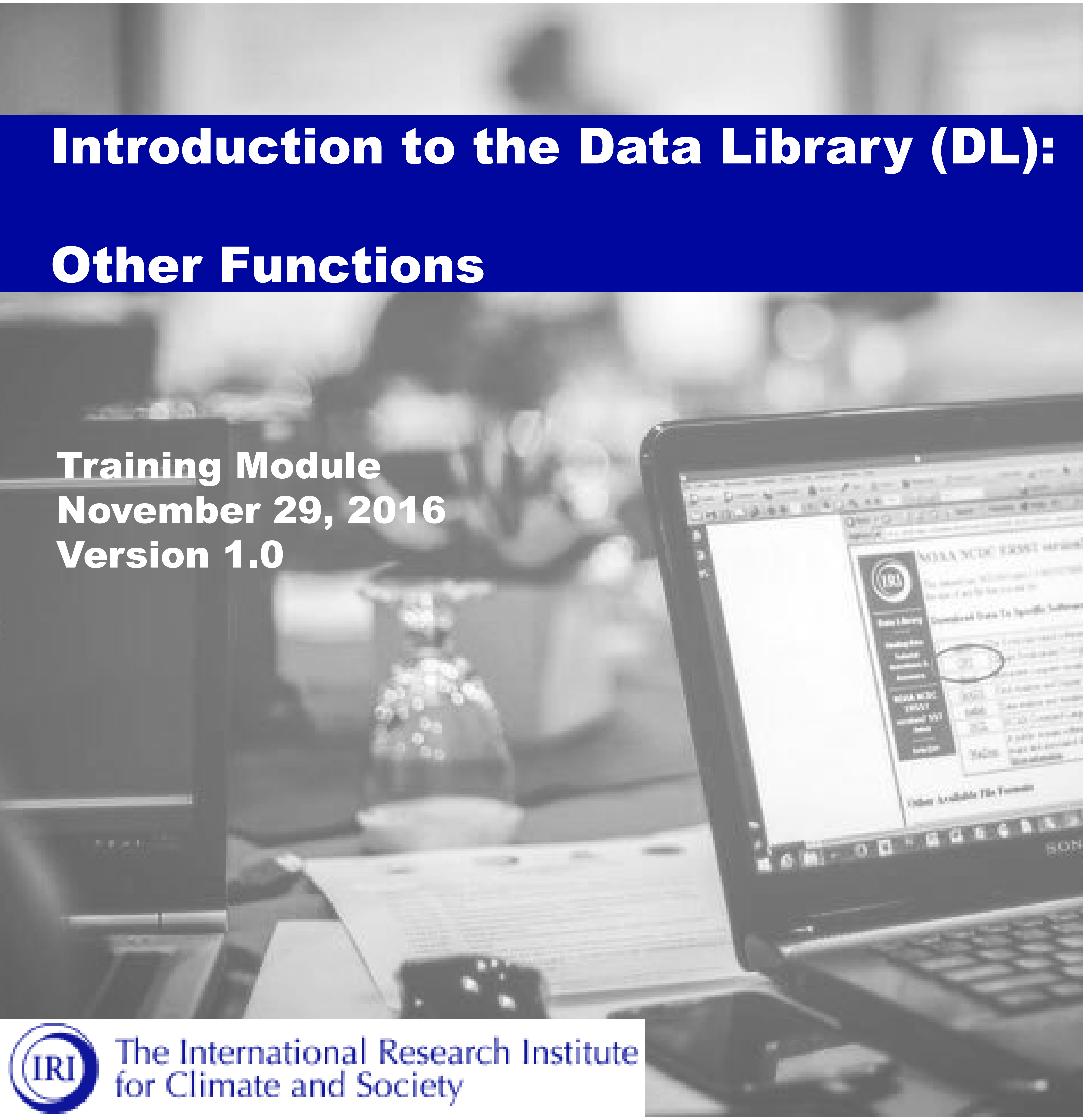
4. DRAFT

Introduction to the Data Library (DL): Other Functions



Training Module
November 29, 2016
Version 1.0

The International Research Institute for Climate and Society

International Research Institute for Climate and Society (IRI), (2016). Introduction to the Data Library (DL)- Other Functions. November 29, Version 1.0. Palisades: IRI.

CONTENTS

CHAPTER ONE

INTRODUCTION TO THE DATA LIBRARY (DL) - OTHER FUNCTIONS

1.1 Introduction

The IRI Climate Data Library is a library of datasets. By library we mean a collection of datasets, collected from various sources, designed to make them more accessible for the library's users (Bluementhal, 2004). For this module we will be expanding on how the users can select other function which will refine the data selection to obtain the desired information. These other functions include, adding, subtracting averaging etc. Traditional GIS platforms are now widely used by planners and decision makers in society. However, they are highly-focused on geospatial capabilities and have limited functionality for temporal analysis. Without information on the latter, meaningful inference about the causation of disease outbreaks is impossible (Jacquez 2000). Furthermore, many tools are unable to readily process the vast quantities of space-time data associated with, for example, the outputs of a global climate model. The IRI Climate Data Library overcomes the limitations imposed by GIS platforms by being based on a much more general multi-dimensional data model that includes both space and time dimensions. All datasets, including GIS features (such as points, lines, and polygons) are geo-located and temporally referenced in a uniform framework.

1.2 Overview

How are the functions categorized?

The Functions have 3 possible features:

- Standardized index
- Coefficient of variation on a variable
- A correlation index [T] with a variable [X Y T]

1.3 Access

The IRI Data Library can be accessed with the following links:

- Worldwide: http://iridl.ldeo.columbia.edu/
- Chile: http://www.climatedatalibrary.cl/
- Venezuela: http://datoteca.ole2.org/
- Uruguay: http://dlibrary.snia.gub.uy/
- Rwanda: http://maproom.meteorwanda.gov.rw/
- Ethiopia: http://www.ethiometmaprooms.gov.et:8082/

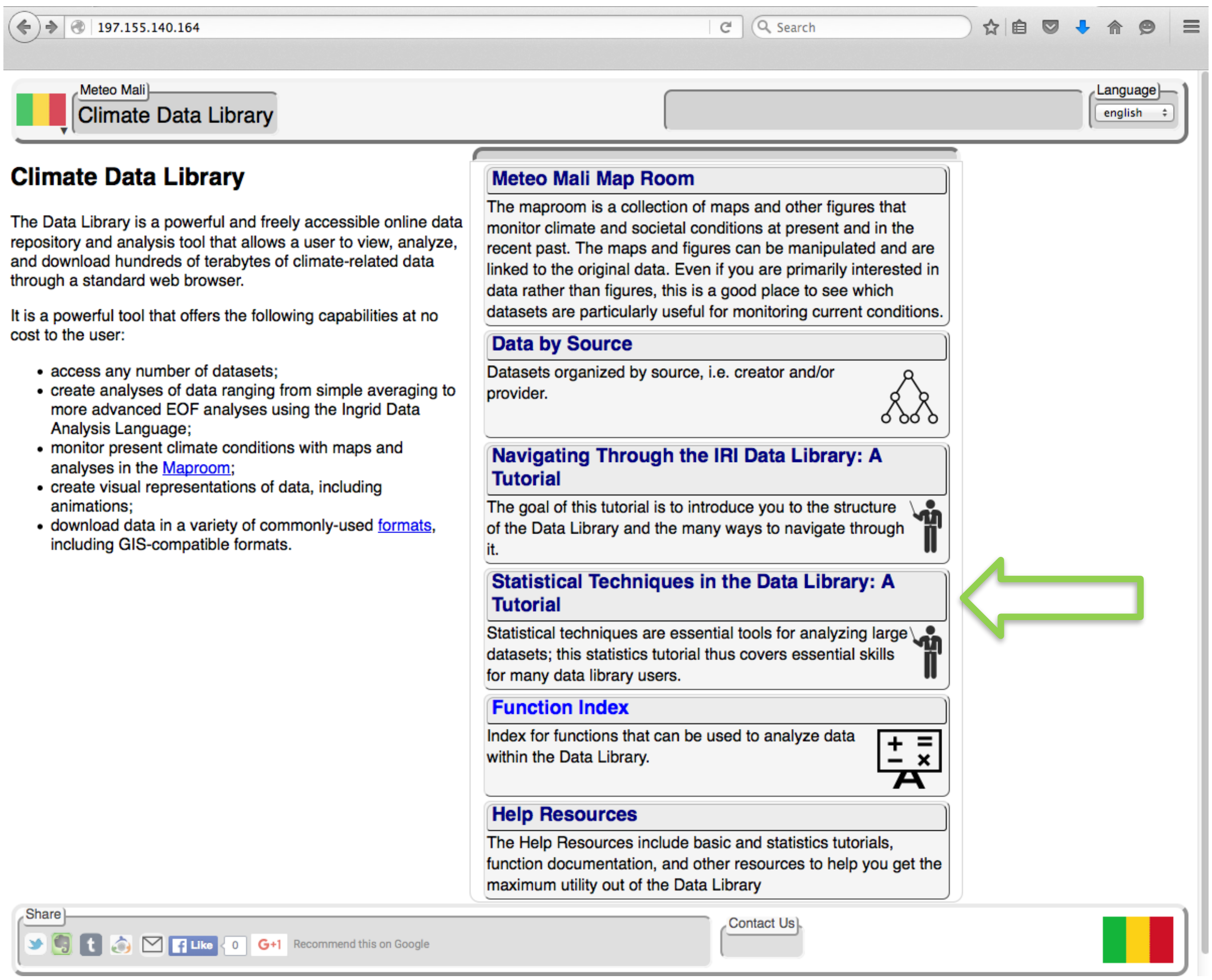


Fig. 1.1: Accessing Functions from dta Library: Statistical Techniques in Data Library - A Tutorial

- Tanzania: http://maproom.meteo.go.tz/
- Mali: http://197.155.140.164/
- Ghana: http://maps.meteo.gov.gh:89/
- Zambia: http://41.72.104.142/
- Madagascar: http://map.meteomadagascar.mg/
- Peru: http://ons.snirh.gob.pe/
- Niger: http://cradata.agrhymet.ne/
- Kenya (KMD): http://kmddl.meteo.go.ke:8081/
- Kenya (ICPAC): http://digilib.icpac.net/

1.4 Calculation of a Standardized Index

When calculating the standardized index, there are a number of functions that can be used as seen from the formulas in Figure 1.2.

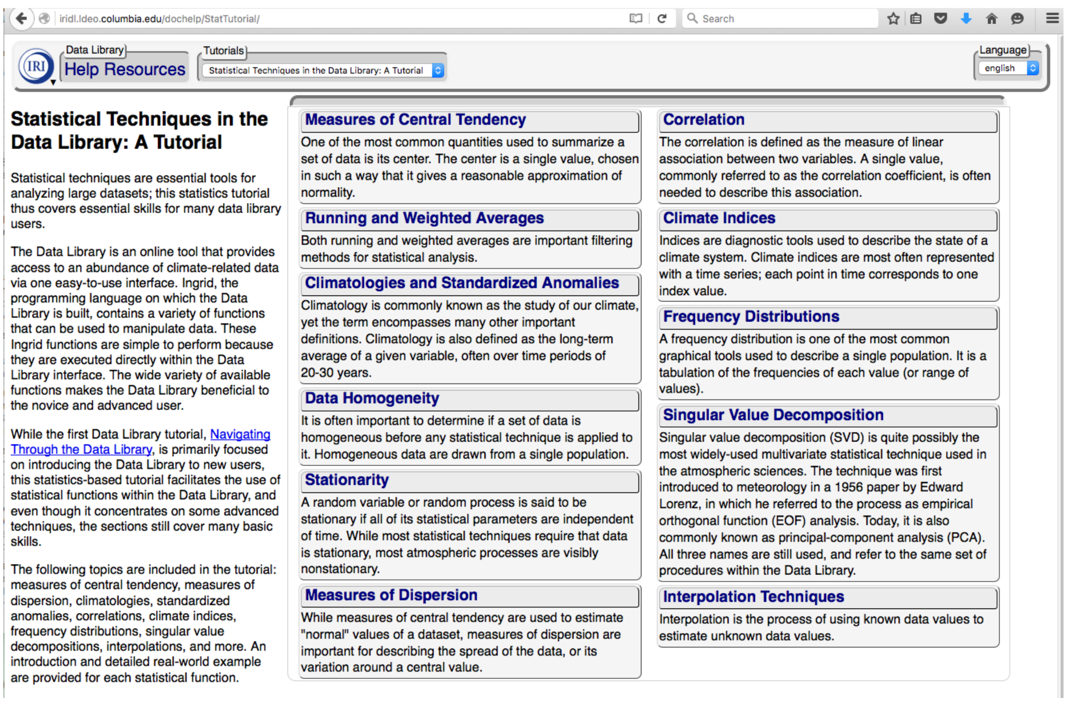


Fig. 1.2: Calculating standardized index

Hence the functions to use used are:

- average
- rmsaover

1.4.1 Example

Figure 1.3, displays the monthly weather station precipitation data from the GHCN. From this data we will be calculating the average standardized series on all stations as seen in Figure 1.4.

$$X = \frac{1}{m}\sum\frac{(x-\mu)}{\sigma}$$

$$\mu = \frac{1}{n}\sum x \qquad\qquad \sigma = \sqrt{\frac{1}{n-1}\sum\left(x-\overline{x}\right)^2}$$

Fig. 1.3: Monthly weather station data from GHCN

Firstly the average of each station needs to be calculated (b) as well as the standard deviation of each station (c). From here, subtract the mean (f) and divide by the standard deviation (j) on the time series of each station. Finally calculate the average standardized series on all stations (Figure 1.4). The functions utilized are: * dup

- exch
- add
- sub
- mul
- div

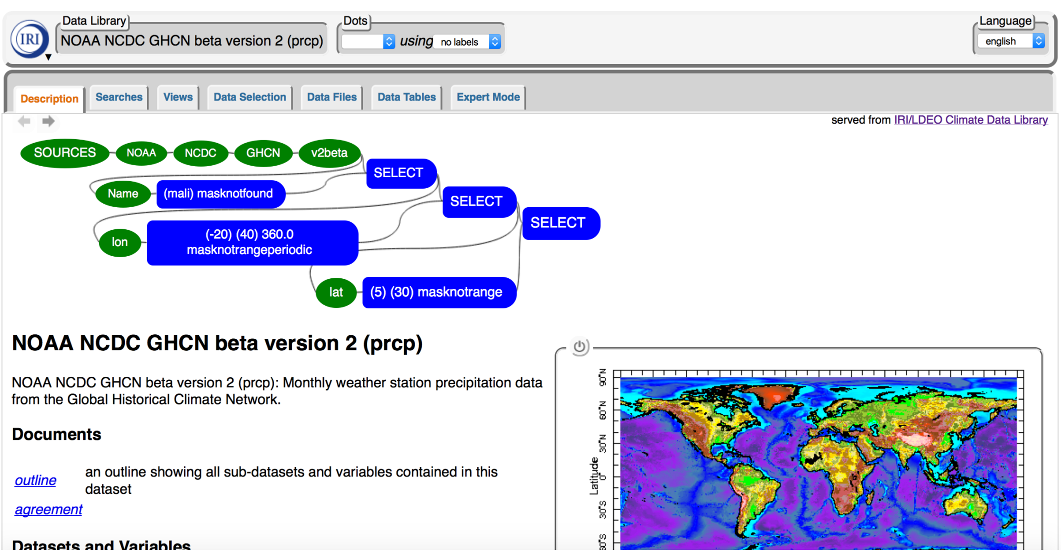


Fig. 1.4: Calculating average standardized index Example

In the computation of mean, standard deviation, coefficient of variation … These are specific uses of two generic functions:

average rmsaover

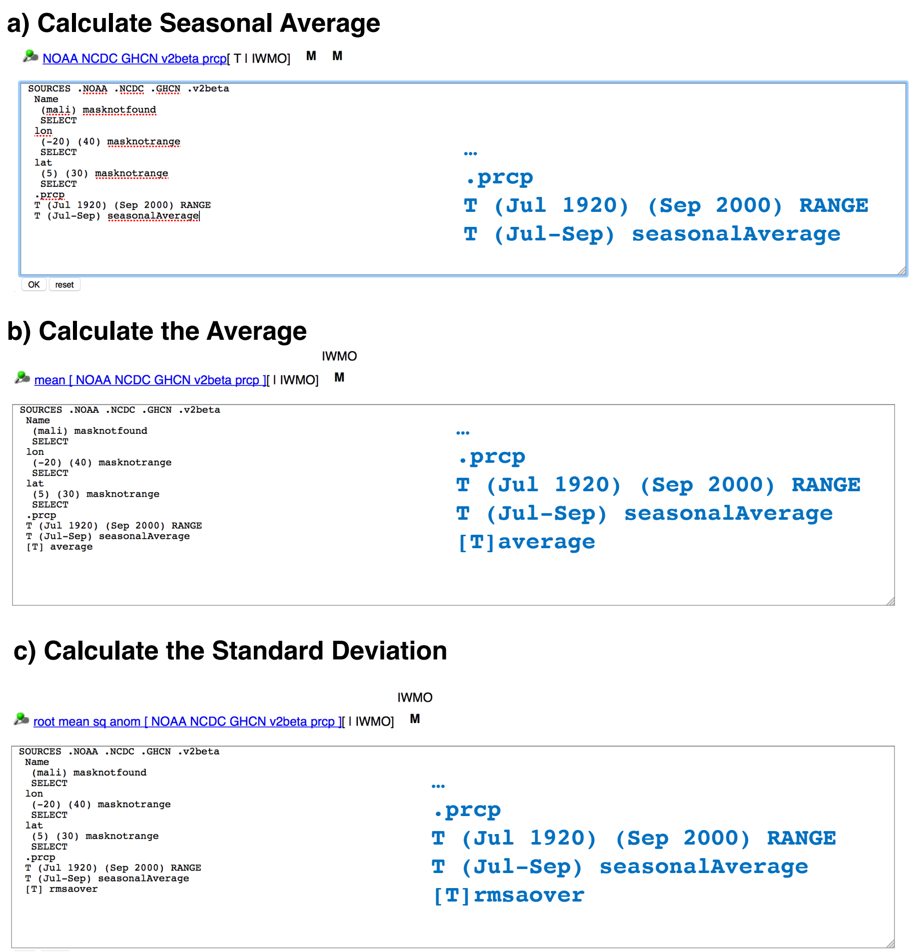


Fig. 1.5: Calculating average standardized index Example (steps: a-c)

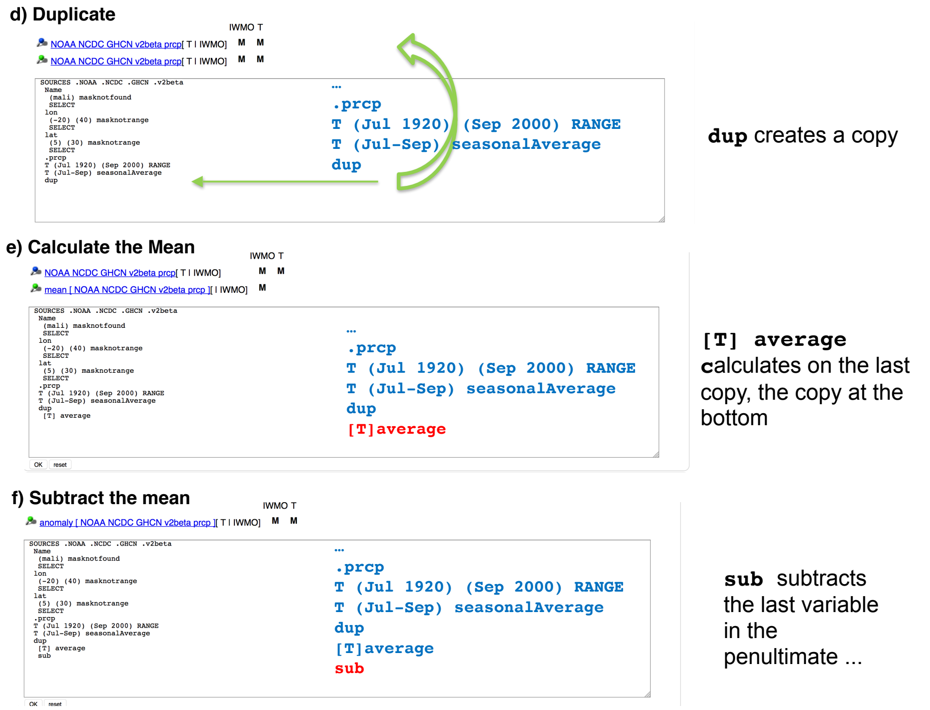


Fig. 1.6: Calculating average standardized index Example (Steps: d-f)

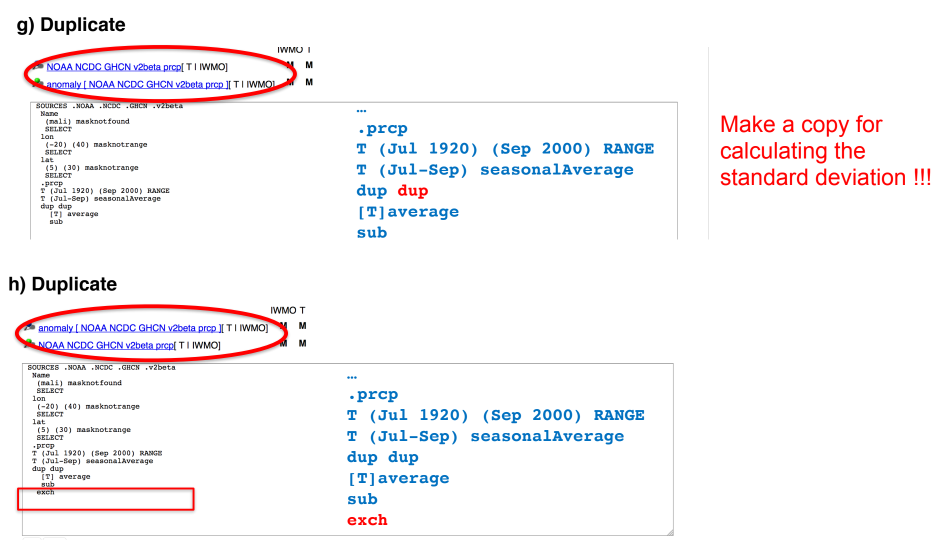


Fig. 1.7: Calculating average standardized index Example (Steps: g-h)

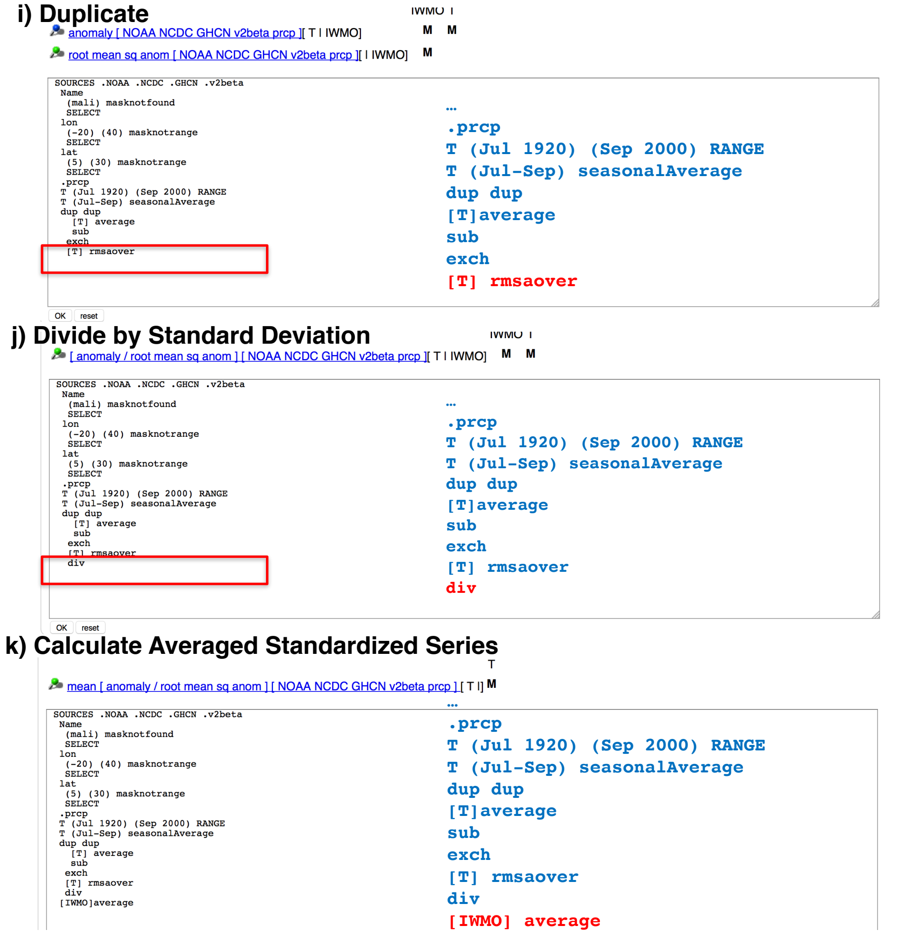


Fig. 1.8: Calculating average standardized index Example (Steps: i-j)

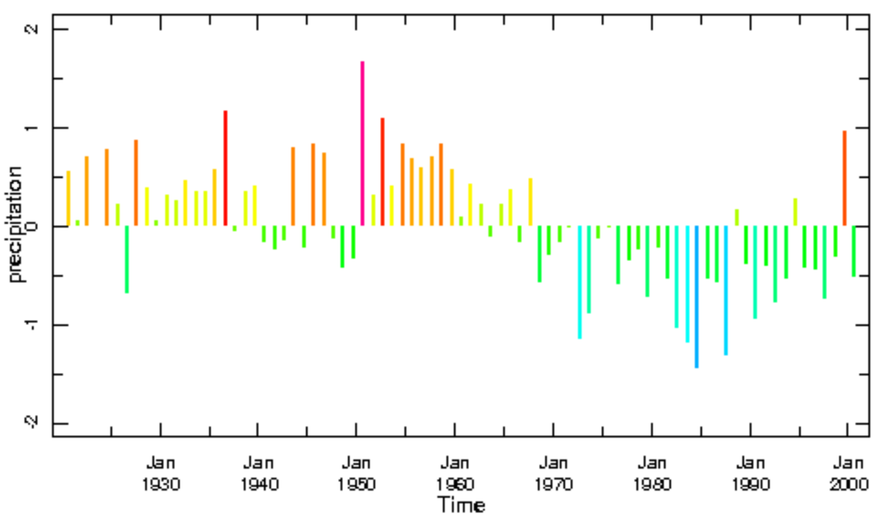


Fig. 1.9: Calculating average standardized index Example's Result

When applied to the time (T) grid use:

[T] average [T] rmsaover

1.5 Calculating a Coefficient of Variation on a Variable

The function to calculating coefficient of variation on a variable is the figure below (Figure 1.10). When there is a mesh of variables the following functions are useful to compute the coefficient: * average * rmsaover

$$\frac{\sigma}{\mu}$$

Fig. 1.10: Calculating the Coefficient

1.5.1 Example

If we go into the Meteo Mali Data Library and get the merged station-satellite rainfall data (Figure 1.11), the functions can be changed on expert mode just as Figure 1.12.

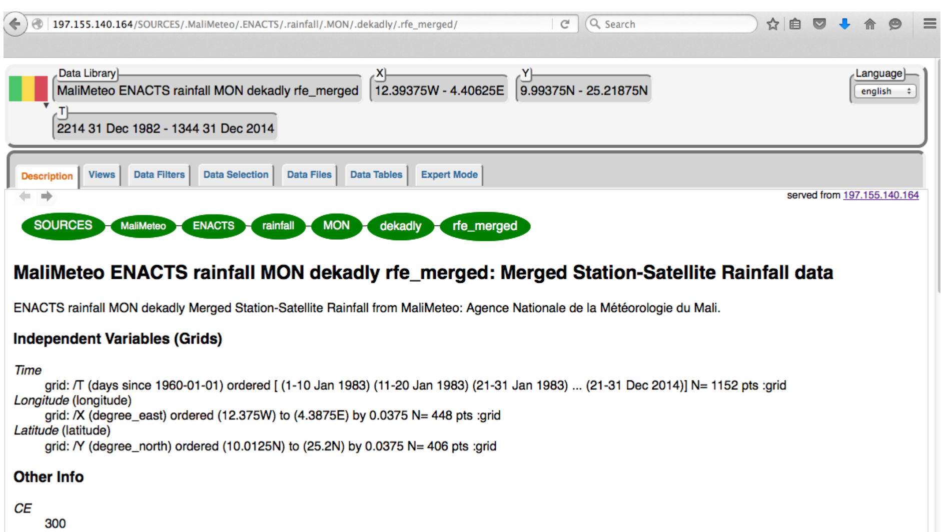


Fig. 1.11: Obtaining Merged Station-Satellite Rainfall Data from Meteo Mali Data Library

The result would then be obtain as shown in Figure 1.13.

```
SOURCES .MaliMeteo .ENACTS .rainfall
      .MON .dekadly .rfe_merged

dup

[T] rmsaover

exch

[T] average

div
```

Fig. 1.12: Adding the Rainfall Coefficient calculating functions to Expert Mode

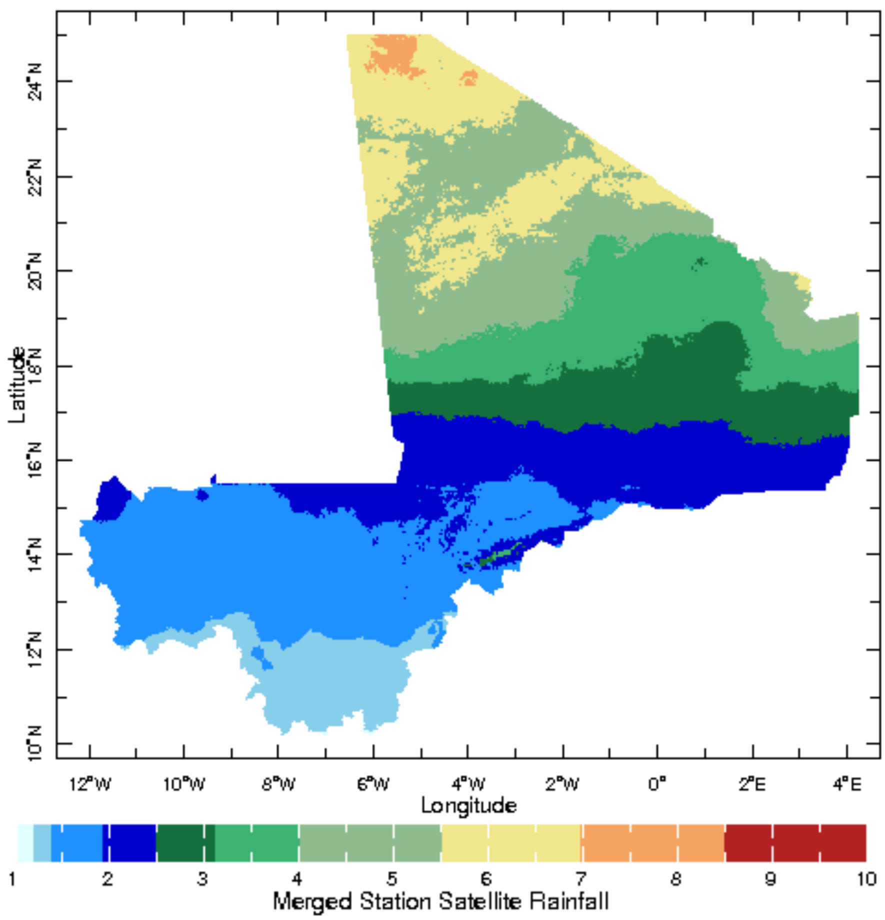


Fig. 1.13: Results of the Calculated Coefficient of the Merged Station Satellite Rainfall in Mali

1.6 Calculating a Correlation

Another function is calculating the correlation of an index [T] with variable [X Y T]. For this example we will use the index or [T] as indicies that are representative of the Niño 3.4 region (Figure 1.14):

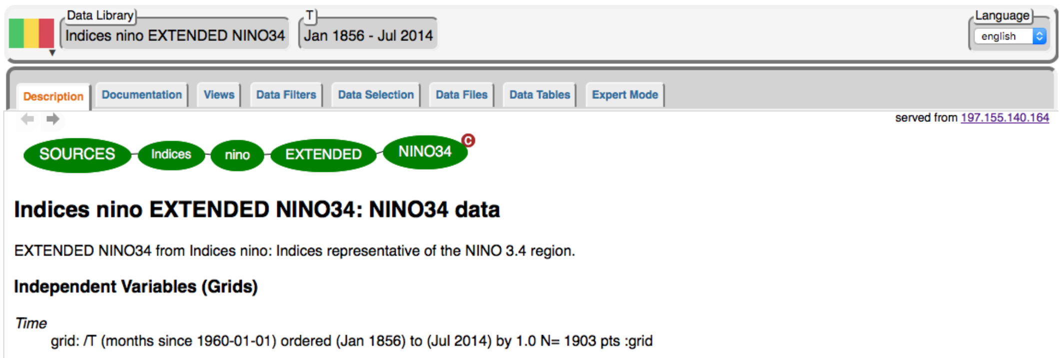


Fig. 1.14: Niño 3.4 region indicies

And as for the variable/ [X Y T], the merged station satellite rainfall data (Figure 1.15):

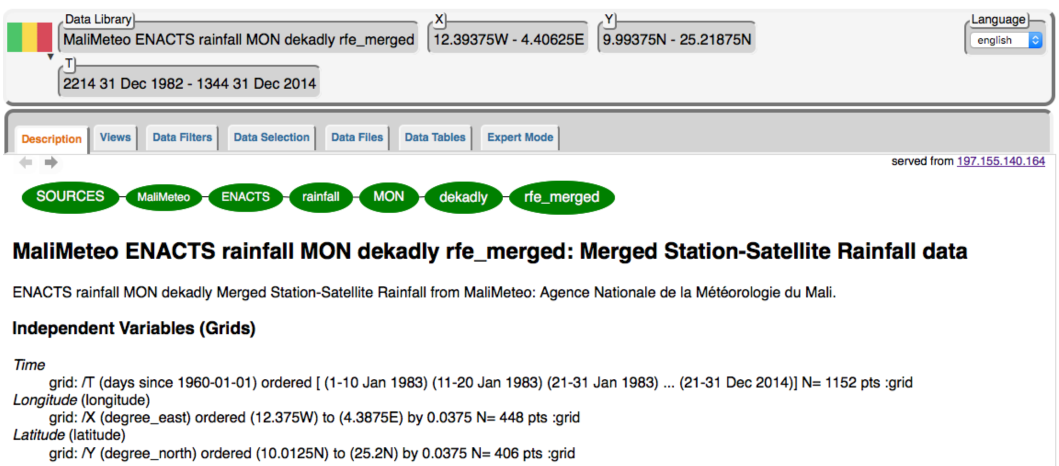


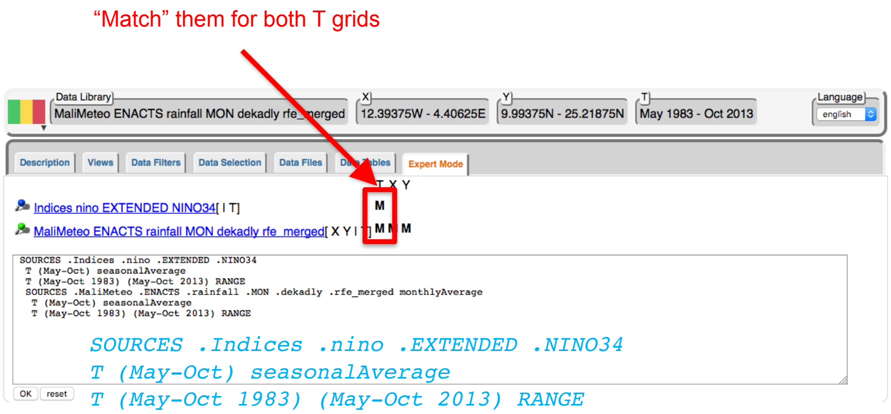
Fig. 1.15: Merged Station Satellite Rainfall Data as the Variable

When proceeding to find the correlation, it is important to compute simultaneous correlation over time, between the average of May-Oct of the two quantities and importantly, must make sure that the T grid is the same for both. Refer to Figure 1.16.

From here the user can go into the Function Documentation Tab and proceed to choose the "correlate" function (Figure 1.17)

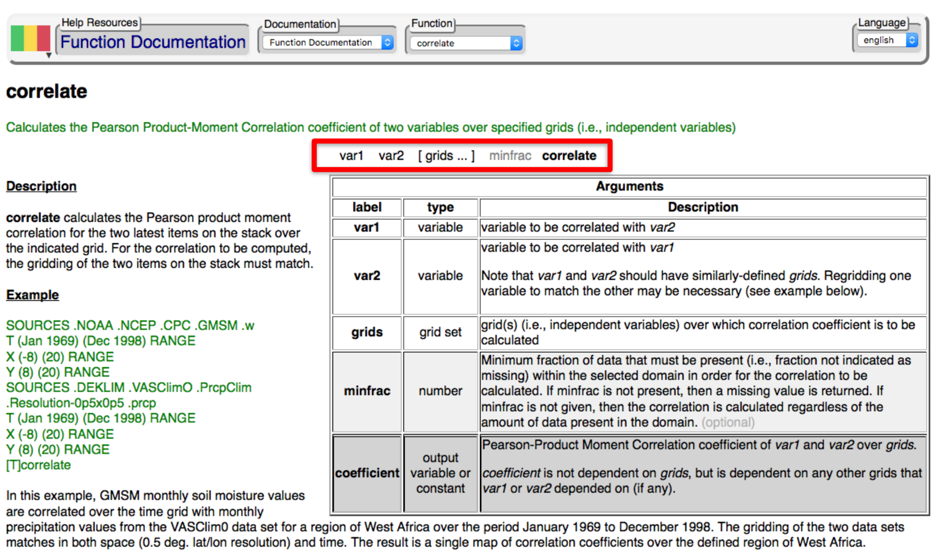
Hence the expect mode needs to look like Figure 1.18 for this example.

The result would then be obtain as shown in Figure 1.19.




```
SOURCES .Indices .nino .EXTENDED .NINO34
T (May-Oct) seasonalAverage
T (May-Oct 1983) (May-Oct 2013) RANGE
SOURCES .MaliMeteo .ENACTS .rainfall .MON .dekadly .rfe_merged
monthlyAverage
T (May-Oct) seasonalAverage
T (May-Oct 1983) (May-Oct 2013) RANGE
```

Fig. 1.16: Expert Mode: make sure T grid is the same in both


Help Resources: Function Documentation | Documentation: Function Documentation | Function: correlate | Language: english

correlate

Calculates the Pearson Product-Moment Correlation coefficient of two variables over specified grids (i.e., independent variables)

var1 var2 [grids ...] minfrac **correlate**

Description

correlate calculates the Pearson product moment correlation for the two latest items on the stack over the indicated grid. For the correlation to be computed, the gridding of the two items on the stack must match.

Example

SOURCES .NOAA .NCEP .CPC .GMSM .w
T (Jan 1969) (Dec 1998) RANGE
X (-8) (20) RANGE
Y (8) (20) RANGE
SOURCES .DEKLIM .VASClimO .PrcpClim
.Resolution-0p5x0p5 .prcp
T (Jan 1969) (Dec 1998) RANGE
X (-8) (20) RANGE
Y (8) (20) RANGE
[T]correlate

Arguments		
label	**type**	**Description**
var1	variable	variable to be correlated with *var2*
var2	variable	variable to be correlated with *var1* Note that *var1* and *var2* should have similarly-defined *grids*. Regridding one variable to match the other may be necessary (see example below).
grids	grid set	grid(s) (i.e., independent variables) over which correlation coefficient is to be calculated
minfrac	number	Minimum fraction of data that must be present (i.e., fraction not indicated as missing) within the selected domain in order for the correlation to be calculated. If minfrac is not present, then a missing value is returned. If minfrac is not given, then the correlation is calculated regardless of the amount of data present in the domain. (optional)
coefficient	output variable or constant	Pearson-Product Moment Correlation coefficient of *var1* and *var2* over *grids*. *coefficient* is not dependent on *grids*, but is dependent on any other grids that *var1* or *var2* depended on (if any).

In this example, GMSM monthly soil moisture values are correlated over the time grid with monthly precipitation values from the VASClimO data set for a region of West Africa over the period January 1969 to December 1998. The gridding of the two data sets matches in both space (0.5 deg. lat/lon resolution) and time. The result is a single map of correlation coefficients over the defined region of West Africa.

Fig. 1.17: Function Documentation: “correlate”

var1 var2 [grids …] minfrac correlate >>>

```
SOURCES .Indices .nino .EXTENDED .NINO34
T (May-Oct) seasonalAverage
T (May-Oct 1983) (May-Oct 2013) RANGE
SOURCES .MaliMeteo .ENACTS .rainfall .MON .dekadly .rfe_merged
monthlyAverage
T (May-Oct) seasonalAverage
T (May-Oct 1983) (May-Oct 2013) RANGE
[T] correlate
```

Fig. 1.18: Expert Mode featured

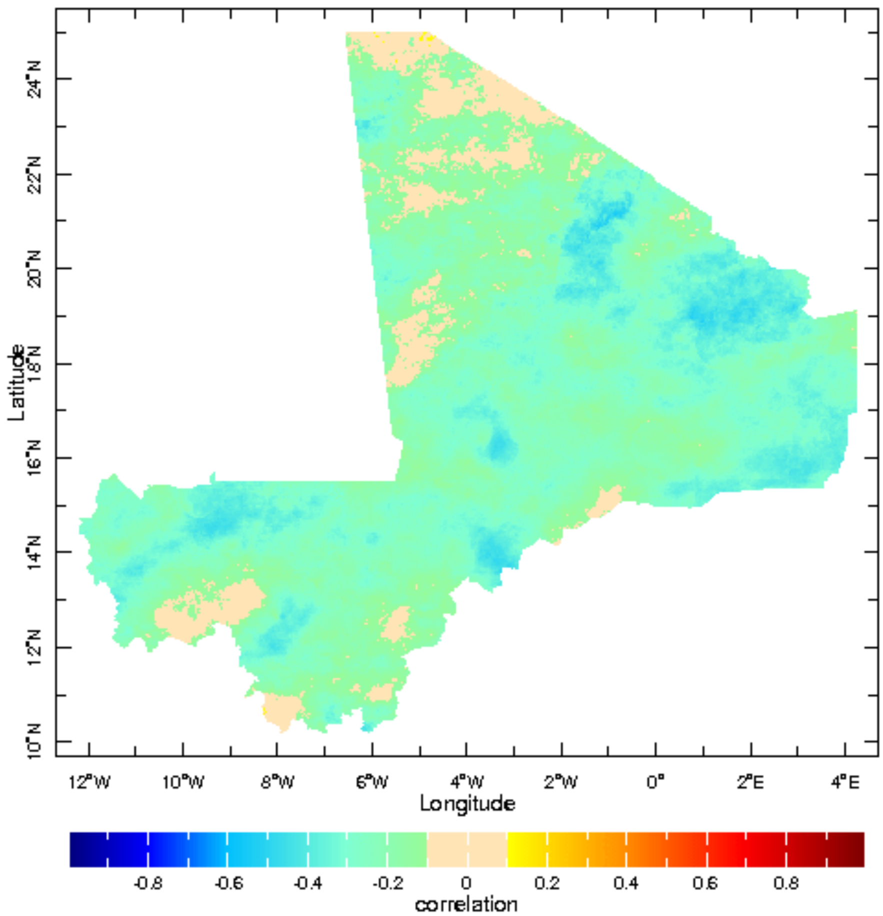


Fig. 1.19: Results of the Calculated Correlation between Niño 3.4 region indices and the Merged Station Satellite Rainfall in Mali

1.7 Summary

From this module the user is expected to have knowledge on how to select functions that allow averaging, subtracting dividing etc, to the selection and obtain a visualization of the results desired.

1.8 Quiz

Please answer the following questions using the IRI Data Library

Q1. What are the possible features of functions?

Q2. What are the functions used to calculate the standardized index?

Q3. How do you make a duplicate ?

Q4. When correlating and index with a variable what must be made sure of T grid?

1.8.1 Quiz - Answers

A1. * Standardized index

- Coefficient of variation on a variable
- A correlation index [T] with a variable [X Y T]

A2. average and rmsover.

A3. Duplicate or make copies using ‘dup’.

A4. T grid must be the same for both the index and the variable when computing the correlation.

1.9 Reference(s)

-