Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

R tools user meeting

23/03/2020

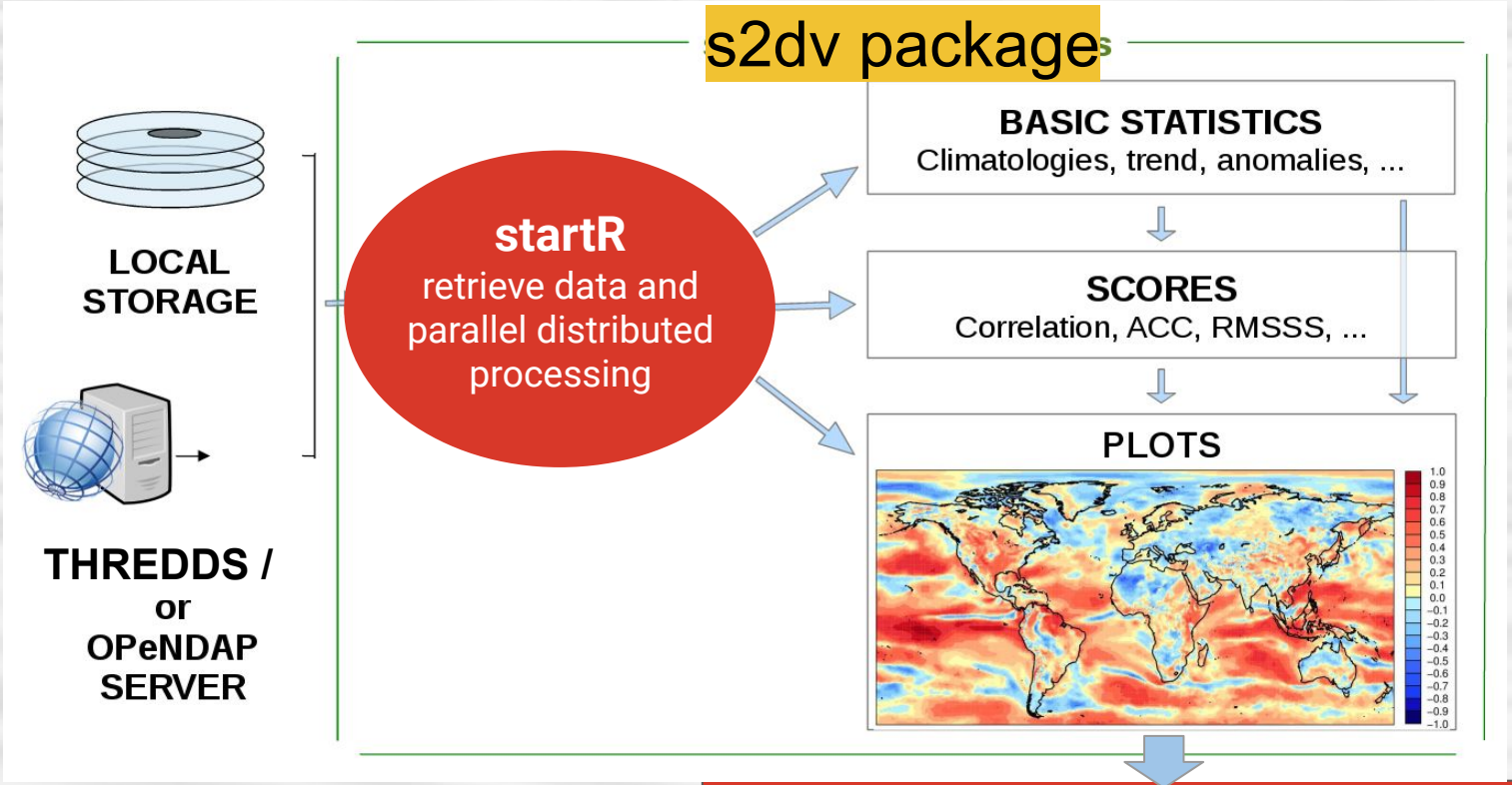
An-Chi Ho and Núria Pérez-Zanón

In-house packages status

PACKAGE	LAST RELEASE	NEXT RELEASE	#ISSUES
• **s2dverification**	October 2019	One more release & deprecated	86
• **s2dv**	March 2020	Summer	0
• **CSTools**	February 2020	May 2020	15
• **ClimProjDiags**	January 2020	-	2
• **startR**	April 2020	TBD	26
• **easyNCDF**	May 2017	March 2020	2
• **multiApply**	September 2019	-	4

ACTION
Review your issues and branches!

R tool scheme



New package!!
s2dv

Introduction to s2dv

s2dv is the advanced version of s2dverification. Most of the functions are with the same name and functionality as previous, but modified to be compatible with startR (i.e., flexible dimension structure/work with named dimension/etc.) Beside, it contains some new formatting functions to facilitate the data-processing.

For now, it only contains part of the s2dverification functions, but we will include the rest in the near future.

More info

- GitLab: https://earth.bsc.es/gitlab/es/s2dv
- CRAN: https://cran.r-project.org/web/packages/s2dv/index.html

s2dv functions (v0.0.1)

Data retrieval and formatting

Load
Reorder
InsertDim
LeapYear
ToyModel

Basic statistics

Clim
Eno
MeanDims
Season
Trend

Skill score

Corr
Regression
RMS
RMSSS

Configuration

ConfigApplyMatchingEntries
ConfigEditDefinition
ConfigEditEntry
ConfigFileOpen
ConfigShowSimilarEntries
ConfigShowTable

Plotting

AnimateMap
ColorBar
PlotClim
PlotEquiMap
PlotLayout
PlotMatrix
PlotSection
PlotStereoMap

New Applied multiApply regime

s2dv vs. s2dverification: Data dimension

s2dverification::Clim

@param var_exp

Model data: **c(nmod/nexp, nmemb/nparam, nsdates, nltime)** up to **c(nmod/nexp, nmemb/nparam, nsdates, nltime, nlevel, nlat, nlon)**.

1. 4 to 7 dimensions required
2. Fixed order and structure

s2dv::Clim

@param exp

A named numeric array of experimental data, with at least two dimensions **'time_dim'** and **'dat_dim'**.

1. Only two dimensions required (pre-paired method, no ensemble mean)
2. Flexible order and structure
3. Named dimensions required

s2dv vs. s2dverification: Named dimension

s2dverification::Season

```
Season <- function(var, posdim = 4, monini, moninf, monsup) {
```

s2dv::Season

```
Season <- function(data, time_dim = 'sdate', monini, moninf, monsup,
                   method = mean, na.rm = TRUE, ncores = NULL) {
```

News on easyNCDF and CSTools

CST_SplitDim Function to Split Dimension

This function split a dimension in two. The user can select the dimension to split and provide indices indicating how to split that dimension or dates and the frequency expected (monthly or by day, month and year). The user can also provide a numeric frequency indicating the length of each division.

CST_MergeDims / MergeDims Function to Merge Dimensions

This function merges two dimensions of the array data in a 's2dv_cube' object into one. The user can select the dimensions to merge and provide the final name of the dimension. The user can select to remove NA values or keep them.

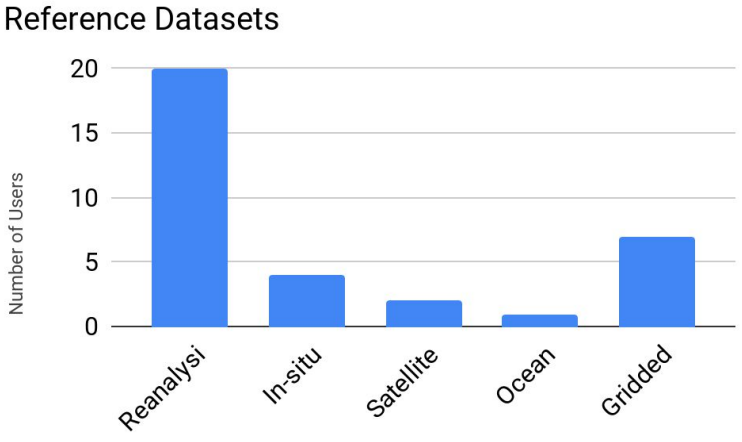
---

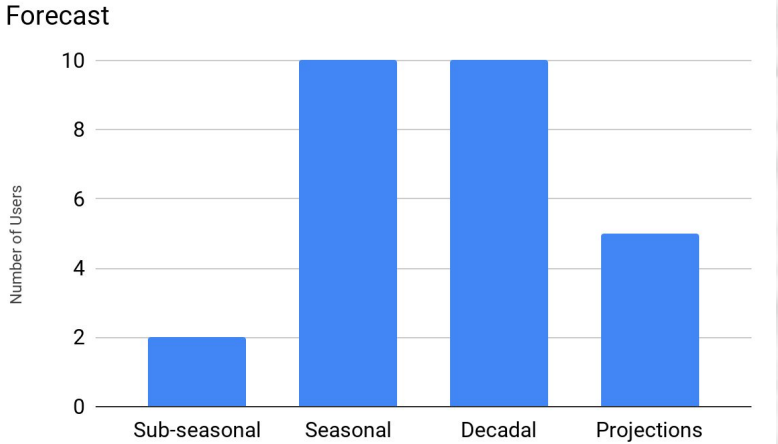
easyNCDF v0.0.4 on CRAN
easyNCDF v0.0.7 in GitLab
this version fixes the reordering problem in startR

easyNCDF v1.0.0 on CRAN and GitLab
documentation in roxygen2 format
when startR is installed, the latest version will be downloaded directly from CRAN.

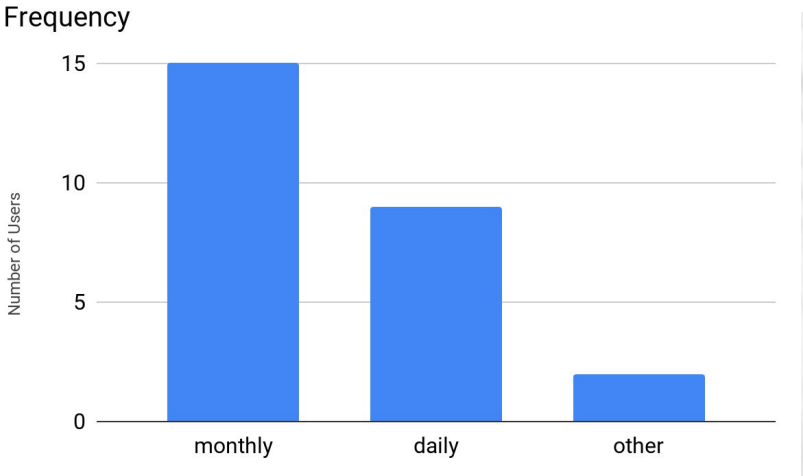
Individual R users meeting outcome

Individual users R meetings:



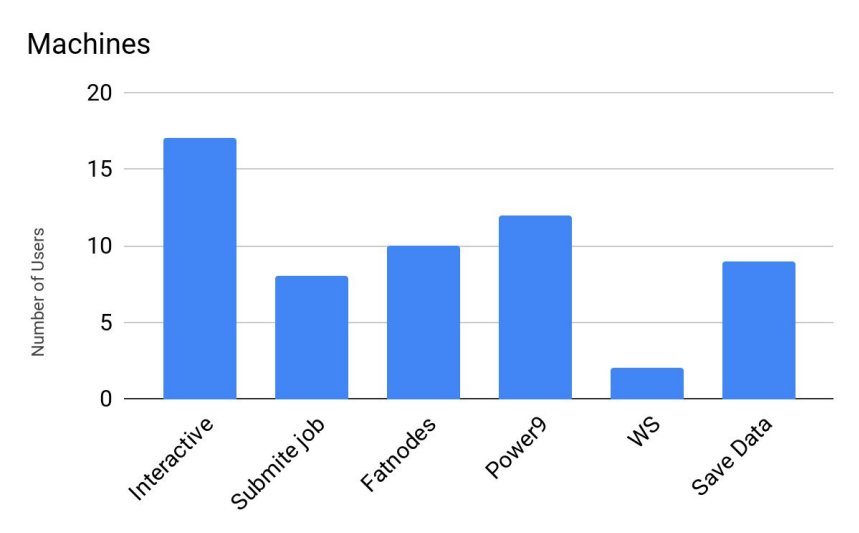


Which data do you use?



Other ∈ {Weekly, Hourly}

Individual users R meetings



Questions we will skip:

- Who is not using GitLab?
- Who has not read startR documentation?

Training session: 15 people ask for a tutorial on **creating their own function** to use in startR/multiApply or specific for startR.

startR users: 7 users are already using startR while 8 want to learn how to use it.

Individual users R meetings

Tasks already done:

• Removing old wiki from startR	(Vero's suggestion)	startR
• FAQ for moving boundaries in PlotEquiMap	(Roberto's suggestion)	s2dv
• FAQ for EC-Earth experiment path for startR	(Danila's suggestion)	startR
• PlotEquiMap + ShapeFiles	(Chou development)	s2dv
• Persistence forecast	(Deborah's development)	s2dv
• Testing dimensions after compute	(Carlos's tests)	startR
• FAQ for choosing/avoiding a node	(Jaume's suggestion)	startR
• … …	**thanks for your help!**	

Todo list:

• Find error in Power 9 when job fails	(Andrea's suggestion)	startR
• FAQ for returns_var	(many of you)	startR
• Polar region usecase using startR	(Ruben's suggestion)	startR

Individual users R meetings

How to load **irregular** data with **startR**

A: List of single grid points (issue #44)

```
lats <- c(51.30760, 51.97100, 52.10000, 54.01486)
lons <- c(4.519800, 4.926000,- 0.416700, 6.587639)

data <- Start(time = 'all', ensemble = 'all',
              sdate = '19991101',
              latitude = values(lats),
              longitude = values(lons),
              var = 'sfcWind',
              dat = ' 'path/$var$_f6h/$var$_$sdate$.nc',
              pattern_dims = 'dat',
              synonims = list(latitude=c('lat', 'latitude'),
                              longitude=c('lon', 'longitude')),
              return_vars = list(time = NULL,
                                 latitude = 'dat',
                                 longitude = 'dat'),
              retrieve = TRUE)
```

B: Irregular grids (issue #5)

```
repos_path <-
'/esarchive/exp/ecearth/i00k/monthly_mean/heatc/o
hc_2d_avg_0-300m_i00k_$sdate$_fc0-4_*.nc'

heatc <- Start(repos = repos_path,
               sdate = 'first',
               var = 'heatc_sl',
               ensemble = 'all',
               x = 'all',
               y = 'all',
               time = 'all',
               return_vars = list(var_names = NULL),
              var_var = 'var_names')
```

Individual users R meetings: Testing s2dv

We have detected users (or potential users) of different functions in s2dv.

We need your help!

Here are some steps for testing a function:

1. Check documentation
2. Modify the inputs of the function
3. Compare s2dverification vs s2dv by running:
   s2dverification::function_name
   s2dv::function_name

s2dverification document
https://cran.r-project.org/web/packages/s2dverification/s2dverification.pdf
s2dv document https://cran.r-project.org/web/packages/s2dv/s2dv.pdf

Question: MeanDims default na.rm = TRUE (s2dverification) or FALSE (mean version)

status	FUNCTION	REVIEWER
-	Corr.R	Juan
-	Eno.R	Danila
-	InsertDim.R	Bala
Done	MeanDims.R	Jaume
-	RMS.R	Chou
-	RMSSS.R	Andrea
-	Regression.R	Carlos
-	Reorder.R	Ferran
-	Season.R	Francesc
-	Trend.R	Nicola

Issue Discussion

startR - How to select a region?

```
res <- Start(
        dat = path_exp,
        var = 'psl',
        member = 'all',
        sdate = '19821201',
        time = indices(1),
        latitude = values(list(lats.min, lats.max)),
        latitude_reorder = Sort(),
        longitude = values(list(lons.min, lons.max)),
        longitude_reorder = CircularSort(0, 360),
        return_vars = list(latitude = 'dat',
                           longitude = 'dat',
                           time = NULL),
                           retrieve = F)
```

NOTICE!!!

The new startR version includes the fixes for lat/lon selecting & reordering problem. The previous script might be affected. Check your scripts before using.

See FAQ How-to-11

startR - How to select a region? - latitude

```
res <- Start(
        dat = path_exp,
        var = 'psl',
        member = 'all',
        sdate = '19821201',
        time = indices(1),
        latitude = values(list(lats.min, lats.max)),
        latitude_reorder = Sort(),
        longitude = values(list(lons.min, lons.max)),
        longitude_reorder = CircularSort(0, 360),
        return_vars = list(latitude = 'dat',
                           longitude = 'dat',
                           time = NULL),
                           retrieve = F)
```

Desired latitude:

(1) North to south (big to small value)
-> Sort(decreasing = TRUE)

(2) South to north (small to big value)
-> Sort()

- Independent from the original latitude order
- Independent from the transform grid order (if transform is applied)

For example, to read latitude from north to south...

Current	New version
latitude = values(list(20, 10)),	latitude = values(list(10, 20)),
latitude_reorder = Sort(),	latitude_reorder = Sort(decreasing = TRUE),

See FAQ How-to-11

startR - How to select a region? - longitude

lon.min (start point)

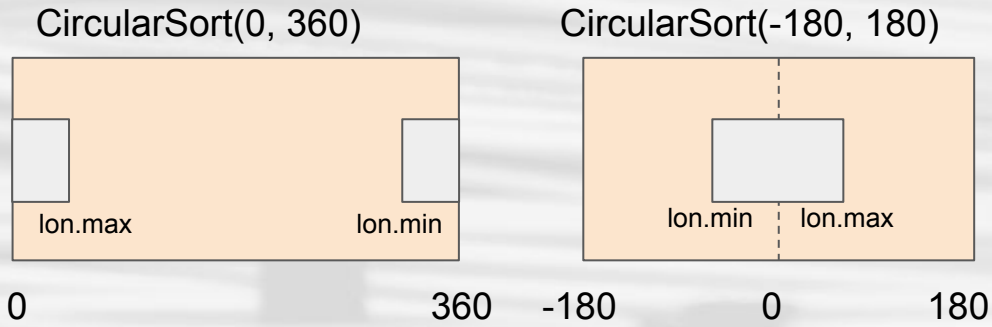
lon.max (end point)

Longitude range (independent of original file definition)

with or without transform

Requested range:	[10, 20]	[20, 10]	[-10, -20]	[-20, -10]	[-10, 10]	[20, -10]	[330:350]	[350:370]
CircularSort(0, 360)	[10:20]	[0:10, 20:360]	[0:340, 350:360]	[340:350]	[0:10, 350:360]	[20:350]	[330:350]	[0:10, 350:360]
CircularSort(-180, 180)	[10:20]	[-180:10, 20:180]	[-180:-20, -10:180]	[-20:-10]	[-10:10]	[-180:-10, 20:180]	[-30:-10]	[-10:10]

```
res <- Start(
        dat = path_exp,
        var = 'psl',
        member = 'all',
        sdate = '19821201',
        time = indices(1),
        latitude = values(list(lats.min, lats.max)),
        latitude_reorder = Sort(),
        longitude = values(list(lons.min, lons.max)),
        longitude_reorder = CircularSort(0, 360),
        return_vars = list(latitude = 'dat',
                           longitude = 'dat',
                           time = NULL),
                           retrieve = F)
```



See FAQ How-to-11

Anything else?

Extra Slides
Packages Description

Individual users R meetings

Simplified startR workflow

1. **Will your analysis decrease the size of your data?**

   yes → try startR

   no → look for another strategy and try startR

Example data dimensions

model member sdate ftime lat lon

similar to s2dverification structure

2. **Which dimensions are used in your computation?**

   e.g.: regrid → lat and lon

   e.g.: climatology → member and sdate

3. **Which dimensions can you use for chunking?** The remaining ones

   e.g.: regrid → member, ftime, sdate

   e.g.: climatology → ftime, lat, lon

ClimProjDiags

Introduction to ClimProjDiags

https://earth.bsc.es/gitlab/es/ClimProjDiags

- ➔ It was created during the **MAGIC C3S contract**
- ➔ It is used in **ESMValTool recipes**
- ➔ It was **named** during a department **R user meeting**
- ➔ It is following the **multiApply paradigm** for compatibility with startR:
    - basic function working in the minimum number of dimensions
    - a wrapper function using Apply* on the basic function to work in N-dimensional arrays with named dimensions

* or other functions of the apply family (e.g. apply, lapply, sapply, mapply, aaply)

Introduction to ClimProjDiags

- ➔ It contains **13 functions**: 8 to compute indices and 5 for formatting
- ➔ It has documentation in **roxygen2** format
- ➔ It contains **4 vignettes** showing the computation of indicators:
  - Anomaly agreement
  - Diurnal temperature range indicator
  - Extreme indices t90p, t10n, rx5days, cdd, wx
  - Heat and coldwaves duration
- ➔ It is published on CRAN

https://cran.r-project.org/package=ClimProjDiags

Introduction to ClimProjDiags

Formatting functions

DailyAno(): Daily anomalies
SeasonSelect(): Selects a season from daily data for multidimensional arrays
SelBox(): Selects spatial region
Subset(): Subsets an N-dimensional array (along, indices, drop parameters)
WeightedMean(): Calculate spatial area-weighted average

Computing Indicators

AnoAgree(): Percentage of anomalies which agrees with the sign of the mean anomaly
Climdex(): Wrapper to compute ETCCDI* climate change indices
CombineIndices(): Combine weighted indices
DTRIndicator(): Diurnal temperature range indicator
DTRRef(): Diurnal temperature range
Extremes(): Sum of spell lengths exceeding daily threshold
Threshold(): Daily thresholds based on quantiles
WaveDuration(): Heat and cold waves duration

*Expert Team on Climate Change Detection and Indices

CSTools

Introduction to CSTools

https://earth.bsc.es/gitlab/es/CSTools

- ➔ It is being developed in **MEDSCOPE** ERA4CS project
- ➔ It is integrating functions from different institutions
- ➔ Its name emerged in a department **R user meeting**
- ➔ It is following the **multiApply paradigm** for compatibility with startR:
  - ○ *CST_function()* working in s2dv_cube objects (OOP)
  - ○ *function()* working in N-dimensional array with named dimensions

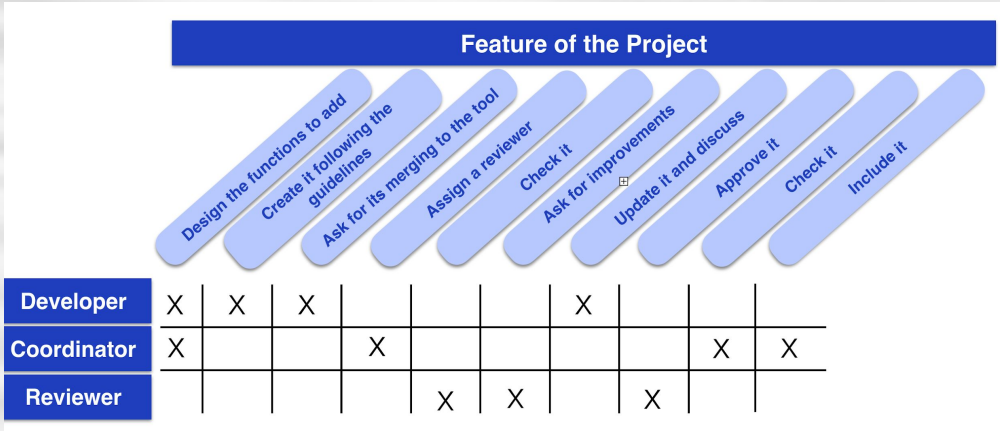
* or other functions of the apply family (e.g. apply, lapply, sapply, mapply, aaply)
*Expert Team on Climate Change Detection and Indices

Introduction to CSTools

- Currently on CRAN **v1.0.1** with **14 functions, 3 data samples and 3 vignettes**:
  - Multi-model Skill Assessment
  - Multivariate RMSE
  - RainFARM

https://cran.r-project.org/package=CSTools

- Working in the next release (November 2019) with 25 functions
- Development Guidelines:
  - Roles
  - Workflow
  - Branching strategy



Introduction to CSTools

Basic functions	Correction	Downscaling	Evaluation
CST_Load	CST_BiasCorrection	CST_Analogs	CST_MultivarRMSE
CST_Anomaly	CST_Calibration	Analogs	CST_MultiMetric
CST_SaveExp	CST_QuantileMapping	CST_RainFARM	CST_MultiEOF
CST_SplitDim	CST_BEI_Weighting	CST_RFSlope	
CST_MergeDims	BEI_Weights	CST_RFWeights	
s2dv_cube	BEI_PDFBest*	RainFARM	
as.s2dv_cube	CST_CategoricalForecast	RFSlope	

Ploting functions

PlotMostLikelyQuantileMap

PlotForecastPDF

PlotCombinedMap

Future future functionalities: ADAMONT, SMOP and Dynamical Bias Correction, RainFARM for Temperature downscaling, Weather Regimes, PlotTriangles

*RainFARM: Rainfall Filtered Autoregressive Model