AMAP – METRADICA project – March 2023

gecevar R package
GEt Climatic and Environmental VARiables

Jeanne CLEMENT[1] Pierre GUILLAUMONT[1] Ghislain VIEILLEDENT[1]

[1] **Cirad** UMR AMAP

Outline

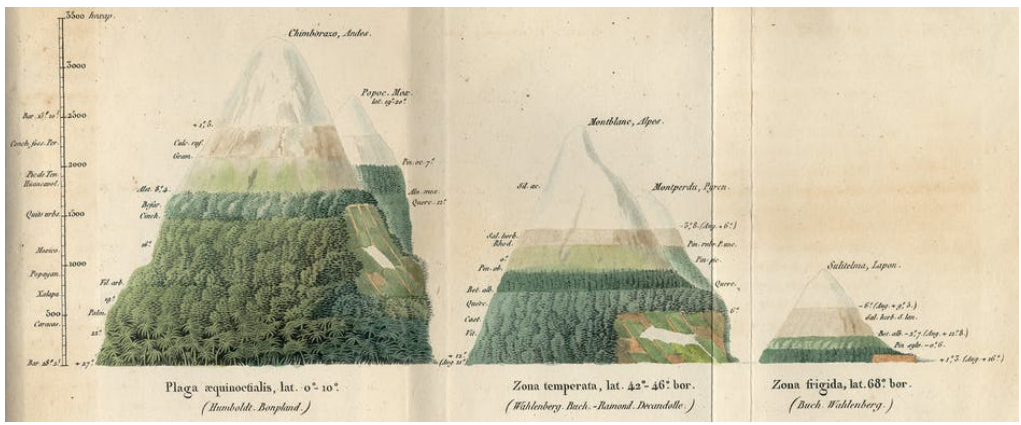
1. Introduction
   - Context
   - Existing software
   - Objectives
2. gecevar R package
   - Name and website
   - Functionalities
   - Variables
   - Specificities
3. Case-studies
   - Examples of computation time
   - French Guiana

Outline

1. Introduction
   - Context
   - Existing software
   - Objectives
2. gecevar R package
   - Name and website
   - Functionalities
   - Variables
   - Specificities
3. Case-studies
   - Examples of computation time
   - French Guiana

Context

- Ecology is the study of the relationships among living organisms and their physical environment.
- Use of spatial environmental and climatic layers is inevitable in ecology.
- (NB : Here, environmental variable = any environmental variable which is not climatic).
- Study of the link between environment/climate and : species occurrence, species demography, individual traits, community characteristic (e.g. diversity, productivity, biomass, community weighted mean), etc.



Data online

- Many global climatic and environmental data are available online.
- Climate : WorldClim, Chelsa.
- Elevation : SRTM, protected areas : WDPA, population : WorldPop, soil : SoilGrids, forest cover : Global Forest Change or Tropical Moist Forests, etc.
- Data are usually available at various resolution (SRTM : 90m), (WorlClim, Chelsa : $\geq$ 30 arc sec).
- Building a data-set from all these source that can be used in ecological studies is challenging.

Time consuming and repetitive tasks

Time consuming

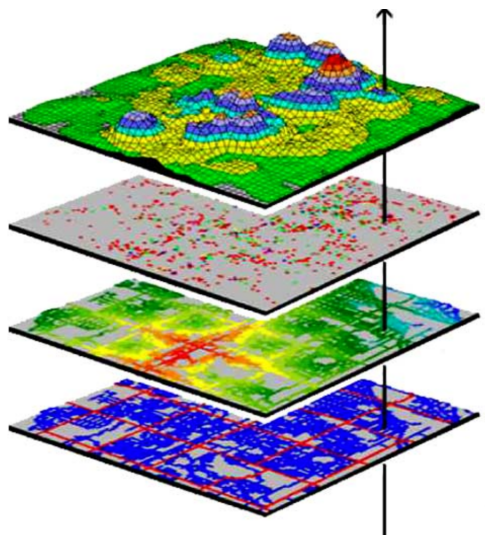
- Many environmental variables (topography, soil, climate, etc.).
- Environment and climatic data are spread in different databases.
- Long computations if region is large and resolution is high.
- Downloading rasters at global scale can take a lot of time.

Repetitive

- We usually repeat this work for every region we are working on.
- For future climate data : repetition for each global climate model, scenario (SSP), and period (eg. 2055, 2085).

Technically challenging steps

- Several technical geoprocessing steps : crop to an extent, resample on a given grid (extent and resolution), combine data in one raster.
- Some variables are missing and must be computed (number of dry months or average from climatic models).
- Intensive computations if region is large and resolution is high (raster might not fit in memory).
- Imply the use of technical geoprocessing software : R packages (terra, star, sf), gdal, GRASS GIS, Google Earth Engine.



Existing software

Many different software

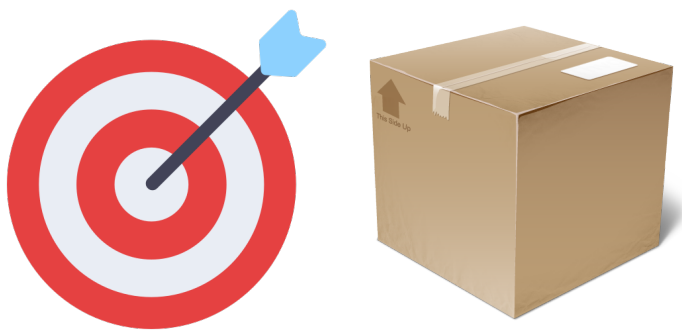
- Climate : `geodata` (WorldClim 2.1), `climate` (weather station), `raster` (WordlClim 1.4), `climateR`
- WDPA : `geodata`, `wdpar`, `worldpa`
- SRTM : `geodata`, `elevatr`
- OSM : `geodata`, `osmextract`

Inconvenients

- Specific data only or specific region only.
- Full download of global maps or tiles (time consuming).
- No post-processing of downloaded data (resampling on a grid, raster stack).
- Missing variables.

Objectives

- Provide an R package to ease the creation of a dataset with environmental and climatic variables for any particular region.
- With easy-to-use and well documented functions.
- Using efficient code and tools to fasten the creation of the dataset.
- $\rightarrow$ Aims of the `gecevar` R package that will be presented.

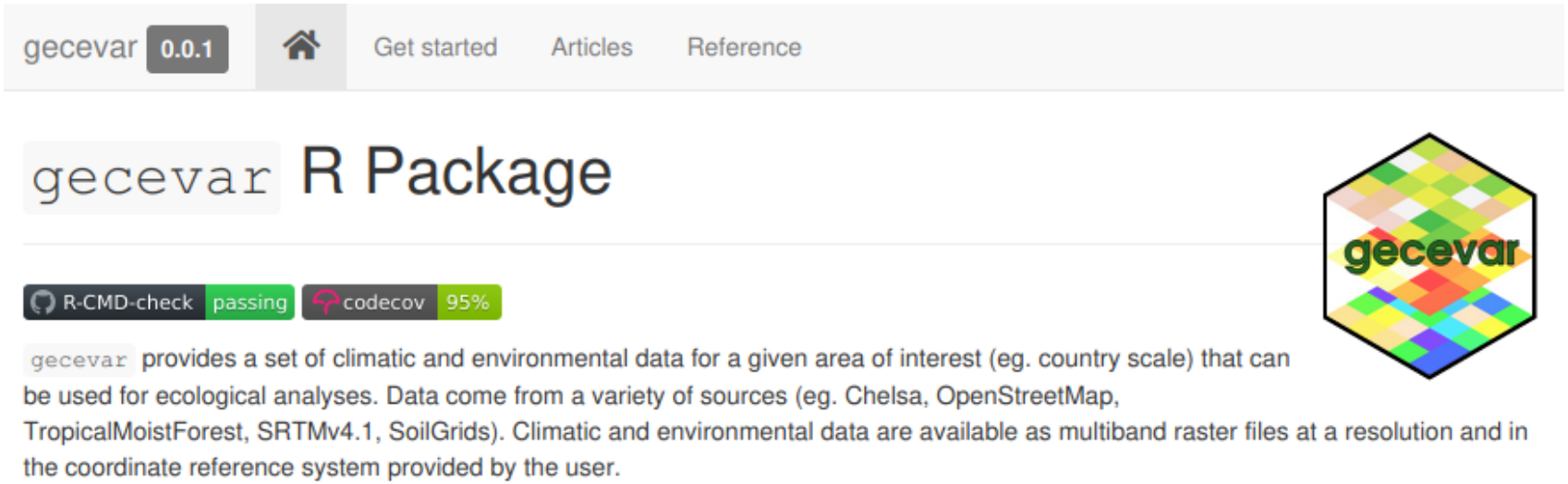


Outline

1. Introduction
   - Context
   - Existing software
   - Objectives
2. gecevar R package
   - Name and website
   - Functionalities
   - Variables
   - Specificities
3. Case-studies
   - Examples of computation time
   - French Guiana

Name and website

- `gecevar` : GEt Climatic and Environmental VARiables.
- Website : https://ecology.ghislainv.fr/gecevar



Functionalities

Functions

- `get_env_variables` : Raster file with 13 environmental variables.
- `get_chelsa_current` : Raster file with 107 variables from Chelsa describing current climate (1981–2010).
- `get_chelsa_future` : Raster file with 81 variables from Chelsa describing future climate (for each GCM, SSP, and period).

Input

By default, the user has to provide only one of the following :

- Country ISO code.
- Shapefile with polygons delimiting the region of interest.
- Extent (xmin, ymin, xmax, ymax) of the region of interest.

Environmental variables

Sources

Various sources : SRTM, SoilGrids, Tropical Moist Forests, OpenStreetMap, WDPA.

Topography

Elevation, Aspect, Slope, Roughness.

Other environmental variables

Solar irradiance, Soil, Forest cover, Distances to forest, sea, road, town, water, Protected areas.

Current climatic variables

Source

Chelsa (chelsa-climate.org)

Bioclimatic variables

- The monthly min, max and mean temperature and monthly precipitation (48 variables).
- The 19 commonly used bioclimatic variables (11 from temperature and 8 from precipitation).

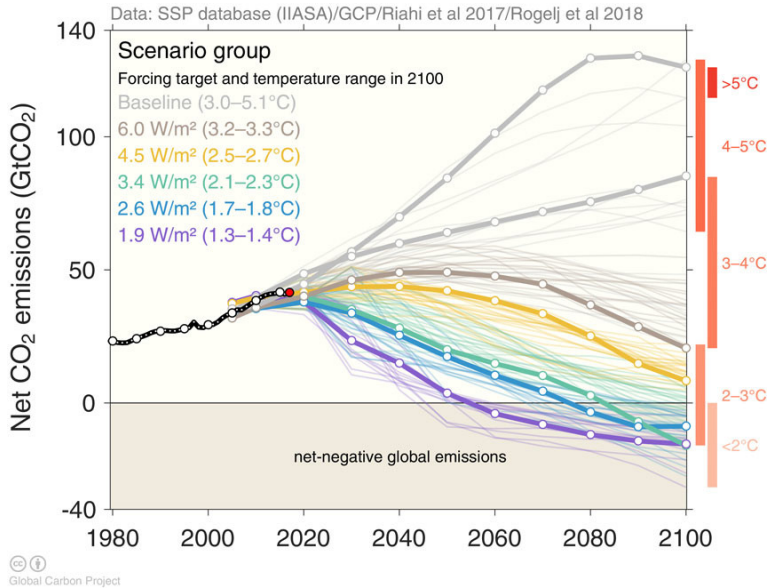
Other climatic variables

- Cloud area fraction, PET, Climatic water deficit (CWD), Number of dry months (NDM).
- For PET and derived variables (CWD, NDM) : Penman-Monteith or Thornthwaite equation.

Future climate

Source

- Chelsa (chelsa-climate.org)
- GCMs : GFDL-ESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR, MRI-ESM2-0, UKESM1-0-LL.
- SSPs : Shared Socio-economic Pathways (126, 370, 585).
- Periods : 2041–2070, 2071–2100.



Future climate

- Most of the climatic variables for each GCM, SSP, and period of time.
- Exception : no Penman-Monteith PET and derived variables.
- Averages from the 5 GCMs (ensemble model).

Software used

- Heavy use of GDAL (`gdalwarp`, `gdalbuilvrt`, `gdal_translate`, `gdaldem`, `gdal_proximity`).
- GRASS GIS software for solar irradiance, (TWI).
- `osmextract` R package for extracting OSM data.
- `terra`, `stars` and `sf` R packages for manipulating spatial objects.

Cloud optimized GeoTIFFs (COGs)

COGs

- Download of large global rasters available online can be very long.
- Here we make use of COGs : Cloud Optimized GeoTIFFs.
- A Cloud Optimized GeoTIFF (COG) is a regular GeoTIFF file, aimed at being hosted on a HTTP file server, with an internal organization that enables more efficient workflows on the cloud
- Download of a sample of the global raster.
- Use of the `gdal_translate` function in GDAL which allows using the `/vsicurl` virtual file system as input.

Some ressources on COGs

- https://www.cogeo.org/
- https://trac.osgeo.org/gdal/wiki/CloudOptimizedGeoTIFF
- https://forestatrisk.cirad.fr/notebooks/cog.html

Outline

1. Introduction
   - Context
   - Existing software
   - Objectives
2. gecevar R package
   - Name and website
   - Functionalities
   - Variables
   - Specificities
3. Case-studies
   - Examples of computation time
   - French Guiana

Examples of computation time

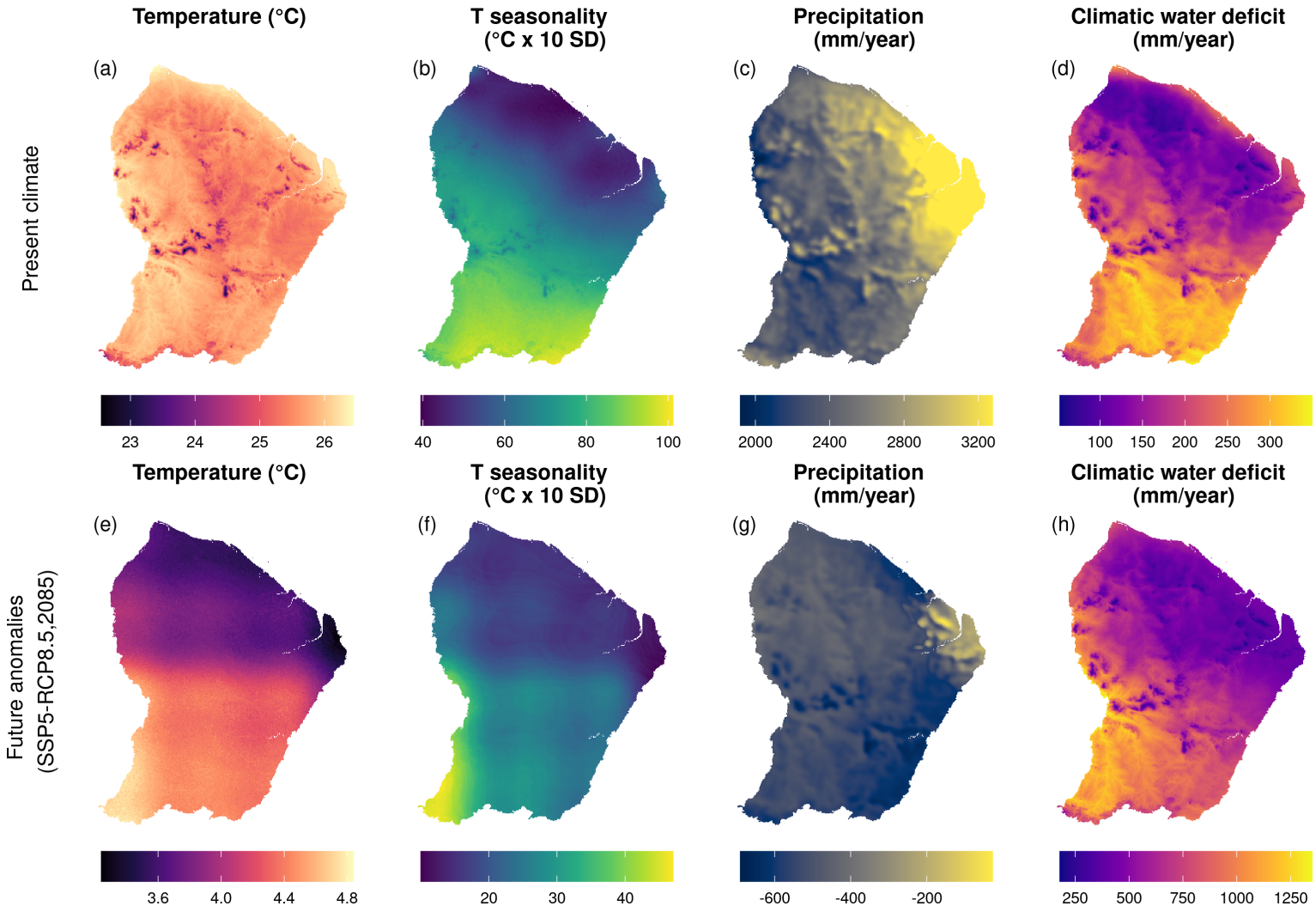
Table 1 – Area size

Area	raster size	nb. of cells
Madagascar	1673 x 875	1,463,875
New Caledonia	689 x 1921	1,323,569
French Guiana	444 x 445	197,580

Table 2 – Computation time for 1km resolution

Area	env	clim current	merge files	clim future
Madagascar	13min	25min	6min	1h15min
New Caledonia	5min	9min	8min	47min
French Guiana	3min	6min	2min	25min

French Guiana



. . . Thank you for attention . . .

https://ecology.ghislainv.fr/presentations

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

cirad

AMAPlab

ceba