7. DRAFT

Introduction to the Data Library (DL):

Percentile [ranking], Composites

Training Module
November 29, 2016
Version 1.0

International Research Institute for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

International Research Institute for Climate and Society (IRI), (2016). Introduction to the Data Library (DL)- Percentile [ranking], Composites. November 29, Version 1.0. Palisades: IRI.

CONTENTS

CHAPTER

ONE

INTRODUCTION TO THE DATA LIBRARY (DL) - PERCENTILE [RANKING], COMPOSITES

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 calculate percentile and composites and obtain the desired information. 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

What will be working on?

For this manual we will be working on two examples:

- How to compute the climatological value that corresponds to a specified percentile
- How to use the "classify" function to composite according to an SST index

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/

- 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 How to calculate the xx-th percentile - the quantity that corresponds to the percentile

The functions in the function documentation we will be using are:

- [datarank] which replaces the value with its ranking (Figure 1.1).

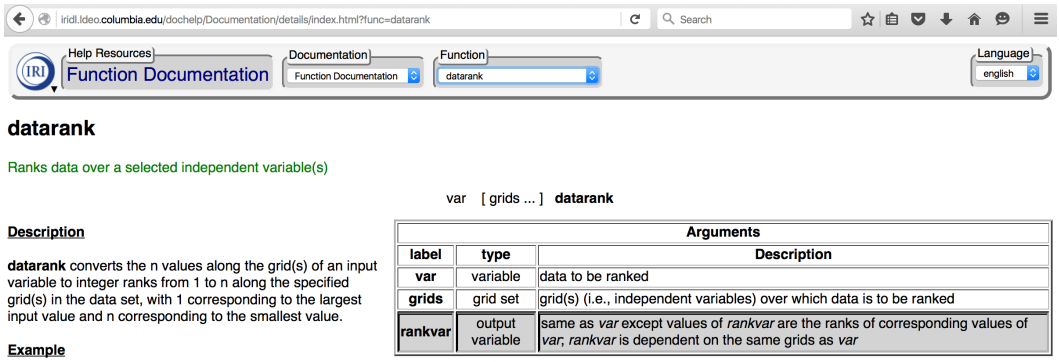

Fig. 1.1: [datarank] function

- [percentileover] which replaces the value with its corresponding percentile (Figure 1.2).
- Other functions which will be useful are the mask and flag family of functions:

mask – maskgt, maskge, masklt,maskle, maskrange, masknotrange flag – flaggt, flagge, flaglt, flagle (Figure 1.3)

Mask - values that satisfy the condition are replaced with “missing value” [NaN]

Flag - values that satisfy the condition are replaced by 1, those that do not satisfy the condition by 0

We use mask and flag functions because functions such as average, do not take into account “missing value”.

1.4.1 Example 1

For this example we will be looking at the average rainfall in the Sahel, Jul-Aug-Sep from 1901 to the year 2000. The general expert mode will show:

SOURCES .UEA .CRU .TS2p1 .monthly .prcp

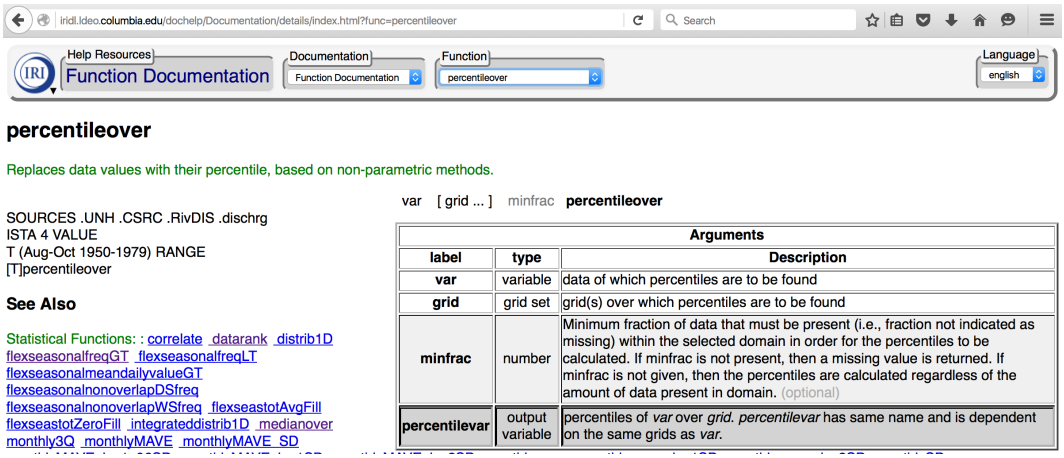

Fig. 1.2: [percentileover] function

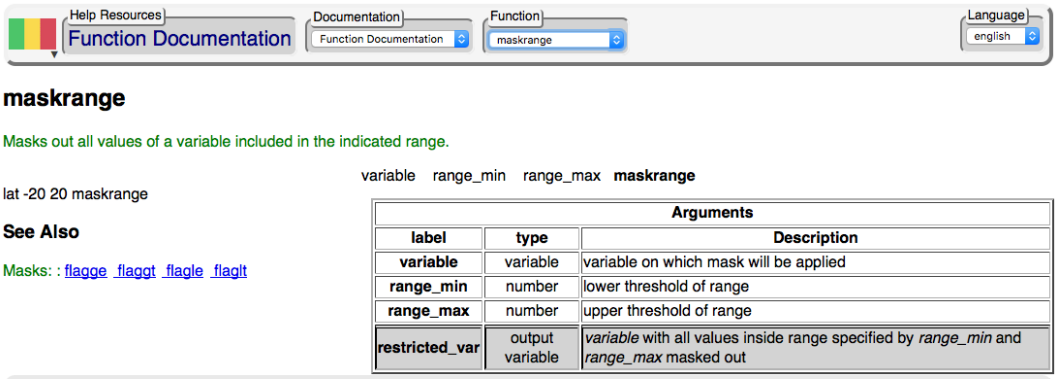

Fig. 1.3: Mask and Flag functions

X -20 40 RANGE

Y 12 18 RANGE

T (Jul 1901) (Sep 2000) RANGE

T 3 boxAverage

T 12 STEP

From here we have to look at how to calculate the 33-th percentile - the quantity that corresponds to the percentile. And the following points become the relevant steps:

- Identify the year that corresponds to the percentile
- Find the corresponding value

1.4.2 Identify the year that corresponds to the percentile

When identifying the year, the use can use the [percentileover] function as a mask function. Hence the expert mode function will be included and should be as follows:

SOURCES .UEA .CRU .TS2p1 .monthly .prcp

X -20 40 RANGE

Y 12 18 RANGE

T (Jul 1901) (Sep 2000) RANGE

T 3 boxAverage

T 12 STEP

[T] percentileover

0.33 0.34 masknotrange

1.4.3 The Results

The product of what has been done from the previous work will show the "missing values" except for the year that corresponds to the desired percentile (1966) which is seen in Figure 1.4.

If we multiply this variable by the values of the original variable, we only get the value that corresponds to the percentile (Before we have to transform 0.333 into 1). So the following should be on the expert mode:

SOURCES .UEA .CRU .TS2p1 .monthly .prcp

X -20 40 RANGE

Y 12 18 RANGE

T (Jul 1901) (Sep 2000) RANGE

T 3 boxAverage

T 12 STEP

dup

[T] percentileover

0.33 0.34 masknotrange

0 mul 1 add

Jul-Sep 1956	
Jul-Sep 1957	
Jul-Sep 1958	
Jul-Sep 1959	
Jul-Sep 1960	
Jul-Sep 1961	
Jul-Sep 1962	
Jul-Sep 1963	
Jul-Sep 1964	
Jul-Sep 1965	
Jul-Sep 1966	0.3333333
Jul-Sep 1967	
Jul-Sep 1968	
Jul-Sep 1969	
Jul-Sep 1970	
Jul-Sep 1971	
Jul-Sep 1972	
Jul-Sep 1973	
Jul-Sep 1974	
Jul-Sep 1975	
Jul-Sep 1976	
Jul-Sep 1977	

Fig. 1.4: The results with the "missing values" http://iridl.ldeo.columbia.edu/SOURCES/.UEA/.CRU/.TS2p1/.monthly/.prcp/X/-20/40/RANGE/Y/12/18/RANGE{[}X/Y{]}average/T/(Jul 1901)(Sep 2000)RANGE/T/3/boxAverage/T/12/STEP/dup[T]percentileover/0.33/0.34/masknotrange/

mul [T] average

We duplicate so we can apply the mask of multiplying with the average. And should give use the result (Figure 1.5) showing the 33th percentile of average rainfall in the Sahel in mm/month for the months of July - September from 1901 to 2000.

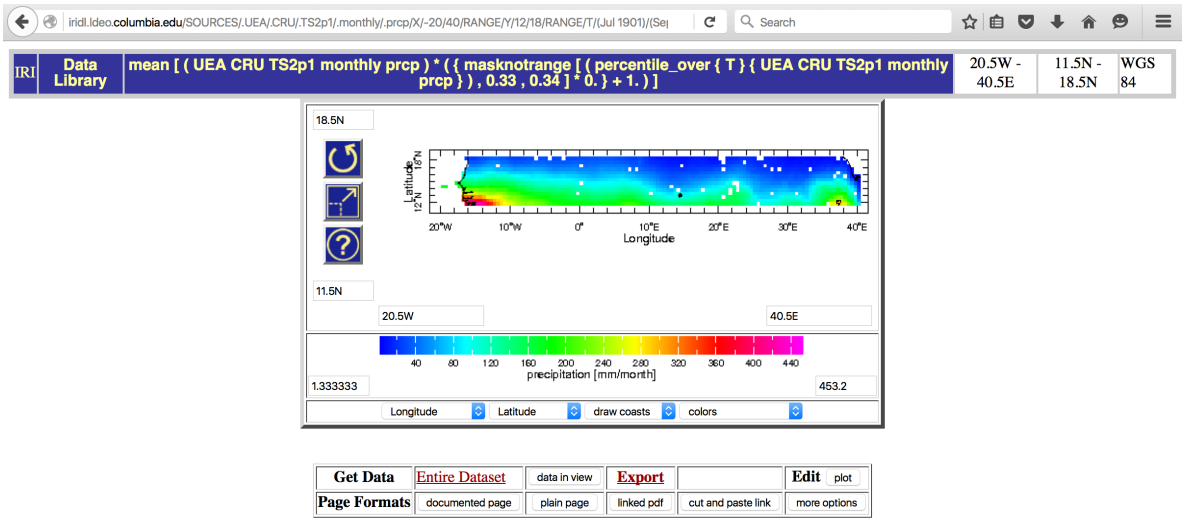

Fig. 1.5: 33th percentile of average rainfall in the Sahel in mm/month for the months of July - September from 1901 to 2000.

1.4.4 Example 2

This example will be exploring how to calculate the average over a selection of years, for instance indexed by a predicator (e.g., SST index). To do so the function to keep in mind is [classify] seen in Figure 1.6.

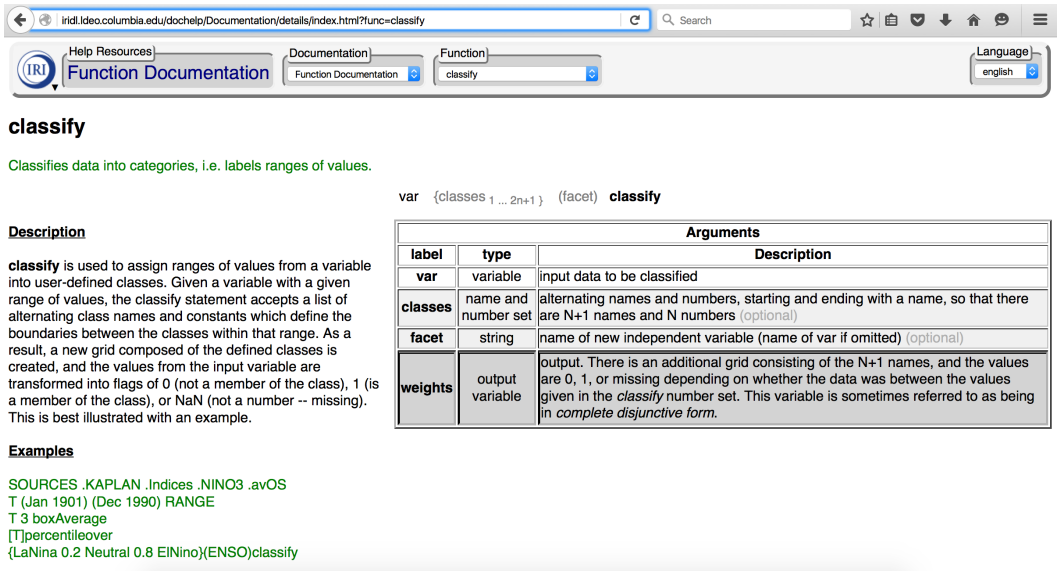

Fig. 1.6: [classify] function http://iridl.ldeo.columbia.edu/dochelp/Documentation/details/index.html?func=classify

For this example please go into the Climate Forecasting Maproom of Mali and refer to the steps in Figure 1.7.

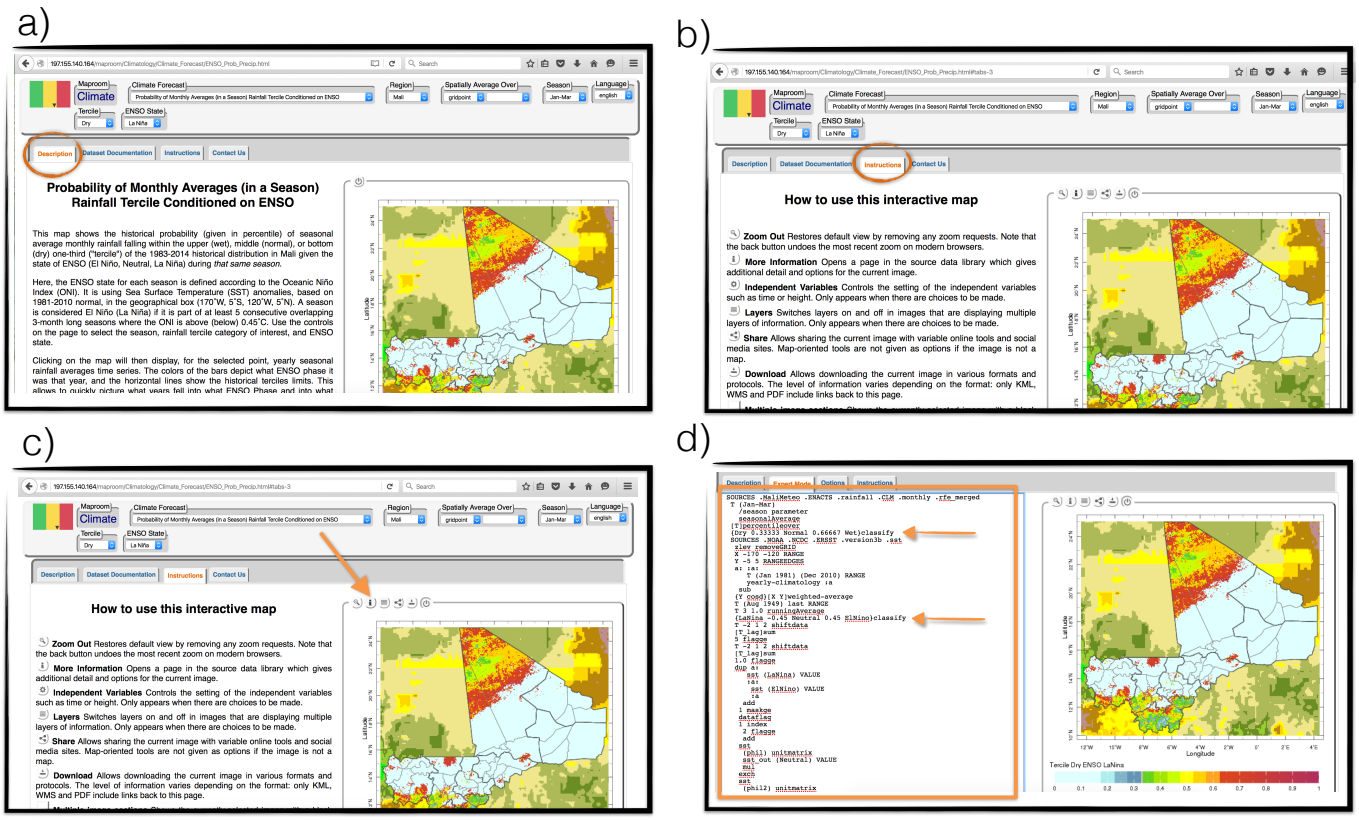

Fig. 1.7: Calculating the average over a selection of years (a) Look at the 'description' tab (b) 'Instructions' tab gives an overview of the buttons (c) Go into 'More Information' and go into (d) 'Expert mode" http://197.155.140.164/maproom/Climatology/Climate_Forecast/ENSO_Prob_Precip.html

Under 'Expert Mode' there will be the [classify function is displayed as: var {classes1 . . . classesN+1} (facet) classify (Seen in Figure 1.7 (d)).

1.4.5 Example 3

For this third example we will be constructing the average of the years of La Niña or El Niño, or indexed by another SST index. So there will be two datasets:

Dataset 1 - Rainfall Anomalies on the JAS season

SOURCES .MaliMeteo .ENACTS .rainfall .CLM .monthly .rfe_merged

T (Jul-Sep) seasonalAverage

dup

[T] average sub

SST anomalies in the Nino3.4 region

SOURCES .NOAA .NCDC .ERSST .version3b .sst

zlev removeGRID

X -170 -120 RANGE

Y -5 5 RANGEEDGES

a: :a: T (Jan 1981) (Dec 2010) RANGE

yearly-climatology :a

sub {Y cosd}[X Y]weighted-average

T (Jul-Sep) seasonalAverage

So to get the compute the average of the years the classify function is added and should read as follows:

SOURCES .MaliMeteo .ENACTS .rainfall .CLM .monthly .rfe_merged

T (Jul-Sep) seasonalAverage

dup

[T] average sub

SOURCES .NOAA .NCDC .ERSST .version3b .sst

zlev removeGRID

X -170 -120 RANGE

Y -5 5 RANGEEDGES

a: :a: T (Jan 1981) (Dec 2010) RANGE

yearly-climatology :a

sub

{Y cosd}[X Y]weighted-average

T (Jul-Sep) seasonalAverage

{LaNina -0.45 Neutral 0.45 ElNino} (enso_phase) classify

This says that the values of the SST index less than -0.45 are classified “La Niña”, those between -0.45 and 0.45 “neutral” and greater than 0.45 as “El Niño”. Therefore the [classify] function will add a match/dimension (look at Figure 1.8)

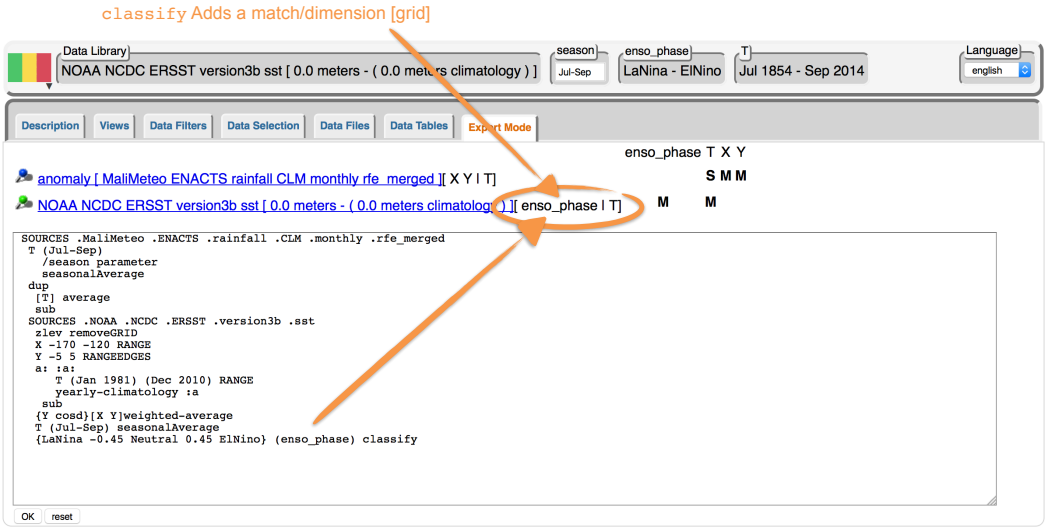

Fig. 1.8: [classify] function adding a match/dimensions

To proceed we will use La Niña phase as an example hence the expert mode should have enso_phase (LaNina) VALUE mul [T]average added to it just as follows (also look at Figure 1.9):

SOURCES .MaliMeteo .ENACTS .rainfall .CLM .monthly .rfe_merged

T (Jul-Sep) seasonalAverage

dup [T] average sub

SOURCES .NOAA .NCDC .ERSST .version3b .sst

zlev removeGRID

X -170 -120 RANGE

Y -5 5 RANGEEDGES

a: :a: T (Jan 1981) (Dec 2010) RANGE

yearly-climatology :a

sub

{Y cosd}[X Y]weighted-average

T (Jul-Sep) seasonalAverage

{LaNina -0.45 Neutral 0.45 ElNino} (enso_phase) classify

enso_phase (LaNina) VALUE

mul

[T]average

1.4.6 The Resulting Composite

The resulting composite of the La Niña years is seen in Figure 1.9 below:

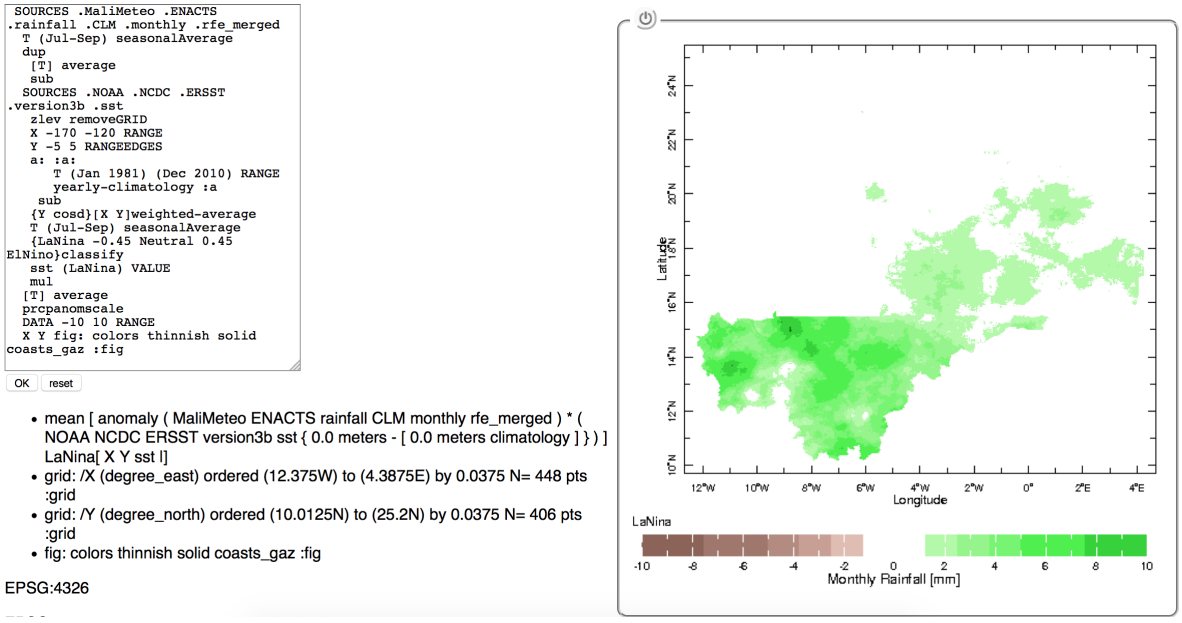

Fig. 1.9: Composite Results of La Niña years

1.4.7 Side note

This also shows that the user can build an index on any variable as seen in Figure 1.10.

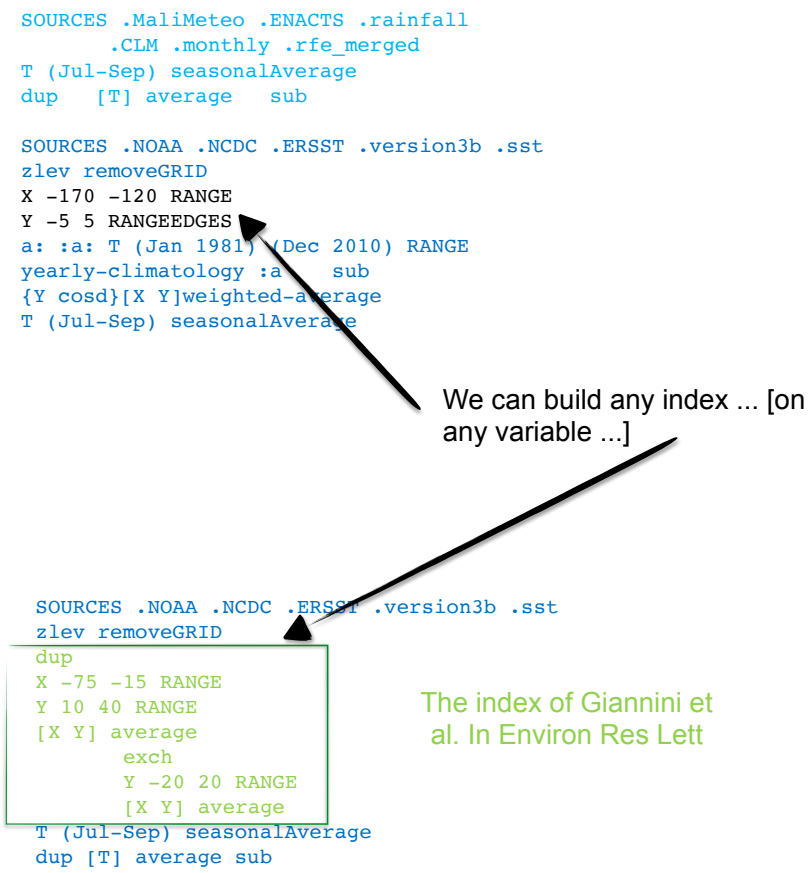

Fig. 1.10: Building and index on any variable (a) variables (b) building an index on the variable

1.5 Summary

From this module the user is expected to have knowledge on how to obtain a percentile and compute composites.

1.6 Quiz

Please answer the following questions using the IRI Data Library

Q1. What does the [classify] function do?

Q2. What doe the [datarank] function do?

Q3. What does the [percentileover] function do?

1.6.1 Quiz - Answers

A1. [classify] function classifies data into categories, i.e. labeled ranges of values.

A2. [datarank] function ranks data over a selected independent variable(s).

A3. [percentileover] function replaces data values with their percentile, based in non-parametric methods.

1.7 Reference(s)

-