JRC Technical Notes

Software description: Regional frequency analysis of climate variables (REFRAN - CV)

Cesar Carmona-Moreno, Eduardo Eiji Maeda, Juan Arevalo, Marco Giacomassi, Paolo Mainardi

EUROCLIMA

EUR 24947 EN - 2011

JRC
EUROPEAN COMMISSION

Institute for Environment and Sustainability

The mission of the JRC-IES is to provide scientific-technical support to the European Union's policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
Cesar Carmona – Moreno
Address: TP 440 via Fermi 2749, 21027 ISPRA , Italy
E-mail: cesar.carmona-morena@jrc.ec.europa.eu
Tel.: 0039 0332 78 9654
Fax: 00 39 0332 78 9960

http://www.jrc.ec.europa.eu/

JRC 66809

EUR 24947 EN
ISBN 978-92-79-21322-9
ISSN 1831-9424
doi:10.2788/74447

Luxembourg: Publications Office of the European Union, 2011

Printed in Italy

Contents

BACKGROUND

This software is to be developed in the context of the EUROCLIMA project. EUROCLIMA is a cooperation program between the European Union and Latin America with a special focus in knowledge sharing on topics related to socio-environmental problems associated with climate change.

The overall goals of the EUROCLIMA initiative are:

- Development of tools to reduce people's vulnerability to the effects of climate change in conjunction with the fight against poverty;
- Reduction of social inequalities, especially those induced by climate change issues, facilitating social sustainable development;
- Reduction of the socio-economic impacts of climate change through cost-efficient adaptations, capable of generating sub-regional and regional synergies;
- Reinforcement of regional integration dialogue with the aim of setting up a permanent consultation mechanism for a joint review of shared goals.

The specific objective of the project is to improve knowledge of Latin American decision-makers and the scientific community on problems and consequences of climate change, particularly in view of integrating these issues into sustainable development strategies.

In order to achieve these goals, it is crucial, for both policy makers and researchers, to understand climate variability at local-regional-continental scales. In this context, the software described in this document represents an initial effort to gather and process climate data available in Latin America, in order to produce concise and clear information about the variability of key climatic variables, such as precipitation and temperature.

GENERAL CONCEPT

The software will have as a general objective to process time series of data from ground stations (initially precipitation and temperature) in order to generate products in the form of spatially-explicit maps. However, the software will be able to process any other time series of environmental spatial data (vegetation, NDVI, evapotranspiration, FAPAR, …).

The contractor will developed the software and will provide also a user and installation manual of the software. The contractor will deliver a fully automatic installation module working in multi-platform environments. The software will be developed under the OPEN SOURCE principles. The software will be developed by the contractor in close collaboration with CAZALAC (Chile), CIIFEN (Equator), UNAL-IDEA (Colombia), TEM (Mexico), INSMET (Cuba) and other Latin American institutions to be defined. These Latin-American institutions will largely contribute to the detailed specification of the software, the design phase, the user validation phase and the in-site implementation.

The main aspect characterizing this software is the use of statistics called L-moments to estimate the probability distribution function of climate variables. The L-moments are similar to other statistical moments, but with the advantage of being less susceptible to the presence of outliers and performing better with smaller sample sizes.

For a random variable X, the first four L-moments are given by the following equations:

$\lambda_1 = E[X]$
$\lambda_2 = E[X_{2:2} - X_{1:2}] / 2$
$\lambda_3 = E[X_{3:3} - 2X_{2:3} + X_{1:3}] / 3$
$\lambda_4 = E[X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}]/ 4$

For convenience, the second, third and forth L-moments are often presented as L-moment ratios:

$\tau_2= \lambda_2/ \lambda_1$
$\tau_3= \lambda_3/ \lambda_2$
$\tau_4= \lambda_4/ \lambda_2$

The 1st L-moment (L-mean) is identical to the conventional statistical mean. The 2nd L-moment (L-cv) measures a variable's dispersion or the expected difference between two random samples. The 3rd and 4th L-moment (L-skewness and L-kurtosis) are measures relating to the shape of the samples distribution. The L-skeweness quantifies the asymmetry of the samples distribution and the L-kurtosis measures whether the samples are peaked or flat relative to a normal distribution.

The data processing will be functionally divided in six modules. The outputs of each module will be partially or entirely used as input for the following module. The modules will be an integrated part of the software, but they will have the ability of running independently, that is to say, the user will have to possibility of running any module at any time, as long as the user have the necessary input dataset; but he will have also the possibility to run all the different modules in a unique run using a dataset of default parameters.

The first module has the objective of checking the raw dataset for error and formatting the climate records into a standard format for the next module. The second module aims to cluster the dataset of ground stations with similar climatic characteristics, forming the so called "homogeneous regions". In the third module, a probability distribution function is defined for each homogeneous region, in order to characterize the precipitation/temperature frequencies observed in the stations belonging to that group. After the distribution functions for each station is defined, it is necessary to interpolate this information for regions without a ground station. The parameters necessary for this interpolation are defined in the forth module and used in the fifth module to construct L-moments maps. Finally, in the sixth module, the L-moment maps are used to assess climate variability through a variety of informative maps.

The framework of the data processing procedures is illustrated in Figure 1. A detailed description of each module will be described in separated sections.

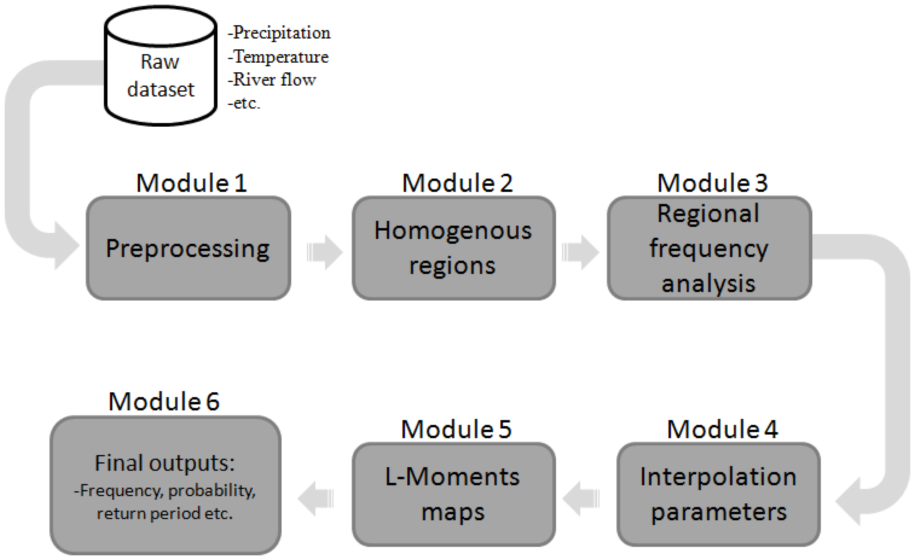

Figure 1. General structure of the software functionalities.

#---

Every time a Module is executed, the software will generate a log-file containing additional parameters defined by the user, name of the input/output files, date and time of the execution. Furthermore, errors and warning messages will also be stored in the log-files. The software will have also the possibility of running on the base of a parameter file directly read by the module interface.

The log-file name will be as follow: L-moments-<day_of_the_year>-<hh:mm>

However, additional options should be provided so the user can change the default names of the log-files.

#---

The user should also have the option of running all the modules in "one click", using default or user defined parameters included in a parameter file. Using the GUI of the software, the user will define the path for reading the "parameter" file. This "parameter file" will contain the parameters of a module to be run or all the parameters needed by the software for running all the modules. A "default parameter" file will also be developed for giving the user the possibility of running all the modules of the software.

#---

The general software concept is to obtain a user friendly interface, which will run in the background algorithms developed for R.

The script examples provided in this document are not exhaustive, meaning that not all software capabilities are mentioned in the examples. Also, not all procedures written in the examples are necessarily part of the software.

The examples used in this document were extracted from: Nuñez, J. 2011. RSARFLM v.1. Regional Frequency Analysis L-moments R script. Water Center for Arid and Semiarid Zones of Latina America and the Caribbean.CAZALAC.La Serena,Chile

R script example: Loading necessary R packages

```
#Module 1: System setup
#---------------------------------------------------------------------------
#Install packages
install.packages("lmom")
install.packages("lmomRFA")
install.packages("nsRFA")
install.packages("raster")
install.packages("rgdal")
install.packages("sp")
install.packages("DEoptim")
install.packages("sqldf")
install.packages("tcltk")

# Load packages
library(lmom)
library(lmomRFA)
library(nsRFA)
library(raster)
library(rgdal)
library(sp)
library(DEoptim)
library(sqldf)
library(tcltk)
#PASO 3: Select working directory
WF<-tk_choose.dir(getwd(), "Choose a suitable folder")
setwd(WF)
#---------------------------------------------------------------------------
```

Module 1 – Load data and preprocessing

Module 1 will perform a quality check in the dataset to verify potential bad values associated with data measurement errors. This module will also be responsible for formatting the dataset provided by the user into a standard format to be used by the following module. The methods used for the quality check are:

- Homogeneity check using double mass curve analysis (WMO, 1994);
- Stationality check, using linear regression analysis;
- and autocorrelation test, using the Lag-1 test for serial independence (Wallis et al., 2007).

The result of the quality check assessment will be presented for the user. Next, the user will have the option of performing a simple data imputation procedure (missing values replaced by mean, mode or nearest neighbour values) and, if desired, perform the quality check again.

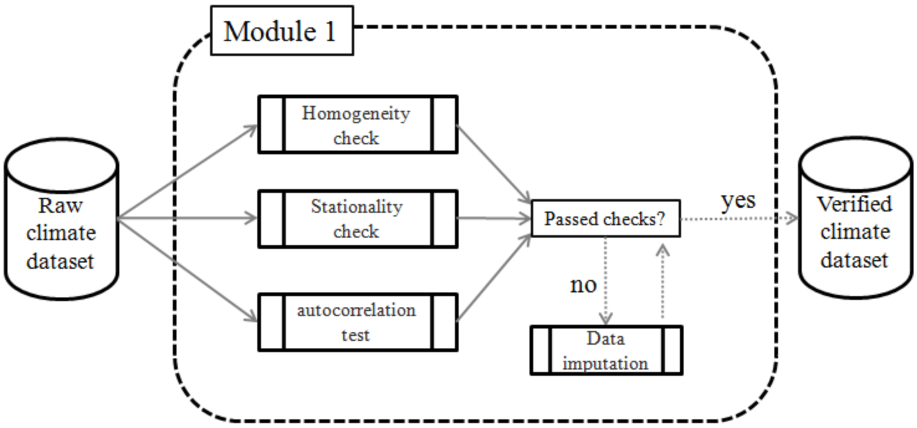

Figure 2. Module 1 data flow.

Inputs [format]	Outputs [format]
• Raw precipitation and temperature datasets [.xls, .xlsx or .csv]	• Number of missing records [on screen] • Number of error records [on screen] • Number of fixed records [on screen] • Verified dataset [xls, .xlsx or .csv] • Possibility to save a summary of the results in .txt or .csv

The user will have the option of providing the input dataset in two formats.

a) Format provided by the Global Historical Climatology Network (GHCN).

b) User defined structure.

The data GHCN has the advantage of providing thousands of temperature and precipitation stations around the globe, with a standard format of data files. Each data file (".dly" format) contains information about the country where the station is located, ID, year, month and a detailed specification of the records. A description of this dataset can be found in the following address:

http://www1.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt

Furthermore, GHCN provides simplified data inventory files, with location, time series length and ID for each station.

When the user defined option is chosen, the user will have to provide basic information necessary to read the files:

- File type (.xls, .txt, .dat, .csv, bsq, bil)
- Separator (<space>, <,>, …)
- Initial row/ Initial column
- Null value
- Initial and Final dates

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	id_estacion	Anio	enero	febrero	marzo	abril	mayo	junio	julio	agosto	septiembı	octubre	noviembr	diciembre
2	st-c-00001	1969	0	0	0	0	2	5	0	7	0	0	0	0
3	st-c-00001	1970	0	0	5	0	0	0	20	0	0	0	0	0
4	st-c-00001	1971	4	0	0	0	0	19	4	1	16	0	0	0
5	st-c-00001	1972	0	0	0	0	3	64	1	8	3	0	0	0
6	st-c-00001	1973	0	0	0	3	0	29	6	0	0	0	0	0
7	st-c-00001	1974	0	0	0	0	0	32	0	0	4	0	0	0
8	st-c-00001	1975	0	0	0	0	40	6	5	0	0	0	0	0
9	st-c-00001	1976	0	0	0	0	NULL	0	3	25	0	0	0	0
10	st-c-00001	1977	0	0	0	1	5	6	21	19	0	0	0	0
11	st-c-00001	1978	0	0	0	0	0	0	108	0	2	0	2	0
12	st-c-00001	1979	0	0	0	0	0	0	0	3	3	0	0	0
13	st-c-00001	1980	0	1	0	50	0	15	40	0	0	12	0	0
14	st-c-00002	1899	0	0	0	0	0	0	0	0	0	0	0	0
15	st-c-00002	1900	0	0	0	0	0	0	0	0	0	0	0	0
16	st-c-00002	1901	0	0	0	0	0	0	0	0	0	0	0	0
17	st-c-00002	1902	0	0	0	0	0	0	0	0	0	0	0	0
18	st-c-00002	1903	0	0	0	0	0	0	15	0	0	0	0	0
19	st-c-00002	1904	0	0	0	0	0	0	0	1	0	0	0	0

Figure 3. Example of input data provided by user.

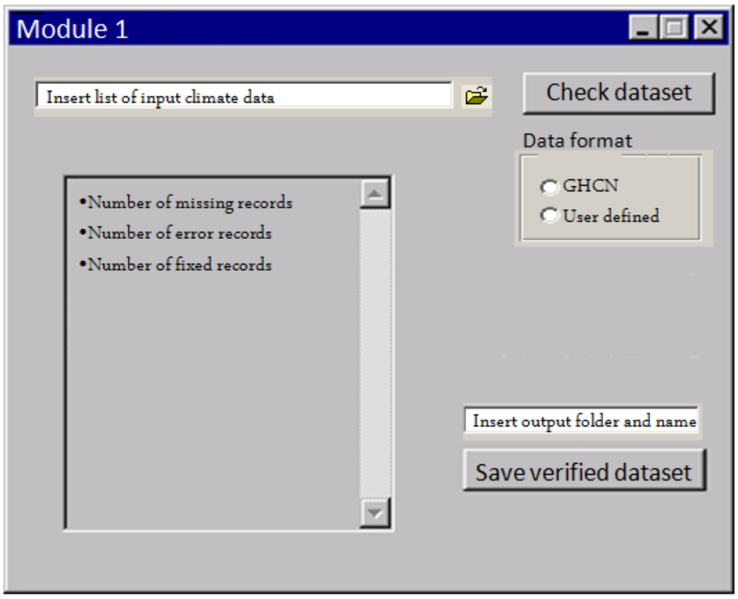

Figure 4. Draft concept of Module 1 GUI

R script example: Module 1

```
#Loading data and Preprocessing
#--------------------------------------------------------------------------

# Example case 1:Import datasets from a website (Cazalac)
BaseDatosNNNRegistros<-
read.table(url("http://www.cazalac.org/documentos/atlas_sequias/chilean_cas
e_example/BaseDatosNNNRegistros.csv"), header=TRUE,
sep=";",na.strings="NA")
BaseDatosNNNEstaciones<-
read.table(url("http://www.cazalac.org/documentos/atlas_sequias/chilean_cas
e_example/BaseDatosNNNEstaciones.csv"), header=TRUE,
sep=";",na.strings="NA")

# Example case 2: Files saved on computer
#BaseDatosNNNRegistros <- read.csv("BaseDatosNNNRegistros.csv",
sep=";",na.strings="NA")
#BaseDatosNNNEstaciones <- read.csv("BaseDatosNNNEstaciones.csv",
sep=";",na.strings="NA")

# This is an example of data screening for valid records. A more elaborated
data screening needs to be implemented in order to be used with a large
range of datasets
EstacionesOriginales<-as.factor(BaseDatosNNNRegistros[[1]])
NumeroEstacionesOriginales<-nlevels(EstacionesOriginales
PPNNN<-na.omit(BaseDatosNNNRegistros)# Use only complete records
EstacionesCompletas<-as.factor(PPNNN[[1]])
NumeroEstacionesCompletas<-nlevels(EstacionesCompletas)# Number of stations
with complete dataset
#--------------------------------------------------------------------------
```

Module 2 – Defining homogeneous regions

The second module has the objective of clustering stations into homogenous groups. A homogeneous group is defined by stations which data, after rescaling by the at-site mean, can be described by a common probability distribution. The user will have the option of choosing among different methodologies.

Index based approaches:

The user will have the possibility of defining a certain number of groups and/or the range of values for each group. The software will have also the possibility of proposing an automatic range of values based on the number of clusters defined by the user (equal distribution range of values.

Some examples follow:

a- Seasonal Index (SI): User will have the option of defining the number of groups, for example, 5 groups divided from 0 to 1 (0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8,0.8-1) but user also will have (as software option) the possibility of defining the range of values for each group. A default number of groups will be presented for the user in the beginning of the operation.

b- Julian Mean Day (JMD): User will have the option of defining the number of groups divided between the minimum and maximum values of the dataset. The software will have the option of suggesting an optimum number of groups.

c- Mean Annual Precipitation (MAP): User will have the option of defining the number of groups, divided between the minimum and maximum values. The software will have the option of suggesting an optimum number of groups.

Map based approaches:

The user will have also the possibility of entry a spatial map (i.e. in a standard image format compatible with ENVI formats: .shp, bil, bsq, …). Each pixel will represent a cluster number. The software will cross the image with the geographical coordinates of the Meteorological stations for defining the belonging group-cluster.

-Holdridge map: The maps will be provided by the user. The user will have to identify the name of the map attribute with which the groups will be associated.

-NDVI classification: Map provided by the user. The number of classes will be defined by the user.

Statistical methods:

If this option is chosen by the user, the software will perform a statistical clustering analysis using the following methodologies: K-means, Agglomerative Hierarchical, Univariate, Maximum Likelihood, TBD). The software will provide outputs (TBD) and charts (TBD) that will allow the user to confirm.

Additional methods to be defined

The software will include, for each method, a help button with a brief description of the technique. After performing the clustering, the homogeneity of each sub-region is to be confirmed using the H1 heterogeneity measure of Hosking and Wallis (1997) (as implemented in the 'regtst' function in R).

Each homogeneous group represents a series of records from many stations. The final product of this module should be a single file, in which the records of several homogeneous groups are stored. This can be done in the format of an R "list" file (as implemented in the 'list' function in R) and exemplified in Figure 5.

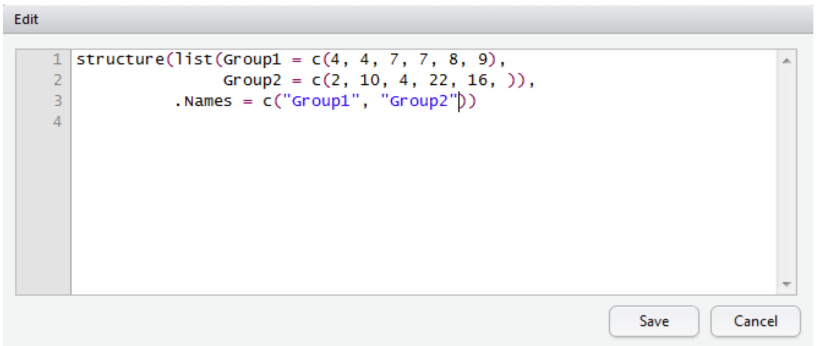

Figure 5. Example of a file structure for storing the records of many homogeneous groups into a single file.

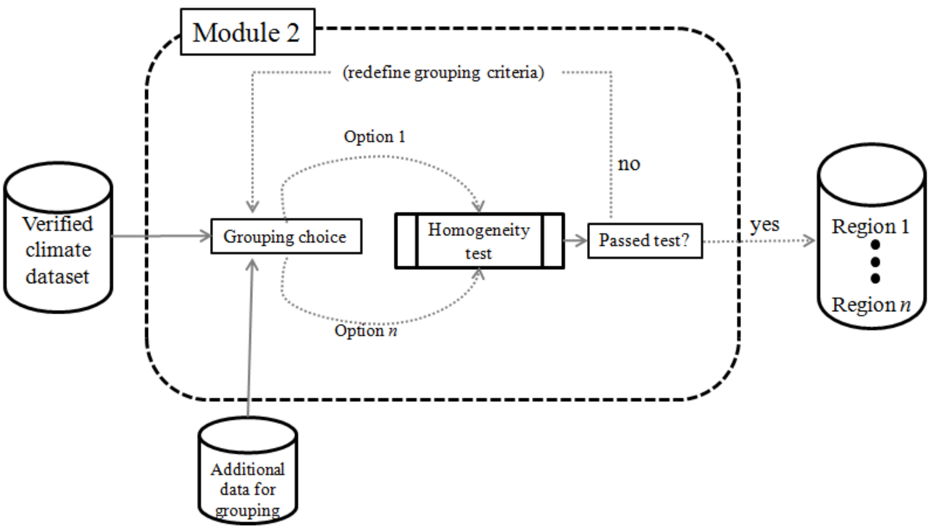

Figure 6. Module 2 data flow.

Inputs [format]	Outputs [format]
• Verified dataset [.xls, .xlsx or .csv] • Additional maps to create homogenous regions [Geotiff, .img, Esri Grid]	• Results of the heterogeneity test [on screen, possibility to save in .txt or .csv] • File with the clustered dataset for each group [.xls, .xlsx or .csv, the file will only be saved after the user is satisfied with the discordancy test]

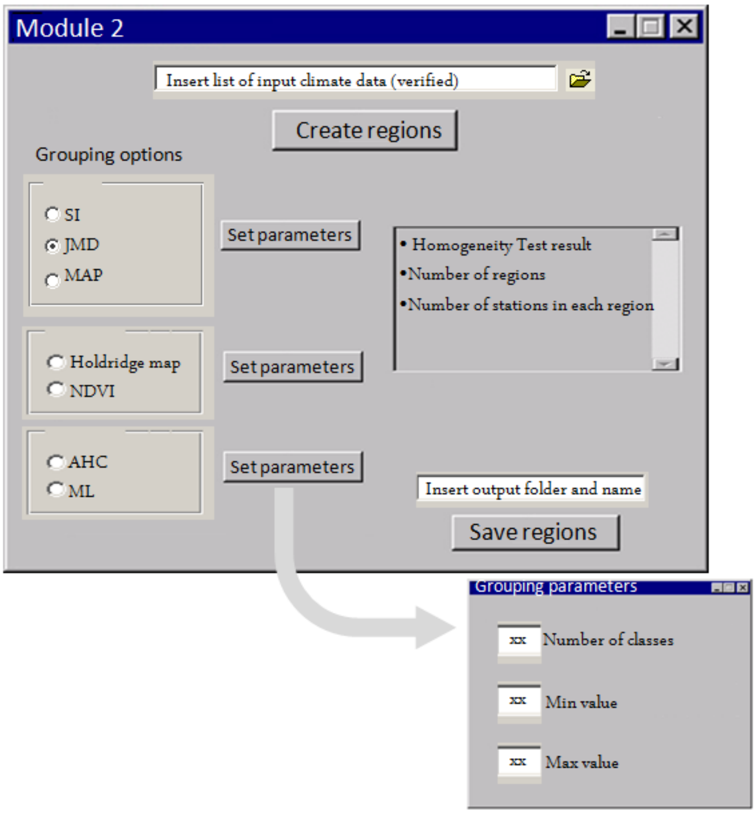

Figure 7. Draft concept of Module 2 GUI

R script example: Module 2

```
#---------------------------------------------------------------------------
#Module 2: Creating homogeneous regions
#---------------------------------------------------------------------------
# First, some variables necessary for defining the homogeneous regions are
calculated from the datasets

LluviaAnual<-PPNNN[3:14] # Calculate annual precipitation
```

```
L<-length(PPNNN[[1]]) # Obtain the longitude of the records

SumaLluviaAnual<-matrix(rowSums(LluviaAnual),nrow=L,ncol=1)

#Start stationarity index (SI) and Mean Julian Day (MJD) calculation
x<-matrix(0,nrow=L,ncol=12)
y<-matrix(0,nrow=L,ncol=12)

angulo_corregido<-matrix(0,nrow=L,ncol=1)
Mes<-seq(1:12)
DiaJuliano<-seq(15,345,30)
DiaJulianoAng<-DiaJuliano*2*pi/365
for (i in 1:L) {
  for (j in 1:12) {
    x[i,j]<-PPNNN[i,(j+2)]*cos(DiaJulianoAng[j])
    y[i,j]<-PPNNN[i,(j+2)]*sin(DiaJulianoAng[j])
}
}
xcos<-matrix(rowSums(x),nrow=L,ncol=1)
ysin<-matrix(rowSums(y),nrow=L,ncol=1)
angulo<-atan(ysin/xcos)

for (k in 1:L) {
  if (xcos[k]>0&ysin[k]>0) angulo_corregido[k]<-angulo[k] else if
(ysin[k]>0&xcos[k]<0) angulo_corregido[k]<-angulo[k]+pi else
angulo_corregido[k]<-angulo[1]+pi*2
}
JMD<-(angulo_corregido*365)/(2*pi)
SI<-sqrt(xcos^2+ysin^2)/SumaLluviaAnual
# End of stationarity index (SI) and Mean Julian Day (MJD) calculation


BaseDatosNNNIntermedia<-cbind(PPNNN,SumaLluviaAnual,SI,JMD)

# Starts calculation of Average values for each station
SI_por_Estacion<-
as.matrix(tapply(BaseDatosNNNIntermedia[[16]],BaseDatosNNNIntermedia[[1]],m
ean,na.rm=TRUE))
hist(SI_por_Estacion)

PMA_por_Estacion<-
as.matrix(tapply(BaseDatosNNNIntermedia[[15]],BaseDatosNNNIntermedia[[1]],m
ean,na.rm=TRUE))
hist(PMA_por_Estacion)

JMD_por_Estacion<-
as.matrix(tapply(BaseDatosNNNIntermedia[[17]],BaseDatosNNNIntermedia[[1]],m
ean,na.rm=TRUE))
hist(JMD_por_Estacion)

LR_por_Estacion<-
as.matrix(tapply(BaseDatosNNNIntermedia[[15]],BaseDatosNNNIntermedia[[1]],l
ength))
hist(LR_por_Estacion

id_estacion<-levels(EstacionesCompletas)# Identify stations to be used
```

```
BaseDatosIndices<-
cbind(id_estacion,SI_por_Estacion,PMA_por_Estacion,JMD_por_Estacion,LR_por_
Estacion)
colnames(BaseDatosIndices)[2]<-'SIMedio'
colnames(BaseDatosIndices)[3]<-'PMA'
colnames(BaseDatosIndices)[4]<-'JMDMedio'
colnames(BaseDatosIndices)[5]<-'LR'

BaseConsolidadaNNN<-
merge(BaseDatosNNNEstaciones,BaseDatosIndices,by.x="id_estacion",by.y="id_e
stacion")
BaseConsolidadaNNN_sin_NA<-na.omit(BaseConsolidadaNNN) # Eliminate stations
with missing data. In the software, the user will have to decide in the
beginning which stations he will want to eliminate or not

#Create a general database
BaseCompletaNNN<-merge(BaseConsolidadaNNN_sin_NA,BaseDatosNNNIntermedia,
by.x = "id_estacion", by.y = "id_estacion")
write.csv(BaseCompletaNNN, file = "BaseCompletaNNN.csv",row.names=FALSE)

###Update the database###
remove(BaseCompletaNNN)
BaseCompletaNNN <- read.csv("BaseCompletaNNN.csv")# Load updated database

#CREATE HOMOGENEOUS REGIONS

#In this example the regions are created based on fixed criteria. In the
software, the criteria should be define by the user (although default
options should be available)
#The fixed criteria of the example are:
# Grouping by average SI into five groups (0-0.2, 0.2-0.4, 0.4-0.6,0.6-
0.8,0.8-1)
# After, in each SI group, the stations are separate by MJD (30 days group)
# After, the statios are separated by Mean annual precipitation (MAP)
Region1<-sqldf("select id_estacion, SumaLluviaAnual from BaseCompletaNNN
where PMA between 50 and 159 and LR>15")
Region1_dat<-Region1["SumaLluviaAnual"][,]
Region1_fac<-factor(Region1["id_estacion"][,])
Reg1<-split(Region1_dat,Region1_fac)# Con esto separo los registros según
la estación

Region2<-sqldf("select id_estacion, SumaLluviaAnual from BaseCompletaNNN
where PMA between 160 and 227 and LR>15")
Region2_dat<-Region2["SumaLluviaAnual"][,]
Region2_fac<-factor(Region2["id_estacion"][,])
Reg2<-split(Region2_dat,Region2_fac)

Region3<-sqldf("select id_estacion, SumaLluviaAnual from BaseCompletaNNN
where PMA between 227 and 261 and LR>15")
Region3_dat<-Region3["SumaLluviaAnual"][,]
Region3_fac<-factor(Region3["id_estacion"][,])
Reg3<-split(Region3_dat,Region3_fac)

Region4<-sqldf("select id_estacion, SumaLluviaAnual from BaseCompletaNNN
where PMA between 261 and 306 and LR>15")
Region4_dat<-Region4["SumaLluviaAnual"][,]
Region4_fac<-factor(Region4["id_estacion"][,])
```

```
Reg4<-split(Region4_dat,Region4_fac)

Region5<-sqldf("select id_estacion, SumaLluviaAnual from BaseCompletaNNN
where PMA between 306 and 396 and LR>15")
Region5_dat<-Region5["SumaLluviaAnual"][,]
Region5_fac<-factor(Region5["id_estacion"][,])
Reg5<-split(Region5_dat,Region5_fac)

Region6<-sqldf("select id_estacion, SumaLluviaAnual from BaseCompletaNNN
where PMA between 396 and 463 and LR>15")
Region6_dat<-Region6["SumaLluviaAnual"][,]
Region6_fac<-factor(Region6["id_estacion"][,])
Reg6<-split(Region6_dat,Region6_fac)

Region7<-sqldf("select id_estacion, SumaLluviaAnual from BaseCompletaNNN
where PMA between 463 and 566 and LR>15")
Region7_dat<-Region7["SumaLluviaAnual"][,]
Region7_fac<-factor(Region7["id_estacion"][,])
Reg7<-split(Region7_dat,Region7_fac)

Region8<-sqldf("select id_estacion, SumaLluviaAnual from BaseCompletaNNN
where PMA between 566 and 1215 and LR>15")
Region8_dat<-Region8["SumaLluviaAnual"][,]
Region8_fac<-factor(Region8["id_estacion"][,])
Reg8<-split(Region8_dat,Region8_fac)

# Example for choosing a particular station
#RegionXX <- sqldf("select * from BaseCompletaNNN where id_estacion=='st-
nnn-0001'")
# Example to choose all stations, except one
#Regionzz<- sqldf("select * from BaseCompletaNNN where id_estacion!='st-
nnn-0001'")
# Reference: Halekoh et al, 2010. Handling large(r) datasets in R.
http://genetics.agrsci.dk/~sorenh/misc/Rdocs/R-largedata.pdf

BaseRegiones<-list(Reg1,Reg2,Reg3,Reg4,Reg5, Reg6, Reg7,Reg8)# create a
list with all regions
#--------------------------------------------------------------------------
```

Module 3 – Regional frequency analysis

This module performs the Regional Frequency Analysis (RFA) using the homogeneous regions, by selecting the probability distribution function for each homogeneous group.

The selection of the best function is based on the Z|DIST| goodness-of-fit test described by Hosking and Wallis (1997). This statistic is already implemented in R through the same command used to obtain the homogeneity statistics (‘regtst’).

After the best distribution is defined according to the Zdist test result, the user will have the option of visualizing a popup window with a summary of the Region.

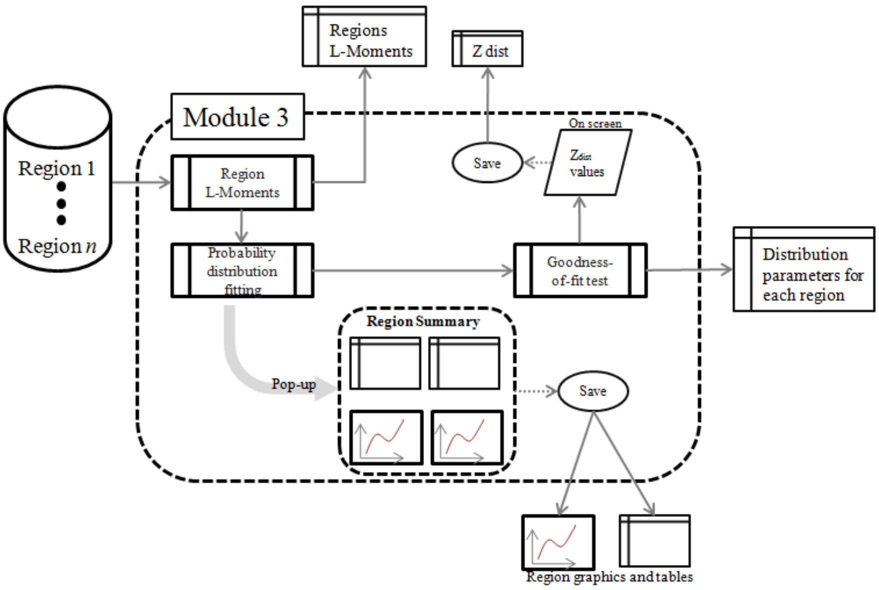

Figure 8. Module 3 data flow.

Inputs [format]	Outputs [format]
• File with the clustered dataset for each homogeneous group [.xls, .xlsx or .csv]	• Table with Z|DIST| values for each group[on screen, possibility to save in .txt or .csv] • Parameters of the best-fit distribution [on screen AND saved in .csv or software specific format] • Regions L-Moments [.csv or software specific format] • Group summary – Opens popup window with the summary of the selected homogeneous group -Figure with L-moment ratio diagram; -Table with the group info (e.g. number of stations, number of records etc

	[on screen, possibility to save in .jpeg or .tif]

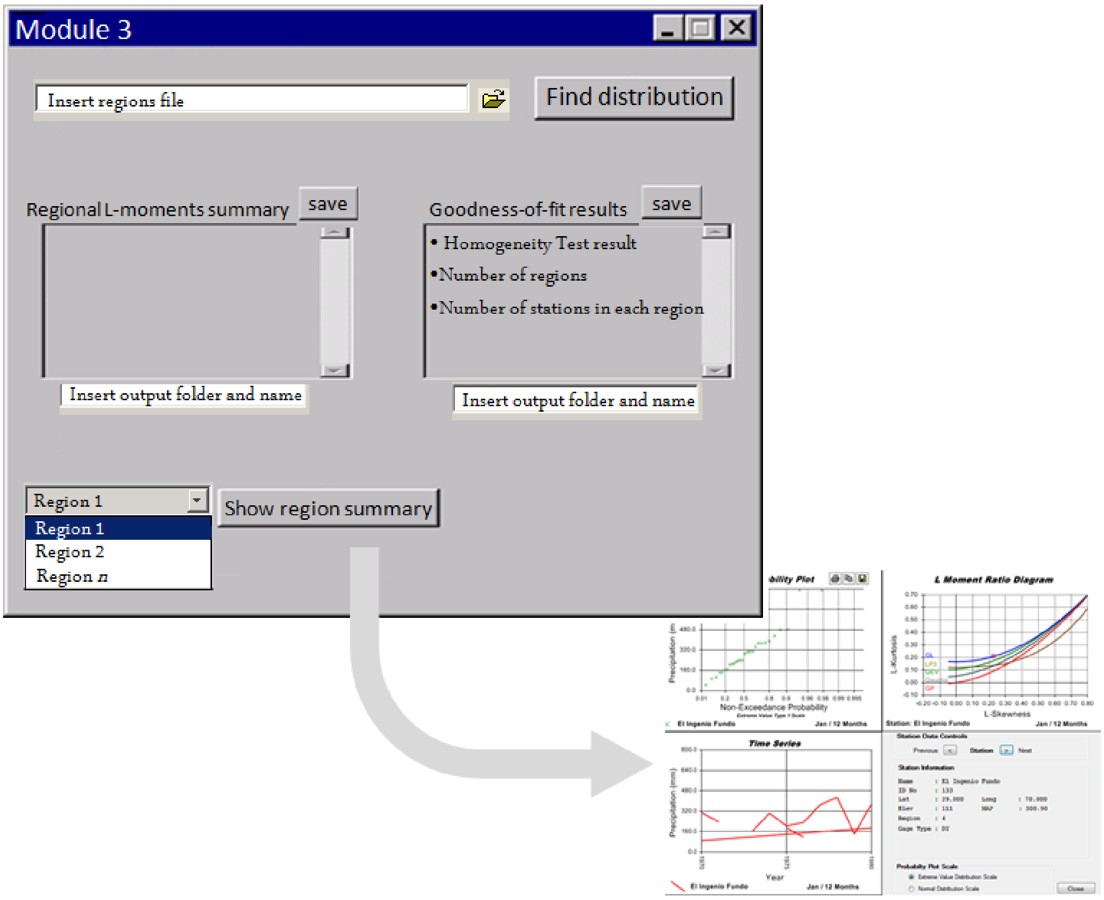

Figure 9. Draft concept of Module 3 GUI

R script example: Module 3

```
#----------------------------------------------------------------------------
#Module 3: REGIONAL FREQUENCY ANALYSIS
#----------------------------------------------------------------------------
# DECLARATION OF VARIABLES TO STORE RESULTS
Regiones<-length(BaseRegiones)
ResultadosSummaryStatistics<-array(0,dim=c(100,7,Regiones))# Maximum 100
years of data,statistics,regions
ResultadosSummaryStatisticsRegData<-array(0,dim=c(150,7,Regiones))#(Maximum
150 years of data,statistics,regions)
ResultadosRlmoments<-array(0,dim=c(5,Regiones))#5= Regional L-moments
ResultadosARFD<-array(0,dim=c(100,Regiones))#100= Maximum number of
stations by region
ResultadosARFH<-array(0,dim=c(3,Regiones))# 3= Homogeneity index H1,H2,H3
ResultadosARFZ<-array(0,dim=c(5,Regiones))# 5= Number of probability models
to calculate the goodness-of-fit(glo, gev, gno, pe3, gpa)
```

```
Resultadosrfitdist<-array(0,dim=c(1,Regiones))# 1=One adjustment by region
Resultadosrfitpara<-array(0,dim=c(5,Regiones))#5= number of Wakeby
parameters
ResultadosRegionalQuantiles<-array(0,dim=c(19,Regiones))# 19=Maximum number
of quantiles to be calculated
ResultadosRMAP<-array(0,dim=c(1,Regiones))# 1= One annual medium
precipitation value by region

# L-Moments based on the Regional Frecuency Analysis
for (z in 1:Regiones) {
par(mfrow=c(1,2))
SummaryStatistics<-regsamlmu (BaseRegiones[[z]]) #Calculates the L-moments
for the different variables stored in the dataset columns [first:last].
Values should be changed depending on the dataset

SummaryStatisticsRegData<-as.regdata(SummaryStatistics)
lmrd(SummaryStatisticsRegData)# Creates the L-moments ratios diagram
Rlmoments<-regavlmom(SummaryStatisticsRegData)# Calculates the L-moments
for each region with the analyzed stations
lmrdpoints(Rlmoments, type="p", pch=22, col="red" )#adds the regional L-
moments (red points) to the L-moments ratios diagram
ARF<-regtst(SummaryStatisticsRegData, nsim=1000)# Calculates some
statistics for the different regions including the homogeneity test and
goodness of fit for different distributions models.

#Stored discordancy, homogeneity and goodness of fit.
a<-length(BaseRegiones[[z]])
ResultadosRlmoments[1:5,z]<-Rlmoments
ResultadosARFD[1:a,z]<-ARF$D # To store discordancy
ResultadosARFH[1:3,z]<-ARF$H # To store homogeneity measures
ResultadosARFZ[1:5,z]<-ARF$Z # To store goodness of fit

# SELECTION AND ADJUSTMENT OF THE PROBABILITY MODEL DISTRIBUTION
rfit<-regfit(SummaryStatisticsRegData, "pe3") #This command line is used to
specify and adjust the probability distribution model.
## in this example, the pe3 distribution was used because it resulted in
the best goodness of fit result. The softaware should be able to recognize
the best distribution and automatically apply this distribution in the
analysis

RegionalQuantiles<-regquant(seq(0.05, 0.95, by=0.05), rfit)# Calculates
regional quantiles for different cumulative probabilities
# The following three lines generate a quantile graph
rgc <- regqfunc(rfit)# Calculates the Regional Growth Curve
rgc(seq(0.05, 0.95, by=0.05))
curve(rgc, 0.01, 0.99, xlab="Non-exceedence Probability, F", ylab="Growth
Curve")
Resultadosrfitdist[z]<-rfit$dist # Identifies the distribution used
Resultadosrfitpara[1:3,z]<-rfit$para # Shows the results of the parameters
for the adjusted distribution
ResultadosRegionalQuantiles[1:19,z]<-RegionalQuantiles # For each region
"z" we store the results
ResultadosRMAP[z]<-
weighted.mean(SummaryStatisticsRegData[[3]],SummaryStatisticsRegData[[2]])
# It calculates medium precipitation for each region
}# End of cycle for
#---------------------------------------------------------------------------
```

Module 4 – Interpolation parameters

In Module 3, the L-moments are defined for each station. In order to create spatially-explicit maps, this information needs to be interpolated to areas where no stations are available in the region. This procedure is done through a relationship between the L-moments and the Mean Annual Precipitation (MAP). This module will definite the parameters of the curves defining this relationship which will be used to create L-moment maps in Module 5. The user will be able to choose among three options for finding the interpolation parameters:

- Minimization through DEoptim
- Minimization through NLM (Non-linear Minimization)
- Minimization through NLS (Non-linear Squares)

When defining the curve parameters, the software will also provide graphics L-moments vs MAP. The user will have the option of saving these graphics in .tif, .tiff, .png or .jpeg coding the geographical coordinates when possible (geotif data format, for instance).

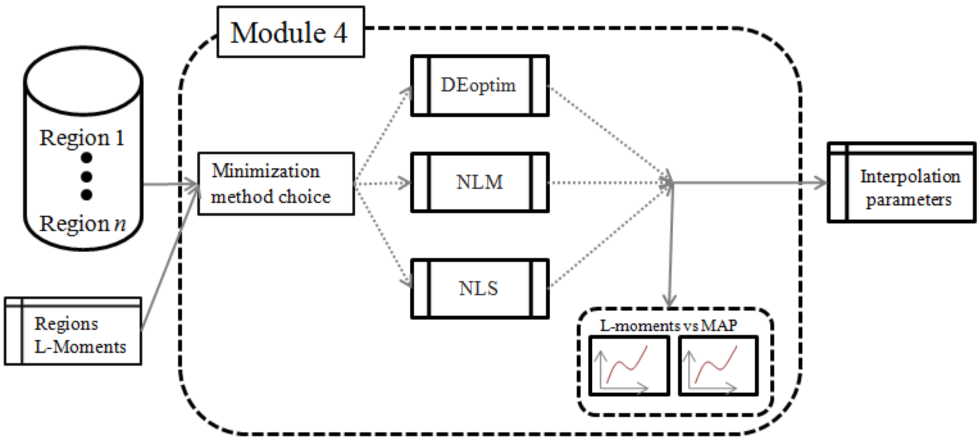

Figure 10. Module 4 data flow.

Inputs [format]	Outputs [format]
• Regions L-Moments [.csv] • File with the clustered dataset for each homogeneous group [.xls, .xlsx or .csv] • Method for interpolation [defined by user]	• interpolation parameters [.csv or software specific format] • Graphic L-moment vs MAP [on screen, possibility to save in .jpeg or .tif]

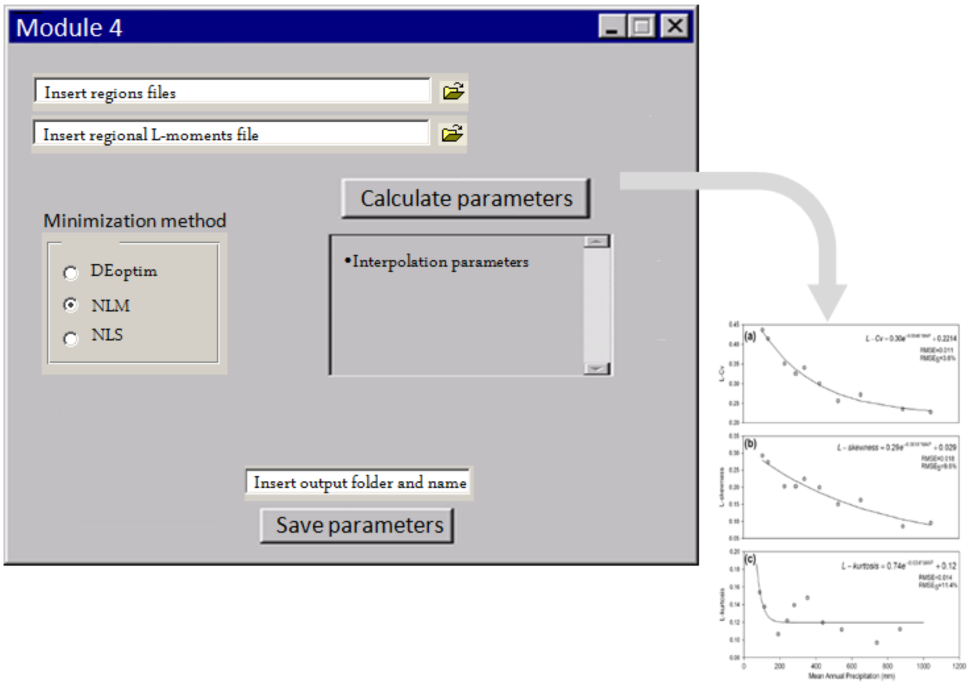

Figure 11. Draft concept of Module 4 GUI

R script example: Module 4

```
#----------------------------------------------------------------------------
#Module 4: ADJUSTMENT FUNCTION FOR THE L-MOMENTS VS ANUAL MEDIUM
PRECIPITATION
#----------------------------------------------------------------------------
# DECLARATION OF VARIABLES
RLCV <-   ResultadosRlmoments[2,]
RLSkewness<-ResultadosRlmoments[3,]
RLKurtosis<-ResultadosRlmoments[4,]
RMAP<-as.numeric(ResultadosRMAP)
MAPvsLCV <- data.frame(RMAP,RLCV)
MAPvsLSkewness<- data.frame(RMAP,RLSkewness)
MAPvsLKurtosis<- data.frame(RMAP,RLKurtosis)

# OPTION ADJUSTMENT 1: Minimization using DEoptim.
PMediaAnual<-RMAP
LCVOBS<-RLCV
LCVEST<-function(p) p[1]*exp(p[2]*PMediaAnual)+p[3]
fun<-function(p) sum((LCVOBS-LCVEST(p))^2)
ss <- DEoptim(fun, lower=c(0,-0.1,0), upper=c(0.3,0,0.2),
control=list(trace=FALSE))
paLCV <- ss$optim$bestmem
paLCV

LSkOBS<-RLSkewness
LSkEST<-function(p) p[1]*exp(p[2]*PMediaAnual)+p[3]
fun<-function(p) sum((LSkOBS-LSkEST(p))^2)
```

```
ss <- DEoptim(fun, lower=c(0,-0.1,0), upper=c(0.3,0,0.2),
control=list(trace=FALSE))
paLSk <- ss$optim$bestmem
paLSk

LKurtOBS<-RLKurtosis
LKurtEST<-function(p) p[1]*exp(p[2]*PMediaAnual)+p[3]
fun<-function(p) sum((LKurtOBS-LKurtEST(p))^2)
ss <- DEoptim(fun, lower=c(0,-0.1,0), upper=c(0.3,0,0.2),
control=list(trace=FALSE))
paLKurt <- ss$optim$bestmem
paLKurt
#..........................................................................
# OPTION ADJUSTMENT 2: Optimization using NLS command (Non-linear Squares)
nlsfitLCV <- nls(RLCV~A*exp(B*RMAP)+C,data=MAPvsLCV, start=list(A=paLCV[1],
B=paLCV[2], C=paLCV[3]))
nlsfitLSkewness <- nls(RLSkewness~A*exp(B*RMAP)+C,data=MAPvsLSkewness,
start=list(A=paLSk[1], B=paLSk[2], C=paLSk[3]))
nlsfitLKurtosis <- nls(RLKurtosis~A*exp(B*RMAP)+C,data=MAPvsLKurtosis,
start=list(A=paLKurt[1], B=paLKurt[2], C=paLKurt[3]))
pp<-seq(min(RMAP),max(RMAP),length=100)
plot(RMAP, RLCV, xlim=c(min(RMAP),max(RMAP)), ylim=c(min(RLCV),max(RLCV)))
lines(pp,predict(nlsfitLCV,list(RMAP=pp)))
plot(RMAP, RLSkewness, xlim=c(min(RMAP),max(RMAP)),
ylim=c(min(RLSkewness),max(RLSkewness)))
lines(pp,predict(nlsfitLSkewness,list(RMAP=pp)))
plot(RMAP, RLKurtosis, xlim=c(min(RMAP),max(RMAP)),
ylim=c(min(RLKurtosis),max(RLKurtosis)))
lines(pp,predict(nlsfitLKurtosis,list(RMAP=pp)))
summary(nlsfitLCV)
summary(nlsfitLSkewness)
summary(nlsfitLKurtosis)
#..........................................................................
# OPTION ADJUSTMENT 3: Minimization through NLM command(Non-Linear
Minimization)

# Aca se presenta alternativa 2 para estimar mejor ajuste.
#fnLCV <- function(p) sum((RLCV - p[1]*exp(p[2]*RMAP)+p[3])^2)
#outLCV <- nlm(fnLCV, p = c(paLCV[1], paLCV[2], paLCV[3]))
#outLCV$estimate

#fnLSkewness <- function(p) sum((RLSkewness - p[1]*exp(p[2]*RMAP)+p[3])^2)
#outLSkewness <- nlm(fnLSkewness, p = c(paLSk[1], paLSk[2],paLSk[3]))
#outLSkewness$estimate

#fnLKurtosis <- function(p) sum((RLKurtosis - p[1]*exp(p[2]*RMAP)+p[3])^2)
#outLKurtosis <- nlm(fnLKurtosis, p = c(paLKurt[1], paLKurt[2],
paLKurt[3]))
#outLKurtosis$estimate
#---------------------------------------------------------------------------
```

Module 5 – L-moments maps

In Module 5 the interpolation parameters will be used to create L-moment maps based on an annual precipitation map provided by the user. The map provided by the user has to have the same units as used for the parameters calculation in Module 4 (e.g. mm/year).

In a general way, the maps to be produced or be read by the software will in any of the most common GIS formats (i.e. Geotiff, img, Esri GRID, bil, bsq, …) and with the same projection and datum as the input maps.

The user will have the option of saving the maps as figure (.tif, geotif, tiff, png or .jpeg), with customized grids, scale, legends and titles.

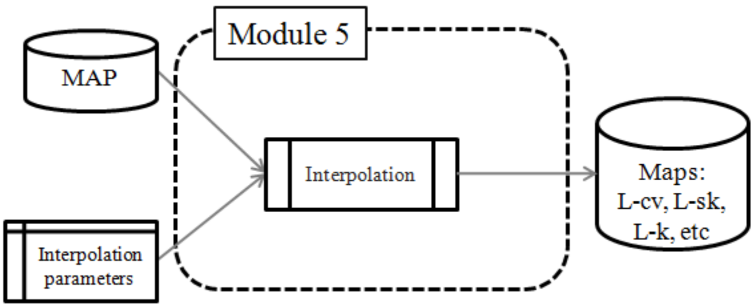

Figure 12. Module 5 data flow.

Inputs [format]	Outputs [format]
• interpolation parameters [.csv or software specific format] • Mean Annual Precipitation map[Geotiff, .img, Esri Grid]	• L-moments maps, 4 first moments [Geotiff, .img, Esri Grid] • -[also possibility to save it in .jpg or tiff directly from the software, with grid, scale, legend and title]

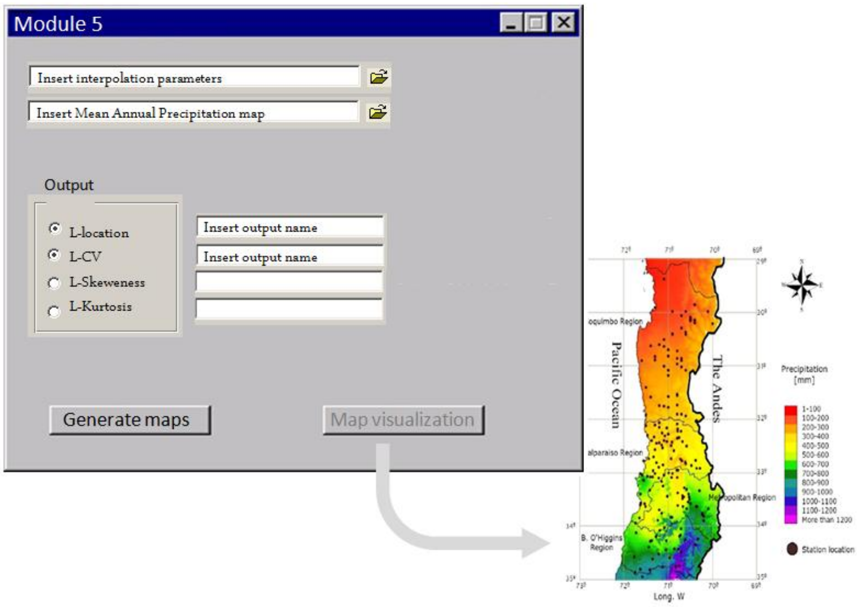

Figure 13. Draft concept of Module 5 GUI

R script example: Module 5

```
#-----------------------------------------------------------------------------
# Module 5: CREATION OF L-moment MAPS
#-----------------------------------------------------------------------------
# IMPORT THEMATIC BASE MAP OF SPATIAL VARIABILITY TO BE USED FOR THE
INTERPOLATION
options(download.file.method="auto")
download.file("http://www.cazalac.org/documentos/atlas_sequias/chilean_case
_example/MapaNNN.tif",destfile=paste(WF,"/",
"MapaNNN.tif",sep=""),mode="wb")
MapaNNN<-readGDAL("MapaChile.tif") # Definition of Thematic base map
r<-raster(MapaNNN)
projection(r) <- "+proj=latlong +ellps=WGS84" # Definition of Geographic
projection
#L-MOMENTS MAPS CALCULATION
LCVmap<-paLCV[1]*exp(paLCV[2]*r)+paLCV[3] #  L-CV map creation based on the
best adjustment coefficients values
LSmap<-paLSk[1]*exp(paLSk[2]*r)+paLSk[3] #  L-skewness map creation based
on the best adjustment coefficients values
LKmap<-paLKurt[1]*exp(paLKurt[2]*r)+paLKurt[3] # L-kurtosis map creation
based on the best adjustment coefficients values
#FORMAT CONVERSION FROM RASTER TO MATRIX TO FACILATE FURTHER CALCULATIONS
R<-as.matrix(r)
J<-as.matrix(LCVmap)
K<-as.matrix(LSmap)
L<-as.matrix(LKmap)
#-----------------------------------------------------------------------------
```

Module 6 – Final map products

Module 6 will provide the final products of the software, that is to say, maps of precipitation frequency, return period, probability etc. The inputs for this module are basically the L-moment maps obtained from Module 5. The user will have the option of calculating all products or just selected maps of the user's interest.

The outputs will be saved in any of the most common GIS formats (i.e. Geotiff, .img, Esri Grid, bil, bsq), and with the same projection and datum as the input L-moment maps. Following the example of Module 5, the user will have the option of saving the maps as figure (.tif, geotif, tiff, png or .jpeg), with customized grids, scale, legends and titles.

The complete list of outputs is to be defined.

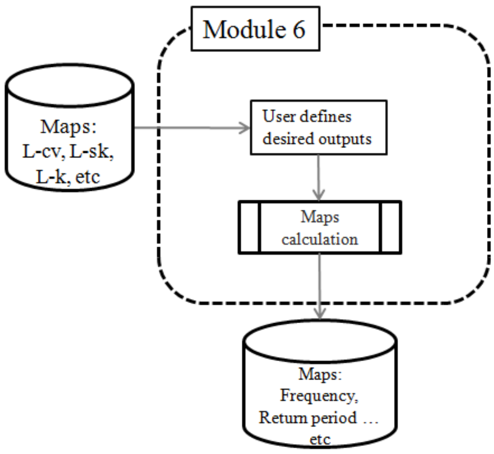

Figure 14. Module 6 data flow.

Inputs [format]	Outputs [format]
• L-moments maps, 4 first moments [Geotiff, .img, Esri Grid] • Outputs and parameters desired by the user (e.g. Non-exceedence probabilities) [defined by user on the software interface]	Outputs on users demand: • Frequency maps • Probability maps • Return period maps [Geotiff, .img, Esri Grid]-[also possibility to save it in .jpg or tiff directly from the software, with grid, scale, legend and title]

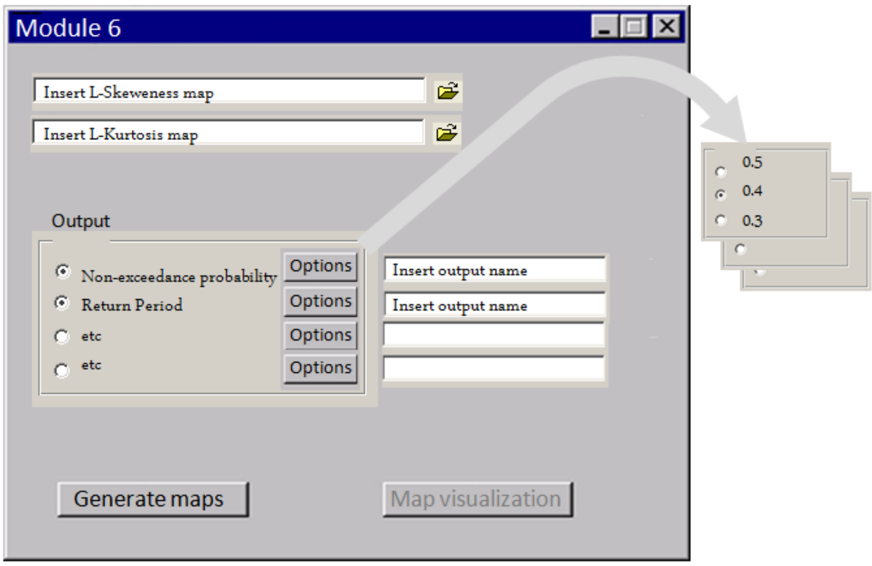

Figure 15. Draft concept of Module 6 GUI

R script example: Module 6

```
#-----------------------------------------------------------------------------
# Module 6: Final products - (return period, frequency etc)
#-----------------------------------------------------------------------------
# CALCULATION OF PARAMETERS FOR THE SELECTED PROBABILITY DISTRIBUTION MODEL
Pearson3<-par.gamma((R/R),J,K) # Command line to generate map parameters
for Pearson distribution based on Viglione (alfa, beta,xi) (R/R is used to
create 1s raster)
#GenPar<-par.genpar((R/R),J,K) #  Command line to generate map parameters
for Generalized Pareto distribution based on Viglione (alfa, beta,xi)(R/R
is used to create 1s raster)
#GEV<-par.GEV((R/R),J,K)# Command line to generate map parameters for
Generalized Extreme Value distribution based on Viglione (alfa, beta,xi)
(R/R is used to create 1s raster)
#LogNorm<-par.lognorm((R/R),J,K)#  Command line to generate map parameters
for LogNormal distribution based on Viglione (alfa, beta,xi) (R/R is used
to create 1s raster)
#GenLogis<-par.genlogis((R/R),J,K)# Command line to generate map parameters
for Generalized Logistic distribution based on Viglione (alfa, beta,xi)
(R/R is used to create 1s raster)
#Kappa<-par.kappa((R/R),J,K,L)#  Command line to generate map parameters
for Kappa distribution based on Viglione (alfa, beta,xi) (R/R is used to
create 1s raster)

# CALCULATION OF FREQUENCY MAPS
#The following command lines are used to create the probality and return
period maps for an specific quantile
Cuantil<-0.4
FreqMap<-F.gamma (Cuantil*(R/R), Pearson3$xi, Pearson3$beta,
Pearson3$alfa)# #Probability map in a matrix format
#FreqMap<-F.genpar (Cuantil*(R/R), Pearson3$xi, Pearson3$beta,
Pearson3$alfa)# Probability map in a matrix format
#FreqMap<-F.GEV (Cuantil*(R/R), Pearson3$xi, Pearson3$beta, Pearson3$alfa)
#Probability map in a matrix format
```

European Commission

EUR 24947 EN – Joint Research Centre – Institute for Environment and Sustainability
Title: Software description: Regional frequency analysis of climate variables
Author(s): Cesar Carmona-Moreno, Eduardo Eiji Maeda, Juan Arevalo, Marco Giacomassi, Paolo Mainardi
Luxembourg: Publications Office of the European Union
2011 – 31 pp. – 21 x 29.7 cm
EUR – Scientific and Technical Research series – ISSN 1831-9424
ISBN 978-92-79-21322-9
doi: 10.2788/74447

Abstract
This document provides the technical description of a software to be developed in the context of the EUROCLIMA project. EUROCLIMA is a cooperation program between the European Union and Latin America with a special focus in knowledge sharing on topics related to socio-environmental problems associated with climate change. The objective of the project is to improve knowledge of Latin American decision-makers and the scientific community on problems and consequences of climate change, particularly in view of integrating these issues into sustainable development strategies. The software described in this document will have as a general objective to process time series of data from ground stations (initially precipitation and temperature) in order to generate products in the form of spatially-explicit maps. However, the software will be able to process any other time series of environmental spatial data. The main aspect characterizing this software is the use of statistics called L-moments to estimate the probability distribution function of climate variables. The L-moments are similar to other statistical moments, but with the advantage of being less susceptible to the presence of outliers and performing better with smaller sample sizes.

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

LB-NA-24947-EN-N

EUROPEAN COMMISSION

Publications Office

ISBN 978-92-79-21322-9

9 789279 213229