ISI regional statistics conference

GHR tools

Streamlining climate and health workflows including data harmonization, modelling and early warning systems

Carles Milà, Research Engineer

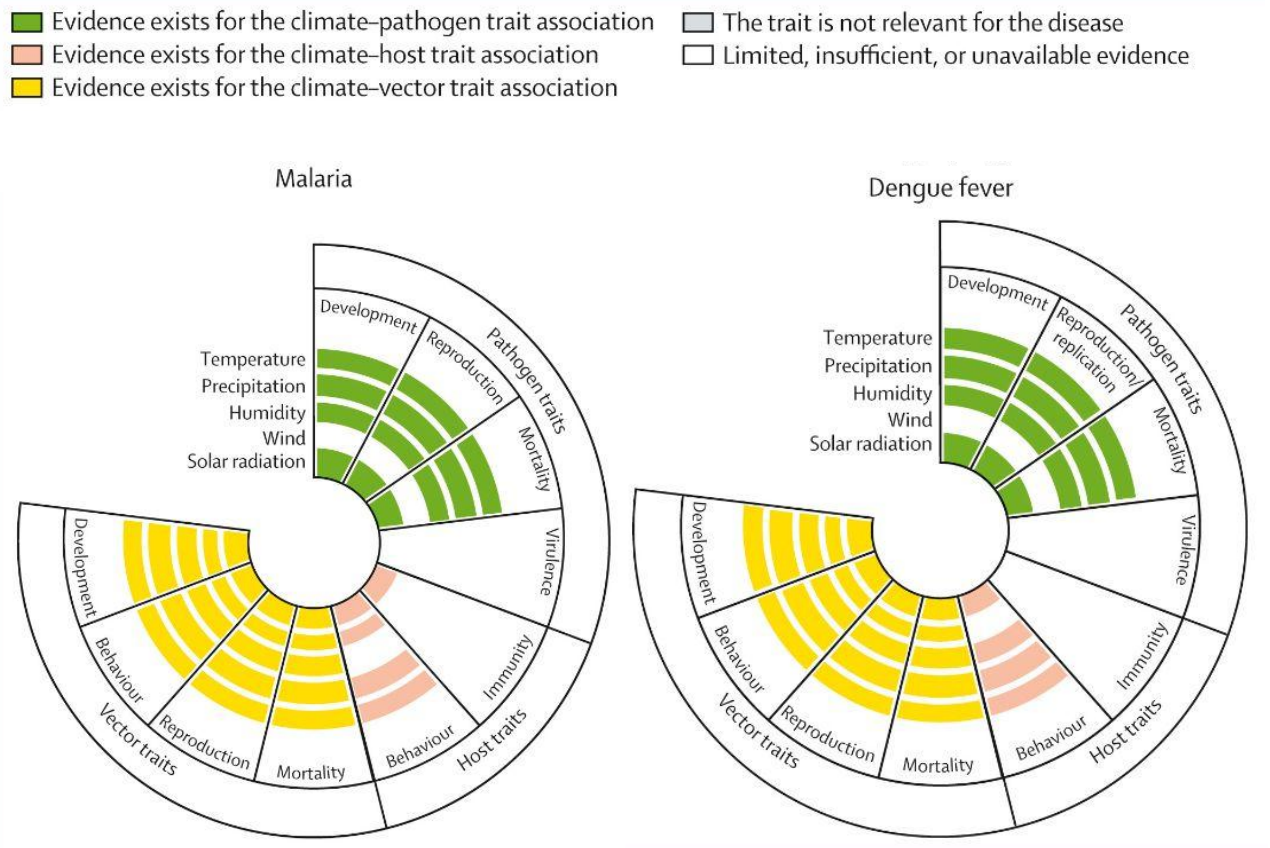
BSC Earth Sciences - Global Health Resilience

carles.milagarcia@bsc.es

ICREA

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Climate variables associated with infectious diseases

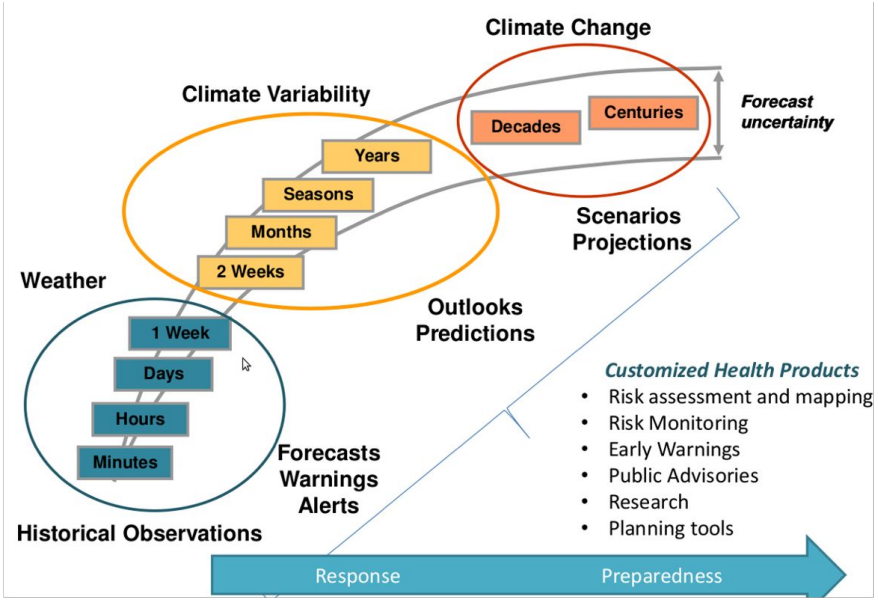


Alcayna *et al*.
Lancet Planetary Health 2025

Early warning systems for climate-sensitive diseases

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Climate-informed early warning systems integrate climate data to predict the risk of an infectious disease outbreak.

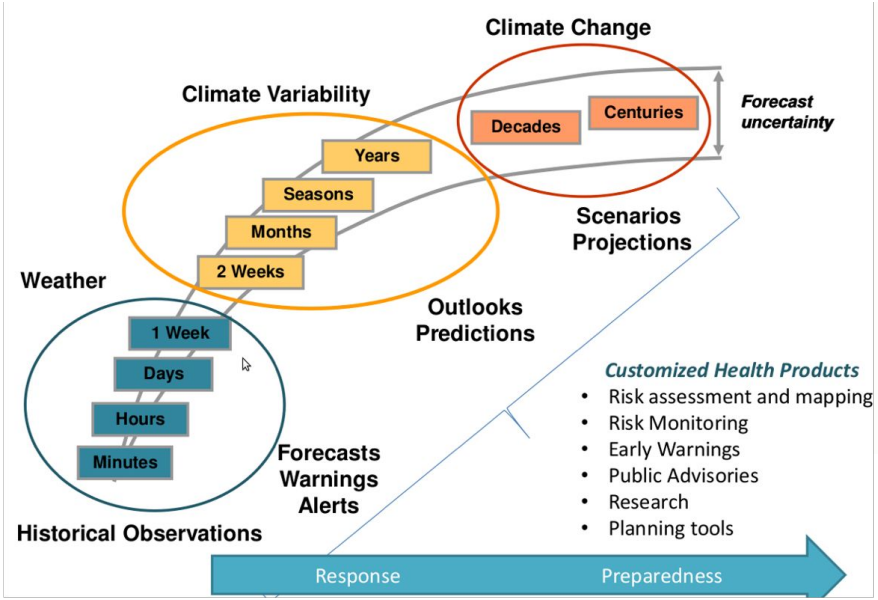


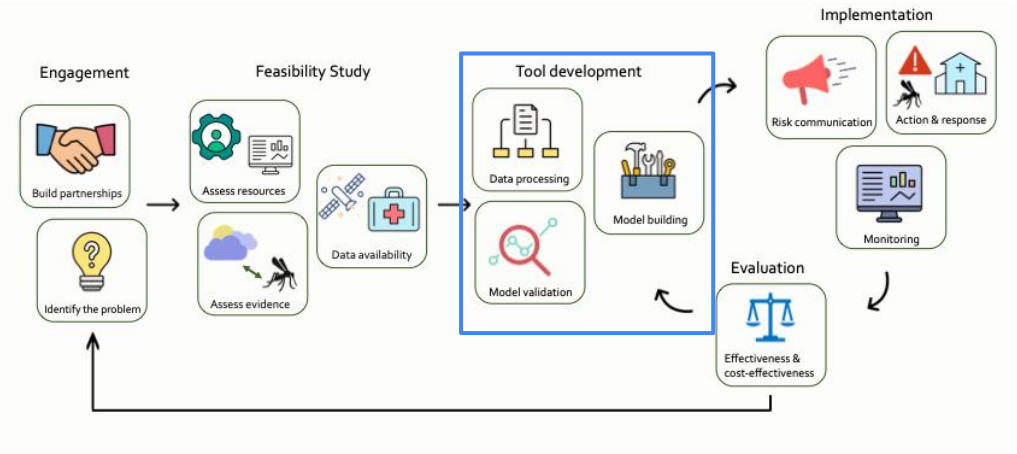
Courtesy: J. Shumake-Guillemot, WMO

Early warning systems for climate-sensitive diseases

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Climate-informed early warning systems integrate climate data to predict the risk of an infectious disease outbreak.

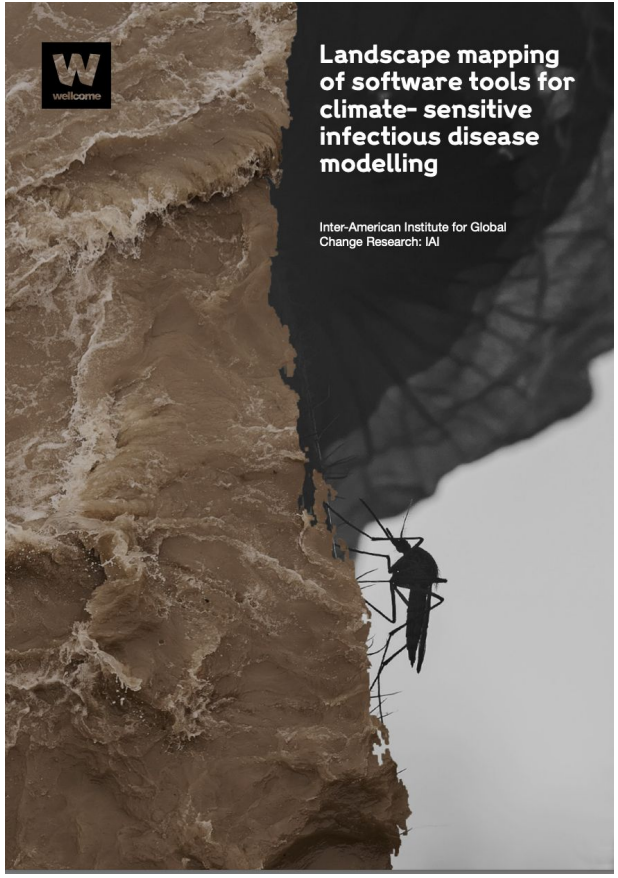




Finch et al 2023

Courtesy: J. Shumake-Guillemot, WMO

Early warning systems for climate-sensitive diseases



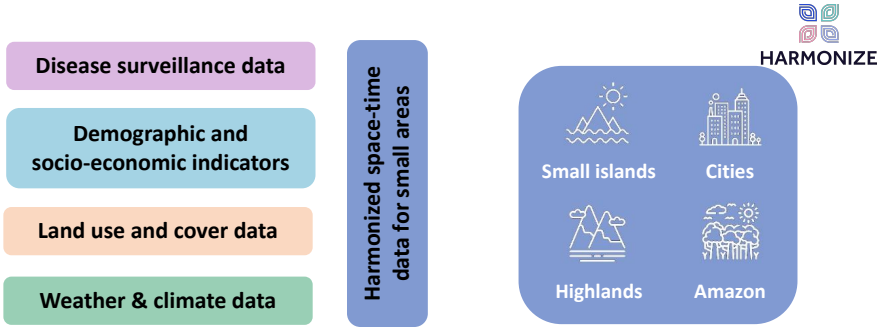
- A review of > 9,000 papers found that only 37 of them contained operational early warning systems.
- Lack of user-friendly, open and reproducible software to sustainably implement modelling tools anywhere in the globe.
- Need to develop open-source and free software for climate-sensitive early warning systems.

About Global Health Resilience (GHR)

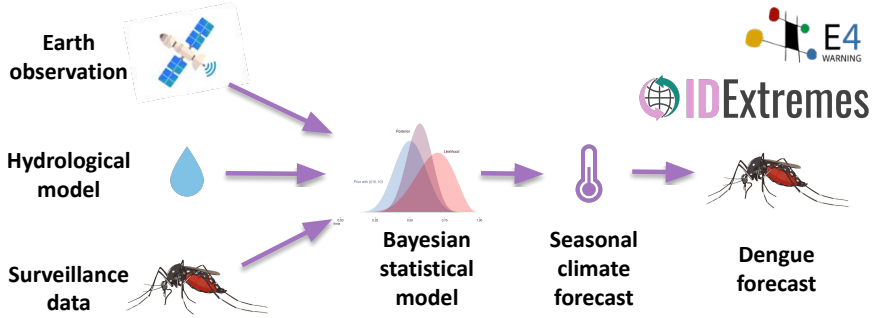
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

We build digital tools for

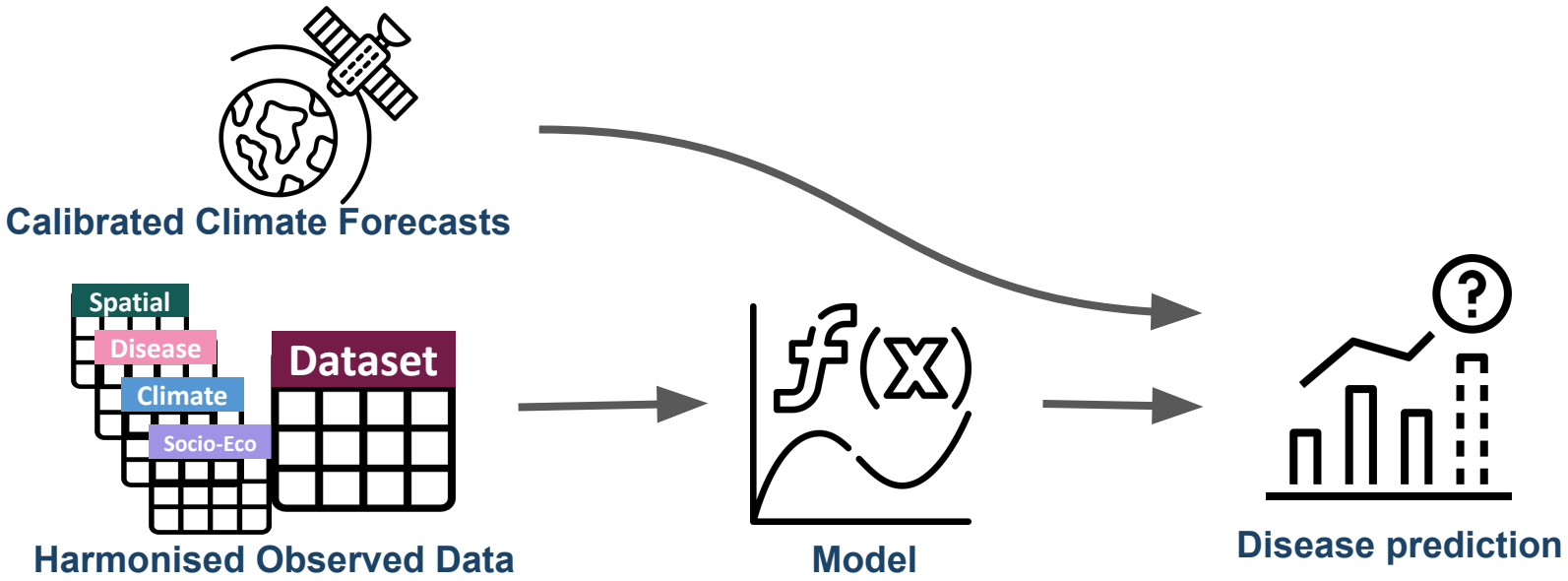
Harmonizing multi-scale spatiotemporal data for health in climate change hotspots



Developing early warning systems for climate-sensitive infectious diseases



Using climate forecasts and models for disease risk prediction

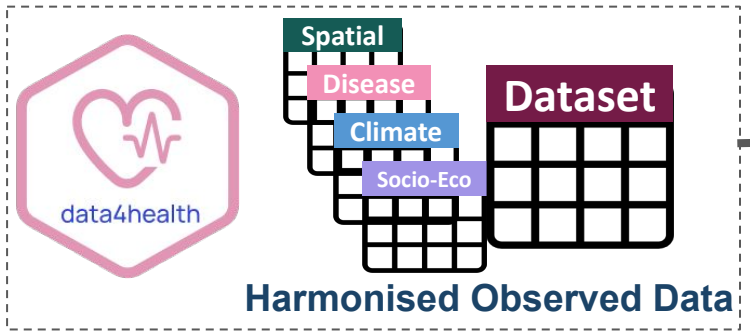


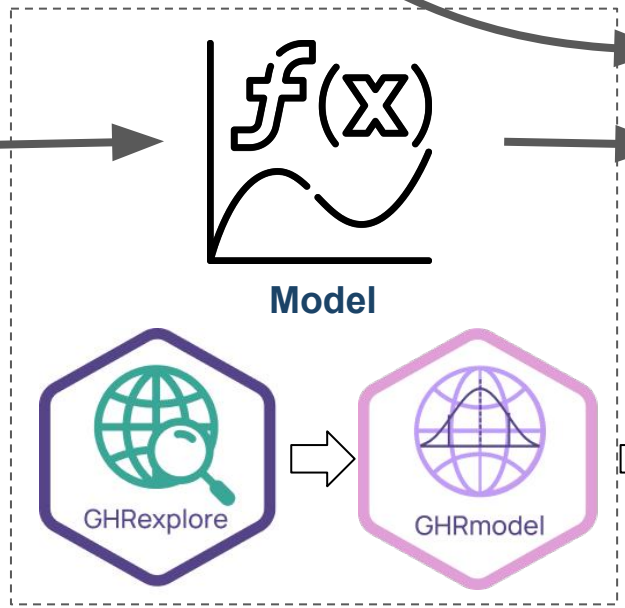
How GHRtools fit into the disease risk modelling and forecasting framework

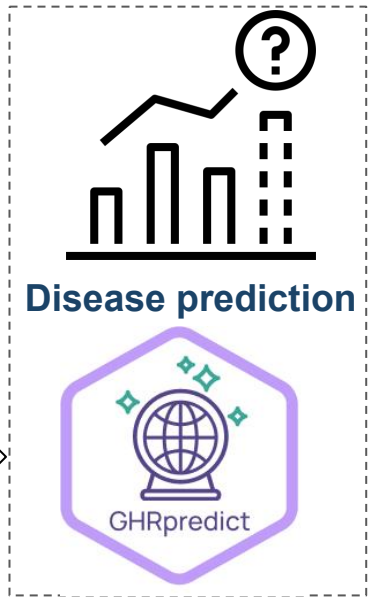
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

HARMONIZE

IDExtremes







Target users

Public health institutions, humanitarian organisations, researchers, meteorological services

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

data4health

An R package to facilitate the access, preprocessing, and aggregation of health data at customized spatiotemporal resolutions.

DANIELA LÜHRSEN
DATA SCIENTIST

DR RAQUEL LANA
RESEARCHER

PROF RACHEL LOWE
LEADING RESEARCHER

Expected release date:
July 2026

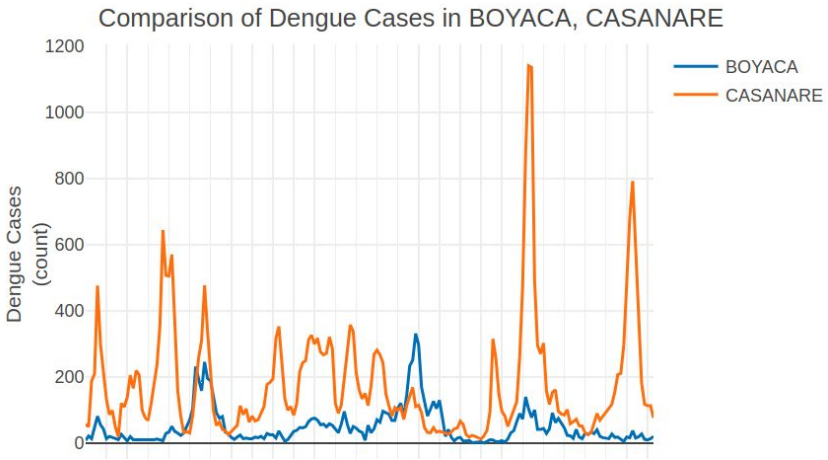
data4health pipeline

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Unprocessed health data

	J	K	M	N	O	P	Q	R	S	T	U	V	W	X
1	DT_CLOS	DT_BIRTH	DT_NOTI	UF_NOTI	MUNI_NOTI	AGE	SEX	PREG	RACE	UF_INF	COPAISINF	MUNI_INF	CLASS	CRIT
2	2015-03-05	1979-07-27	2015-03-02	12	120020	35	F	6	NA	12	1	120033	10	2
3	2015-02-02	2000-09-22	2015-01-22	12	120020	14	F	6	NA	12	1	120033	10	2
4	2015-05-17	2005-05-10	2015-05-17	12	120020	10	F	6	NA	12	1	120020	10	1
5	2015-06-22	1924-04-02	2015-06-12	12	120020	91	F	6	4	12	1	120020	10	1
6	2015-03-23	2000-08-14	2015-03-17	12	120020	14	F	6	NA	12	1	120020	10	2
7	2015-03-19	1997-12-13	2015-03-11	12	120020	17	F	6	4	12	1	120020	10	2
8	2015-03-30	2012-12-23	2015-03-20	12	120020	2	F	6	4	12	1	120020	10	2
9	2015-03-23	1991-10-20	2015-03-18	12	120020	23	F	6	NA	12	1	120020	10	2
10	2015-03-19	1981-04-15	2015-03-12	12	120020	33	F	5	1	12	1	120020	10	2
11	2015-03-05	1958-01-31	2015-03-01	12	120020	57	F	6	NA	12	1	120020	10	2
12	2015-03-05	1995-11-28	2015-02-25	12	120020	19	F	6	4	12	1	120020	10	2
13	2015-03-05	1976-06-01	2015-02-28	12	120020	38	F	6	4	12	1	120020	10	2
14	2015-03-09	1958-06-01	2015-03-05	12	120020	56	F	6	NA	12	1	120020	10	2
15	2015-03-05	1990-09-01	2015-02-26	12	120020	24	F	6	4	12	1	120020	10	2
16	2015-02-13	2003-05-19	2015-02-09	12	120020	11	F	5	4	12	1	120020	10	2
17	2015-02-20	1996-06-19	2015-02-15	12	120020	18	F	5	4	12	1	120020	10	2
18	2015-02-26	1950-02-02	2015-02-20	12	120020	65	F	6	NA	12	1	120020	10	2
19	2015-02-20	2008-12-03	2015-02-13	12	120020	6	F	6	4	12	1	120020	10	2
20	2015-02-24	1950-02-02	2015-02-18	12	120020	65	F	5	4	12	1	120020	10	2
21	2015-02-26	2008-12-03	2015-02-12	12	120020	6	F	6	4	12	1	120020	10	2
22	2015-02-26	1998-07-24	2015-02-21	12	120020	16	F	6	4	12	1	120020	10	2
23	2015-02-13	1961-12-03	2015-02-10	12	120020	53	F	5	4	12	1	120020	10	2
24	2015-02-25	1966-07-09	2015-02-16	12	120020	48	F	5	4	12	1	120020	10	2
25	2015-02-11	2006-07-10	2015-02-05	12	120020	8	F	6	NA	12	1	120020	10	2
26	2015-02-24	1973-09-30	2015-02-19	12	120020	41	F	6	4	12	1	120020	10	2
27	2015-02-10	1981-12-30	2015-01-09	12	120020	33	F	5	4	12	1	120020	10	1
28	2015-02-09	1996-09-14	2015-02-03	12	120020	18	F	6	4	12	1	120020	10	2
29	2015-02-11	1998-08-25	2015-02-06	12	120020	16	F	6	4	12	1	120020	10	2
30	2015-02-02	1997-01-17	2015-01-23	12	120020	18	F	5	4	12	1	120020	10	2
31	2015-02-04	1959-06-14	2015-01-30	12	120020	55	F	6	4	12	1	120020	10	2
32	2015-02-04	1947-06-07	2015-02-02	12	120020	67	F	6	4	12	1	120020	10	2
33	2015-02-04	1977-06-03	2015-01-30	12	120020	37	F	6	NA	12	1	120020	10	2

Clean health data



Load

Load from common file formats (csv, rds, xslx, json, dbf, dbc)

Clean

- Standardize names and values
- Convert dates
- Remove missing or invalid records

Filter

- Select:
  - Date ranges
  - Demographics
  - Numeric thresholds

Aggregate

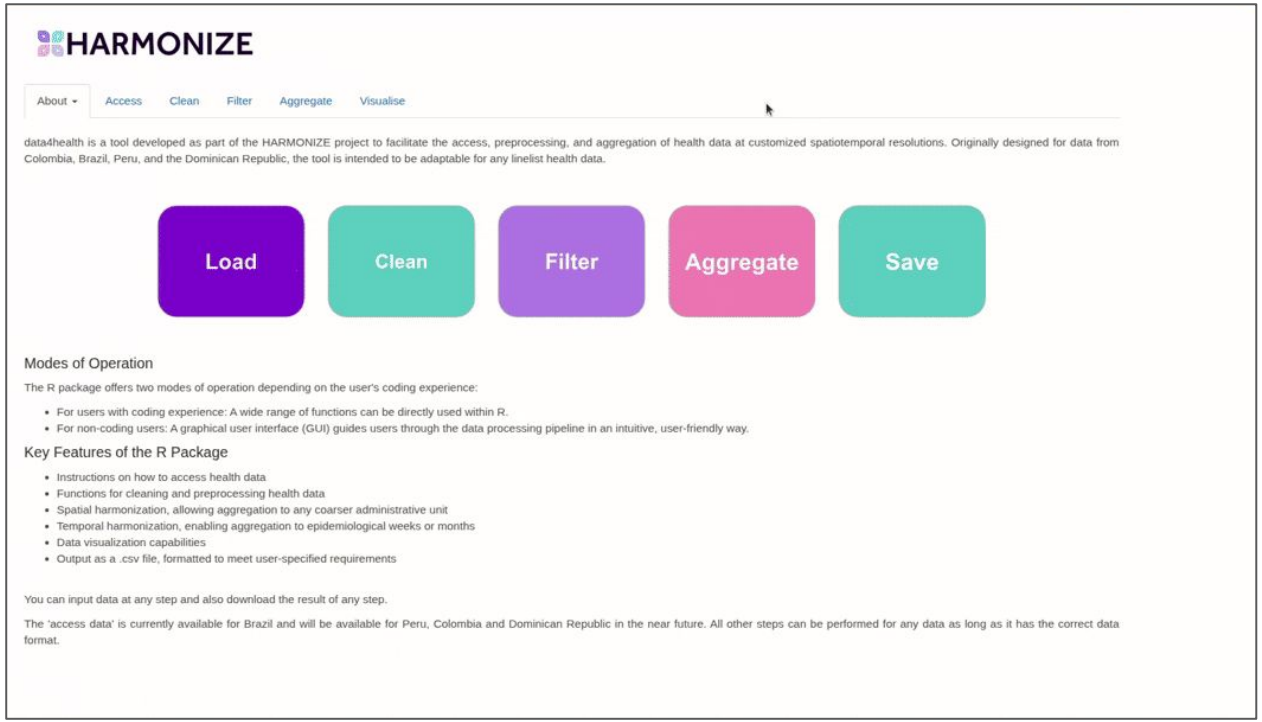
- Summarize cases over:
  - space
  - time
  - other relevant variables

Code-experienced user

```
d4h_load()

d4h_clean()

d4h_filter()

d4h_aggregate()

d4h_save()
```

Non-code-experienced user

```
d4h_ui()
```



clim4health

An R package designed to obtain, transform and export climate data for their use in epidemiological analyses and other types of applications.

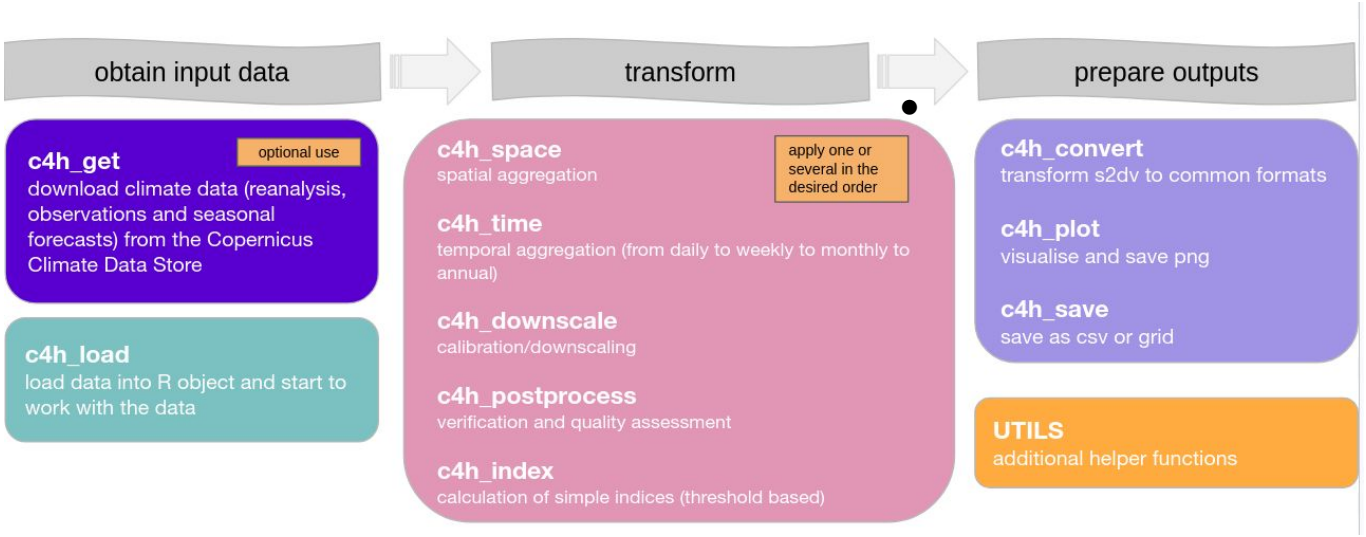
DR EMILY BALL
CLIMATE RESEARCHER

DR CARLES MILA
DATA SCIENTIST

REBECA NUNES
SOFTWARE ENGINEER

PROF RACHEL LOWE
LEADING RESEARCHER

Expected release date:
July 2026



Seasonal forecasts available from Copernicus:

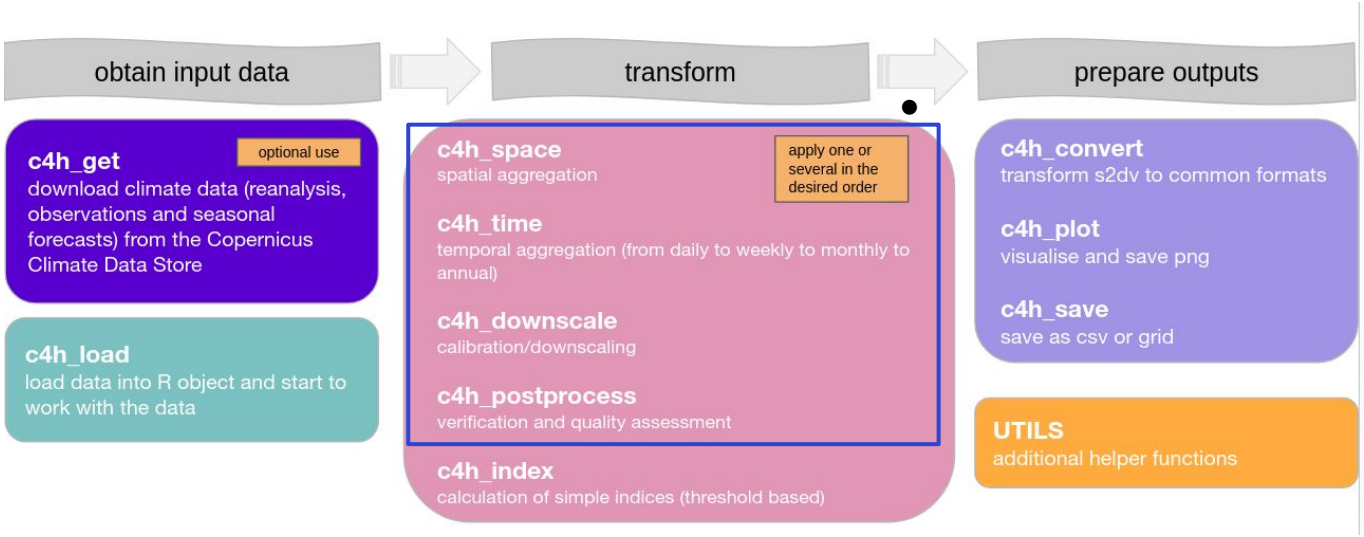
- Spatial resolution: 1° x 1° (~100km)
- System: ECMWF-s5

Reanalysis available from Copernicus:

- ERA5 1979-present 0.25° x 0.25° (~28km)
- ERA5-Land 1950-present 0.1° x 0.1° (~10km)

Ground based stations:

- Point data
- Daily or sub-daily temporal resolution



Seasonal forecasts available from Copernicus:

- Spatial resolution: 1° x 1° (~100km)
- System: ECMWF-s5

Reanalysis available from Copernicus:

- ERA5 1979-present 0.25° x 0.25° (~28km)
- ERA5-Land 1950-present 0.1° x 0.1° (~10km)

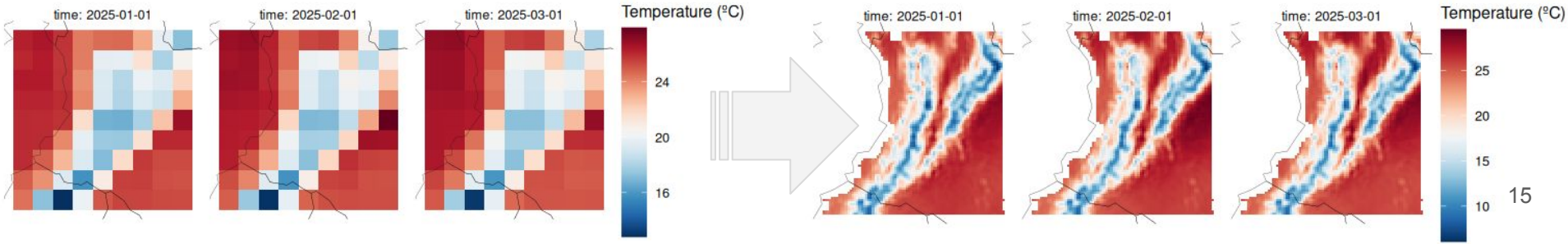
Ground based stations:

- Point data
- Daily or sub-daily temporal resolution

clim4health: downscaling and calibration

We use hindcasts and reanalysis (~10km) to downscale and calibrate seasonal forecasts (~100km). Available methods include:

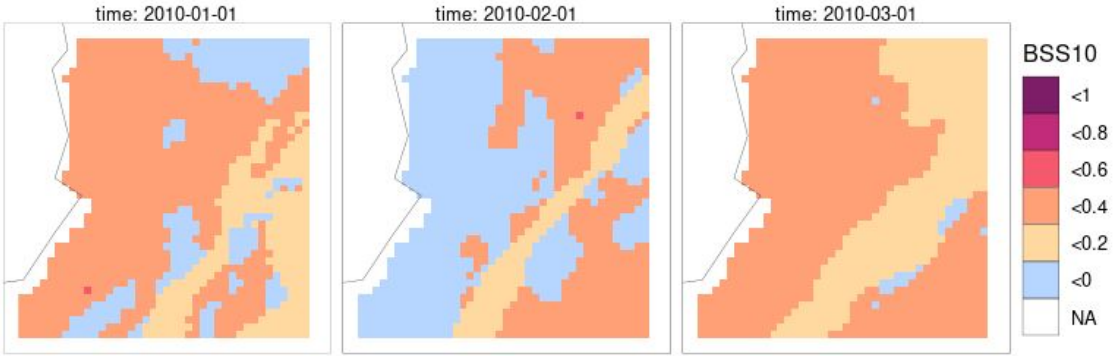
- Interpolation (bilinear, bicubic, nearest neighbour, conservative)
- Interpolation + bias correction (Gaussian, EVMOS, quantile mapping)
- Interpolation + linear regression (cell-wise, focal)
- Analogs (K-NNs)



clim4health: skill assessment

We assess the quality of the seasonal forecasts using hindcasts and reanalysis data and compute skill metrics to compare it to the climatologies:

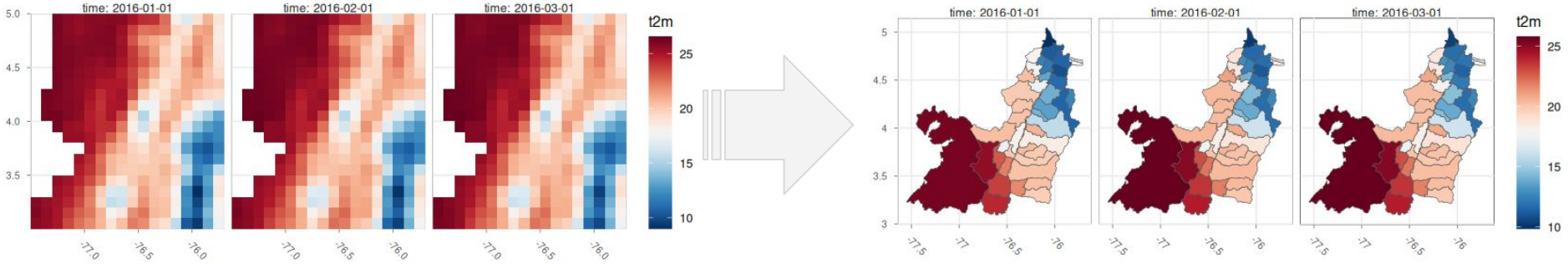
- Brier Skill Score (BSS)
- Ranked probability Skill Score (RSS)
- Continuous Ranked Probability Skill Score (CRPSS)
- + Many others (MSS, MAE, ROC, RMS, AbsBias)



clim4health: spatio-temporal aggregation

We aligned the spatiotemporal resolution of the climate data to the spatio-temporal resolution of the epidemiological data:

- Ensemble aggregations
- Grid-to-polygon aggregations
- Hourly/daily to weekly/monthly aggregations



GHRexplore

An R package for exploratory analysis of temporal and spatio-temporal health data including case counts, incidence rates, and covariates.

DR CARLES MILA
DATA SCIENTIST

DR GIOVENALE MOIRANO
VISITING RESEARCHER

DR ANIA KAWIECKI
RESEARCHER

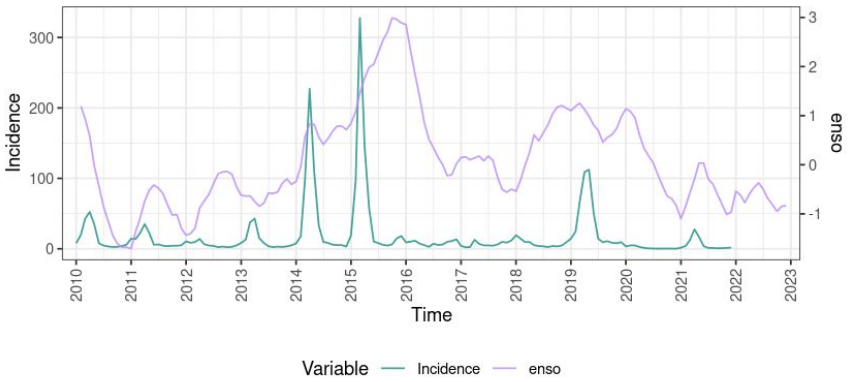
PROF RACHEL LOWE
LEADING RESEARCHER

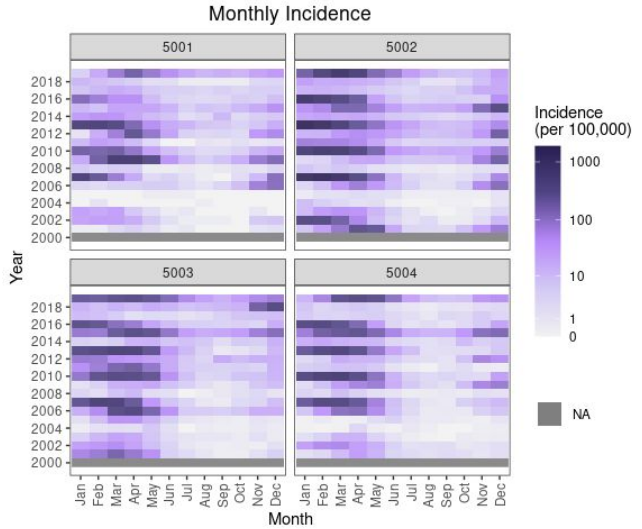
Version 0.2.1 on CRAN

GHRexplore: Available visualizations

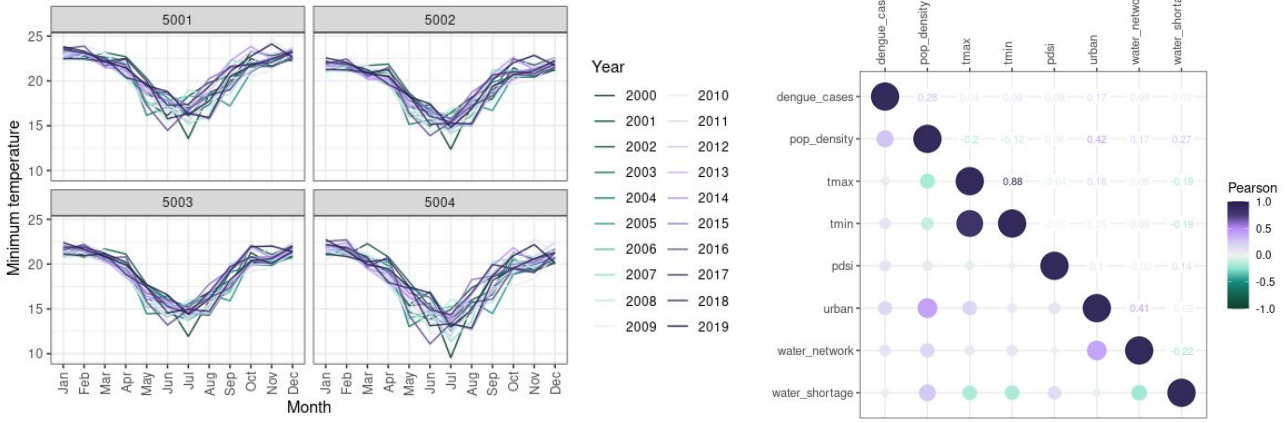
Easy visualization of spatio-temporal disease count, incidence and covariate tabular data:

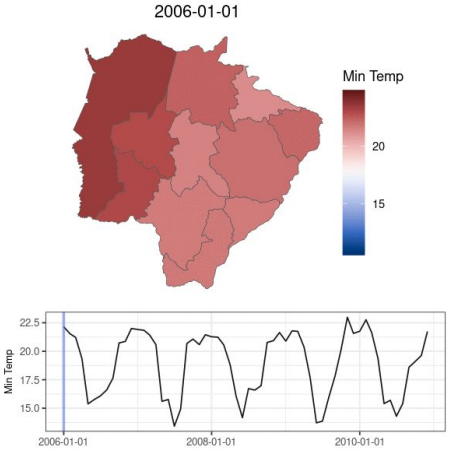
- Explore spatio-temporal structure (seasonality, interannual, spatial patterns)
- Explore relationships with covariates (lags, associations)

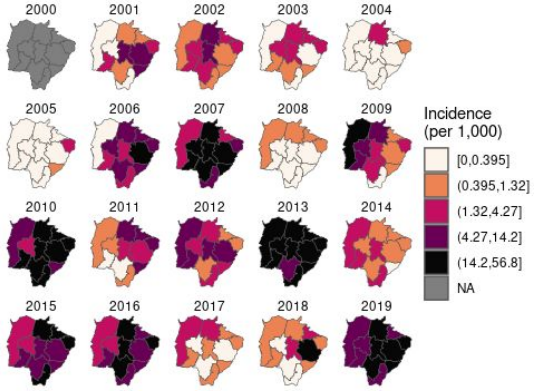


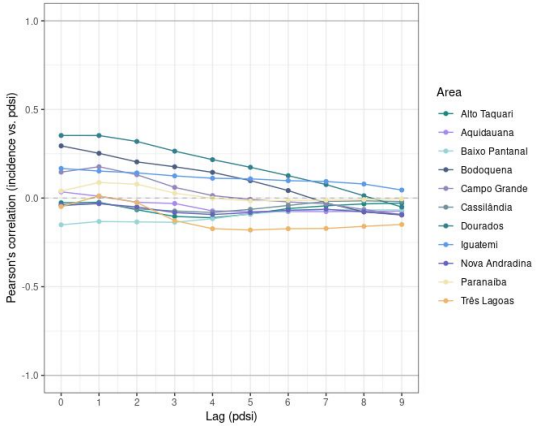


GHRexplore: Available visualizations

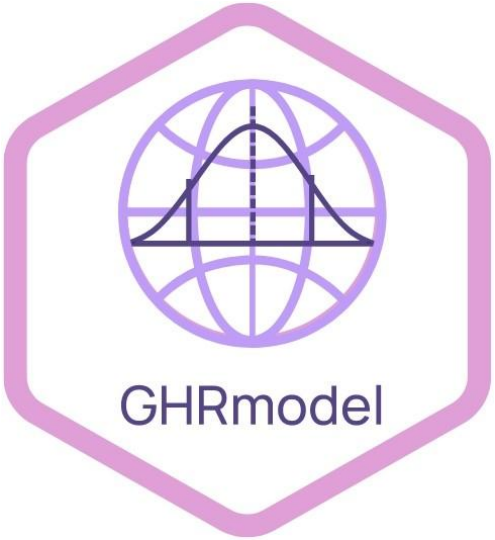








GHRmodel



An R package for modeling health outcomes using Bayesian hierarchical spatio-temporal models with complex covariate effects in the R-INLA framework.

DR GIOVENALE MOIRANO
VISITING RESEARCHER

DR ANIA KAWIECKI
RESEARCHER

DR CARLES MILA
DATA SCIENTIST

PROF RACHEL LOWE
LEADING RESEARCHER

Version 0.1.1 on CRAN

GHRmodel: Modelling framework

$$Y_{i,t} \mid \mu_{i,t}, \kappa \sim NegBin\ (\mu_{i,t}, \kappa)$$

$$log(\mu_{i,t}) = \alpha + log(P_{i,\ a[t]}) + \sum_k \beta_k X_{i,t,k} + \gamma_{i,a[t]} + \delta_{i,m[t]}$$

Explanatory variables:

- Linear
- Non-linear
- Random slopes
- Cross-basis matrices (DLNMs)
- Interactions
- …

Random effects:

- Seasonal
- Interannual
- Spatial
- …

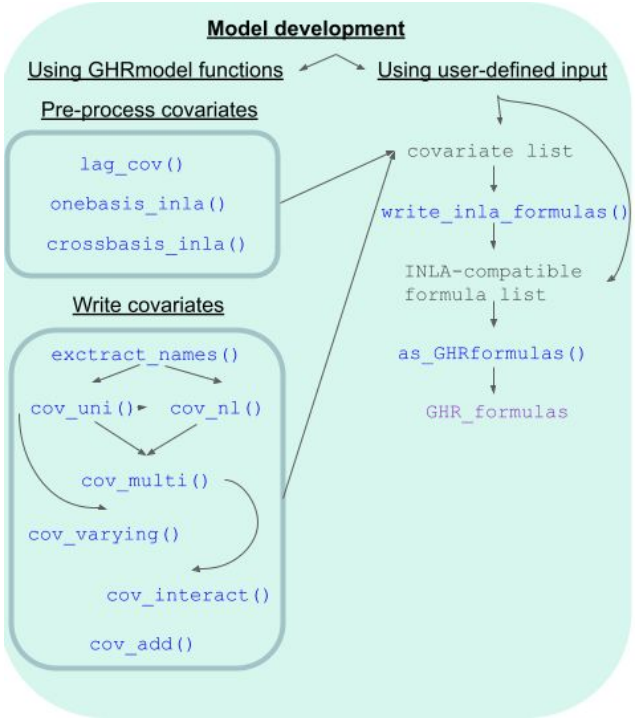
INLA

- Statistical models?
  Transparency
- Bayesian?
  Uncertainty quantification
- INLA?
  Fast Bayesian inference

Transform covariates → Define formulas → Fit models → Model evaluation

GHRmodel: Define a list of formulas

Transform covariates → Define formulas



- Helpers to prepare long lists of formulas to be fitted in INLA including linear and non-linear effects, interactions, and random slopes.
- Helpers to specify complex random effects.
- Prepare one- and cross-basis functions.

GHRmodel: Fit a list of models

Transform covariates → Define formulas → **Fit models**

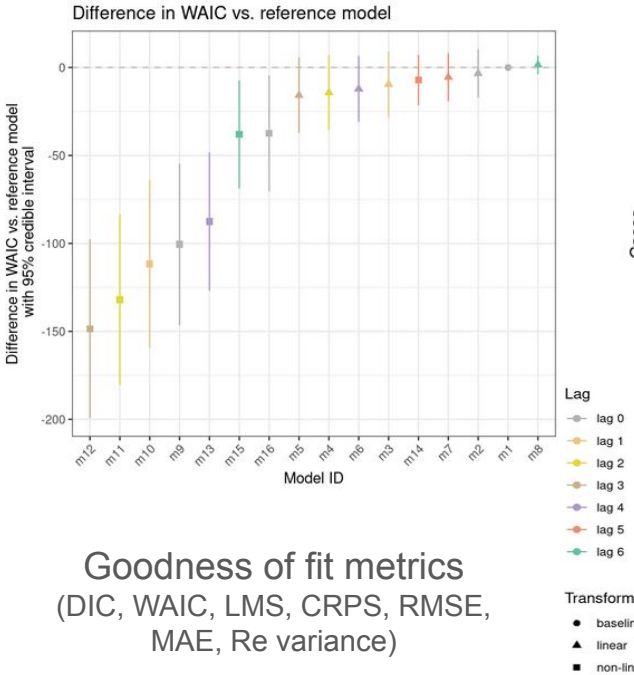
We fit INLA models in batches which we can then easily compare to each other:

```
model_list <- fit_models(
  formulas = formulas_cov_list_ghr, # GHRformulas object
  data = data,                      # Data to fit the models
  family = "nbinomial",             # Negative binomial likelihood
  name = "mod",                     # Label prefix for each model
  offset = "population",            # Offset variable to account for population size
  control_compute = list(
    config = FALSE,                 # Extract from INLA summary.fitted.values
    vcov = FALSE                    # Do not return variance-covariance matrix
  ),
  nthreads = 8                      # Use 8 threads for parallel computation
)
```

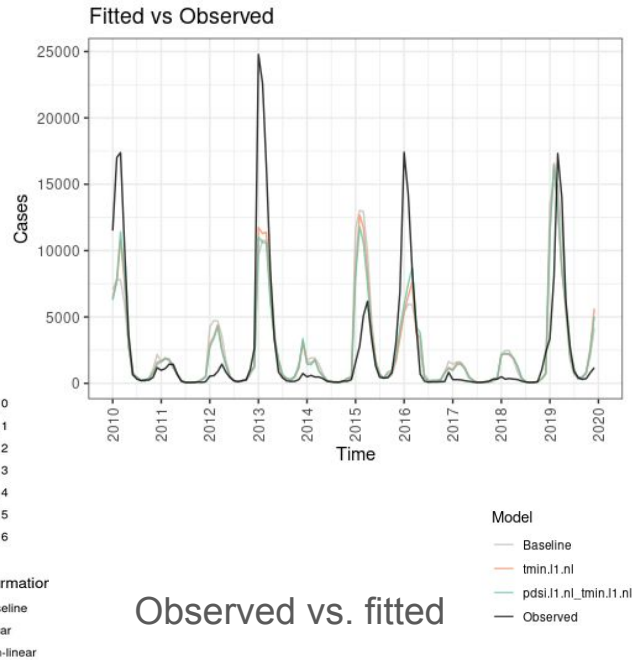
And extract a series of (option) outputs to be used when examining the models

GHRmodel: Evaluation of model fit

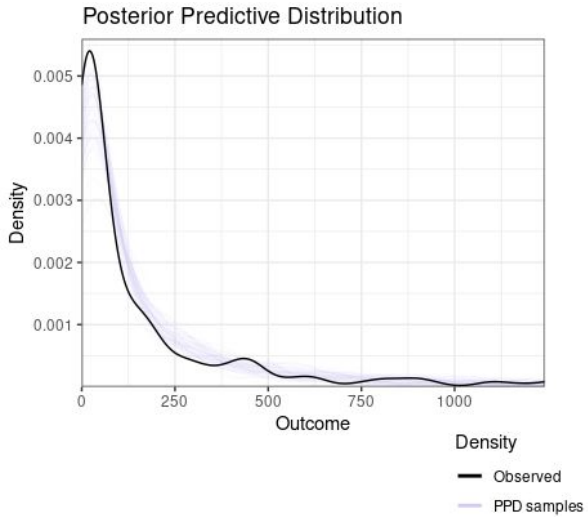
Transform covariates → Define formulas → Fit models → **Model evaluation**



Goodness of fit metrics
(DIC, WAIC, LMS, CRPS, RMSE, MAE, Re variance)



Observed vs. fitted

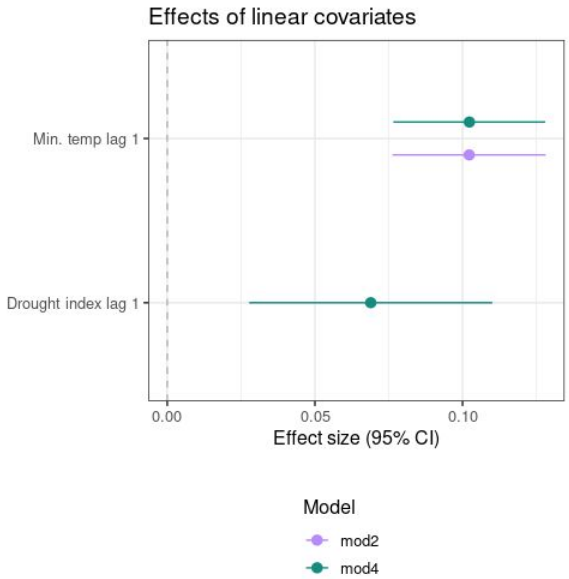


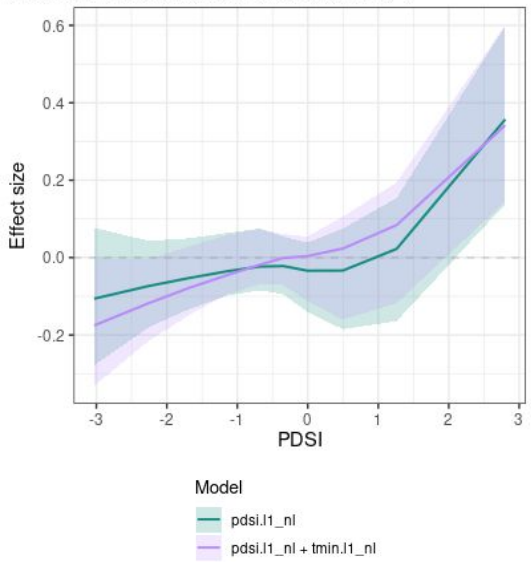
Posterior predictive checks

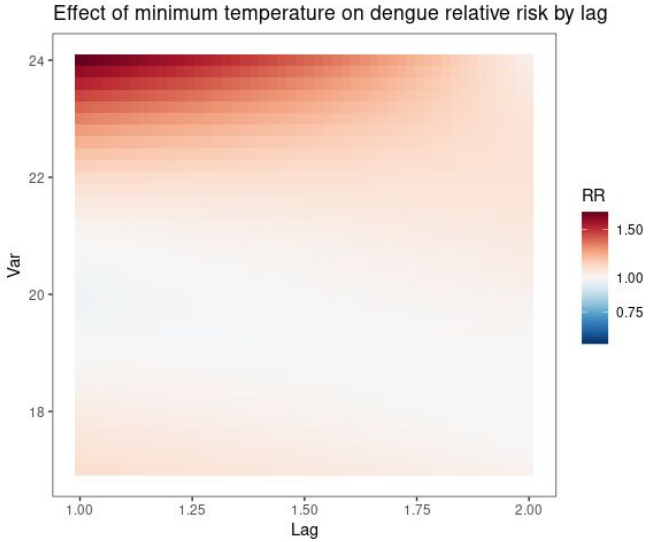
GHRmodel: Evaluation of the posterior estimates

Transform covariates → Define formulas → Fit models → **Model evaluation**

Evaluate effects: linear, non-linear, and DLNMs posterior estimates



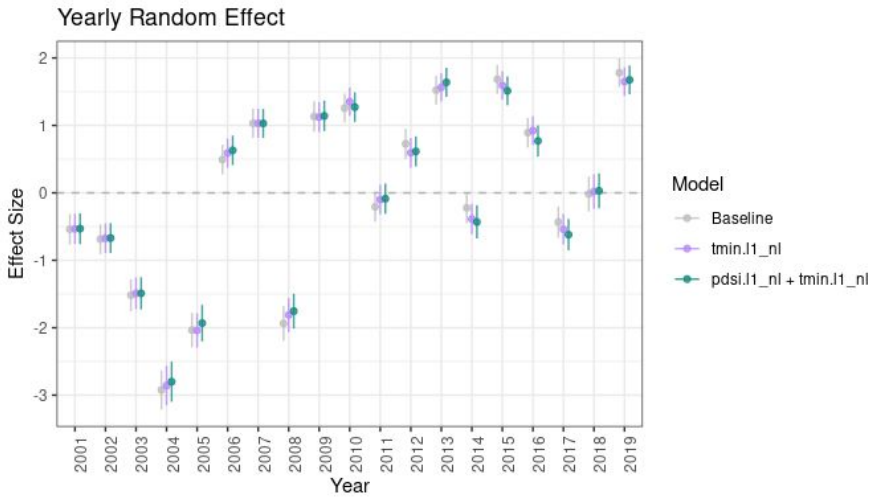


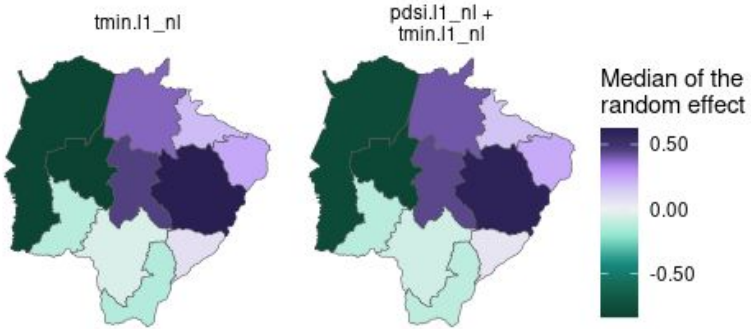


GHRmodel: Evaluation of covariate effects

Transform covariates → Define formulas → Fit models → **Model evaluation**

Evaluate effects: random effects posterior estimates





GHRpredict

An R package designed for computing out-of-sample predictions, enabling cross-validation analysis and the evaluation of predictive performance for INLA models.


DR MARTIN LOTTO
RESEARCHER


DR ANIA KAWIECKI
RESEARCHER


DR CARLES MILA
DATA SCIENTIST


PROF RACHEL LOWE
LEADING RESEARCHER

Expected release date:
December 2026

GHRpredict: Cross-validation

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Cross-validation → Predictive performance

Expanding window cross-validation

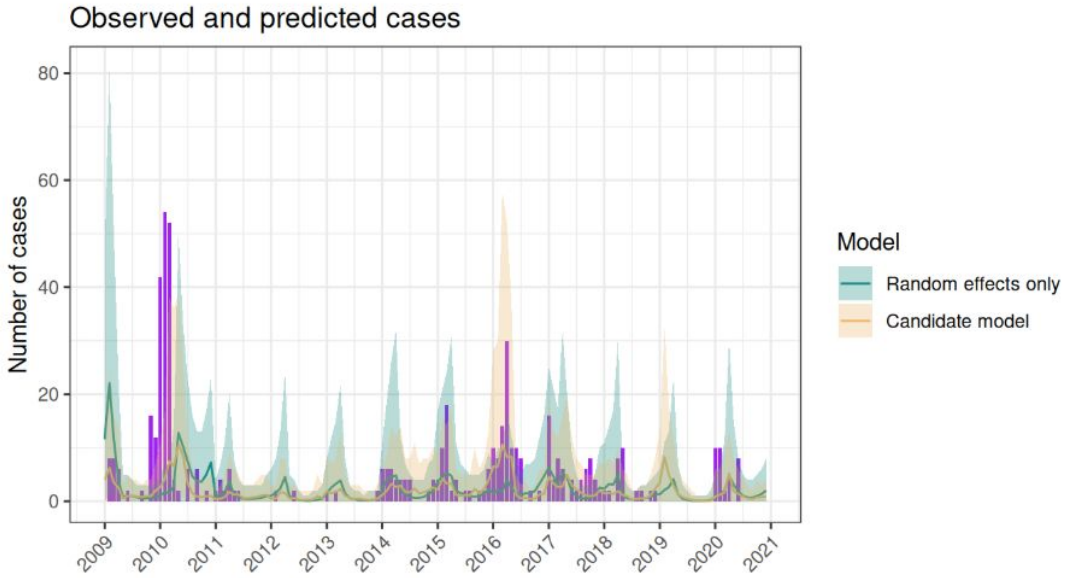
Source: Fletcher et al. (2025)

GHRpredict: Forecast accuracy

Cross-validation → Predictive performance

Computes forecast accuracy metrics for case counts prediction:

- Continuous Ranked Probability Score (CRPS)
- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE).

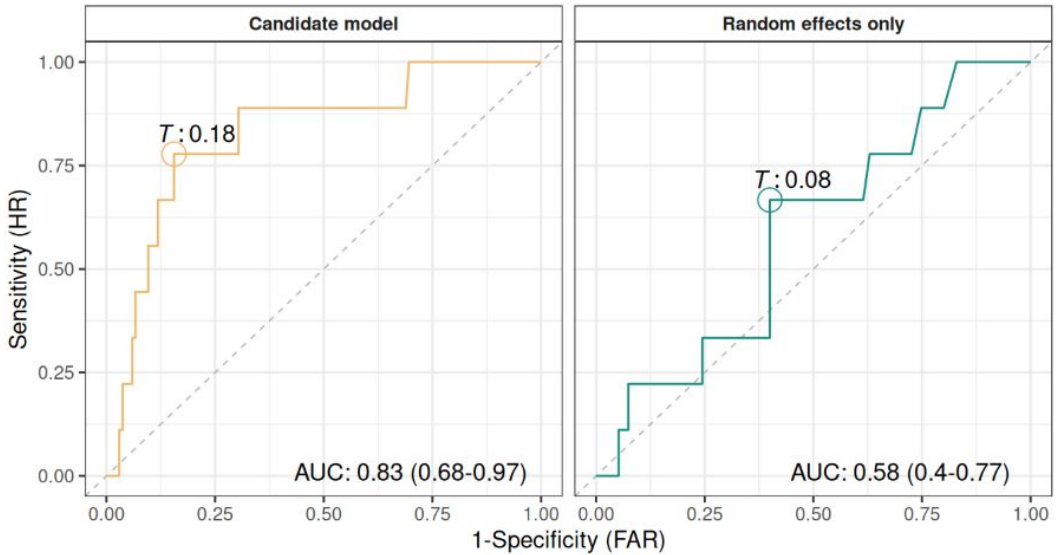


GHRpredict: Outbreak detection ability

Cross-validation → Predictive performance

Computes forecast accuracy metrics for outbreak detection:

- Multiple thresholds to define outbreaks.
- Metrics: ROC, Brier, Confusion matrix
- Optimal trigger outbreak probability based on ROC curves.



GHRtools: outlook

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Obtaining and preprocessing health and climate data

Exploring, modelling and forecasting disease risk

CRAN: July 2026

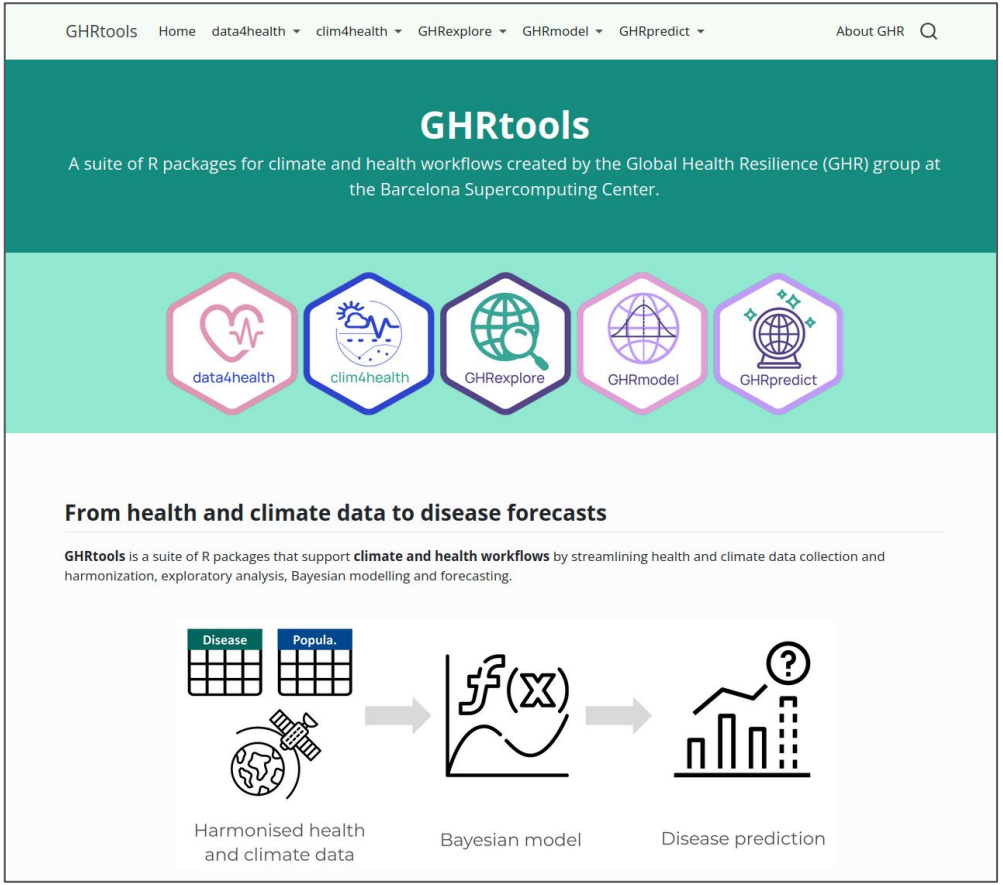
CRAN: July 2026

CRAN: Version 0.2.1

CRAN: Version 0.1.1

CRAN: December 2026

Stay tuned for updates!



https://bsc-es.github.io/GHRtools

- Documentation
- Vignettes and tutorials
- Software manuals

Earth Sciences Department

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Thank you!

Earth Sciences Department

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Global Health Resilience