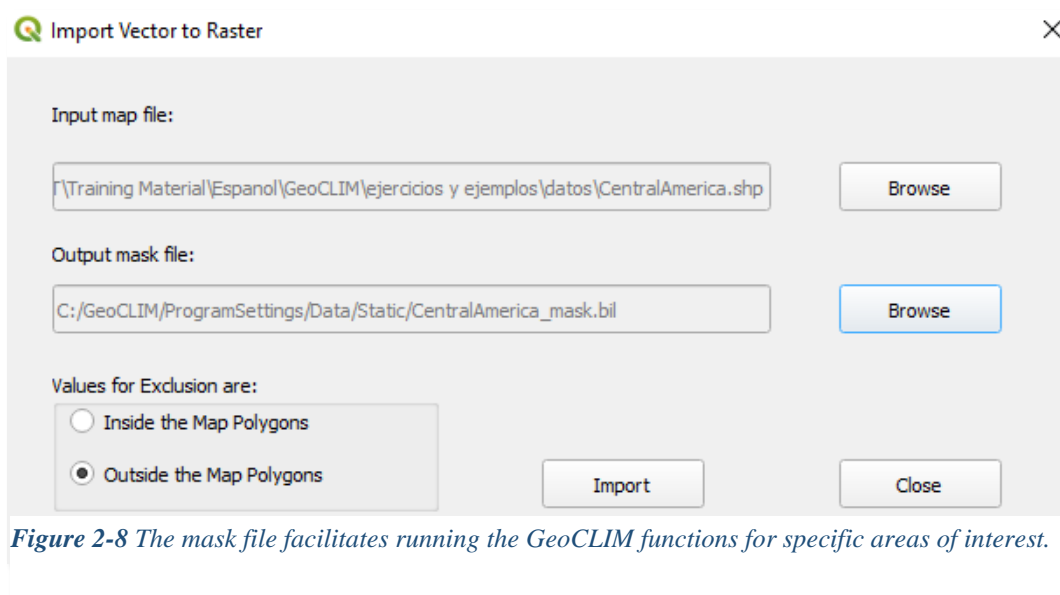
Figure 2-6 *Select the Add/Edit Region icon to define a new region.*

2. Once the form opens, it shows the information for the default region. You could either copy the region by clicking on the Copy Region button and edit the fields with the new information, and save it as the new region, or click on the New button. For example, let's create the region for Central America and the Caribbean including the northern part of South America, see area in black in Figure 2-6 below. To do this step, a shapefile was prepared covering only the area of interest.

3. Region Name: Add the name for the new region and click ok.

4. Set as Default? Select yes to ensure that this region is selected when you open any function.

5. Comments Add any information to help describe the region.

6. Enter the minimum and maximum latitude and longitude for the new region. In our example, we are using the lat/lon for Central America and the Caribbean. 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. For our example we select the shapefile for the region of interest. The tool retrieves the coordinates automatically.

7. Cell Size: Enter the output cell size, in the case of CHIRPS it is 0.05 decimal degrees.

8. Mask File: Browse to `X:\GeoCLIM\ProgramSettings\Data\Static` to select the mask file. If you do not have a mask file, you can create one at this point. Masks are raster images that are used to include only the desired area of interest 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 it. The mask facilitates the execution of the algorithms on the areas where the pixel value=1 while excluding areas where pixel values=0. You can create a mask using a shapefile of the area of interest (i.e., administration unit, watershed, etc.) following the steps below:
    a) Click on the Import Mask from Vector button to open a new form.
    b) Browse to select the shapefile for the area of interest, in our example the shapefile is CentralAmerica.shp, see Figure 2-7.

Figure 2-7 *The mask defines the area of interest, the black zone shows the mask for Central America, the Caribbean and the northern part of South America..*

c) Browse to save the output mask in the `~\GeoCLIM\ProgramSettings\Data\Static` directory. The program gives the name automatically, see Figure 2-8. Make sure that the Outside the Map Polygon is selected.

d) Click Import to create the mask.

Figure 2-8 *The mask file facilitates running the GeoCLIM functions for specific areas of interest.*

9. Map File : select a vector layer in shapefile format depicting polygons related to the region of interest, such as political boundaries, watersheds, etc. The map file is overlaid on the product. A second map file is optional.

10. Click on Save to keep your new region

You are now ready to work. See Figure 2-8 for a complete *Define Region* form and Figure 2-9 for an example of the result (average rainfall) using this region. The image shows the Central America and Caribbean region (lat/lon), the analysis result is limited by the mask.

Figure 2-6 *GeoCLIM runs on specific areas of interest defined by lat/lon and a mask file, called Regions.*

NOTE: Use the Import button, on the data *Definition form*, to import regions used on the Visual Basic version of the GeoCLIM.

Global data: CHIRPS data are produced in two versions, a preliminary version that is available every five days, two days after the end of the pentad (2nd, 7th, 12th, 17th, 22nd, and 27th), and a final version that is released every month, on the middle of the following month, for example: the final data for April is available after May 15th. To conduct monitoring analyses, we combine the final available data and complement it with preliminary data. GeoCLIM provides predefined settings to download final CHIRPS for global, Africa and Central America windows, see link below:

https://data.chc.ucsb.edu/products/CHIRPS-2.0/

To download dekadal global **preliminary** CHIRPS

https://data.chc.ucsb.edu/products/CHIRPS-2.0/prelim/ select the format *.tiff, *.tif or *.bil.

To download dekadal global **final** CHIRPS

https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_dekad/ select the format *.tiff, *.tif or *.bil.

Rainfall, mm
- <=10
- 10 - 50
- 50 - 100
- 100 - 150
- 150 - 200
- 200 - 250
- 250 - 300
- 300 - 400
- 400 - 500
- 500 - 750
- 750 - 1000
- 1000 - 1250
- 1250 - 1500
- 1500 - 2000
- >2000

Figure 2-7 The mask file defines the area of work. In this example, the region and mask cover the area in color.

NOTE: The size of the region **must** be smaller or equal to the size of the climate dataset. Otherwise, the region becomes unavailable when selecting the dataset. In other words, if you do not see your region in the pulldown menu, after selecting your dataset, it may be because the geographic extent of the region is bigger of that of the dataset.

2.6. The GeoCLIM database: fews_tool.sqlite file

Figure 2-8 *You could see the content of the geoclim.sqlite file on the Database pulldown menu on the main QGIS toolbar*

The fews_tools.sqlite files found in the fews_tools directory in the following path `c:\Users\[user]\AppData\Roaming\QGIS\QGIS3\profiles\default\pyt hon\plugins\fews_tools\` holds all the changes mentioned in this chapter such as workspace changes and all the datasets and regions that you add or edit. You can see the content of the fews_tools.sqlite by clicking on the ***Database*** pulldown menu on the main QGIS toolbar. Select the ***DB manager*** and open the ***SpatialLite/ fews_tools.sqlite***. Please explore the content of the file, see Figure 2-8. Remember that all changes must be done through the setting in the GeoCLIM main toolbar. Please backup the file periodically to ensure that you have all the settings saved.**NOTE**: Sometimes the directories are hidden, just go on the file explorer path, and type the next directory. You should be able to see the next directory in the list. Continue selecting directories until you get to Fews_tools.

NOTE: For existing users of QGIS/GeoCLIM for versions before 3.1, replace the geoclim.sqlite file that you have saved into the few_stool directory after installing the new version. Restart QGIS and all the content in the Geoclim.sqlite file will be copied to the fews_tools.sqite file. After that back up only the fews_tools.sqlite file to save your settings.

Chapter 3: Data Management in GeoCLIM

Summary

This chapter examines the data management in GeoCLIM, including the different types of data formats used, downloading data, data availability, and data import/export using archives.

3.1. Data types

GeoCLIM uses four main data types: raster data in BIL format BIL (* .bil) y GeoTIF (*.tif or *.tiff), vector data in shapefile format (*.shp), tables in comma-delimited format (*.csv), and data Archives that
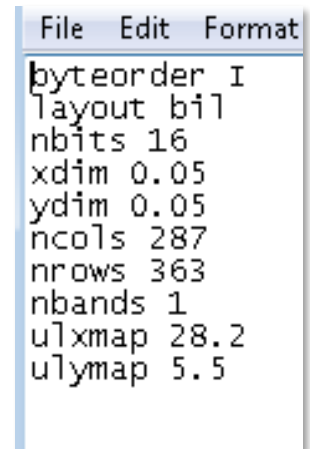
3.1.1. Characteristics of the raster dataset

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.

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

Figure 3-1 *Example of a HDR file.*

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.

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 since precipitation is only positive. However, if you look at the header file for CHIRPS, it uses `signed` values (+ or -) since the nodata value is -9999. In the example in Figure 3-1, the `pixeltype` is missing, so the program assumes that the data is unsigned; in the case of CHIRPS, we would have to add a new line on the header file defining the `pixeltype` as `signedint`. 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
```
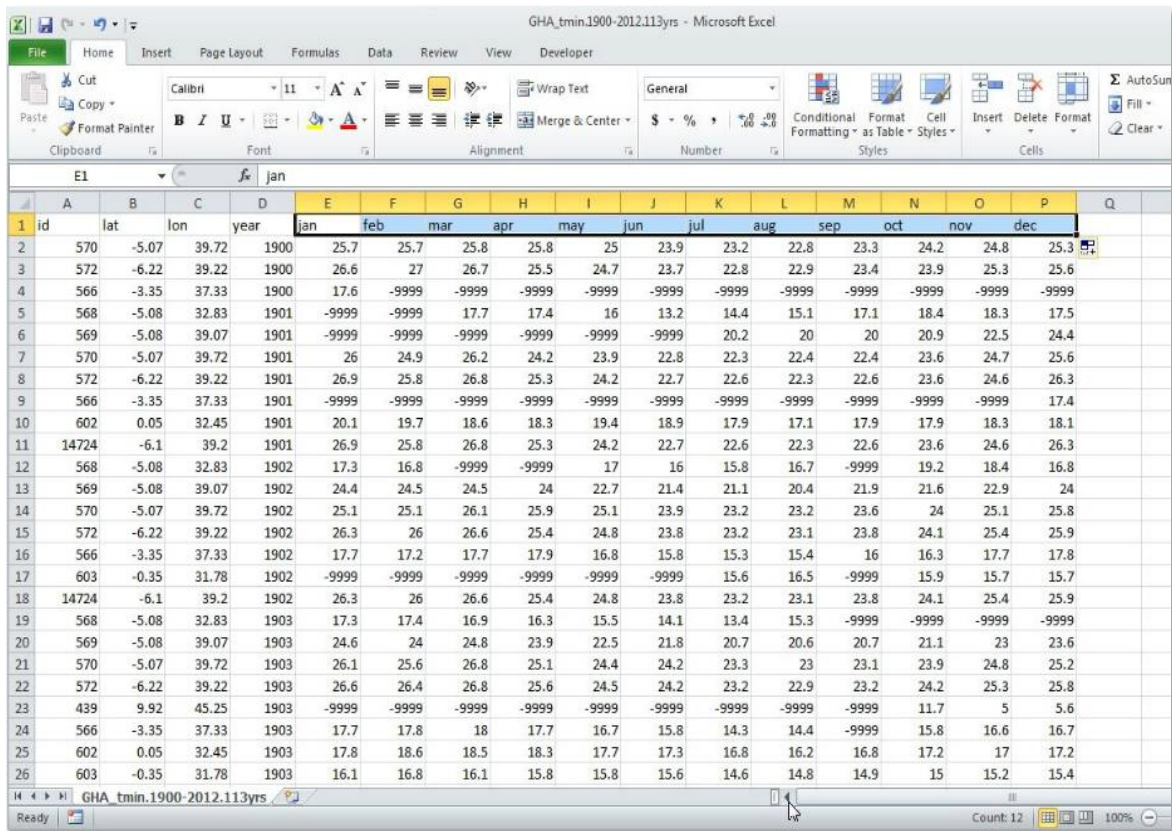
Figure 3-3 *The range of values a dataset could store depends on the nbits*

3.1.2. Vector data

Another type of data used in GeoCLIM is vector data in shapefile format (*.shp). To get more information about how to open, create, or edit shapefiles in QGIS, go to Appendix A.

3.1.3. 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, 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 order, and additional columns are permitted. However, the time-period columns need to be consecutive (Figure 3-4).

Figure 3-4 *GeoCLIM accepts tables in comma delimited format (CSV).*
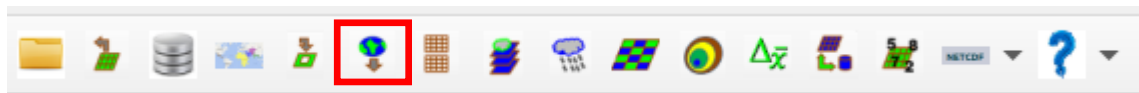
3.1.4. GeoCLIM Archives

A GeoCLIM Archive is a compressed file with the ending '.climdata' that contains a group of files of a specific climatic variable. The Archive contains the necessary information to define the data in GeoCLIM. This is an easy way to save a copy of the dataset, or part of it, that has been defined in GeoCLIM or to share data between users. See section 3.4 for instructions on how to create an Archive.

3.2. Download data

GeoCLIM allows you to download data for those datasets that include FTP information on their definition form. To download data, follow the steps below:

1. Click on the **Download Data by Date** icon on the GeoCLIM toolbar. See red box in Figure 3-5.

Figure 3-5 *To download data using GeoCLIM, make sure that the ftp information is included in the dataset definition. Click on the download icon and complete the form.*

2. Select the dataset that you want. Figure 3-6.
3. The ***Data Extent, Duration and Parameter*** should change accordingly. Figure 3-6.
4. Select the starting and ending date. Figure 3-6.
5. Click ok to start the download. Figure 3-6.

Figure 3-6 *Download data by date for a selected dataset.*

> **Global data:** CHIRPS data are produced in two versions, a preliminary version that is available every five days, two days after the end of the pentad (2nd, 7th, 12th, 17th, 22nd, and 27th), and a final version that is released every month, on the middle of the following month, for example: the final data for April is available after May 15th. To conduct monitoring analyses, we combine the final available data and complement it with preliminary data. GeoCLIM provides predefined settings to download final CHIRPS for global, Africa and Central America windows, see link below:
> https://data.chc.ucsb.edu/products/CHIRPS-2.0/
> To download dekadal global **preliminary** CHIRPS
> https://data.chc.ucsb.edu/products/CHIRPS-2.0/prelim/ select the format *.tiff, *.tif or *.bil.
> To download dekadal global **final** CHIRPS
> https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_dekad/ select the format *.tiff, *.tif or *.bil.

3.3. Data availability/export data

The *View Available Data* function, see red box in Figure 3-7 facilitates the following:

a) Display a list of files in your dataset
b) Display a list of missing files in the timeseries
c) Deletes files in the dataset.
d) Exports the dataset or part of it into a GeoCLIM archive or as NetCDF format.
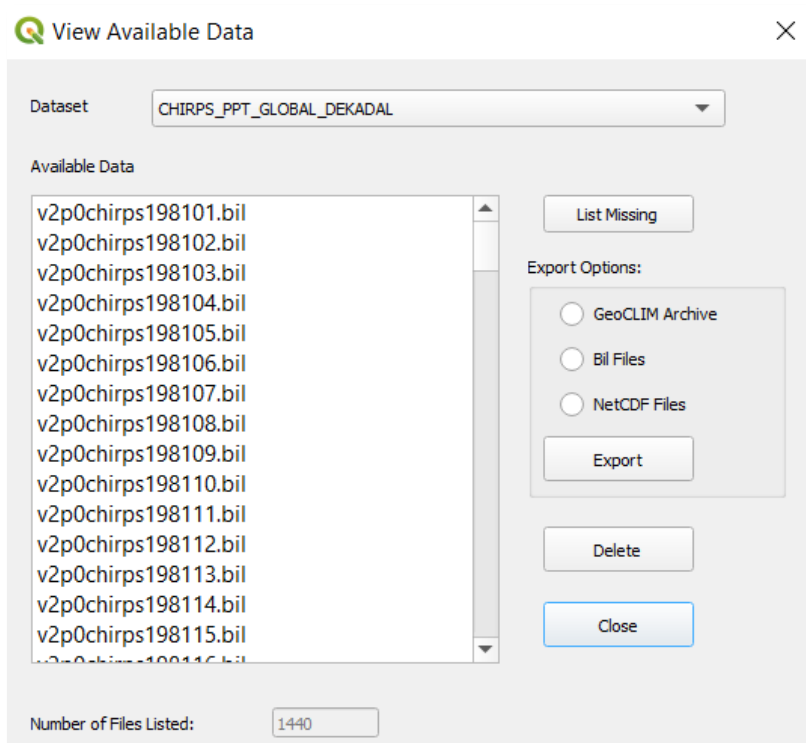
To check the available datafiles in your dataset click on the *View Available Data* icon on the GeoCLIM toolbar.

Figure 3-7 *To find the data available to run the GeoCLIM, click on the View Available Data icon and select the desired dataset.*

The function will open the information for the default dataset. Select the desired dataset from the pulldown menu. The *Available Rainfall Data* tool also allows you to identify the missing data in the time series, Figure 3-8. Click on [List Missing Data] to obtain a list of missing files in the time series. Select the complete or part of the dataset and click on the [Export] button to save the selected files into a data archive that could be shared with other GeoCLIM users. A data archive is a compressed file that includes all the files, selected on the *Available Rainfall Data* tool, together with the data-definition information.

Figure 3-8 *The View available Data tool allows to list the data files in the time-series for a selected climate dataset.*

3.4. Create an Archive

To create an archive, follow the steps below:

1. Open the ***View Available Data*** tool.
2. Select a climate variable.
3. Select the entire or part of the time-series to be added to an archive.
4. Click Export
5. Give a name to the new archive and click OK .
6. Select the destination directory.
7. Click OK . A new file with ".climdata" format is created in the destination directory.

3.5. Importing archives

To import a GeoCLIM archive, follow the steps below:

1. Open the ***Import Climate Data Archives*** tool from the toolbar. See Figure 3-9.
2. Browse to the archive to be imported and click Import.
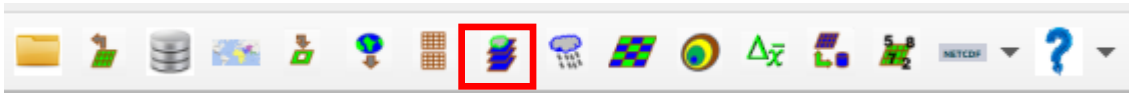
Figure 3-9 *The GeoCLIM allows you to share with other users a complete or part of a dataset using a data archive.*

To ensure that the archive imported correctly, open the ***View Available Data*** function, and click on the dataset pulldown menu and select the new dataset.

Chapter 4: Climatological Analysis

Summary

Figure 4-1 *The Climatological Analysis tool facilitates the calculation of statistics among other functions for climate variables.*

The ***Climatological Analysis*** tool (red box in Figure 4-1) facilitates the calculation of statistics, trends, and frequencies (among others) for rainfall, temperature, and evapotranspiration. The tool uses data that have already been downloaded or imported into 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 a given number of years such as 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 in 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.

4.1. Running climatological Analysis
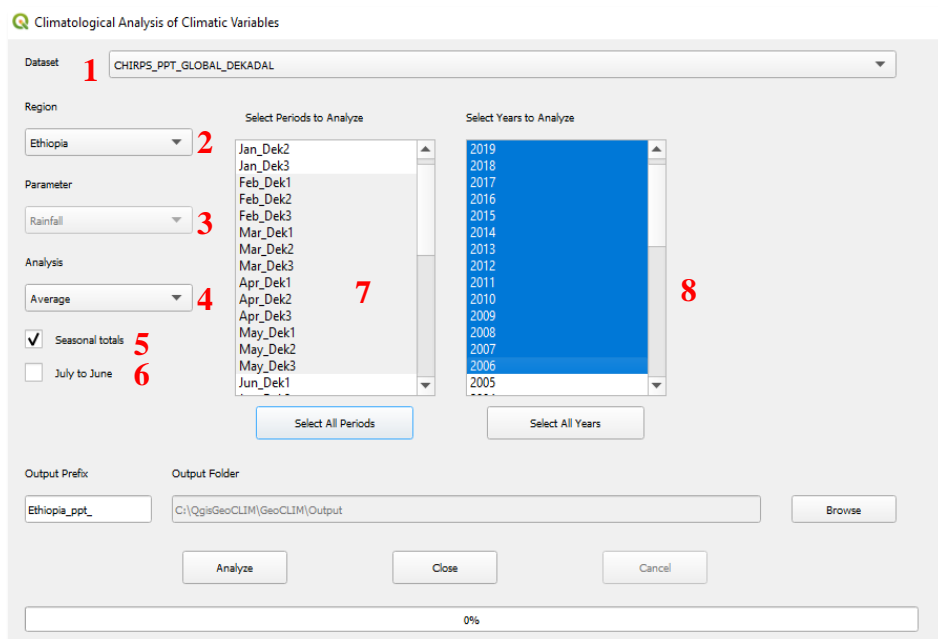
To open the climatological analysis tool, Click on the ***Rainfall Climatological Analysis*** ![icon] icon on the GeoCLIM toolbar (Figure 4-2). To use the tool, follow the steps below:

1. Select the dataset from the |Dataset ∨| pulldown menu, (Figure 4-2 **(1)**). This will make available the regions that are within the geographic domain of the dataset. If your region is not available, make sure that the lat/lon box of your region is inside the domain of the dataset.
2. Select the region of interest (Figure 4-2 **(2)**), (see chapter 2 to set up a region).

> **NOTE:** The number of regions available depend on the geographic extent of the dataset. If your region does not show up after selecting the dataset, make sure that your region is inside the geographic domain of the dataset.
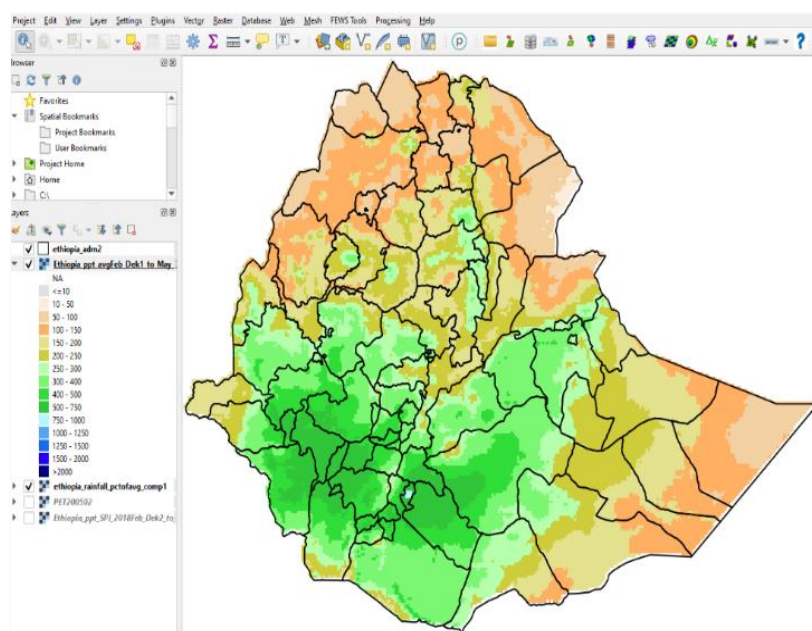
3. The Parameter field is filled automatically with the name of the climate variable (Rainfall, Avg Temperature, Min Temperature, Max Temperature, or Evapotranspiration) depending on the dataset selected, see (Figure 4-2 **(3)**).
4. Select the type of analysis from the |Analysis ∨| menu (Figure 4-2 **(4)**).
5. Check the ■ Add up seasonal totals box as shown in (Figure 4-2 **(5)**), to indicate that the analysis will be done using the seasonal totals.
6. If the season to analyze goes across years, for example, Oct to March, check the ■ July to June Sequence checkbox (Figure 4-2 **(6)**).
7. 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 10 days totals (dekads) (Figure 4-2 **(7)**).
8. Select the years of interest on the right panel (Figure 4-2 **(8)**).
9. (Optional) Modify the |Output Folder| field if you want to save outputs in a different location than the default path.
10. Modify the |Output Prefix| if necessary.

> **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 4-2 The Climatological Analysis tool facilitates the calculation of statistics, trends, SPI among other analysis, using the complete or part of the time series for a season.

The output from this analysis is displayed on the **QGIS** canvas (Figure 4-3). This result is also saved in the output folder together with the seasonal totals for each year, as raster data sets in (*.bil) format.
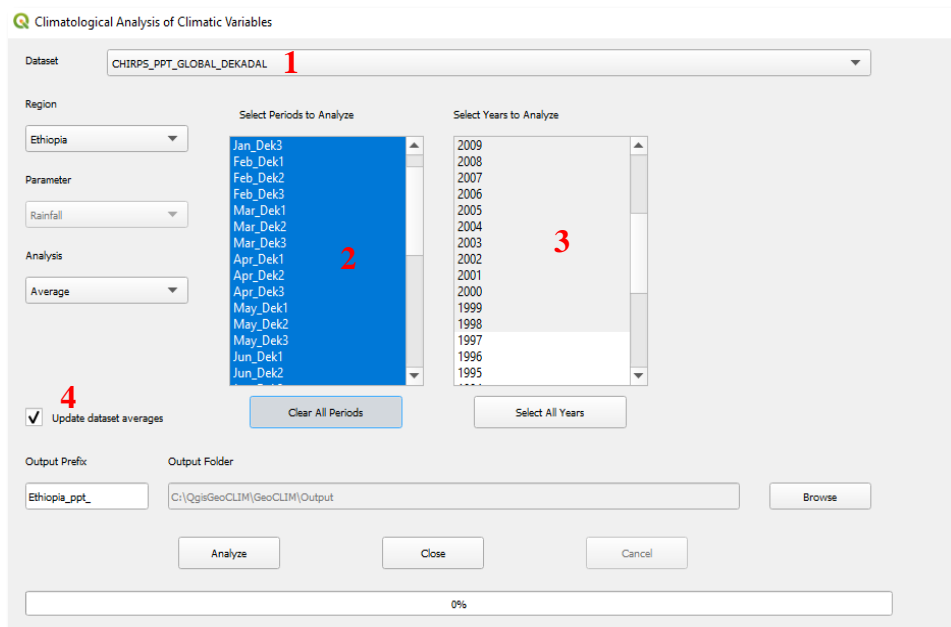
Figure 4-3 Average rainfall for the period Feb dek01-May dek03 1981-2010.

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 the results are displayed on the QGIS canvas.

4.2. Updating the GeoCLIM averages

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

1. Select the Dataset, Figure 4-4**(1)**
2. Select all the periods, Figure 4-4**(2)**
3. Select the years to be used in calculating the average, Figure 4-4**(3)**. Make sure that all the years selected have data for all the periods.
4. Check the ■ Update GeoCLIM Averages box, Figure 4-4**(4)**
5. Click on Analyze. The raster-file outputs are saved in the folder and with the prefix defined on the climate dataset definition.
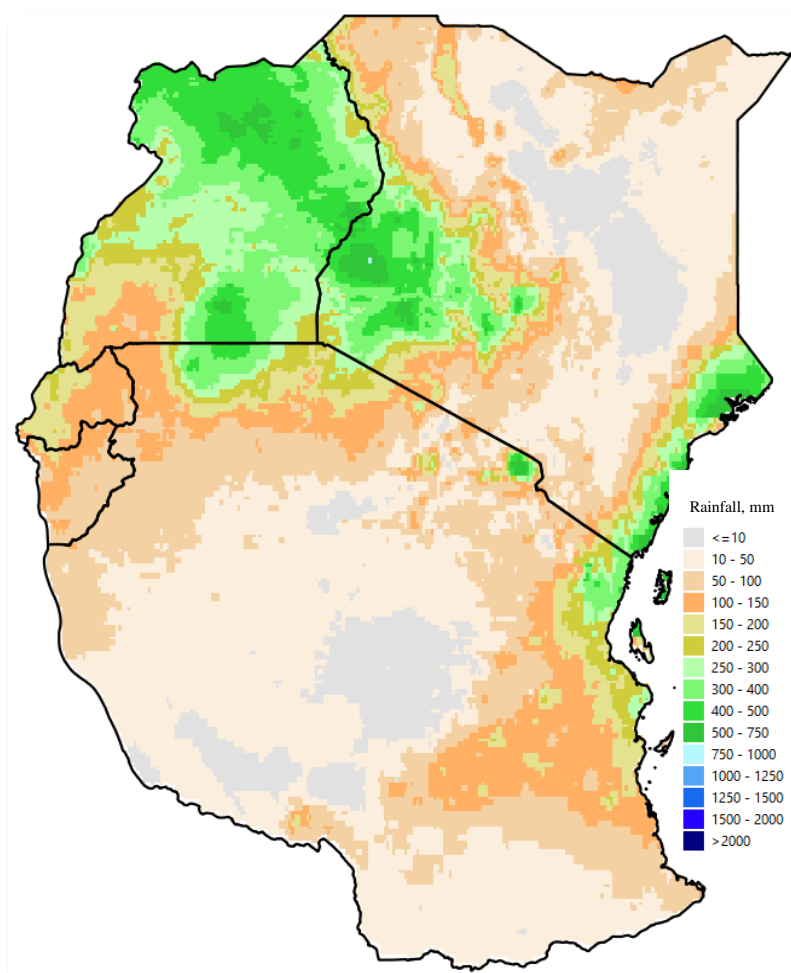
Figure 4-4 *The Climatological analysis tool allows us to calculate the average for each period (month, dekad or pentad) for the dataset.*

NOTE: The ■ Update GeoCLIM Averages option creates the average for the geographic extend of the dataset.

4.3. Analysis Methods

4.3.1. 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 4-5 shows the average rainfall (CHIRPS) 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.

Figure 4-5 Average rainfall (mm) for the May-July season for the years 1981-2013.

To calculate the average, follow the steps below:

1. Start the *Climatological Analysis* tool, as described in section 4.1
2. Select the dataset
3. Select the region
4. Select Average from the analysis methods list
5. Check the ■ Seasonal totals option
6. Select May Dek1 to July dek3 on the left panel and 1981-2013 on the right
7. Click on Analyze to run the tool.

NOTE: When the ■ 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 the average for May dekad 1; 1981-2013, May dekad 2; 1981-2013, etc., until July Dekad 3.

NOTE: A by-product of this process is a seasonal total file in raster format saved for each year in the output directory.
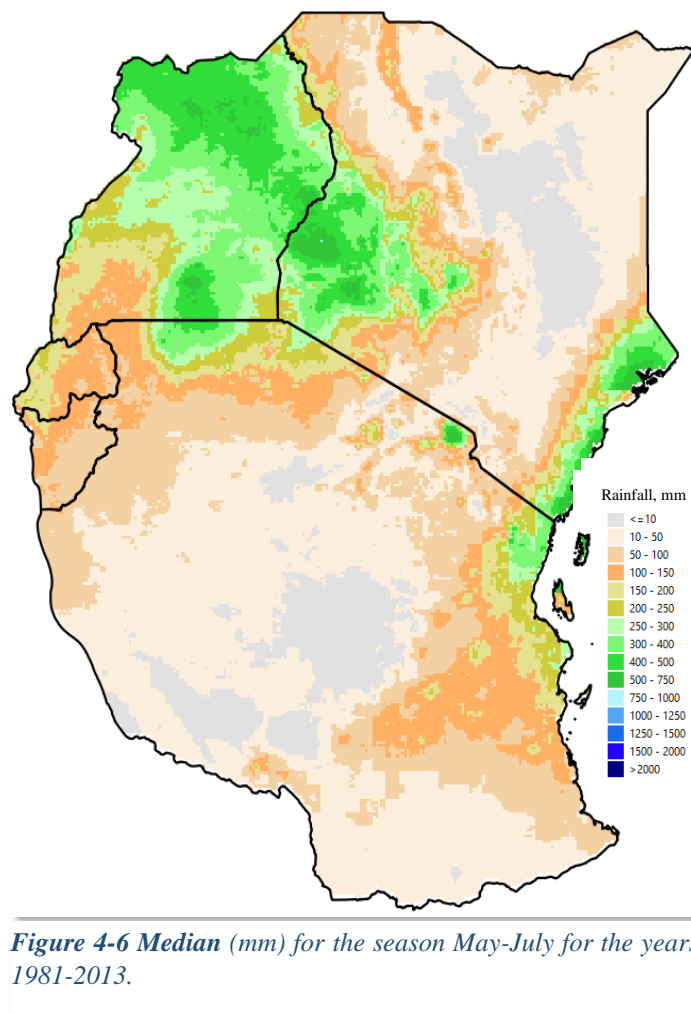
4.3.2. Median

The *Median* analysis method calculates the midpoint value of a frequency distribution for the selected climate variable. Figure 4-6 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 4.1
2. Select the dataset
3. Select the region
4. Select the season on the left panel and the years on the right panel
5. Check the ■ Seasonal totals option
6. Select Median from the analysis methods list
7. Click on Analyze to run the analysis

Figure 4-6 Median *(mm) for the season May-July for the years 1981-2013.*
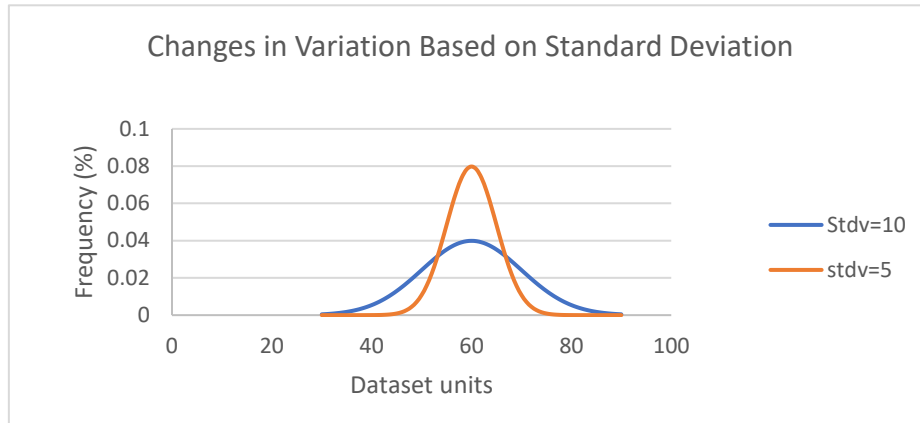
4.3.3. Measuring variability with standard deviation and coefficient of variation

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

4.3.3.1. *Standard deviation*

Figure 4-7 *The distribution of two data sets with same mean and different SD. The orange 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).*

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 4-7). See Figure 4-8**(a)** for an example of an SD product using GeoCLIM.

4.3.3.2. *Coefficient of variation.*

The Coefficient of Variation (CV) is the ratio of the SD to the mean $CV = \left(\frac{SD}{average}\right) * 100$.
Table 4.1.
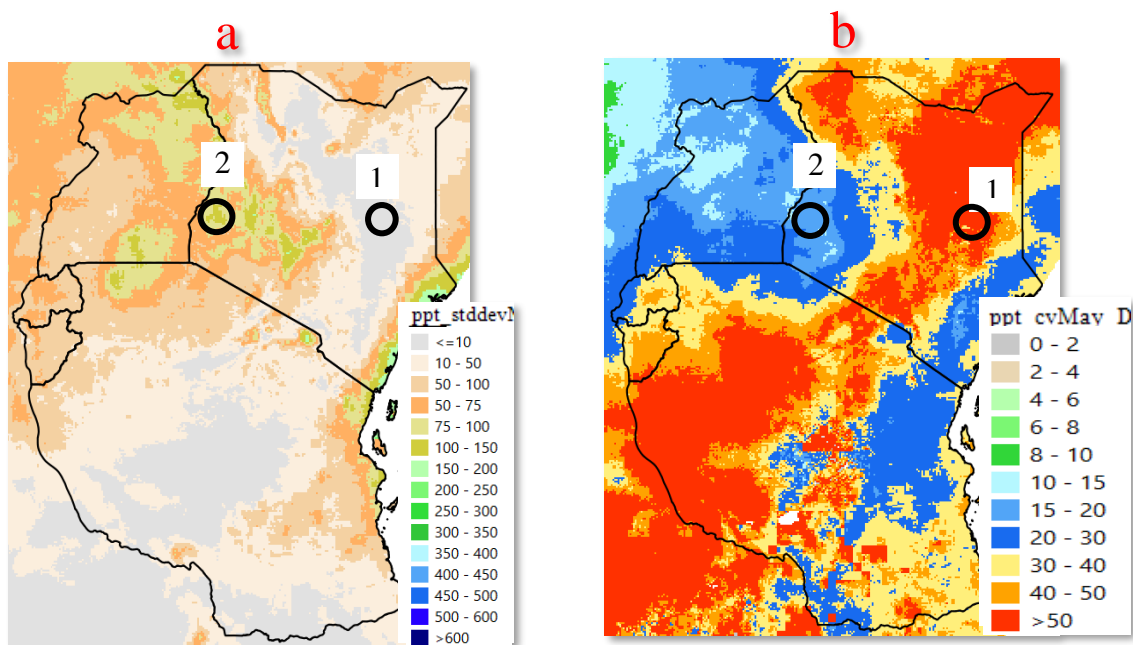
SD	Mean	CV
171mm	721mm	24%

Table 4.1 *The CV is the ratio of the SD over the mean.*

Figure 4-8 **(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 different magnitudes of variation or between regions with different means. Figure 4-8 **(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 4.1
2. Select the dataset
3. Select the region
4. Select the season on the left panel and the years on the right panel
5. Check the ■ Seasonal totals option
6. Select Standard Deviation or Coefficient of Variation from the analysis methods list
7. Click on Analyze to run the analysis

Figure 4-8 *(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.*

4.3.4. 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 4-9 shows the count as 40 (1981-2020) for all pixels, which means there are no missing values in the time-series used in the analysis.
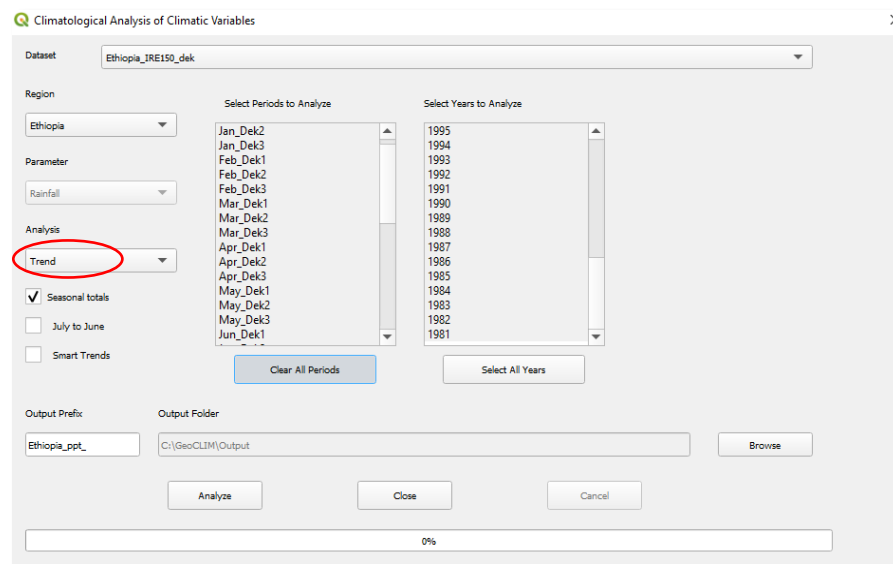
Figure 4-9 *The function counts the number of valid values in the time series. In the example, there is no missing data and there are 40 values.*

To calculate count, follow the steps below:

1. Start the ***Climatological Analysis*** tool, as described in section 4.1
2. Select the dataset
3. Select the region
4. Select the season on the left panel and the years on the right panel
5. Check the ■ Seasonal totals option
6. Select Count from the analysis methods list
8. Click on Analyze to run the analysis

4.3.5. 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 4-10). This function in GeoCLIM produces two maps; one is the slope of the regression which represents the trend, and the other is the coefficient of determination (r-squared, or r2), which represents the strength of the relationship.
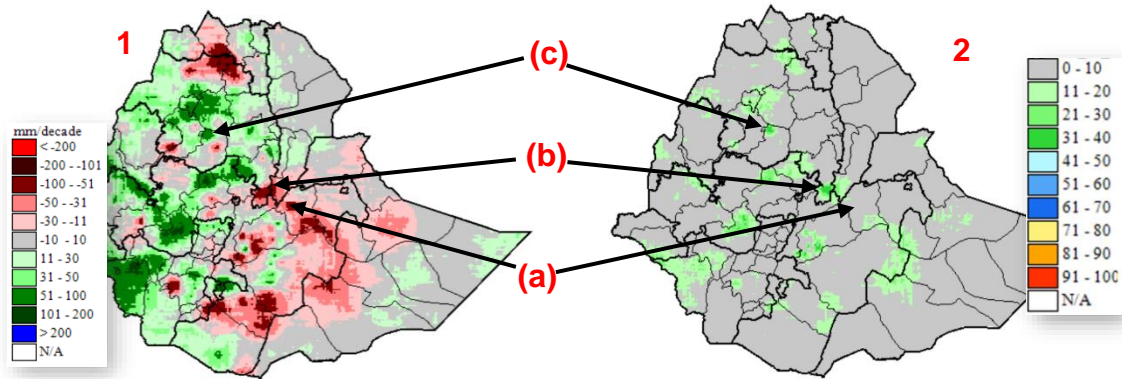
Figure 4-10 *To calculate the trend for a climate variable, select the season, make sure that the* **Seasonal totals** *option is checked, and select the years to be used in the calculation.*

To calculate the trend, follow the steps below:

1. Start the ***Climatological Analysis*** tool, as described in section 4.1
2. Select the dataset
3. Select the region
4. Select the season on the left panel and the years on the right panel
5. Check the ■ Seasonal totals option
6. Select Trend from the analysis methods list
7. Click on Analyze to run the analysis
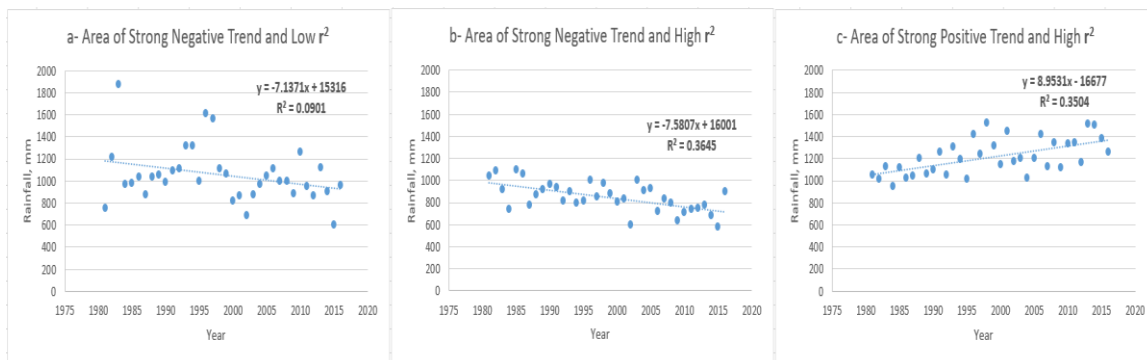
Figure 4-11 shows the results of the Trend analysis method in GeoCLIM for the annual rainfall total in Ethiopia for the period 1981 – 2016, using the Improved Rainfall Estimates (IRE), see chapter 10 on how to create IRE data. Figure 4-11 **(1)** shows the slope of the regression line, or the trend for each pixel in mm/decade of increasing (green-blue) or decreasing (pink-red) rainfall. The legend shows these results per decade (10 years). Figure

4-11 **(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 4-11 *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 4-11 **(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 4-12 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 4-12 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.

Figure 4-12 *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 4.10(1) (2) with different r2.*

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.

4.3.6. Percentiles
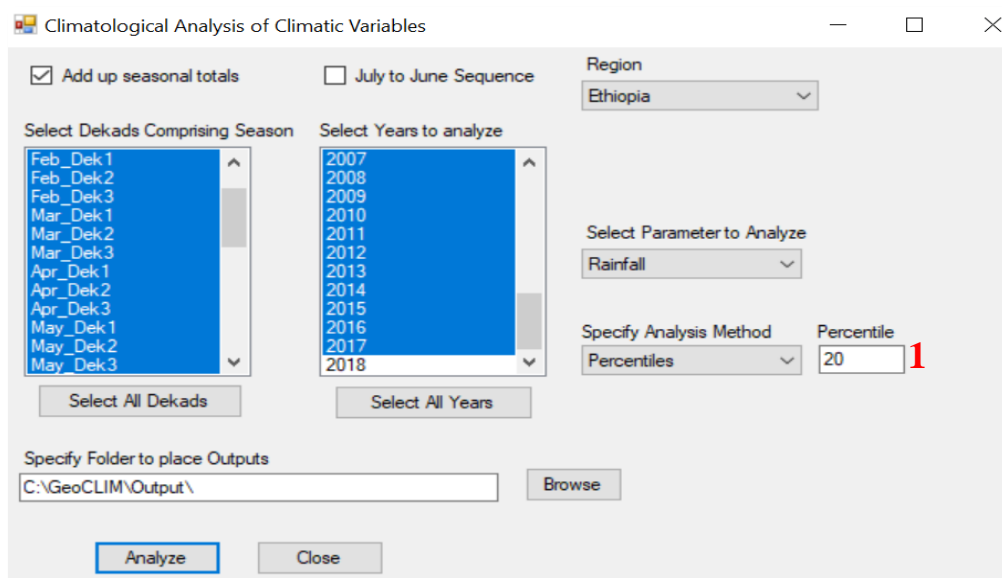
A percentile is a statistic that specifies the value below which a certain percent of observations in a ranked dataset will fall. Percentiles are calculated at breakpoints ranging from 0 to 100. The 0th percentile corresponds to the lowest value. The 100th percentile is the highest. The 50th percentile is the median value. To calculate a percentile value, we first must 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 4.2), 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.
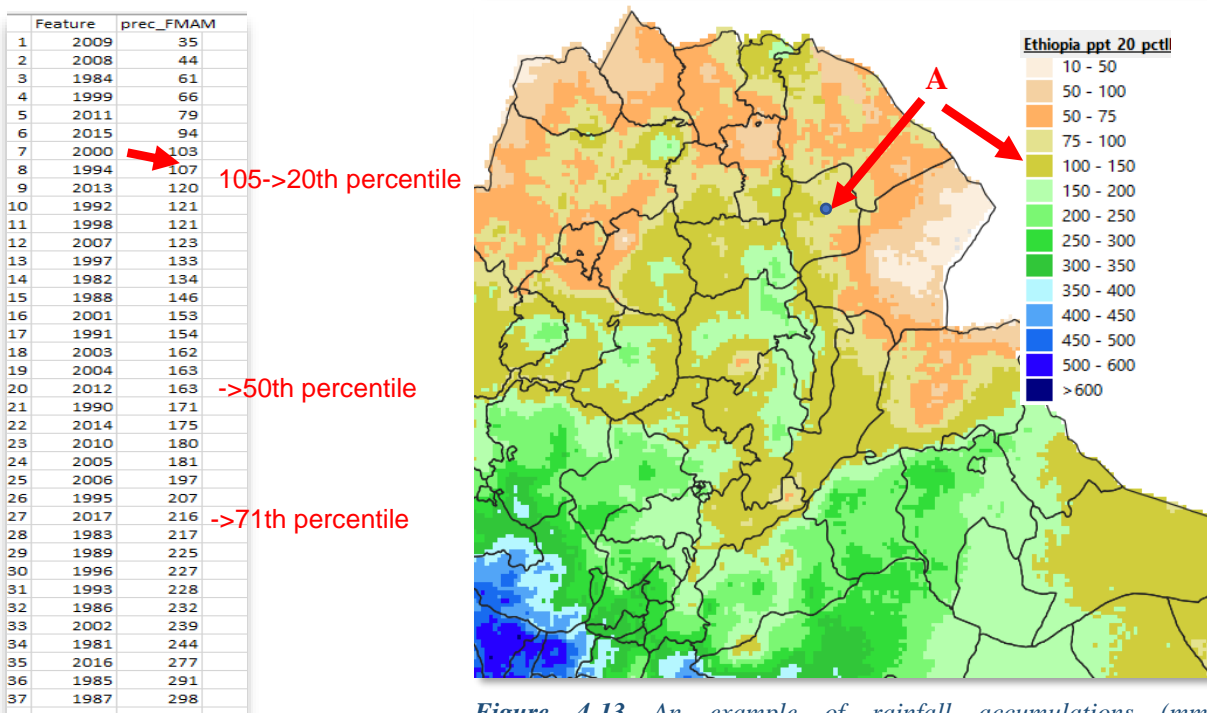
Another use of percentiles is when we have a value, let's say the rainfall total for the FMAM for 2017=216 mm, and we would like to know what percentile that value represents, or how frequent a value like this happens. Using the data in table 4.2 (see Note below on how the data was obtained) and the 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.

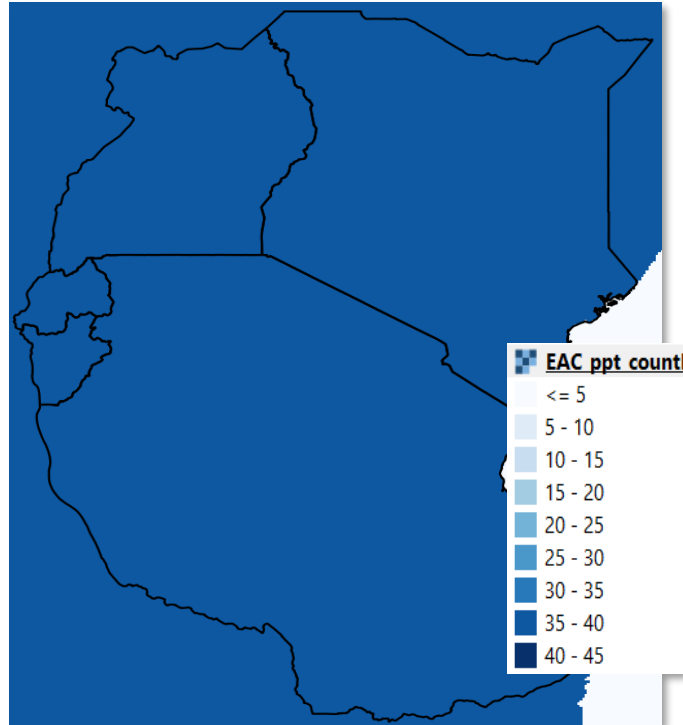To calculate a given percentile for your region of interest, follow the steps below:

1. Start the **Climatological Analysis** tool, as described in section 4.1
2. Select the dataset
3. Select the region
4. Select the season on the left panel and the years on the right panel
5. Check the ■ Seasonal totals option
6. Select **Percentile** from the analysis methods list
7. Enter the percentile rank desired (Figure 4-14 **(1)**)
8. Click on Analyze to run the analysis.

Figure 4-14 The Percentiles method in GeoCLIM produces a raster map with the rainfall value for each pixel corresponding to the percentile rank.

	Feature	prec_FMAM
1	2009	35
2	2008	44
3	1984	61
4	1999	66
5	2011	79
6	2015	94
7	2000	103
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
21	1990	171
22	2014	175
23	2010	180
24	2005	181
25	2006	197
26	1995	207
27	2017	216
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

105->20th percentile

->50th percentile

->71th percentile

Table 4.2 Seasonal total FMAM 1981-2017 for point A in Figure 4-13.

Figure 4-13 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.
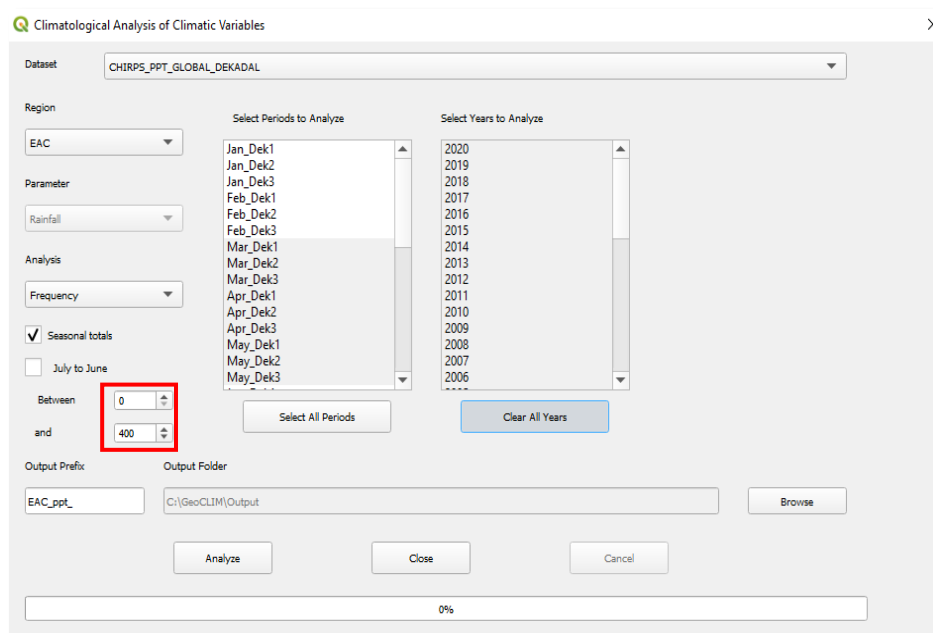
This function in GeoCLIM helps answer questions such as, what are the low/high values (e.g., 15th/90th percentiles) in the time-series? (Figure 14). Table 4.2 shows the time-series for the FMAM seasonal total for the period 1981-2017 for point **(A)** in Figure 4-14. The result of the PERCENTILE.EXC function in Excel shows that the 20th percentile is = 105.

> **NOTE:** Table 4.2 was created using the ***Extract Statistics*** tool for a point shown in Figure 14Figure 4-13 **(A)**.

4.3.7. Frequency

The Frequency analysis method in the GeoCLIM Climatological Analysis module (Figure 4-15) 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, in the period 1981-2020?" 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 4-16 represents the number of events per decade (ten years).
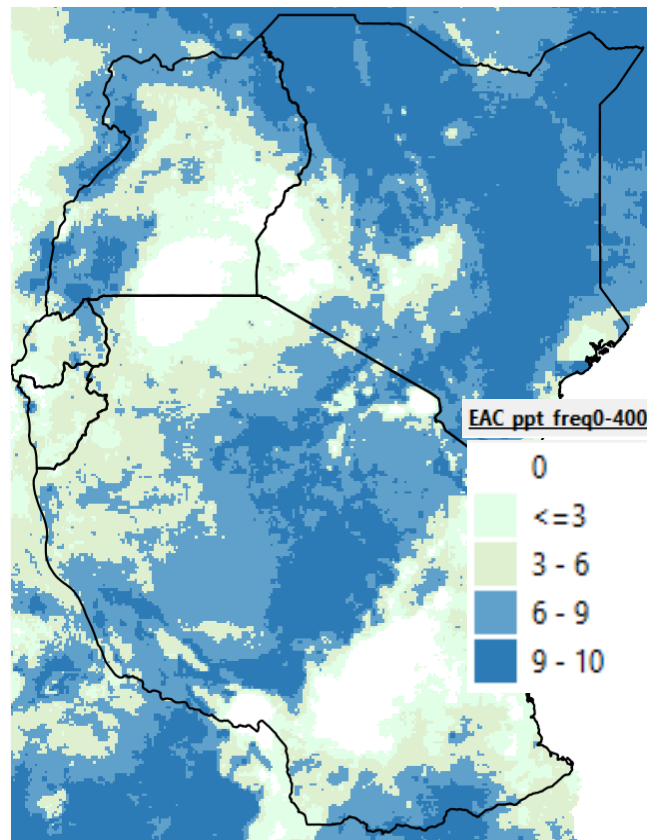
Figure 4-15 *Frequency function allows for the selection of a range of values (red box) and identifies the number of times this range has occurred in the time series.*

To calculate the frequency of a range of values, follow the steps below:

1. Start the ***Climatological Analysis*** tool, as described in section 4.1
2. Select the dataset

3. Select the region
4. Select the season on the left panel and the years on the right panel
5. Check the ■ Seasonal totals option
6. Select *Frequency* from the analysis methods list
7. Fill in the values Between and And to define the range of values
8. Click on Analyze to run the analysis

Figure 4-16 *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.*

4.3.8. Standardized Precipitation Index (SPI)
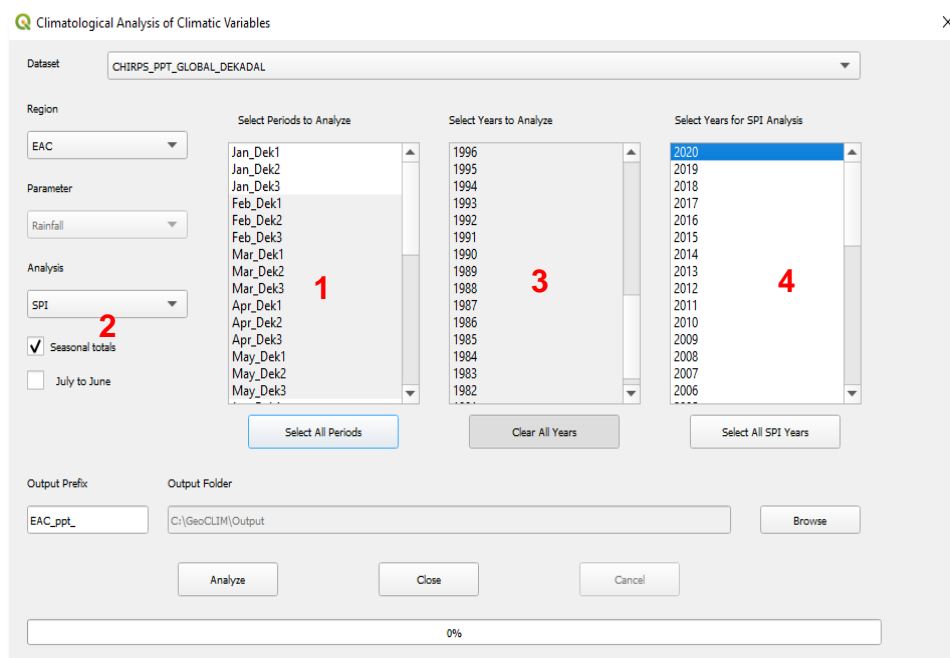
SPI is a drought index that expresses the difference of precipitation from the mean, for a specified time period, in standard deviation units or z-scores. The SPI calculation uses historical data to determine the mean and standard deviation. Since precipitation is, in most cases, not normal for periods less than 12 months, the data must be transformed into a normal distribution. The new distribution of standardized precipitation is linearly proportional to precipitation deficit.

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
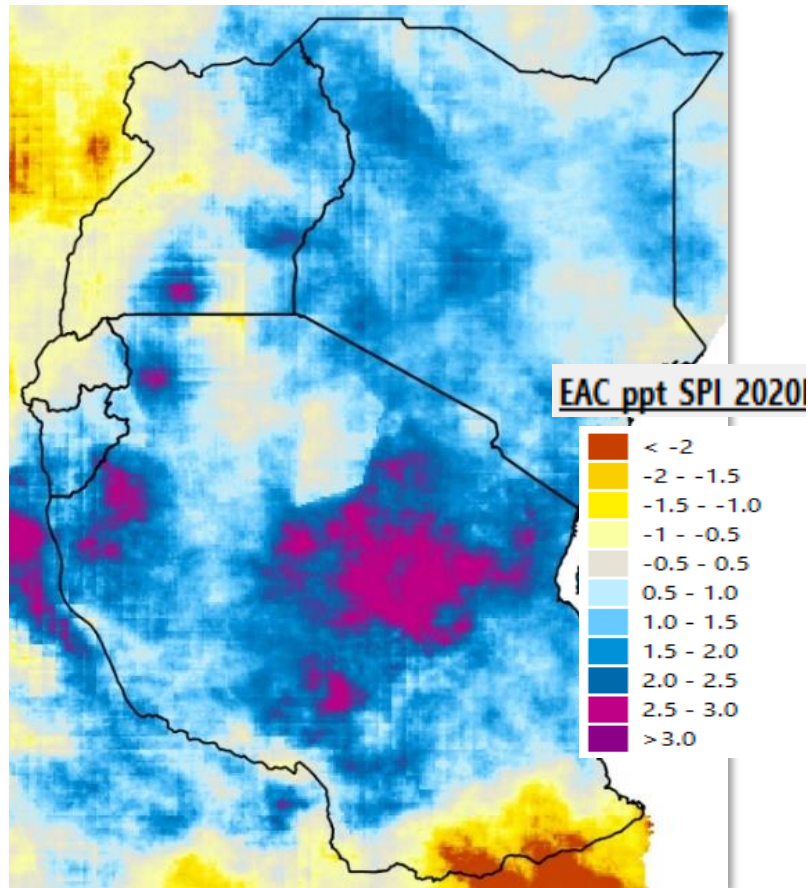
that the observation is roughly a one-in-six dry event and is termed "moderate." An SPI less than -1.5 indicates a one-in-fifteen dry event and is termed "extreme." Values less than -2.0 are typically referred to as "exceptional," indicating that it is in the driest 2% of all events. (Mckee 1993).

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

1. Start the *Climatological Analysis* tool, as described in section 4.1
2. Select the dataset
3. Select the region
4. Select the season (Figure 4-17 (**1**))
5. Check the ■ Seasonal totals option and select *SPI* from the analysis methods list (Figure 4-17 (**2**))
6. Select the group of year to make the time series (Figure 4-17 (**3**))
7. Select a single or a group of years for which SPI is to be calculated (Figure 4-17 (**4**))
8. Click on Analyze to run the analysis

Figure 4-17 *The GeoCLIM allows you to calculate the SPI for a single or multiple years.*

Figure 4-18 SPI Feb-May 2020 (CHIRPS 1981-2020). The GeoCLIM SPI raster output is in units of [SPI * 100], but the legend shows actual SPI values.

Chapter 5: View and Explore Rainfall Summaries

Summary

Figure 5-1 The Rainfall Summaries tool facilitates the calculation of total and anomalies for a season.

When analyzing a rainy season, it is important to estimate how different it is from average.

The **Rainfall Summaries** tool (red box in Figure 5-1) 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`
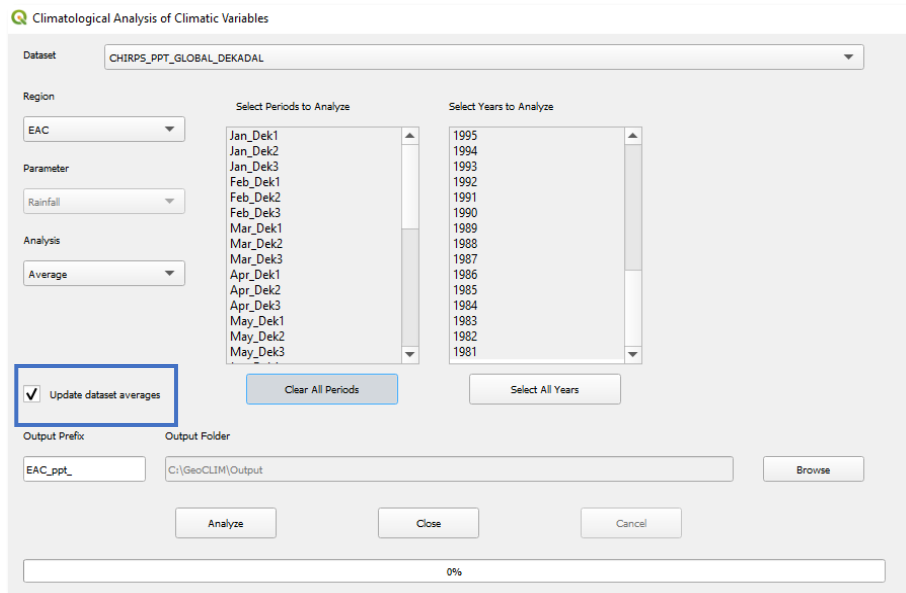
The outputs provide answers to questions such as, "What is the average for the period June dek1 to Aug dek3" and "How different was June dek1- Aug dek3 2012 from average?"

5.1. Requirements

To use the **Rainfall Summaries** tool, the selected climate dataset must have GeoCLIM averages available. These are the averages defined in the dataset definition form, see chapter 2 for more information. If the averages do not exist, a window message shows up, allowing you to open the **Climatological Analysis of Climatic Variables** tool to calculate the averages as follows: (See Figure 5-2).

- Select the dataset
- Select all the periods
- Select the number of years to be used in calculating the averages
- Check the box update dataset averages.

The tool uses the information entered in the data definition form to calculate the averages for the domain of the dataset.

Figure 5-2 *In order to calculate rainfall summaries you need to have averages for all the periods (pentads, dekads or months).*

5.2. Calculate seasonal total and anomalies

Once the averages are available, close the Climatological Analysis tool to go back to the *Rainfall Summaries* tool, Figure 5-3.

1. Select the period of analysis (defined by the From and To date)
2. Select the summaries to be displayed
   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 appendix A for how to edit the legend in QGIS

Figure 5-3 *The Rainfall Summaries function facilitates the calculation of anomalies for a period.*

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

Chapter 6: Climate Composites

Summary
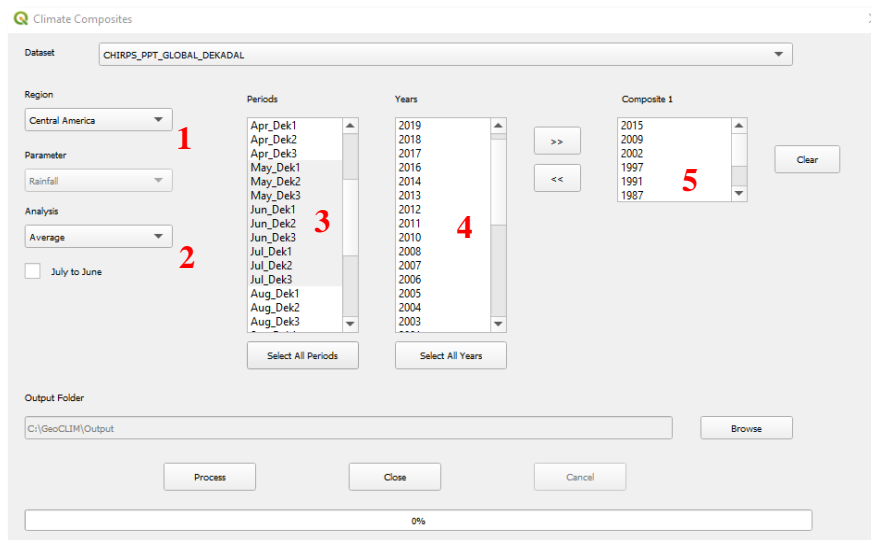
Figure 6-1 *The Climate Composites facilitates the analysis of a season among a group of non-consecutive years or compares the seasonal rainfall performance among two groups of years.*

The ***Climate Composites*** tool (red box in Figure 6-1) facilitates the analysis of a season among a group of 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 (EL Niño), the percent of average, as well as anomalies or standardized anomalies for one or two groups of years using an average baseline period.

6.1. Average

This function calculates the seasonal average for a group of years. 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).
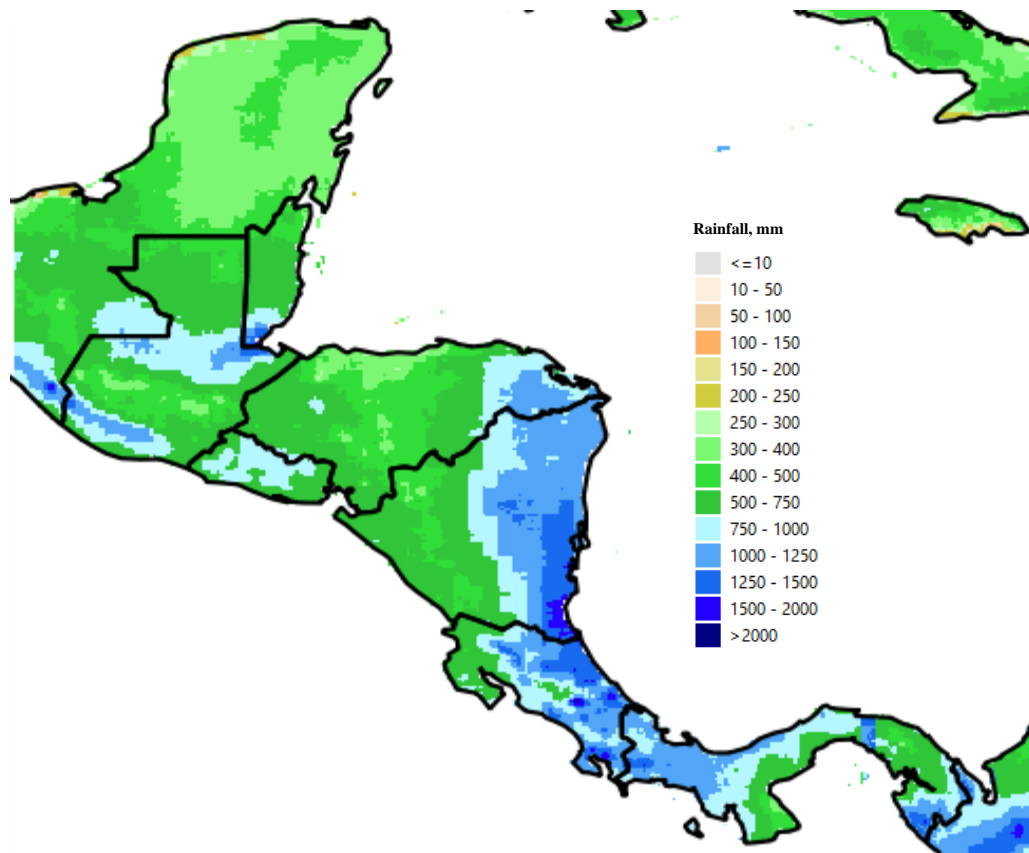
Figure 6-2 *The Composites tool calculates the seasonal average from a group of years, among other functions.*

To calculate the seasonal average for a group of years, follow the steps below:
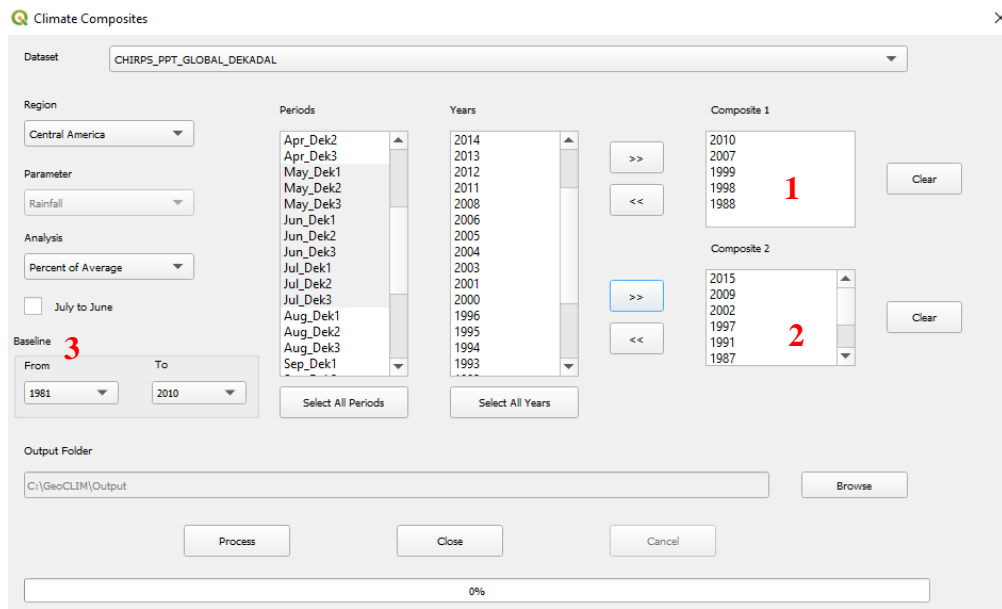
1. Select the region of interest (Figure 6-2 **(1)**)
2. Select the method: In this case, select Average (Figure 6-2 **(2)**)
3. Select the season to be analyzed (Figure 6-2 **(3)**)
4. Select the years to be included for composite 1 (Figure 6-2 **(4)**)
5. Move the selected years to the ***composite 1*** box by clicking the ⧉ button (Figure 6-2 **(5)**).
6. Click on ⧉Process. Figure 6-3 shows the results.

Figure 6-3 *Average rainfall for El Niño years in Central America for the May - July season, composite1.*

6.2. 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 6-4 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 6-4 *The composites function calculates the percent of average for a single group of years or compares two groups of years.*
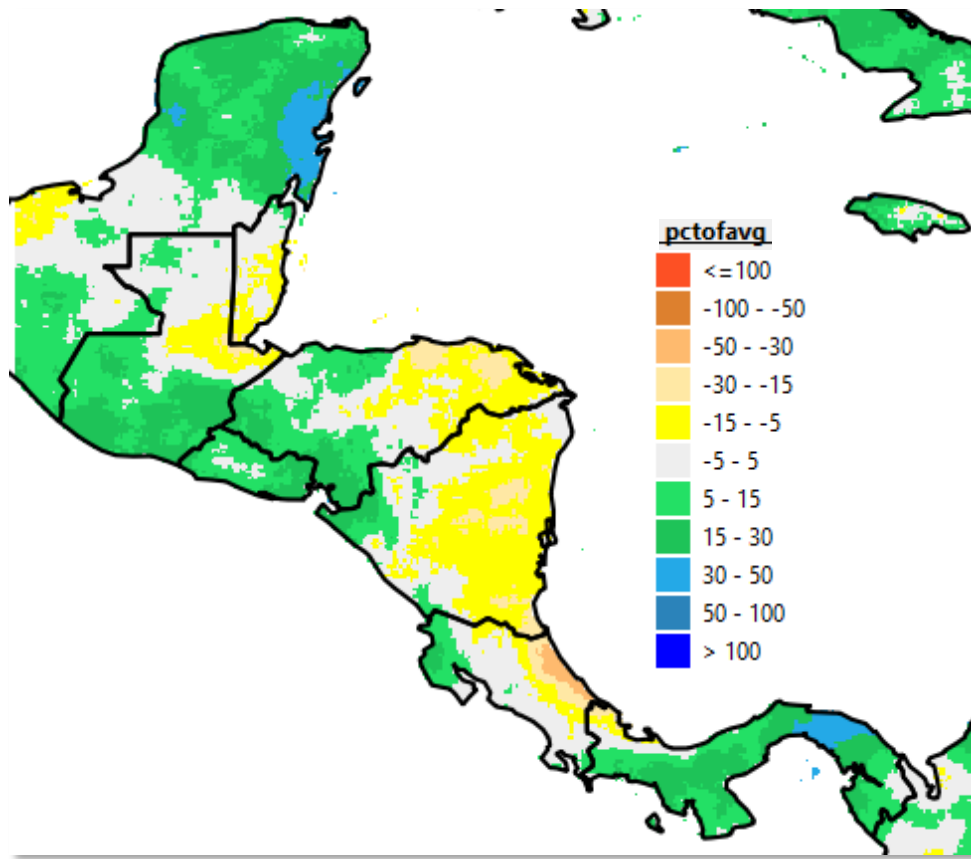
To calculate the percent of average for composite 1 (Eq. 6.1):

$$\text{Pct}_{\text{comp1}} = \left(\frac{\text{average}_{\text{comp1}} + (0.05 * ltavg) + 0.1}{\text{average}_{\text{baseline}} + (0.05 * ltavg) + 0.1} \right) * 100 \qquad (6.1)$$

1. If composite 2 is empty, the program saves pct_comp1 as the final output and displays the product.
2. If composite 2 is **not** empty, the program calculates the difference between the two composites as shown by Eq. 6.2:

$$pct_{comp} = \left(\frac{(average_{comp1} - average_{comp2}) + 0.1}{average_{baseline} + 0.1} \right) * 100 \qquad (6.2)$$

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. 6.2. In our example, Figure 6-5 shows that rainfall is higher during La Niña years in most of the Pacific coast of Central America.

Figure 6-5 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.

6.3. Anomaly: (Applies to composite 1 and composite 2)
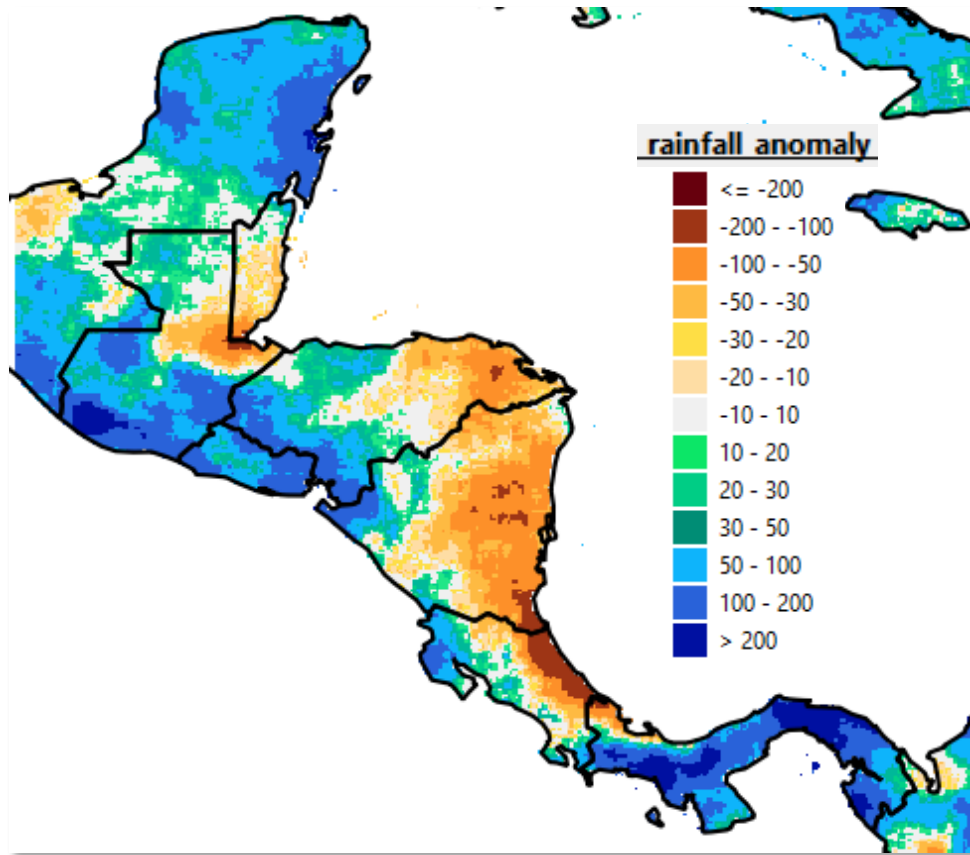
This analysis method calculates the average for each composite and the baseline; then, it calculates the anomaly for each composite (Eq. 6.3).

$$anom_{compN} = average_{compN} - average_{baseline} \qquad (6.3)$$

1. If composite 2 is empty, the $anom_{comp1}$ is saved as the final output and displayed on the QGIS canvas.
2. If composite 2 is not empty, the program calculates the difference between the anomalies of the two composites (Eq. 6.4).

$$anom_{comp} = anom_{comp1} - anom_{comp2} \qquad (6.4)$$

Figure 6-6 shows the results of the calculation of the anomalies for El Niño/La Niña example.

Figure 6-6 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.

6.4. 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 composite 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. 6.5.)
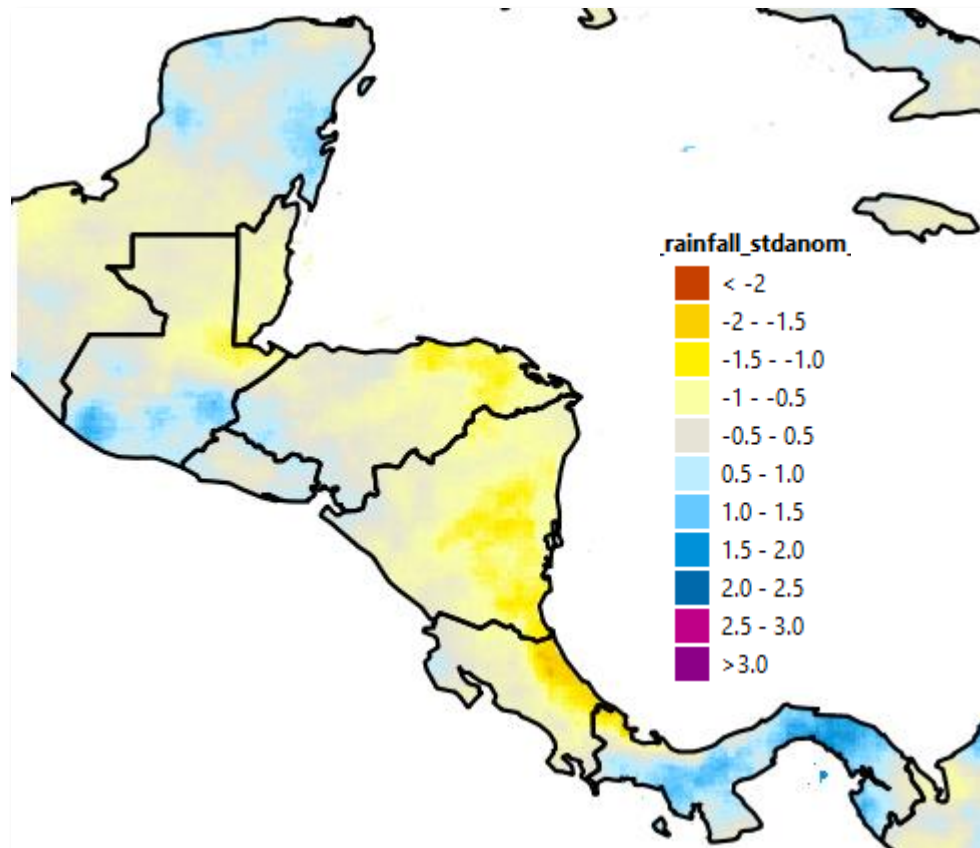
$$stdanom_{compN} = \left(\frac{(average_{compN} - average_{baseline}) + 0.1}{stdev_{available\ years} + 0.1} \right) * 100 \qquad (6.5)$$

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:

$$stdadnom_{comp} = stdanom_{comp1} - stdanom_{comp2} \qquad (6.6)$$

The results in Figure 6-7 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 6-7** This function calculates standardized anomaly, which is the difference in 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 6-7, 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 7: Contour Tool

Summary
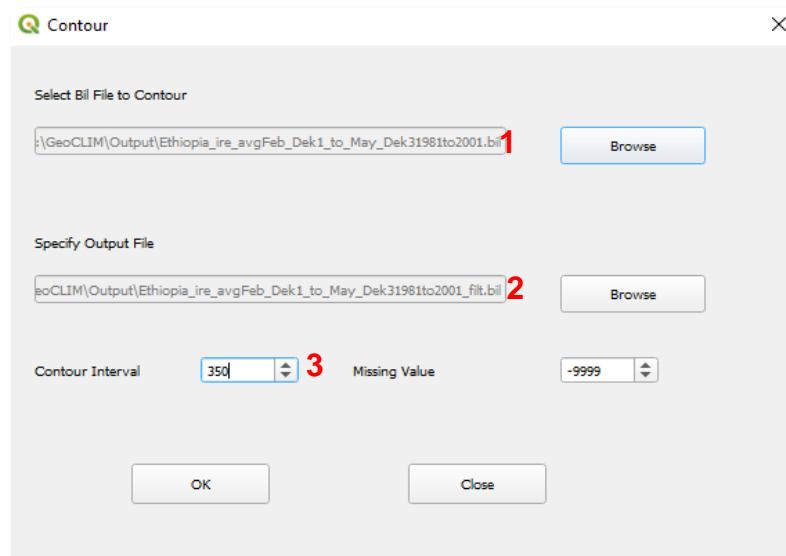
Figure 7-1 *The Make Contours tool delineates areas within a defined interval value in a climate-variable.*

The ***Make Contours*** tool (red box in Figure 7-1) delineates areas within a defined interval value in a climate variable. Analyzing contours from different periods of time helps to identify changes in a variable within a region. For example: we could identify the changes in rainfall-receiving areas with more than 350 mm during the Belg season (Feb-May) in Ethiopia in the last 40 years, by comparing the contours for average rainfall in the period 1981-2001 and 2002-2020.

7.1. Making contours

To run the tool, follow the steps below:
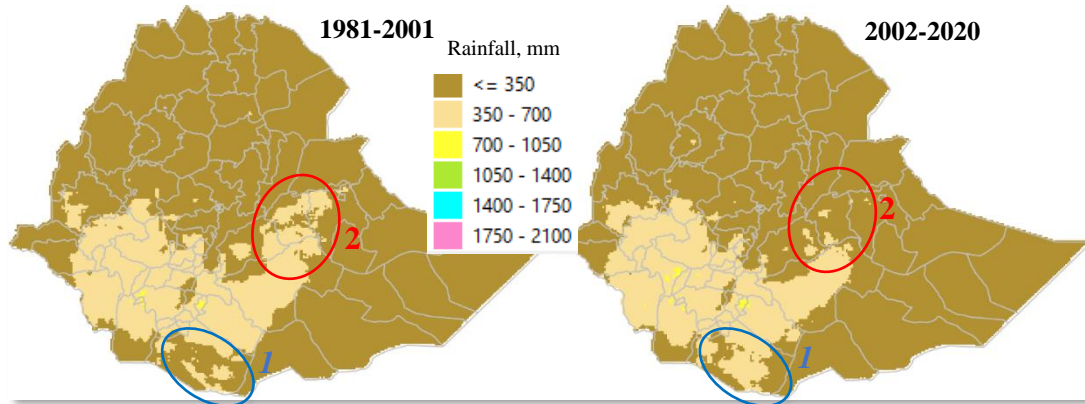
1. Open the ***Make Contours*** tool from the GeoCLIM main toolbar
2. Specify the BIL input file; in this example, we are using the average February-May rainfall season for the period 1981-2001 (Figure 7-2 **(1).**)
3. The tool automatically specifies the output file (Figure 7-2 **(2).**)

Figure 7-2 *GeoCLIM facilitates the delineation of areas within a range of rainfall values.*

4. Select a contour interval value. In this case, 350 for an interval of rainfall of 350 mm (Figure 7-2 **(3).**)
5. Change the missing value if necessary
6. Click OK to run the tool.
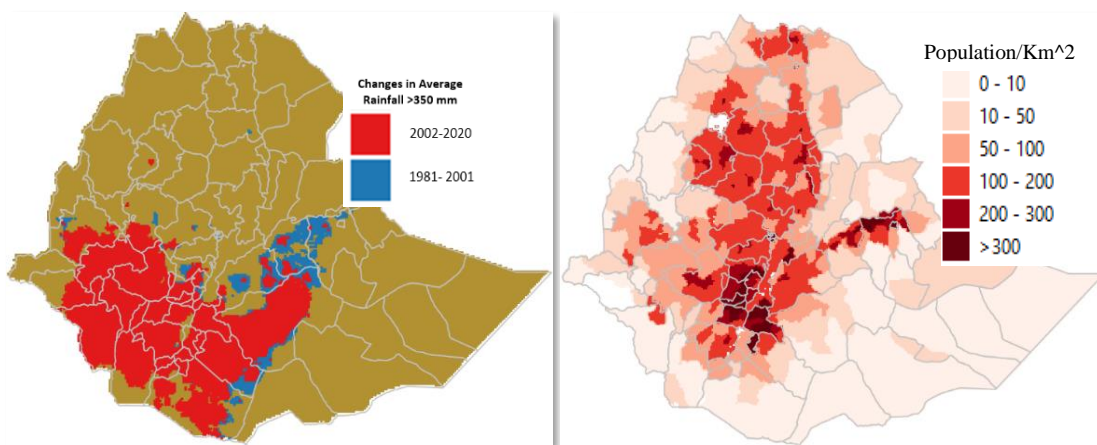
After calculating the contours for both periods of time, the results show an increase/decrease in rainfall in polygons 1/2, respectively in Figure 7-3.

Figure 7-3 *The 350 mm interval of average rainfall for the Feb-May season, during the periods 1981-2001(map on the left), 2002 – 2020 (map on the right), show reduction of rainfall in areas in red*

To further identify the difference between the two periods, we can overlay the areas with more than 350 mm, as shown in Figure 7-4 **(A)**. The image shows that the area that receives more than 350mm is shrinking in the eastern part of the country. We then compare the results to a population map and see that rainfall is decreasing in highly populated areas, Figure 7-4 **(B)**.

The map in Figure 7-4 **(A)** was produced as follows:

Figure 7-4 *Changes in rainfall patterns during the Belg season in Ethiopia. Image A shows the changes in areas that receive > 350mm. Image B shows the population per Km^2.*

1. Use GeoCLIM to calculate average 1981-2001 and 2002 -2020
2. Use the Contours tool at a given interval (350mm for this example.)
3. Use the raster calculator in QGIS to create a binary layer; values greater than $350 = 1$ and values less than $350 = 0$. (layer < 350) $* 0 +$ (layer $>= 350$) $* 1$
4. Use the Raster/conversion/polygonise tool to create polygons from the binary layers
5. Overlay the two resulting polygon shapefiles

Chapter 8: Calculate Long-Term Changes in Averages

Summary
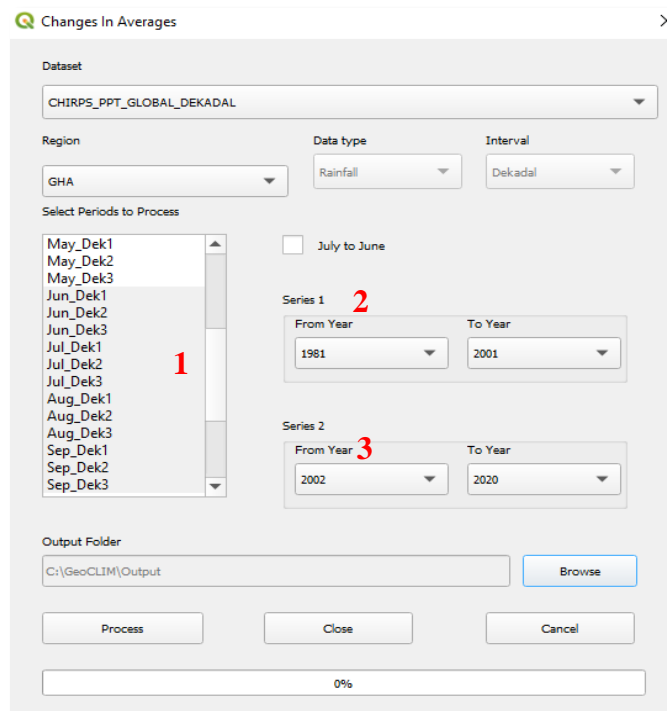
Figure 8-1 *The Long-Term Changes in Averages tool facilitates the comparison of two periods to determine changes in rainfall patterns.*

Another way to estimate changes in climate patterns is by comparing the averages between two periods within a time-series. ***The long-Term Changes in Averages*** tool (red box in Figure 8-1) 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 (difference = group2-group1).

8.1. Calculating changes in averages

To run the tool, follow the steps below:
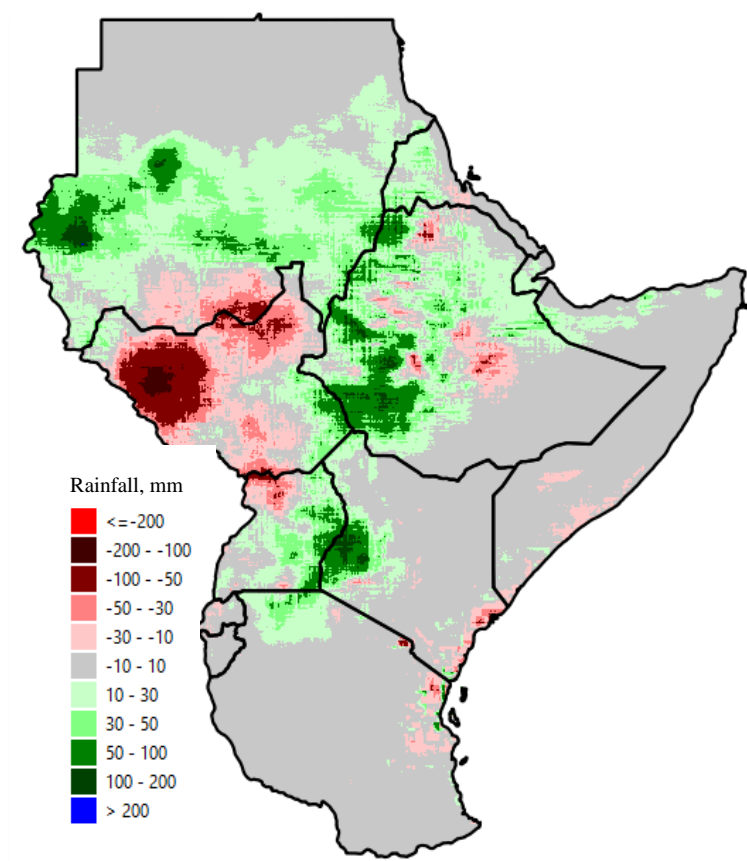
1. Open the ***long-Term Changes in Averages*** tool from the GeoCLIM main toolbar, see Figure 8-1.

Figure 8-2 *The* Long-Term Changes in Averages *tool allows you to identify changes in climate patters by calculating the differences in averages between two periods of time in a time*

2. Select the season to be analyzed, see Figure 8-2 **(1)**
3. In series 1, select the first period of time, see Figure 8-2 **(2)**
4. In series 2, select the second period, see Figure 8-2 **(3)**
5. Click Process to finish
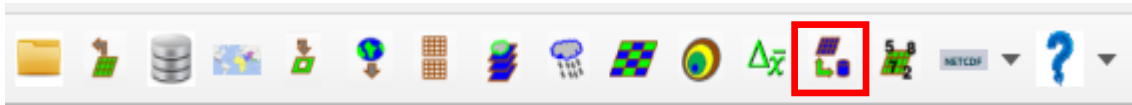
Figure 8-3 shows the output map result of the change in average from the 1981-2001 period (series 1) compared to the 2002-2020 period (series 2), for the June-September season. The result shows areas with increasing (green-blue) and decreasing (pink-red) mm of rainfall during the 2002-2020 period.

Figure 8-3 *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 9: Background-Assisted Station Interpolation for Improved Climate Surfaces (BASIICS)

Summary
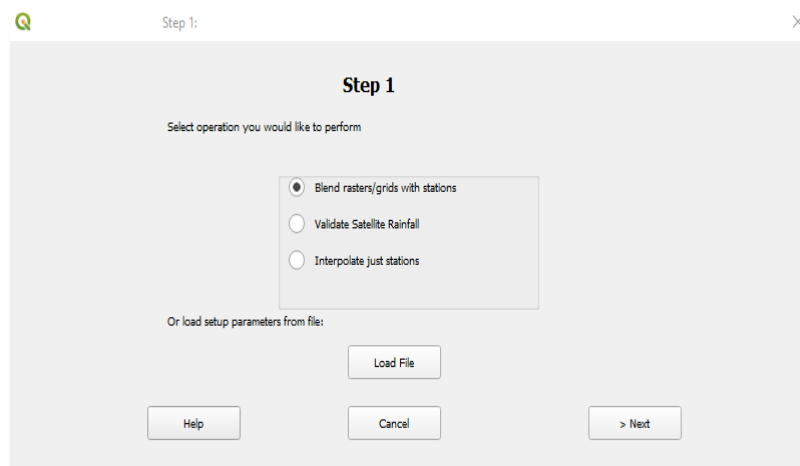
Figure 9-1 the Background-Assisted Station Interpolation for Improved Climate Surfaces, (BASIICS) algorithm in GeoCLIM facilitates the improvement of climate variables by blending raster data with local stations, among other functions.

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. See icon in red box in Figure 9-1.

The BASIICS tool includes the following processes as shown in Figure 9-2:

- Blend climate raster/grids with stations (BASIICS)
- Validate satellite data using ground station values
- Interpolate stations only

Figure 9-2 There are three options available in the BASIICS tool; (1) Blend stations and raster data, (2) Validate Satellite Rainfall and (3) Interpolate Just Stations.

The following three-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 settings to a file so you could use it later to update the improved rainfall times series.

9.1. 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 point values by a blending process. The validation process first extracts 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 results are: 1) A shapefile with the points that were included in the process. 2) A field of the interpolated values. 3) A .csv table that contain the station values, the corresponding grid value together with diagnostic 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:

9.1.1. Step 1: Select the BASIICS option.

1. Click on the BASIICS icon from the GeoCLIM main toolbar to open the dialog box (*Step 1*) (Figure 9-1)
2. Select the ■ Validate Satellite Rainfall option. See Figure 9-2
3. Click on the > Next button to proceed to *Step 2*

9.1.2. Step 2: Dataset and station parameters

Complete the form with raster and station data information. This form is made of 3 sections, (Figure 9-3).

Figure 9-3 Step 2 allows you to enter the raster and station information for the validation.

9.1.2.1. *Section 1: Grid Dataset Name*

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 ∨ pulldown menu.

9.1.2.2. *Section 2: Stations*

This section relates to the station input parameters.

1. The tool assumes that all station data are in a single csv file. Browse to select the file that contains the station data. See an example in Figure 9-4 of the file format. See the Data Management chapter for more information on the format of the table and other file types in GeoCLIM.
2. After selecting the station file, the tool identifies the header row and automatically completes the fields. Make any necessary changes to ensure that all field have the correct specification. When all the specifications are defined, move to section 3.

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 9-4 The CSV table with station data must contain a statin ID, lon, lat, year and a column for each pentad, dekad or month.

9.1.2.3. *Section 3: Outputs*

1. Specify the output prefix for all raster files created with the interpolation of the input stations.
2. Select the output folder.
3. Select the name for the statistics output file.

9.1.3. Step 3: Date parameters

Figure 9-5 Step 3 allows you to select a period to validate and save the settings to use later.
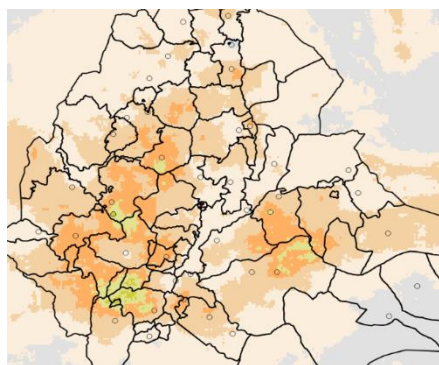
Select the validation period. Figure 9-5.

1. The time interval (e.g., month, dekad, or pentads) for the selected raster dataset is automatically displayed. Select the time range **From** and **To** 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 dekads, see Figure 9-5. And we are validating from (Feb dekad 01) to (May dekad 03), 2020.
2. Save the setup. At this step, you can save the validation settings so you could open it from step 1, edit and reuse it.
3. Click on the Finish button to run the process.
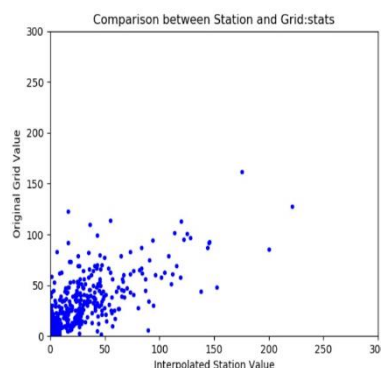
Outputs: The validation process creates the following outputs:

1. A shapefile, for each period, containing all the stations that were used in the process.
2. An interpolated field, for each period, using the IDW process. See Figure 9-6Figure 9-6.
3. A scatterplot showing the satellite rainfall field values against the station values (Figure 9-7).
4. A CSV file with columns containing the metadata for each station together with the station value, the corresponding raster value, and the at-station interpolated value. These at-station interpolated values are produced to improve comparability between the gridded/raster data and the station data. The CSV file includes statistics showing the correlation of the rainfall field and station data (Figure 9-8).

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

Figure 9-6 *The validation process produces an interpolation field together with a shapefile containing all the points included.*

Figure 9-7 *Scatterplot of interpolated station value on X and raster (CHIRPS) value on Y.*

Name	FileName	Long	Lat	StnVal	InterpAtSt	GridVal
TIADIG11	202015	39.447	14.278	3	2.92	18
HAAISH21	202015	42.578	10.757	11	11.02	4
TICHER11	202015	39.767	12.542	0	0.98	22
WODUBI2	202015	41.01	11.723	0	0.39	1

Figure 9-6 *Text file that includes a list of the station value and its corresponding raster value for each date together with statistics describing their relationship.*

9.2. Blend rasters/grids with stations (BASIICS)

The blending algorithm is a methodology designed to combine station values 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 the differences (ratios and anomalies) between the point and the grid value, where these two data are collocated. 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. 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.
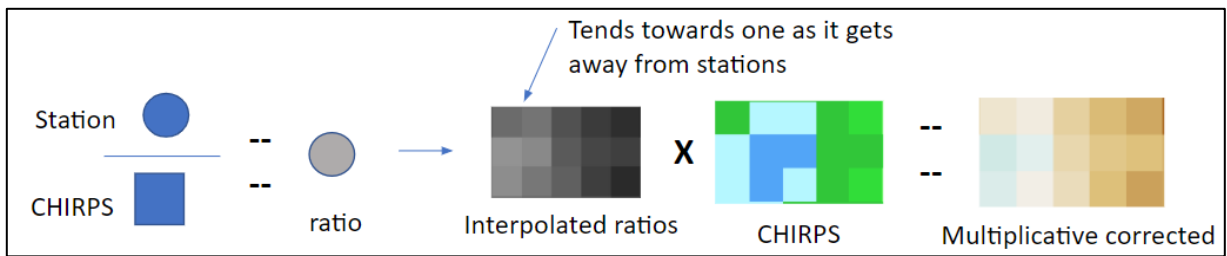
9.2.1. Input data for blending

(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.

9.2.2. The process

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 9.2 for validation.
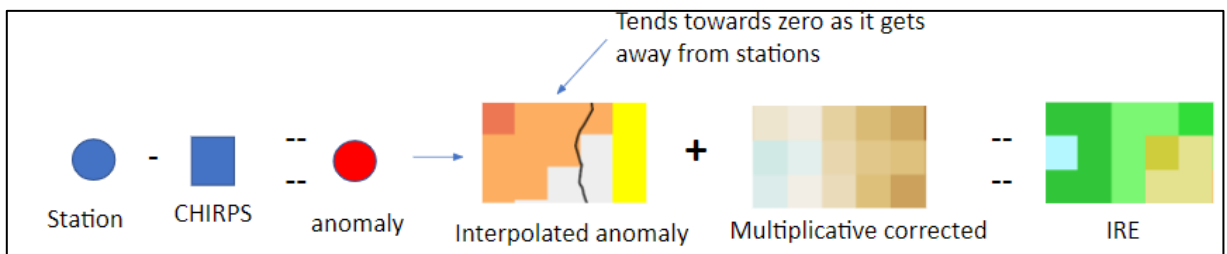
Step2. Multiplicative correction: A ratio is calculated between station and grid values; these ratios are interpolated using a modified IDW method, giving a maximum effective distance to each station. Once the maximum effective distance is reached, the interpolated layer takes the value of 1, (Figure 9-7). 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, take the value of the original raster layer.

Figure 9-7 The first part of the blending process is a multiplicative correction based on the ratio between the station and the raster values.

Step3. Additive correction: The anomaly between the station and the original raster is calculated at each point. The anomalies 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 0.

Step4. The interpolated anomalies are added to the multiplicative corrected rainfall layer from step2 to obtain the Improved Rainfall Estimate (IRE), see Figure 9-8.
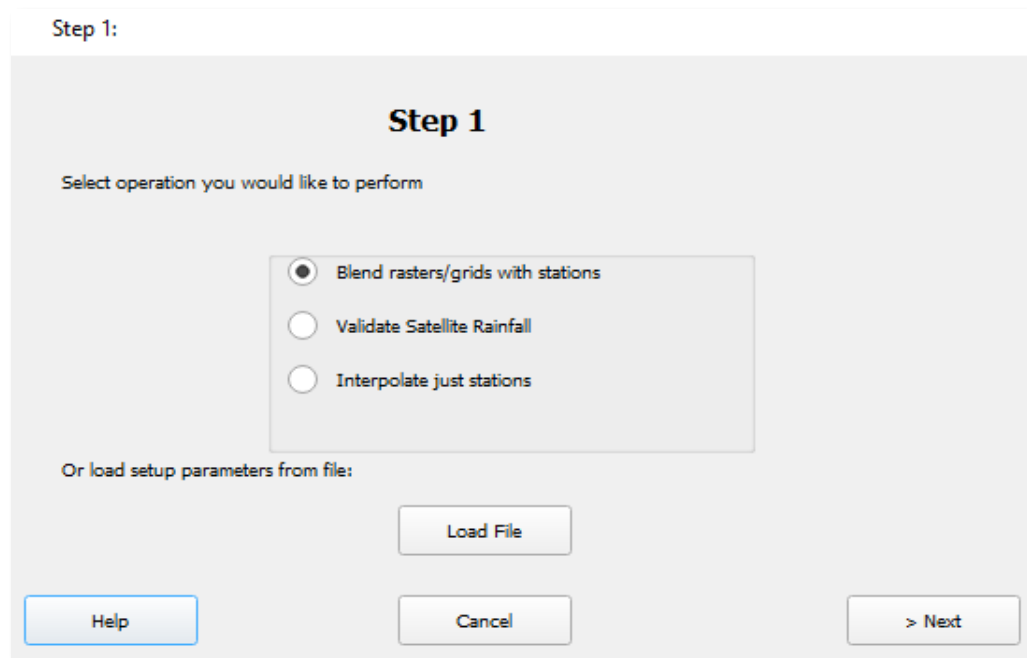
Figure 9-8 A second correction is done based on the anomalies between the station and the original raster values.

9.3. How to create improved rainfall estimates

9.3.1. Step 1: Select BASIICS option

1. Click on the ***BASIICS*** button from the GeoCLIM main toolbar. See Figure 9-1.

2. Select the ■ Blend rasters/grids with stations option. At this point you can click on the Load File button to load previously saved settings or click on the > Next button to start a new blending process. See Figure 9-9.

Figure 9-9 *Select the Blend raster/grids with stations option.*

9.3.2. Step 2: Dataset and stations parameters

Complete the form with raster and station data information. This form is made of 5 sections, (Figure 9-10).

Figure 9-10 *Step 3 of the blending process requires information about the raster data, the stations, the output location, the interpolation parameters, and the geographic domain.*

9.3.2.1. *Section 1: Grid Dataset Name*

This section relates to the raster/grid input parameters. This process allows the improvement of climate datasets that have already been registered in GeoCLIM. To select the climate dataset to be used in the blending process, use the Dataset Name ∨ pulldown menu. In this example we are going to blend CHIRPS dekadal data with stations.

9.3.2.2. *Section 2: Stations*

This section relates to the station input parameters.

3. The tool assumes that all station data are in a single csv file. Browse to select the file which contains the station data. See an example in Figure 9-11 of the file format. The order of the columns is not important, but must include the following:

    a) A unique station identifier ID, in a single column.
    b) A column with longitude in decimal degrees.
    c) A column with latitude in decimal degrees.
    d) A column with year value (yyyy).
    e) A series of consecutive columns for the number of periods (72 for pentads, 36 for dekads, or 12 for months).
    f) Any missing data should be completed with a single Missing Value, for example (-9999).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	ID	lon	lat	year	dek01	dek02	dek03	dek04	dek05	dek06	dek07	dek08	dek09	dek10	dek11
	GOBAHI41	37.4167	11.6	1981	0	0	0	0	0	0	0	0	0	0	0.3
	TIMEKE12	39.5312	13.4705	1981	0	0	0	0	0	0	5.5	4.5	14	0	5.6
	TIILAI11	39.63	13.53	1981	0	0	0	0	0	0	2.4	4.9	5.6	0	11.6

Figure 9-11 *The CSV table with station data must contain a station ID, lon, lat, year and a column for each pentad, dekad or month.*

Once you select the station file, the tool identifies the header row and automatically completes most of the fields. Make any necessary changes to ensure that all field have the correct specification.

9.3.2.3. *Section 3 – Outputs*
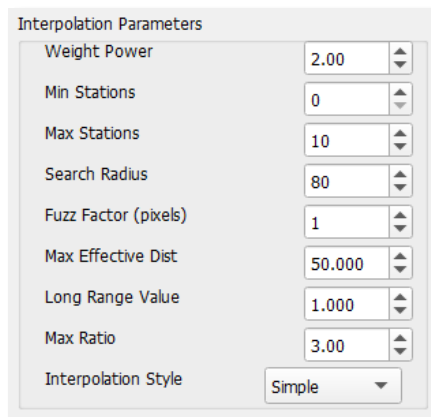
In the third section, you can specify the output directory where to save the blended products. At this point, you have two options: (1) create a new dataset or (2) update an existing dataset.

1. Create a new dataset: This first option allows you to create a new dataset in the correct format so it works with the GeoCLIM functions; for example, you are blending, for the first time, your stations with the historical data of CHIRPS or CHIRP and want to create a new dataset from the results. To do this:
    a) Provide a prefix for the output files.
    b) Browse to the GeoCLIM data repository. For example: **X:**`~\GeoCLIM\ProgramSettings\Data\Climate\new_dataset` where **X:**~ is the path to your GeoCLIM workspace.
    c) The path on the **Statistics Output Filename** field changes automatically when you define the output directory.
    d) Make sure to complete the fields in sections 4 and 5 before continuing. (See sections 4 and 5 for complete explanation of the parameters).

e) Click Next after completing all the fields.
f) A dialog box appears asking ***Do you want to create a new dataset from outputs?***
      i)   Click on Yes.
      ii) Enter a new name with no spaces.
      iii) Select the data type.
      iv) Select the extent of the data. If your region is outside of Africa or Central America, please select global.
      v)   Click OK to move to Step 3.

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) In the **Output folder** field, browse to the existing directory **X**:~\GeoCLIM\ProgramSettings\Data\Climate\existing_datas et where **X**:~ is the path to your GeoCLIM workspace.
   b) The path on the **Statistics Output Filename** field changes automatically when you specify the output directory.
   c) Make sure to complete the fields in sections 4 and 5 before continuing. (See sections 4 and 5 for complete explanation of the parameters).
   d) Click Next after completing all the fields.
   e) A dialog box appears asking ***Do you want to create a new dataset from outputs?***
      i)   Click on No to move to Step 3.

9.3.2.4. *Section 4 - the Blending Process- Interpolation parameters*
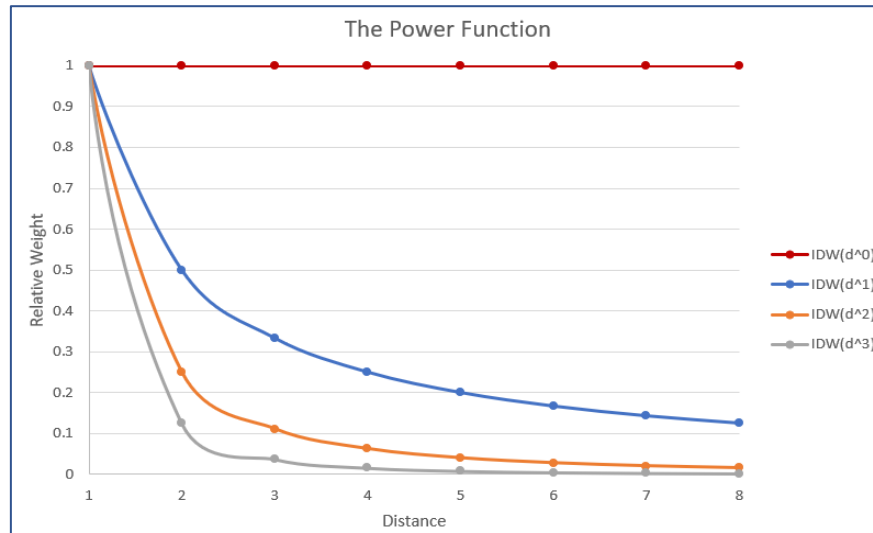
The program offers a set of options to adjust the parameters of the interpolation (Figure 9-12).

Figure 9-12 The blending process includes a series of parameters that you could modify.

Make sure you have a full understanding of the parameters before making any changes, otherwise, leave the default values. See a complete description of the parameters below.

Weight Power (WEIGHTPOWER): The power to which the inverse distance is raised in calculating the weight, indicating how fast the influence of the station decreases as the distance from the point increases. Figure 9-13 shows an example of IDW with different powers.

Figure 9-13 *The power indicates how fast the relative weight decreases as distance increases.*

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.

> **NOTE**: 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.

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 consider the local characteristics of the region to choose a proper value for this parameter. We recommend you test different combination of values for the Max Effective distance and the search radius 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.

9.3.2.5. *Section 5 Region*
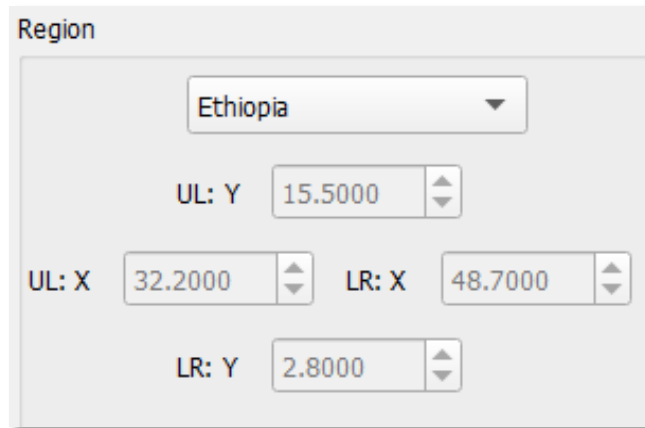
Define Map Limits: Allows you to define the interpolation area (Figure 9-16). Make sure that the area is smaller or equal to the gridded dataset. This area can be defined by using the extent of an existing GeoCLIM Region or other spatial data (raster or vector). This option helps to speed up the interpolation process.

To run the blending process, follow the steps below:

1. Choose the region from the list.

2. Click on Next To move to step 3.

Region

Ethiopia

UL: Y   15.5000

UL: X   32.2000   LR: X   48.7000

LR: Y   2.8000

Figure 9-14 Select the region (basin, admin unit, etc.) for the new data.

9.3.3. Step 3: Date Parameters and saving settings

To save the date parameters and settings follow the steps below (Figure 9-15):
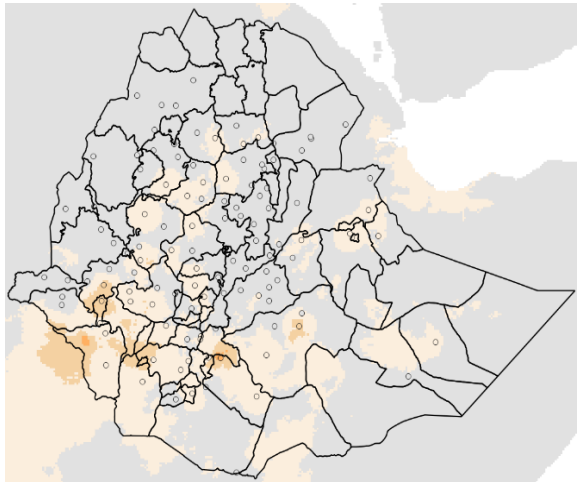
1. The time interval (e.g., month, dekad, or pentads) for the selected raster dataset is automatically displayed.
2. Select the time range **From** and **To** of the data to blend. The time period and time interval are based on the selected climate dataset definition. In this example we are using dekads, see figure 9-17. And we are blending from Feb (dekad 01) to May (dekad 03) 2020.
3. Save the setup. At this step, you can save the blending settings so you could open it from step 1, edit and reuse it.
4. Click on the Finish button to run the process.

Figure 9-15 *Step 3 allows you to define the time range for the blending process and you could also save the setting to use later.*
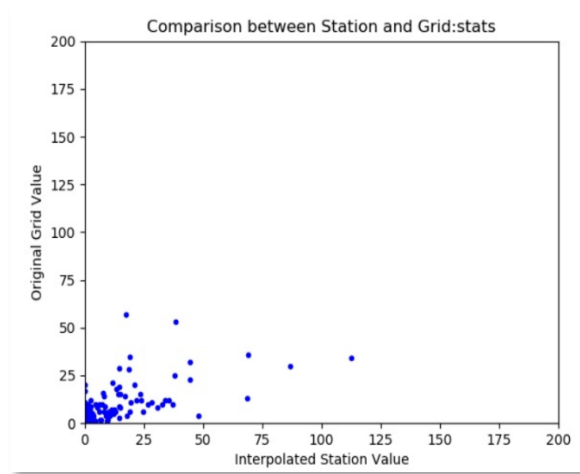
9.4. Outputs: The blending process creates the following outputs:

1. A shapefile, for each period, containing all the stations that were used in the process.
2. The blended field, for each period. See Figure 9-19a.
3. Three scatterplots showing the relationship of the original grid and the station values (Figure 9-19b) for example.
4. A CSV file, (Figure 9-19c) containing the metadata for each station together with the following columns:
   a) Station value
   b) Corresponding raster value
   c) The BASIICS value at station location
   d) Cross-validated BASIICS value. Indicates the BASIICS value at the station location without including that specific station in the process. This value responses to the question, what would be the value at this pixel if the station were not there.

e) Cross validated interpolation only. Pixel value of interpolation of stations only, without including the corresponding station.

Figure 9-19a *BASIICS field with participating stations.*

Figure 9-19b *comparison between station and raster values.*
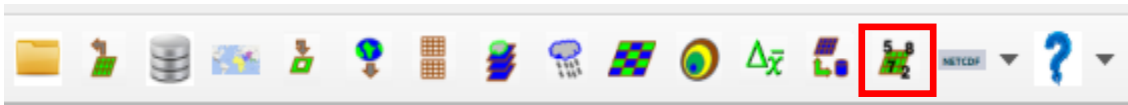
Name	FileName	Long	Lat	StnVal	GridVal	BASIICSval	XValidatedBASIICS	XValidatedNoGrid
SHADDI21	202032	38.748	9.019	0	2		0	0.14	0
GOADET1:	202032	37.493	11.274	0	2		0	0.74	0
HAAISH21	202032	42.578	10.757	0	2		0	2	-9999
SHALEM11	202032	39.033	10.033	0	1		0	0.33	0
WOAMBA	202032	39.217	11.203	0	2		0	0.36	0
SHAMBO2	202032	37.87	8.97	0	0		0	0	0

Figure 9-19c *CSV table with information at station location.*

Chapter 10: Extracting Raster Statistics and Time Series

Summary

The **Extract Grid Statistics** tool (red box in Figure 10-1) 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. The tool allows to calculate (**Average**) the spatial average within each polygon, (**Count**) the number of pixels with a valid value within the polygon, (**Maximum**) the maximum value within each polygon, and so on with each of the other parameters. For example, we can calculate the spatial average of precipitation for each district for each month from 1981 to the present. This produces a CSV table that could be analyzed using Excel.
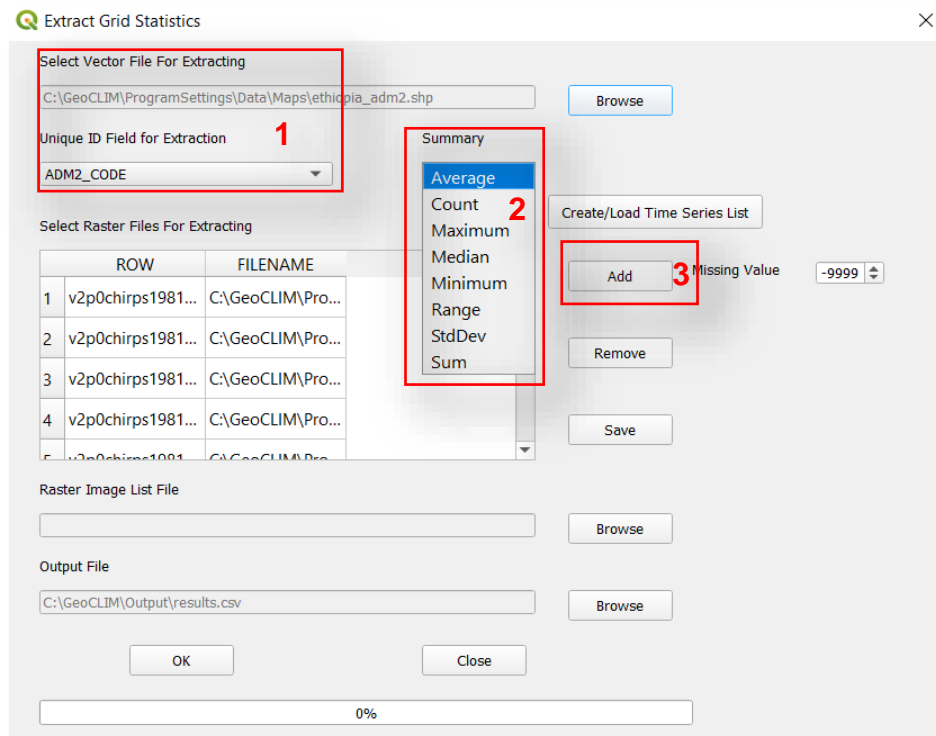
Figure 10-1 *The Extract Grid Statistics tool calculates summary statistics for a polygon (or a set of polygons) from a shapefile, using a raster or set of rasters fields.*

10.1. Extract Statistics

To extract summary statistics for a set of polygons, follow the steps below:
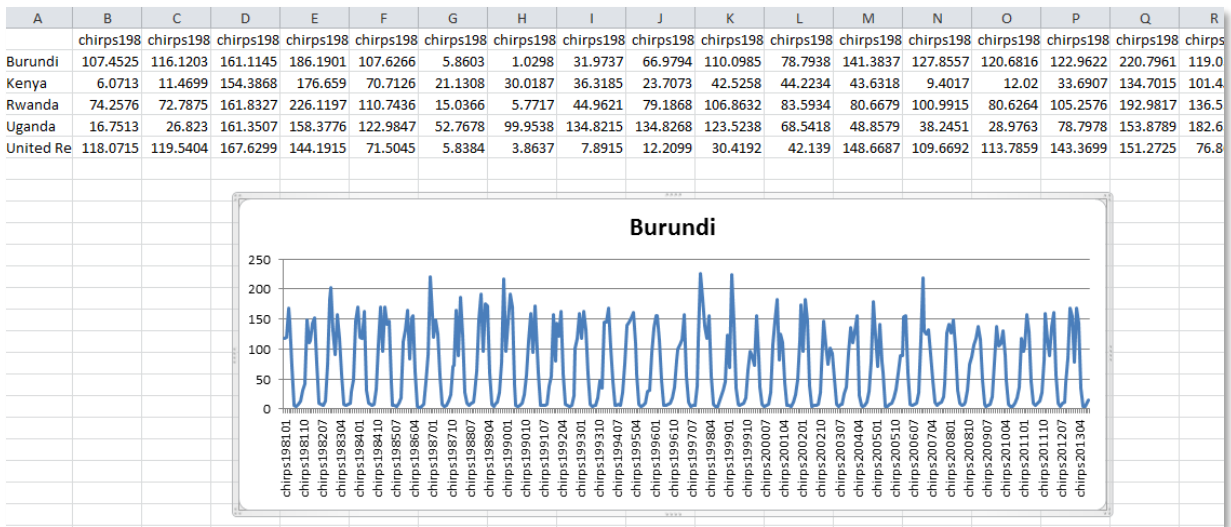
1. Open the **Extract Statistics from Raster Data using Shapefile** tool from the GeoCLIM toolbar.
2. Select a shapefile containing the polygons of interest (e.g., districts) and select a unique ID field (a data field in the shapefile that will uniquely identify each polygon, such as district names) Figure 10-2 **(1)**.
3. Select the type of summary for the pixels within the polygon, see the Summary∨ pulldown menu, Figure 10-2 **(2)**.
4. Select the raster file(s). Figure 10-2 **(3)**: Click on the Add button and browse to the directory where the raster files are located.
5. Select the files to be used in the process.
6. Click Open.
7. Back on the **Extract Grid Statistics** window, specify the output directory if necessary.
8. Click OK.

Figure 10-2 *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.*

10.2. 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 10-3 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 plots, open the CSV file in Microsoft Excel (or another 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	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 10-3 *The resulting table has a row for every polygon and every column represents the summary value for each raster.*
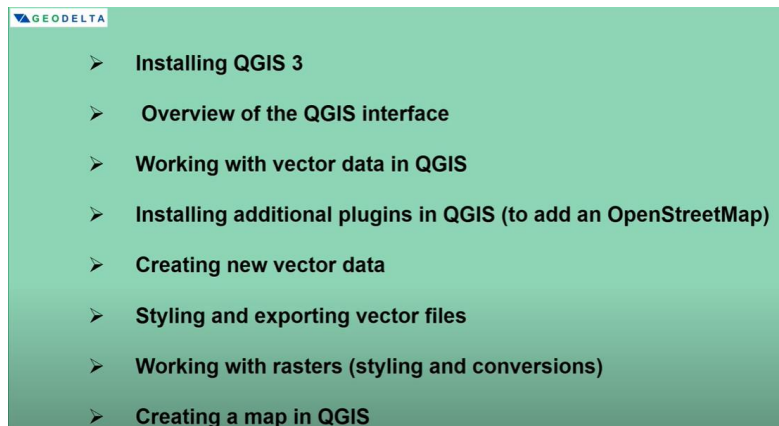
Appendix A

Summary

This section presents basic concepts of QGIS that are useful in the analyzes of climate data using FEWS Tools. Please find links to useful videos that present a variety of topics from uploading data to changing coordinate systems.

1. QGIS Basics
   a) https://youtu.be/NHolzMgaqwE

2. Loading Excel tables into QGIS https://youtu.be/rJwDrnJl4xg
3. Coordinates systems in QGIS https://youtu.be/p71qfALmdII

Acknowledgements

GeoCLIM development was supported by Tamuka Magadzire[1], Cheryl Holen[2], Minxuan Sun[2], Austin Christianson[2], Karthik Vanumamalai[2], and Joshua Sickmeyer[2] in support of the U.S. Agency for International Development's (USAID) Planning for Resilience in East Africa through Policy, Adaptation, Research, and Economic Development (PREPARED) Project, USAID FEWS NET, and Global Climate Change activities, in consultation with Chris Funk[1,3], Greg Husak[1], Gilbert Ouma[4], Ismael Mulama Lutta[5], and Gideon Galu[1]. The documentation was developed by Diego Pedreros [3], Tamuka Magadzire[1] with additional support from Claudia J. Young[2], Libby White [1], and Juliet Way-Henthorne [1]. Project oversight for the development of GeoCLIM was provided by James Verdin[6], Chris Funk[1,3], Mario Rodriguez[1], and James Rowland [3].

1. Climate Hazard Center (CHC), University of California, Santa Barbara

2. Contractor for United States Geological Survey (USGS), EROS Center

3. United States Geological Survey (USGS), EROS Center

4. University of Nairobi, Kenya

5. IGAD Climate Prediction and Application Center (ICPAC)

6. United States Agency for International Development (USAID)

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

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

GeoDelta Labs. (2020, July 16). An Absolute Beginner's Guide to QGIS 3 [Video]. YouTube. https://youtu.be/NHolzMgaqwE

Hart, Hal. (2018, October 23). QGIS Basics #10: Importing Excel Data [Video]. YouTube. https://youtu.be/rJwDrnJl4xg

Hart, Hal. (2019, August 2). QGIS Basic #89: CRS and Layer CRS Manipulation [Video]. YouTube. https://youtu.be/p71qfALmdII