HARMONIZE

clim4health: a new R package to harmonize climate datasets for health impact studies

Emily Ball, Alba Llabrés-Brustenga, Daniela Lührsen, Ania Kawiecki, Raúl Capellán, Carles Milà, Rachel Lowe

12 May 2025

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

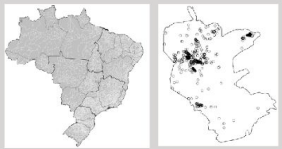
This work was supported by the Wellcome Trust grant number 224694/Z/21/Z.

User inputs

What kind of data?
- Health outcome
- Socio-economic and environmental indicators

Spatial scale
- Shape file (administrative level boundaries - country, regions, states, municipalities, etc)
- Coordinates (points)

Temporal scale
- Time step (weekly, monthly, yearly)
- Start/end date



data4health

Incidence rates | Probable infection

socio4health

Census | Household survey

land4health

Satellite images ⇔ Drone images

clim4health

Met station network
Reanalysis
Seasonal forecasts
⇔ Weather and hydrological sensors

Harmonized space-time data for polygons or points

HARMONIZE Users

Observatório de CLIMA e SAÚDE

INFO DENGUE

MINISTERIO DE SALUD Y PROTECCIÓN SOCIAL

PERÚ Ministerio de Salud

GOBIERNO DE LA REPÚBLICA DOMINICANA
SALUD PÚBLICA

HDX

National/regional | Hotspot | ⇕ Downscaling/upscaling algorithms

Example sources of climate data

Seasonal forecasts available from Copernicus:

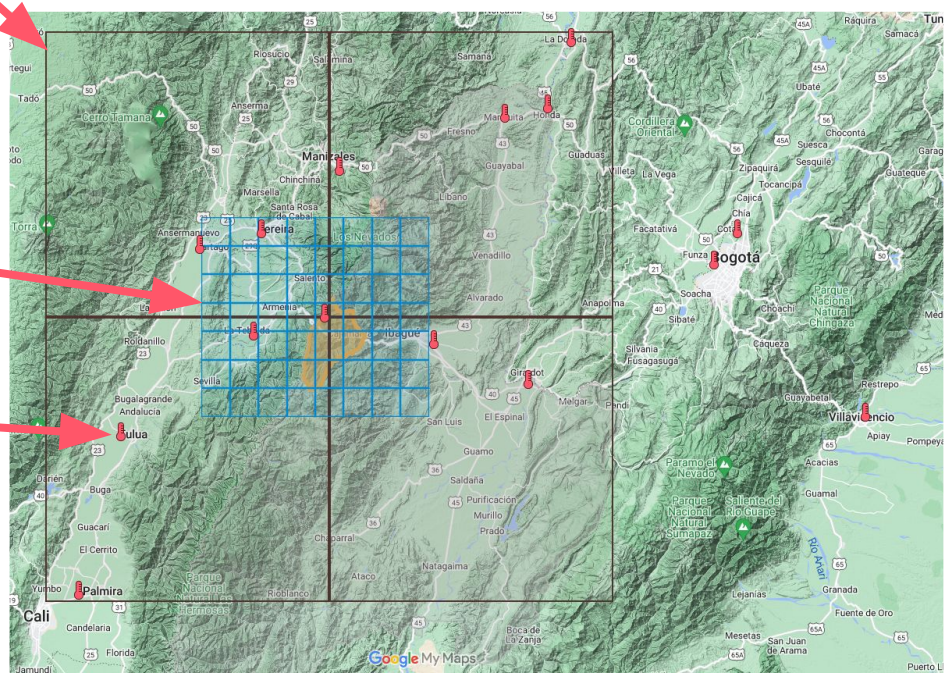
- Spatial resolution: 1° x 1°
- Temporal frequency: 6h, 12h, 24h and monthly statistics
- Hindcast period: 1993-2016
- System: ECMWF-s5

Reanalysis available from Copernicus:

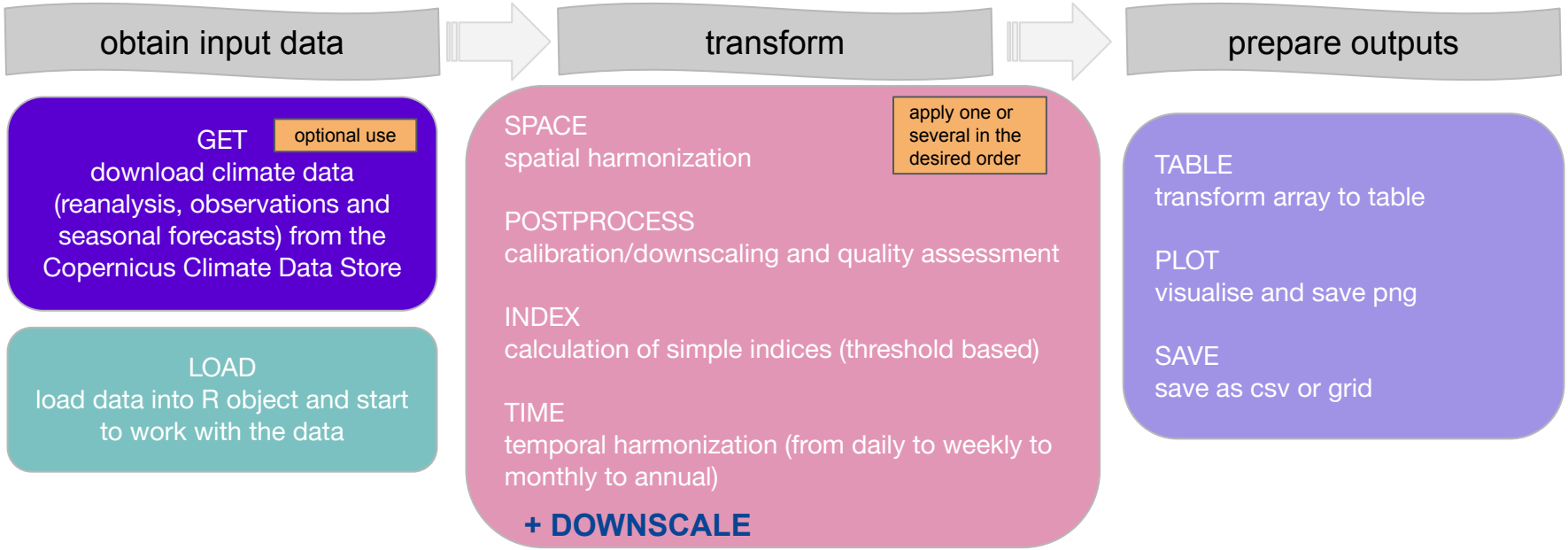
- ERA5 1979-present 0.25° x 0.25°
- ERA5-Land 1950-present 0.1° x 0.1°

Ground-based observations:

- Global Hourly - Integrated Surface Database (ISD): point data, several variables, 1901-present



clim4health structure



For each function, we will also provide detailed vignettes, as well as recommendations for parameter selection

clim4health_get

- Download data from the Copernicus Data Store

- Explore the datasets available
  - datasets
  - variables
  - required inputs
  - how to specify the desired region


```
clim4health_get_help()
# Available climate datasets:
# - reanalysis-era5-land
# - reanalysis-era5-land-monthly-means
# - reanalysis-era5-single-levels
# - reanalysis-era5-single-levels-monthly-means
# - seasonal-monthly-single-levels
```

```
# Download reanalysis data
clim4health_get(pat = pat_api,
                dataset = "reanalysis-era5-land-monthly-means",
                product_type = "monthly_averaged_reanalysis",
                variable = "2m_temperature",
                year = c(2010, 2011, 2012),
                month = c(4, 5),
                area = c(33, -93, -23, -17),
                outname = "era5land")
```

clim4health_load

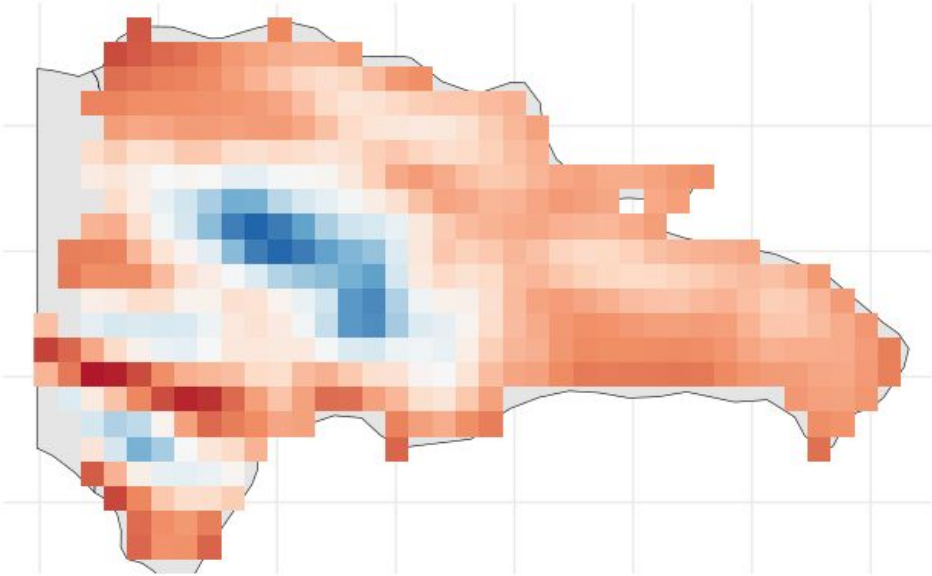
- load NetCDF or csv files
- account for matching dates between reanalysis and hindcasts
- load is the base to establish the dimensions of the array to work with in the rest of the functions

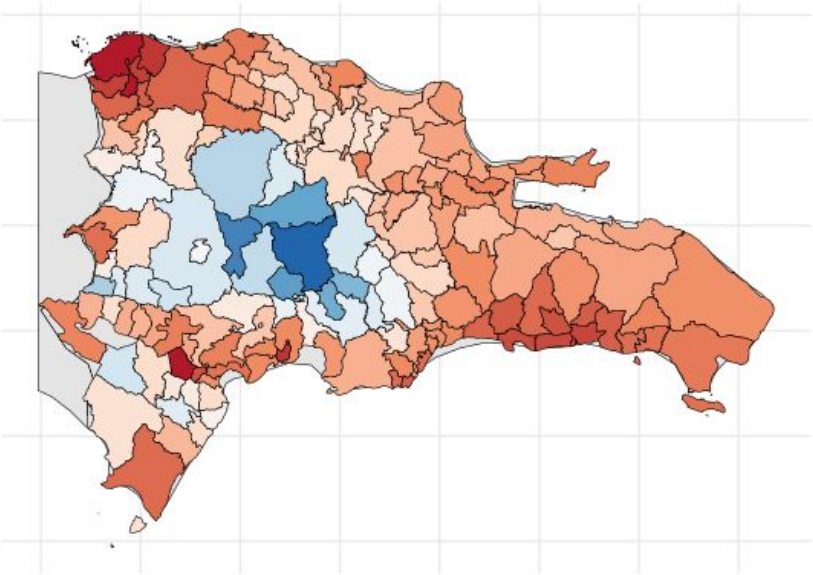
```
class(obs)
#[1] "s2dv_cube"

dim(obs$data)
#  dataset       var      time  ensemble  latitude longitude
#        1         1         6         1       561       761
```

clim4health_space

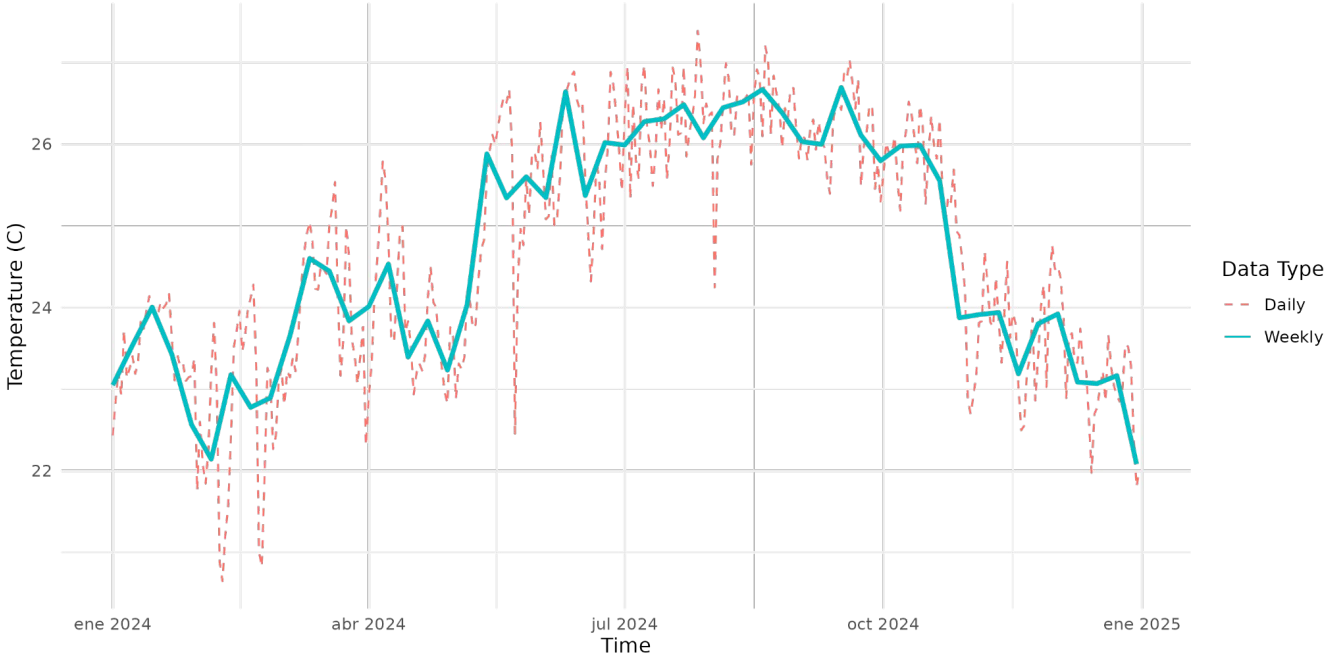
- aggregate gridded climate data to spatial regions





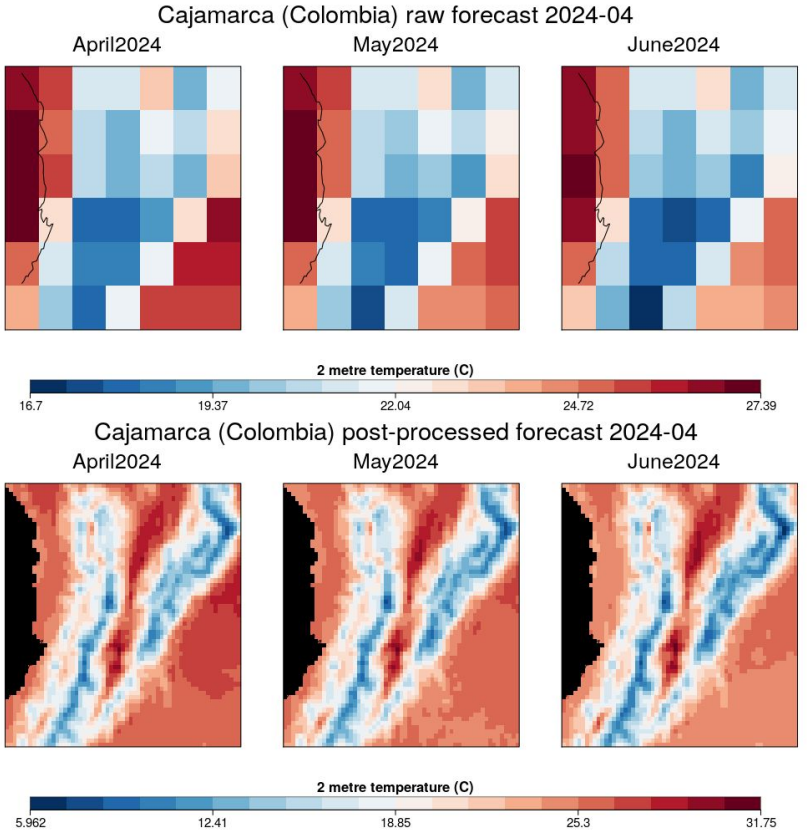
clim4health_time

- aggregate raw climate data to desired temporal frequency
- e.g. daily to weekly data



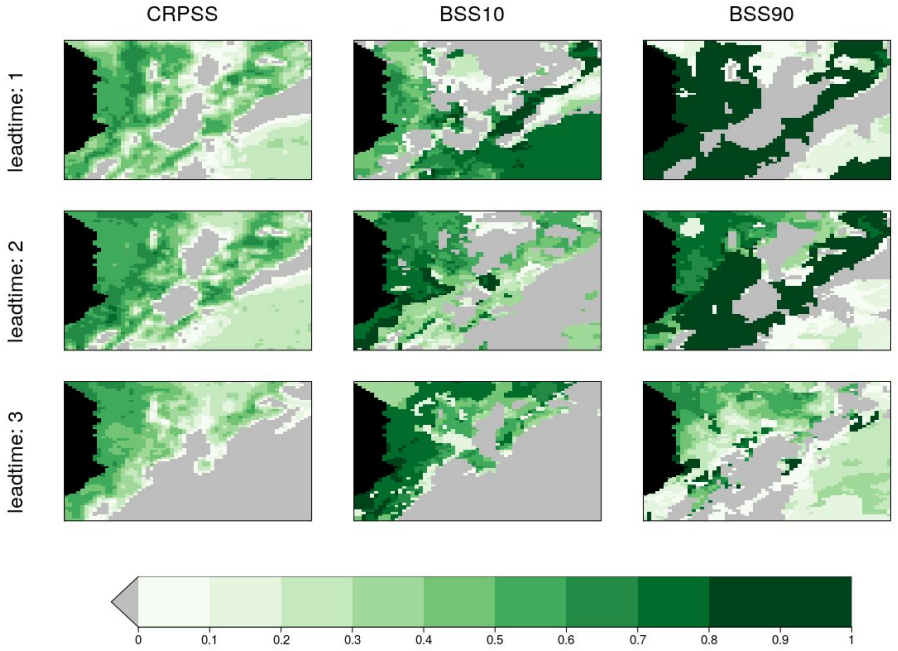
clim4health_downscale

- downscale coarse climate data to a finer grid or to point locations
- plus recommendations of methods for variables and hotspots
- **new function**



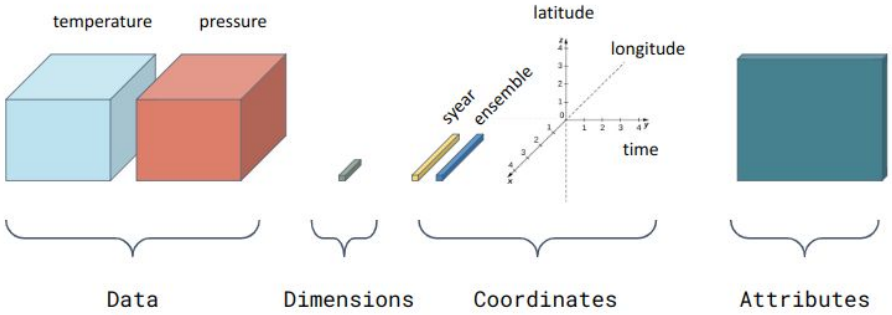
clim4health_postprocess

- calibration and quality assessment of data
- how does the distribution of a hindcast compare to the climatological observations? How skilfully does it predict the:
  - full distribution
  - tails of the distribution
- In progress!



clim4health_table

- output multidimensional climate data with metadata into a tidy dataframe



dataset	var	time	ensemble	latitude	longitude	values
dat1	t2m	2024-04-01	1	33	-93	291.61
dat1	t2m	2024-05-01	1	33	-93	298.08
dat1	t2m	2024-06-01	1	33	-93	301.45
dat1	t2m	2024-04-01	2	33	-93	292.39
dat1	t2m	2024-05-01	2	33	-93	296.74
dat1	t2m	2024-06-01	2	33	-93	299.25

Status: May 2025

clim4health development

finished
started
delayed

	first version	documentation	unit tests	vignette	improvements	final version
clim4health_get	October 2024	November 2024	Feb-July 2025	December 2024	October 2025	October 2025
clim4health_load	October 2024	November 2024	Feb-July 2025	December 2024	October 2025	October 2025
clim4health_spatial	November 2024	November 2024	Feb-July 2025	December 2024	November 2025	November 2025
clim4health_downscale	June 2025	July 2025	August 2025	July 2025	December 2025	March 2026
clim4health_postprocess	May 2025	June 2025	Feb-July 2025	July 2025	December 2025	March 2026
clim4health_index	October 2024	November 2024	Feb-July 2025	December 2024	January 2026	March 2026
clim4health_time	October 2024	November 2024	Feb-July 2025	December 2024	January 2026	March 2026
clim4health_table	November 2024	November 2024	Feb-July 2025	December 2024	January 2026	March 2026
clim4health_plot	August 2025	August 2025	September 2025	September 2025	February 2026	March 2026
clim4health_save	August 2025	August 2025	September 2025	September 2025	February 2026	March 2026

HARMONIZE

Thank you - any questions?

Emily Ball, Alba Llabrés-Brustenga, Daniela Lührsen, Ania Kawiecki, Raúl Capellán, Carles Milà, Rachel Lowe

12 May 2025

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

This work was supported by the Wellcome Trust grant number 224694/Z/21/Z.