

# GEOMED GHR tools

A Digital Workflow for Bayesian Disease Risk Forecasting in R

**Ania Kawiecki, Carles Milà, Martín Lotto Batista, Giovenale Moirano, Emily Ball, Daniela Lührsen, Tilly Alcayna, Sabu Best, Leslie Rollock, Leo Tremblay, Claudia Codeço, Mauricio Santos-Vega, Rachel Lowe**

BSC Earth Sciences - Global Health Resilience

[ania.kawiecki@bsc.es](mailto:ania.kawiecki@bsc.es)



**IDExtremes**



**TACTIC**



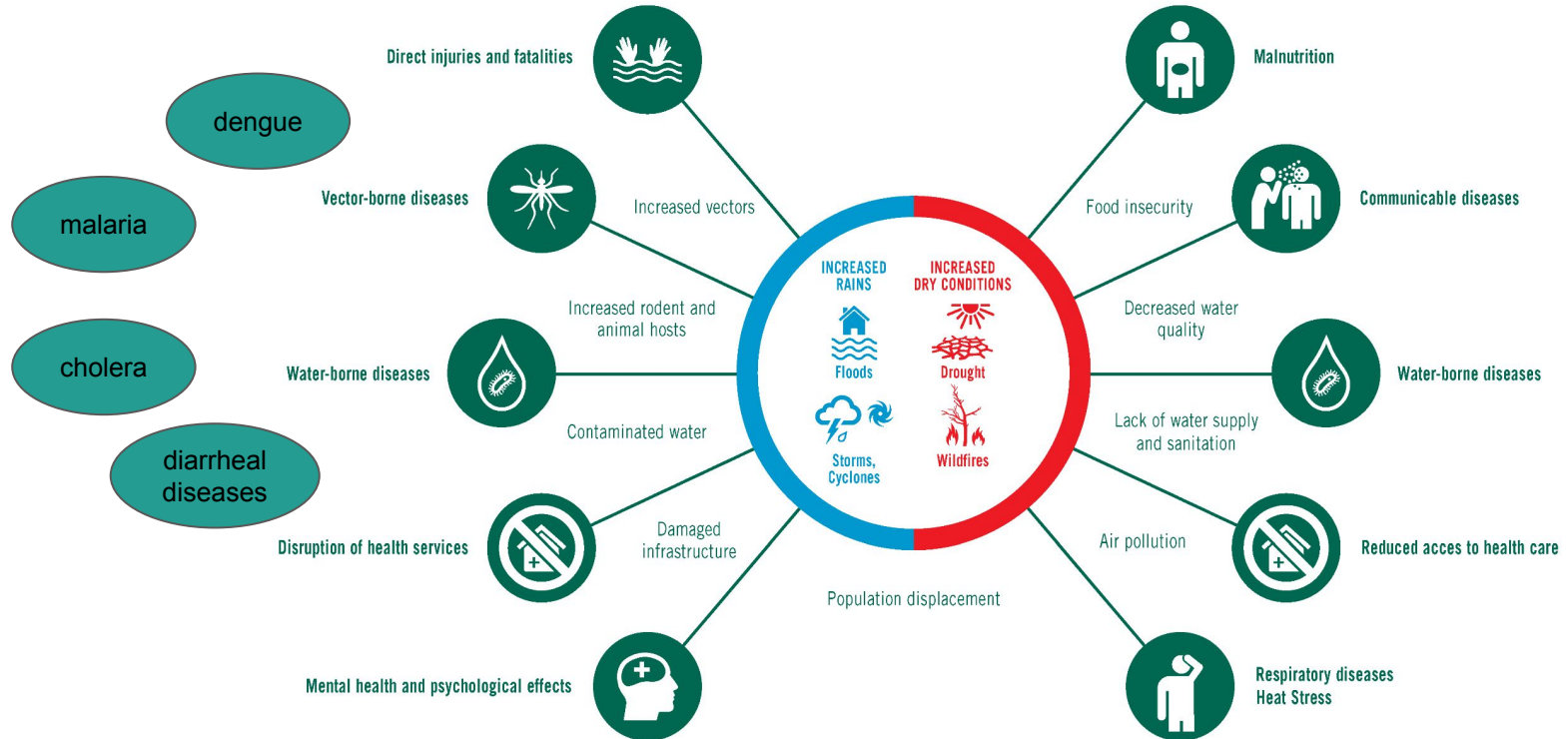
**HARMONIZE**



**Barcelona  
Supercomputing  
Center**

Centro Nacional de Supercomputación

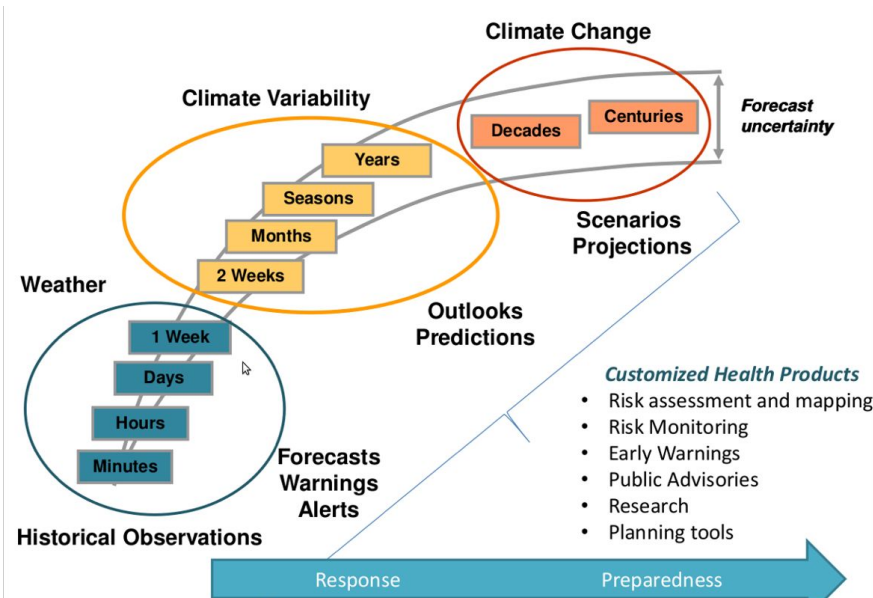
# Heightened health risks related to extreme climatic events



Source: McGregor & Ebi (2018) adapted from WHO (2016)

# Early warning systems for climate-sensitive diseases

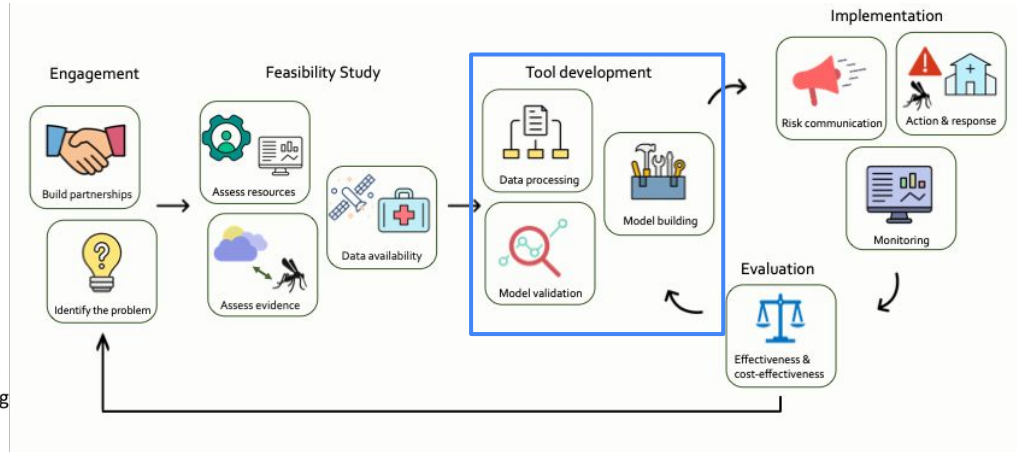
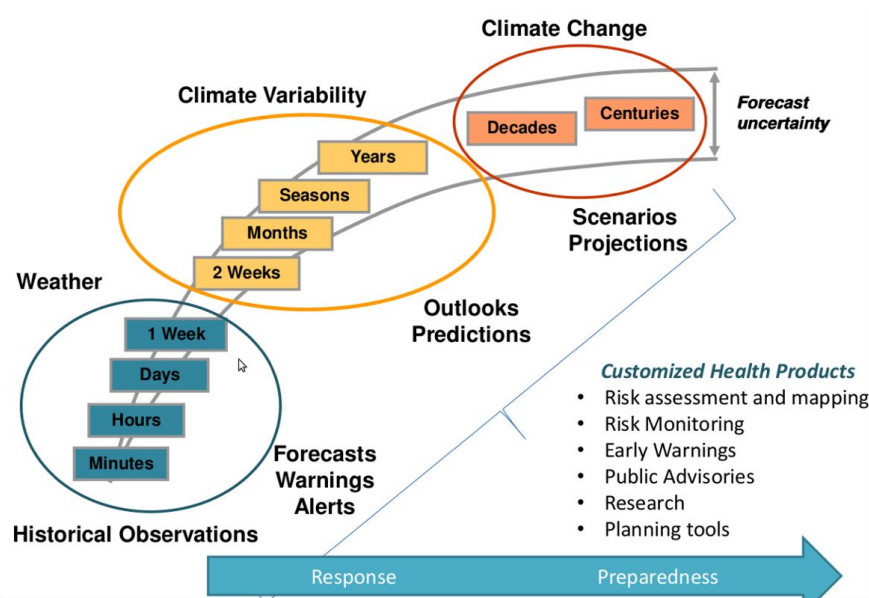
Climate-informed early warning systems integrate climate data to predict the risk of an infectious disease outbreak.



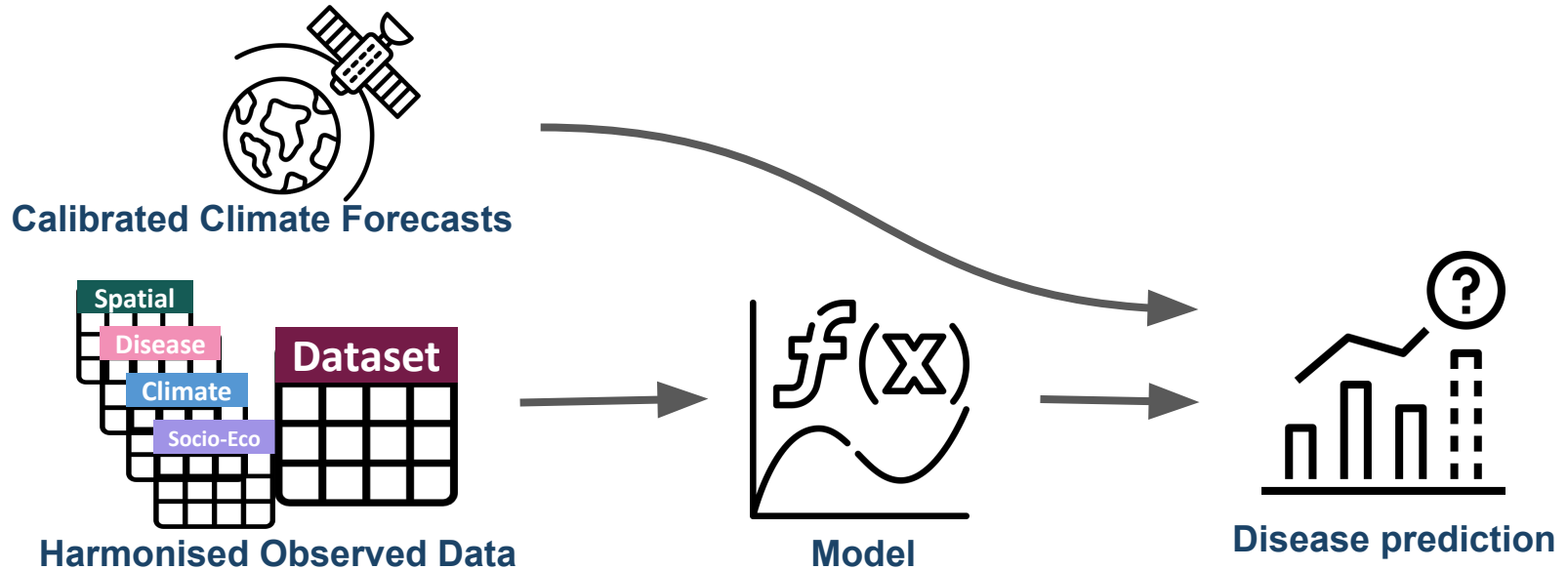
Courtesy: J. Shumake-Guillemot, WMO

# Early warning systems for climate-sensitive diseases

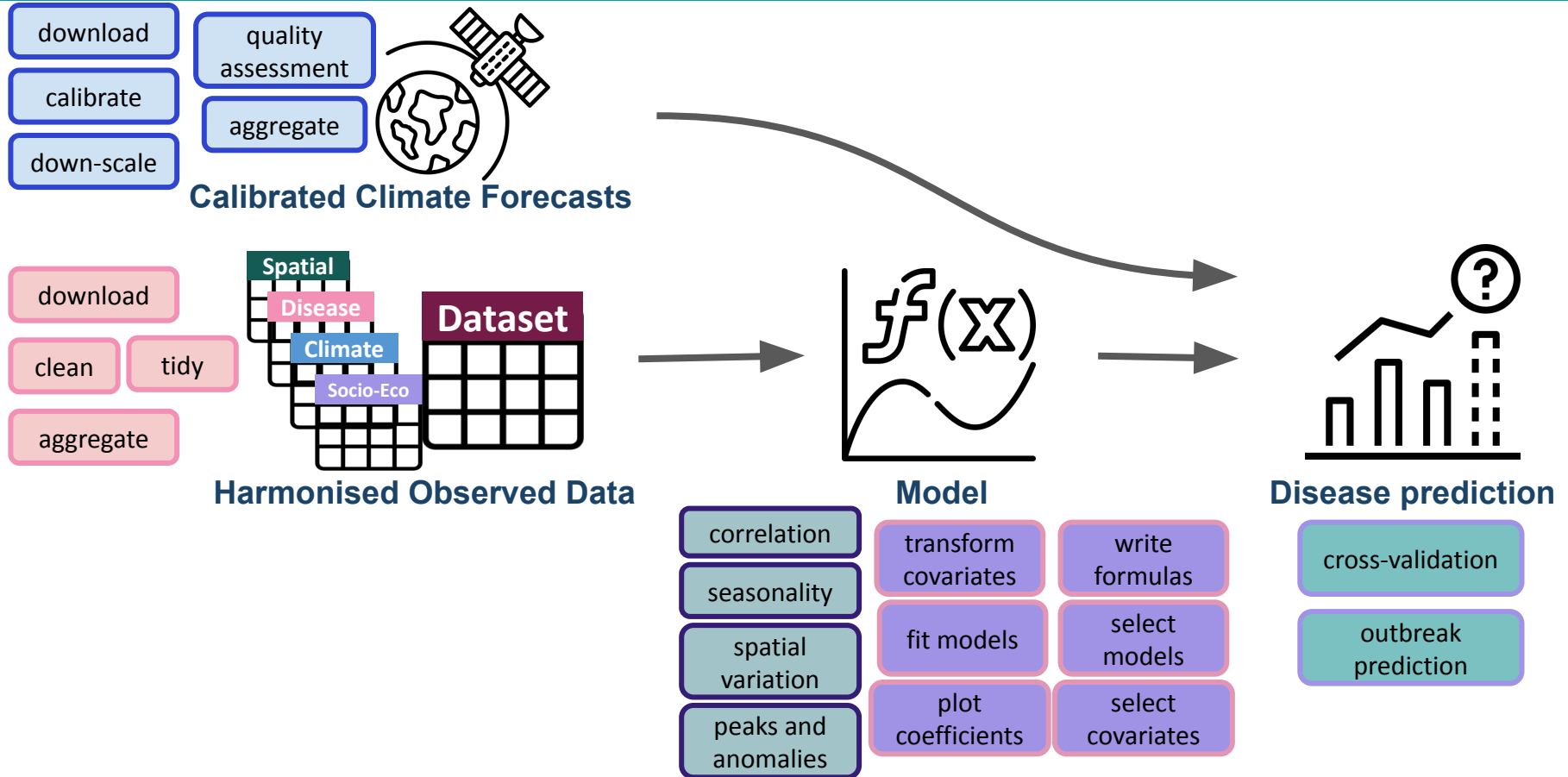
Climate-informed early warning systems integrate climate data to predict the risk of an infectious disease outbreak.



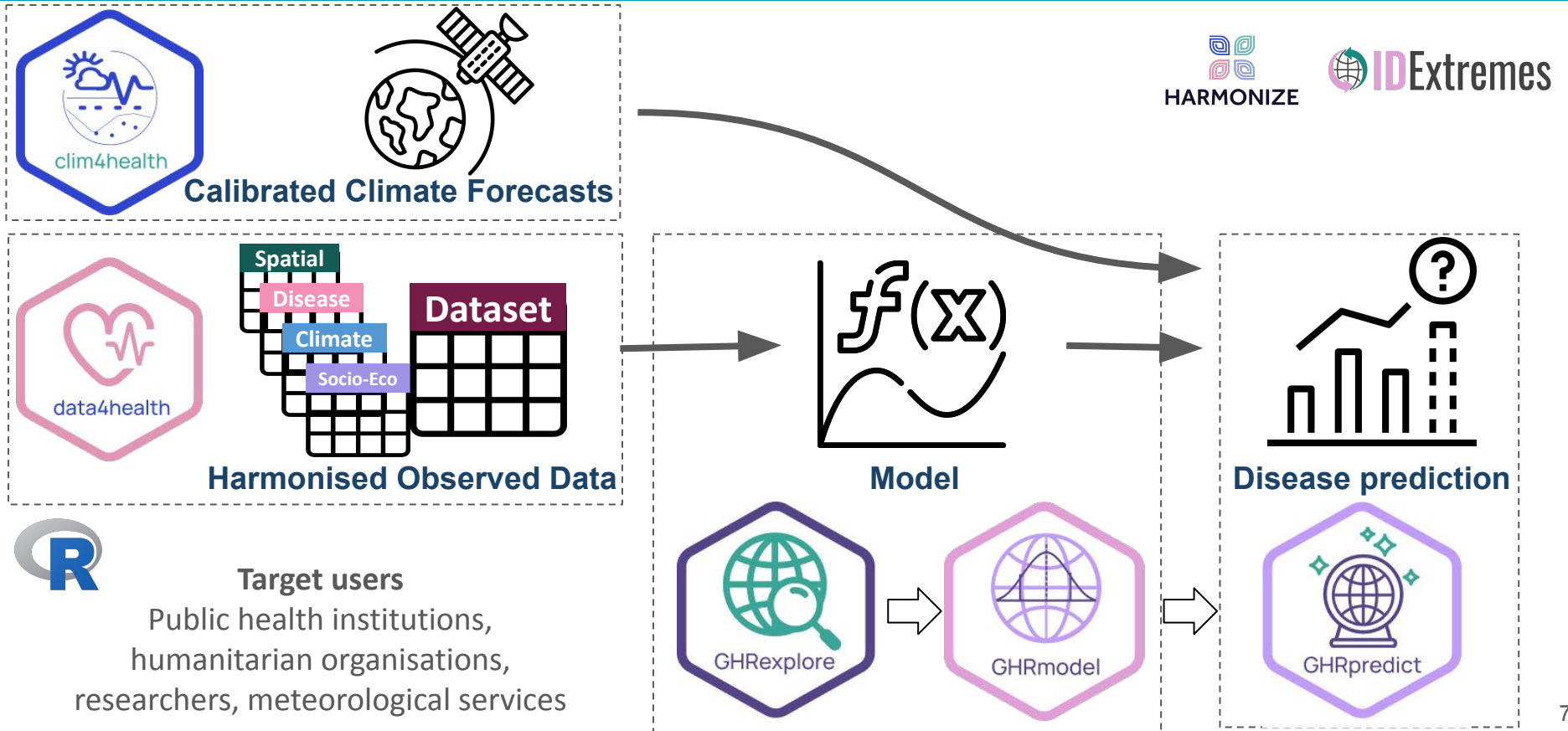
Finch et al 2023



# Using climate forecasts and models for disease risk prediction



# How GHRtools fit into the disease risk modelling and forecasting framework





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

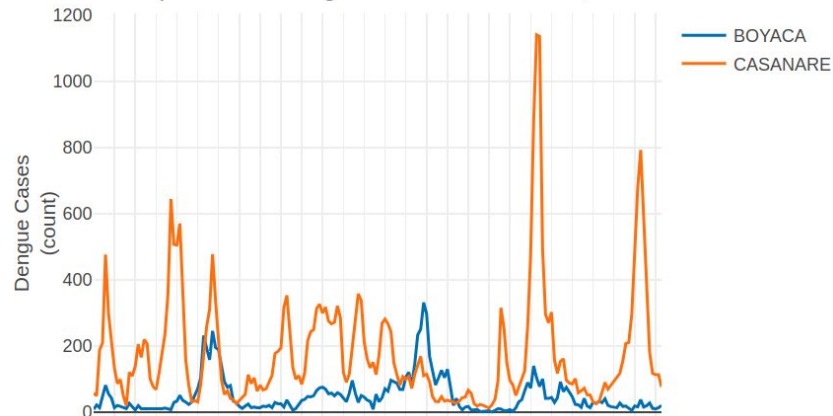
## Unprocessed health data

	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	COPAISIN	MUNI_INF	CLASS	CRIT	
2	2015-03-05	1979-07-27	2015-03-02	12	120020	35F		6NA		12		1	120033	10	2
3	2015-02-02	2000-09-22	2015-01-22	12	120020	14F		6NA		12		1	120033	10	2
4	2015-05-17	2005-05-10	2015-05-17	12	120020	10F		6NA		12		1	120020	10	1
5	2015-06-22	1924-04-02	2015-06-12	12	120020	91F		6	4	12		1	120020	10	1
6	2015-03-23	2000-08-14	2015-03-17	12	120020	14F		6NA		12		1	120020	10	2
7	2015-03-19	1997-12-13	2015-03-11	12	120020	17F		6	4	12		1	120020	10	2
8	2015-03-30	2012-12-23	2015-03-20	12	120020	2F		6	4	12		1	120020	10	2
9	2015-03-23	1991-10-20	2015-03-18	12	120020	23F		6NA		12		1	120020	10	2
10	2015-03-19	1981-04-15	2015-03-12	12	120020	33F		5	1	12		1	120020	10	2
11	2015-03-05	1958-01-31	2015-03-01	12	120020	57F		6NA		12		1	120020	10	2
12	2015-03-05	1995-11-28	2015-02-25	12	120020	19F		6	4	12		1	120020	10	2
13	2015-03-05	1976-06-01	2015-02-28	12	120020	38F		6	4	12		1	120020	10	2
14	2015-03-09	1958-09-01	2015-03-05	12	120020	56F		6NA		12		1	120020	10	2
15	2015-03-05	1990-09-01	2015-02-26	12	120020	24F		6	4	12		1	120020	10	2
16	2015-02-13	2003-05-19	2015-02-09	12	120020	11F		5	4	12		1	120020	10	2
17	2015-02-20	1996-06-19	2015-02-15	12	120020	18F		5	4	12		1	120020	10	2
18	2015-02-26	1950-02-02	2015-02-20	12	120020	65F		6NA		12		1	120020	10	2
19	2015-02-20	2008-12-03	2015-02-13	12	120020	6F		6	4	12		1	120020	10	2
20	2015-02-24	1950-02-02	2015-02-18	12	120020	65F		5	4	12		1	120020	10	2
21	2015-02-26	2008-12-03	2015-02-12	12	120020	6F		6	4	12		1	120020	10	2
22	2015-02-26	1998-07-24	2015-02-21	12	120020	16F		6	4	12		1	120020	10	2
23	2015-02-13	1961-12-03	2015-02-10	12	120020	53F		5	4	12		1	120020	10	2
24	2015-02-25	1969-07-09	2015-02-16	12	120020	48F		5	4	12		1	120020	10	2
25	2015-02-11	2006-07-10	2015-02-05	12	120020	8F		6NA		12		1	120020	10	2
26	2015-02-24	1973-09-30	2015-02-19	12	120020	41F		6	4	12		1	120020	10	2
27	2015-02-10	1981-12-30	2015-01-09	12	120020	33F		5	4	12		1	120020	10	1
28	2015-02-09	1996-09-14	2015-02-03	12	120020	18F		6	4	12		1	120020	10	2
29	2015-02-11	1989-09-25	2015-02-06	12	120020	16F		6	4	12		1	120020	10	2
30	2015-02-02	1997-01-17	2015-01-23	12	120020	18F		5	4	12		1	120020	10	2
31	2015-02-04	1959-06-14	2015-01-30	12	120020	55F		6	4	12		1	120020	10	2
32	2015-02-04	1947-06-07	2015-02-02	12	120020	67F		6	4	12		1	120020	10	2
33	2015-02-04	1977-06-03	2015-01-30	12	120020	37F		6NA		12		1	120020	10	2



## Clean health data

Comparison of Dengue Cases in BOYACA, CASANARE



Load

Load from common file formats (csv, rds, xlsx, 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()
```

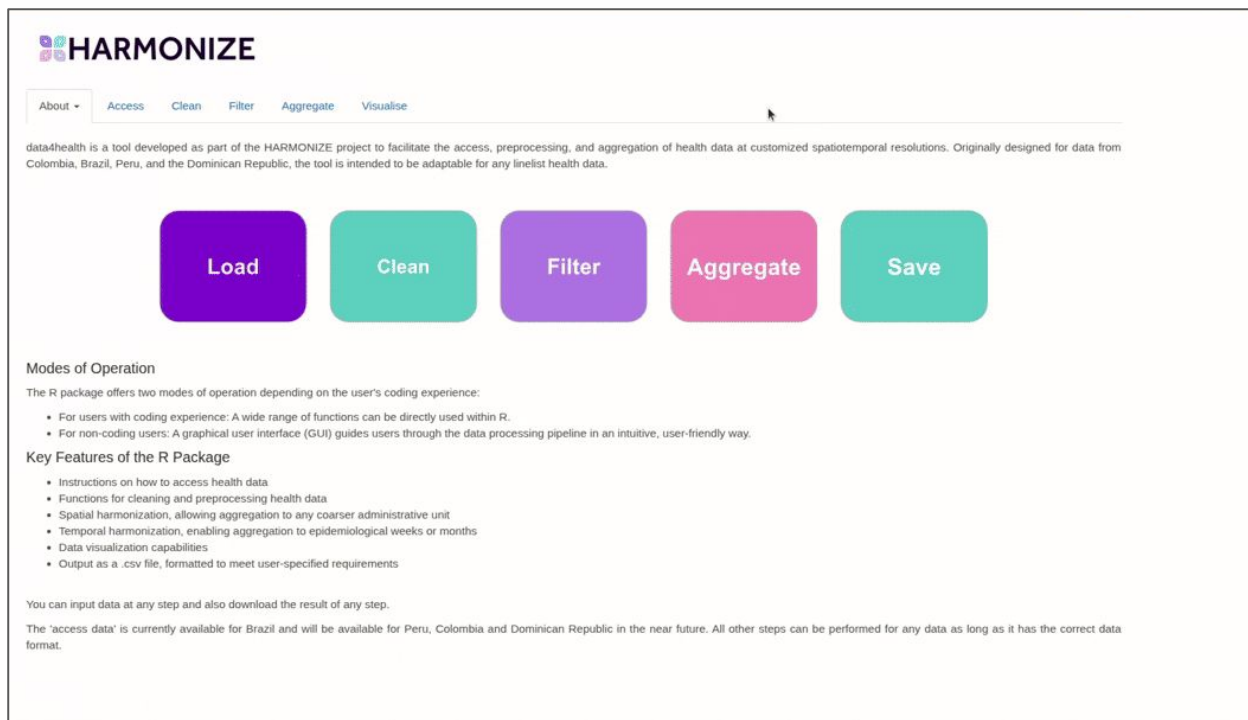
```
d4h_filter()
```

```
d4h_aggregate()
```

```
d4h_save()
```

## Non-code-experienced user

```
d4h_ui()
```



**HARMONIZE**

About | Access | Clean | Filter | Aggregate | Visualise

data4health is a tool developed as part of the HARMONIZE project to facilitate the access, preprocessing, and aggregation of health data at customized spatiotemporal resolutions. Originally designed for data from Colombia, Brazil, Peru, and the Dominican Republic, the tool is intended to be adaptable for any list health data.

**Load** **Clean** **Filter** **Aggregate** **Save**

### Modes of Operation

The R package offers two modes of operation depending on the user's coding experience:

- For users with coding experience: A wide range of functions can be directly used within R.
- For non-coding users: A graphical user interface (GUI) guides users through the data processing pipeline in an intuitive, user-friendly way.

### Key Features of the R Package

- Instructions on how to access health data
- Functions for cleaning and preprocessing health data
- Spatial harmonization, allowing aggregation to any coarser administrative unit
- Temporal harmonization, enabling aggregation to epidemiological weeks or months
- Data visualization capabilities
- Output as a .csv file, formatted to meet user-specified requirements

You can input data at any step and also download the result of any step.

The 'access data' is currently available for Brazil and will be available for Peru, Colombia and Dominican Republic in the near future. All other steps can be performed for any data as long as it has the correct data format.



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

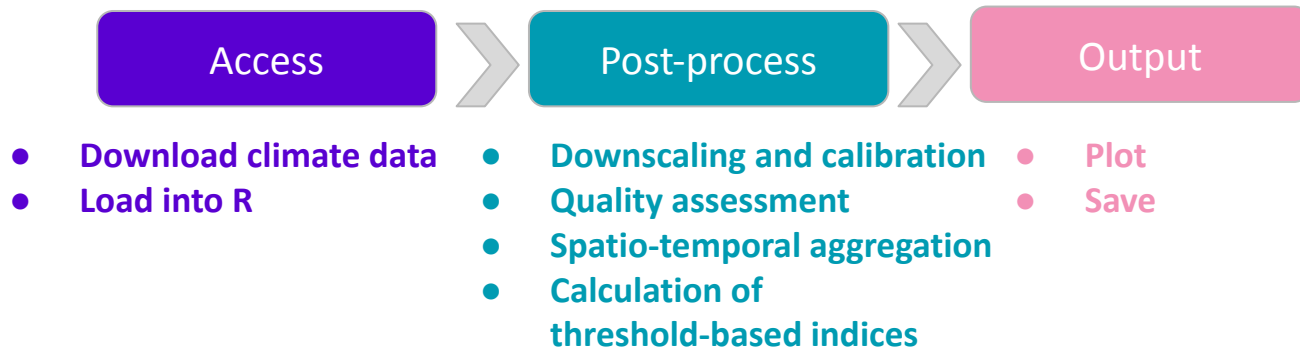


**REBECA NUNES**  
 SOFTWARE ENGINEER



**PROF RACHEL LOWE**  
 LEADING RESEARCHER

Expected release date:  
 July 2026



**Seasonal forecasts** available from Copernicus:

- Spatial resolution:  $1^\circ \times 1^\circ$  (~100km)
- System: ECMWF-s5

**Reanalysis** available from Copernicus:

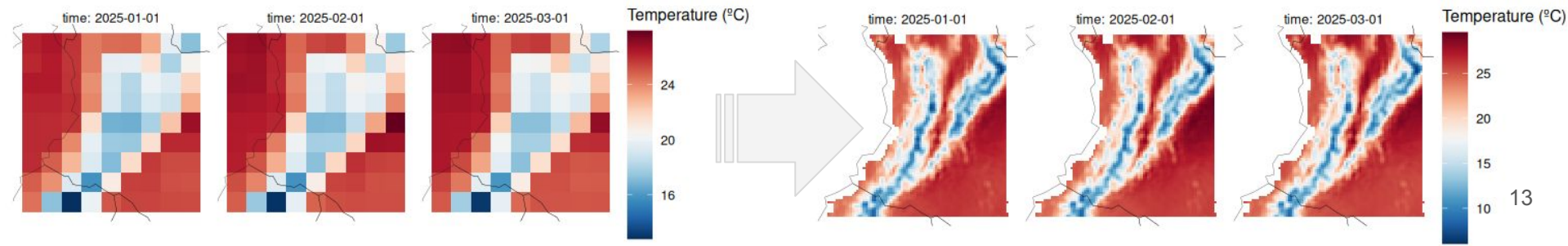
- ERA5 1979-present  $0.25^\circ \times 0.25^\circ$  (~28km)
- ERA5-Land 1950-present  $0.1^\circ \times 0.1^\circ$  (~10km)

**Ground based stations:**

- Point data
- Daily or sub-daily temporal resolution

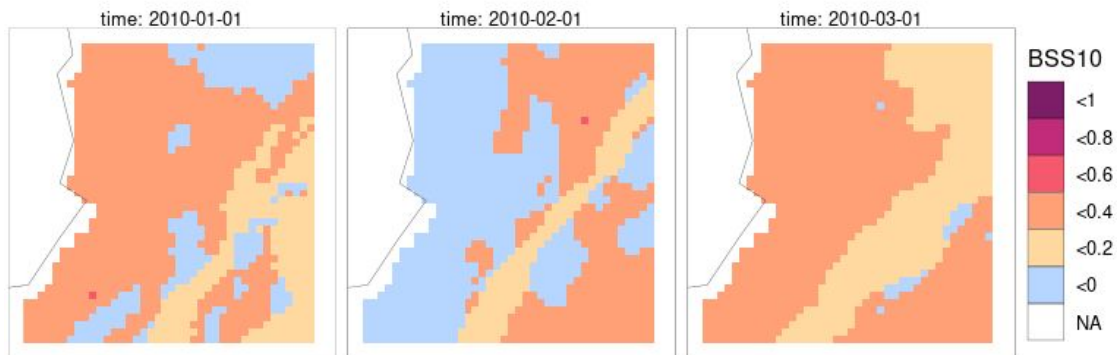
We use hindcasts and reanalysis ( $\sim 10\text{km}$ ) to downscale and calibrate seasonal forecasts ( $\sim 100\text{km}$ ). Available methods include:

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



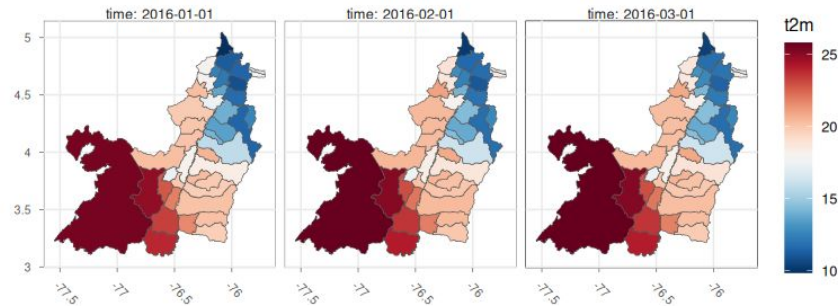
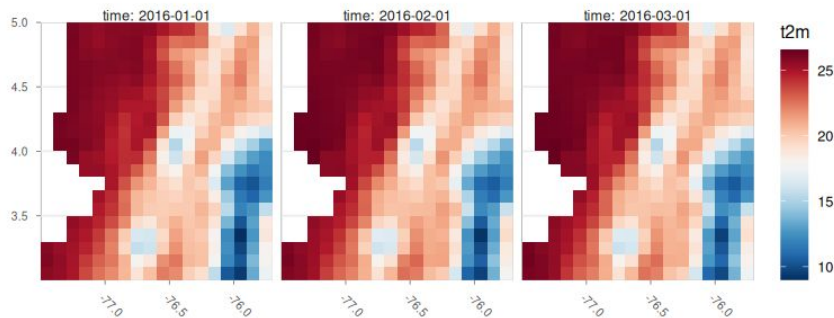
Assess the quality of the seasonal forecasts using hindcasts and reanalysis data and compute skill metrics to compare it to the climatologies:

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



Align 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





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

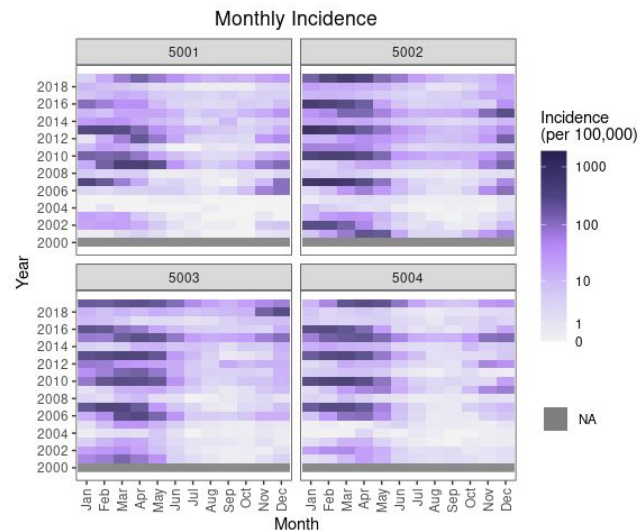
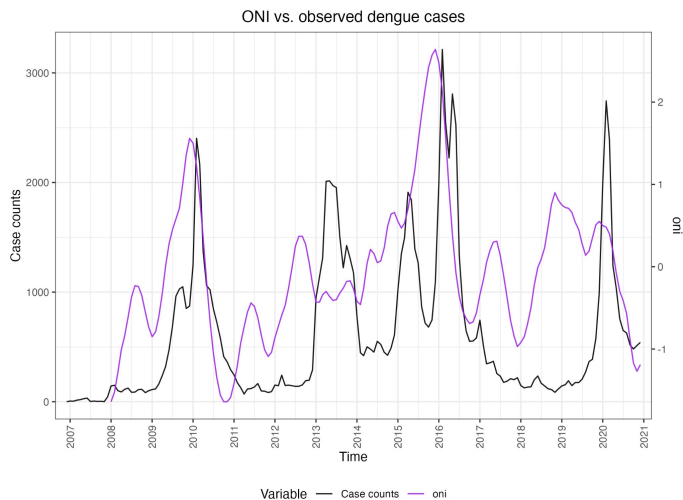


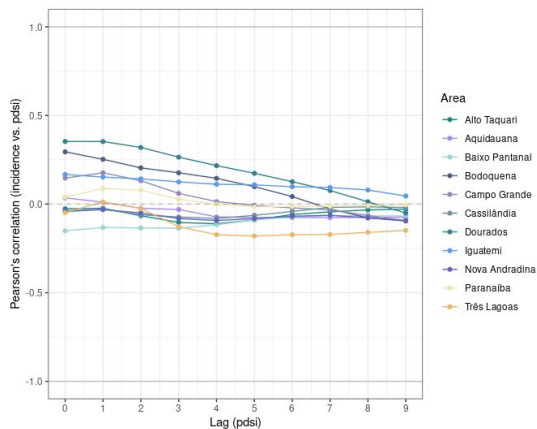
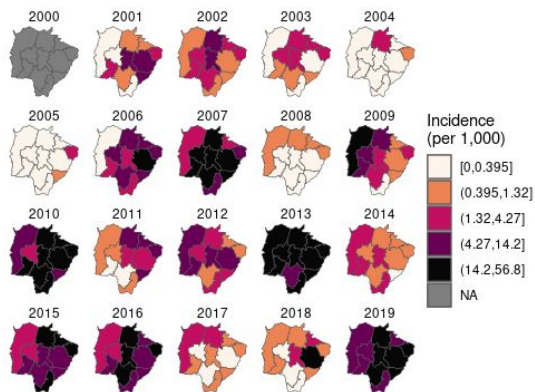
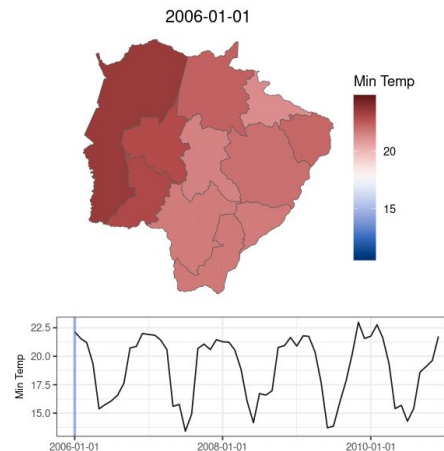
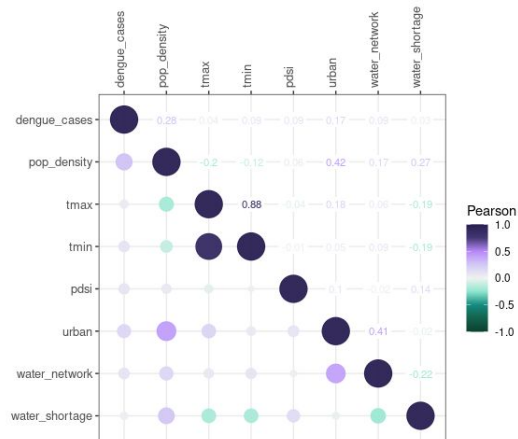
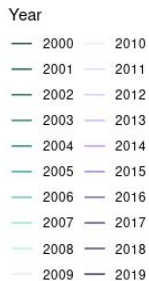
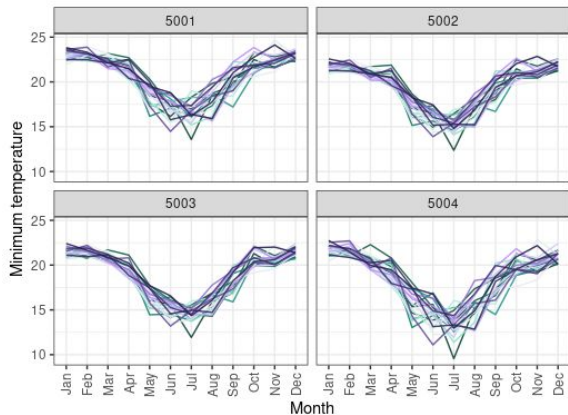
**PROF RACHEL LOWE**  
LEADING RESEARCHER

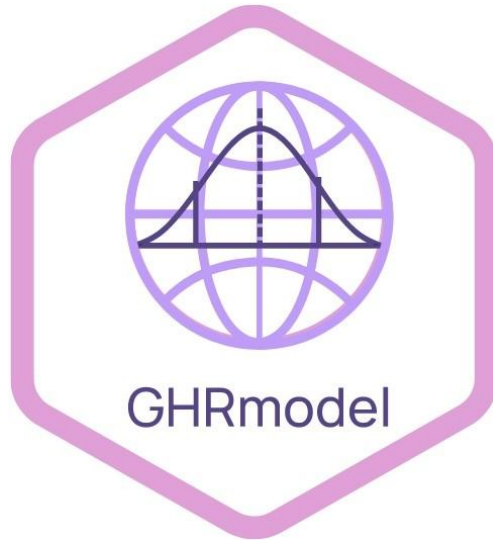
Version 0.2.1 on CRAN

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)







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

$$Y_{i,t} | \mu_{i,p} \kappa \sim \text{NegBin}(\mu_{i,p} \kappa)$$

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

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

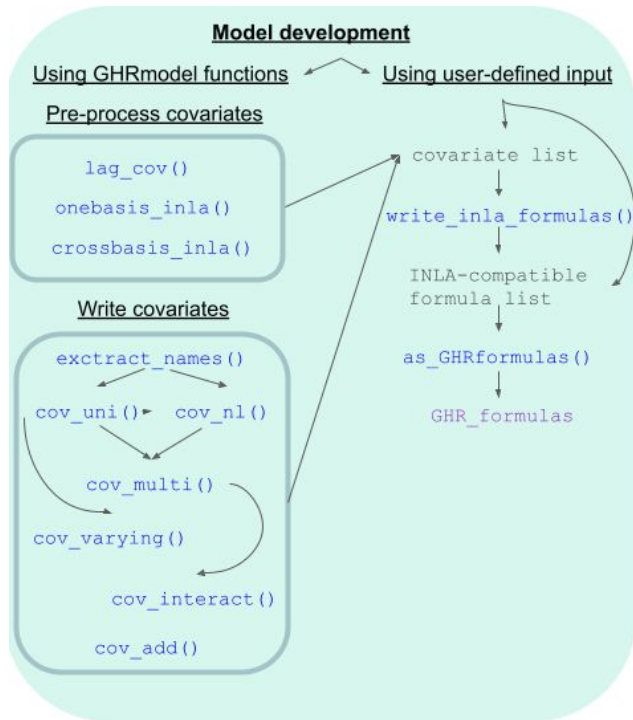
Fit models

Model  
evaluation

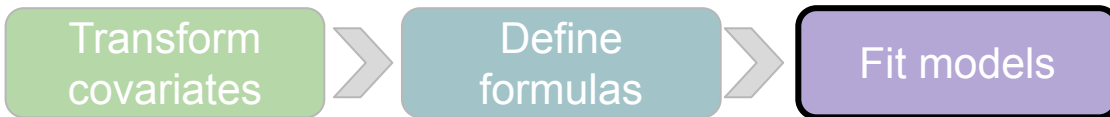
Transform  
covariates



Define  
formulas



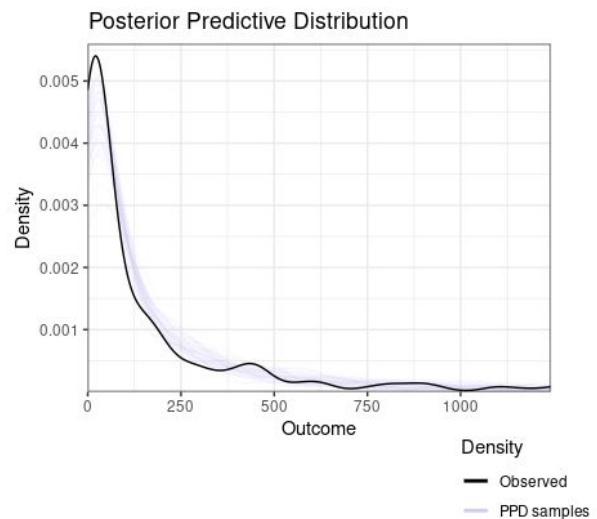
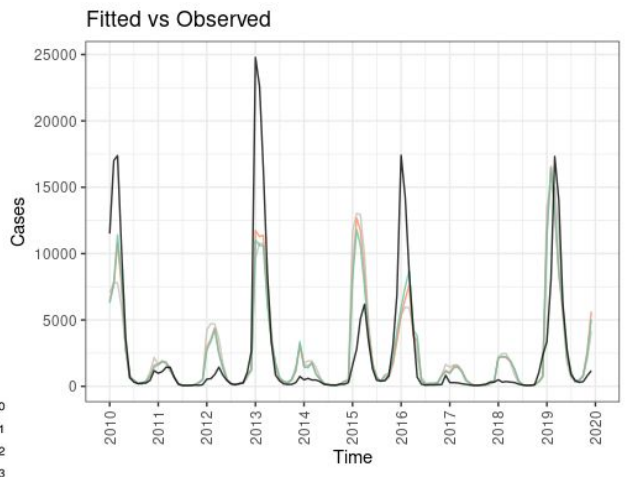
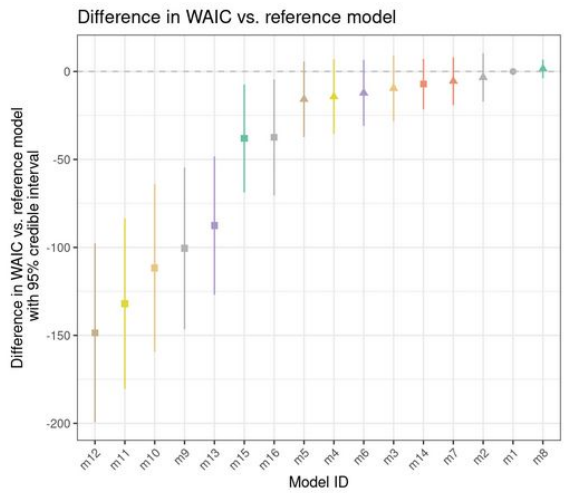
- Helpers to prepare long lists of formulas to be fitted in INLA including linear and non-linear effects, interactions, and random slopes.
- Helpers to specify complex random effects.
- Prepare one- and cross-basis functions.



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 outputs to be used when examining the models



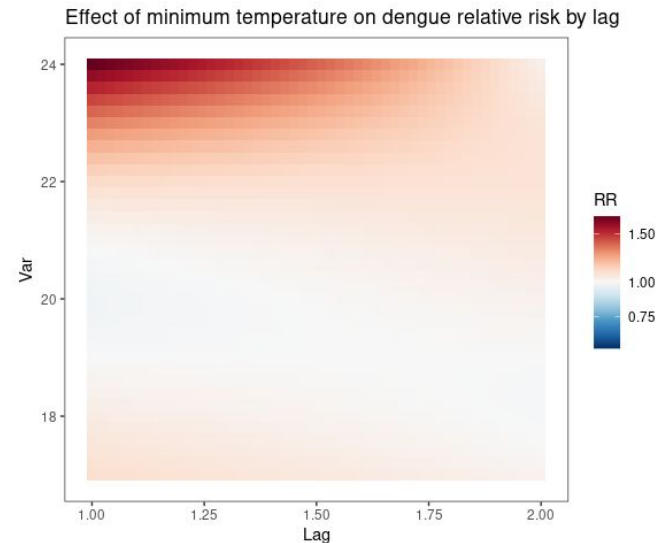
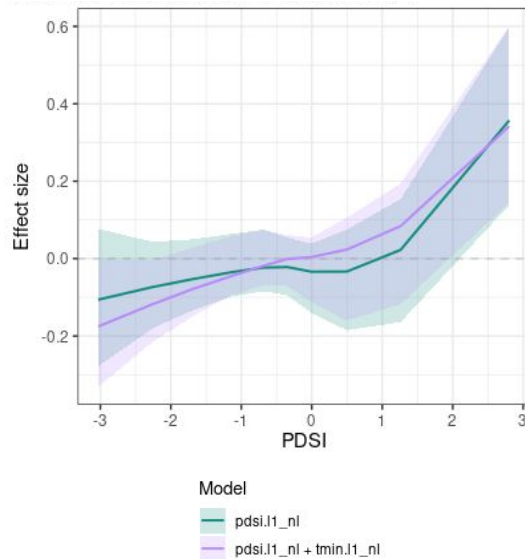
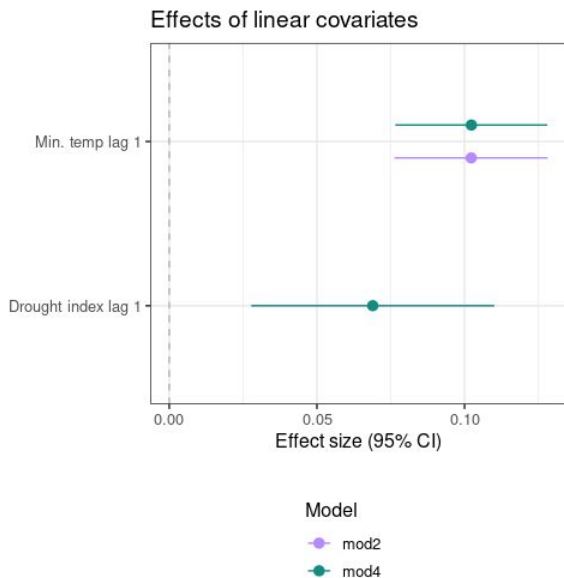
Goodness of fit metrics  
(DIC, WAIC, LMS, CRPS, RMSE, MAE, Re variance)

Observed vs. fitted

Posterior predictive checks

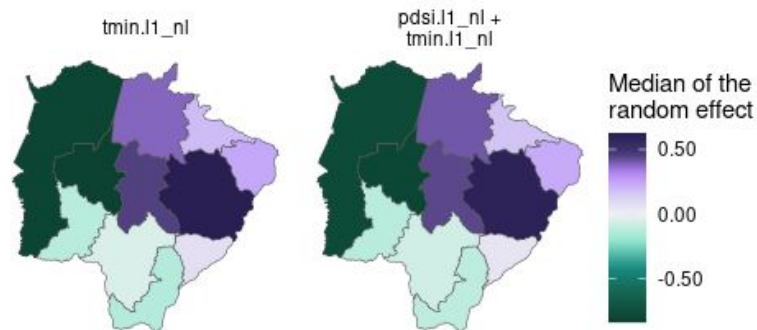
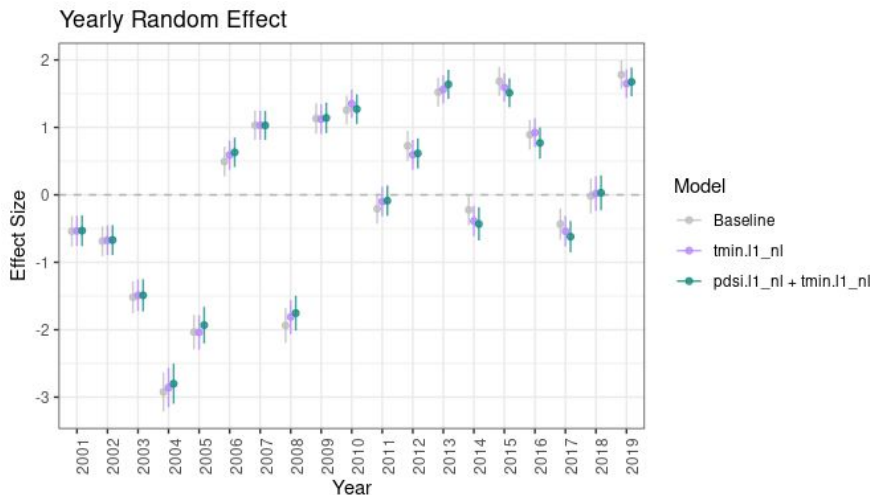


Evaluate effects: linear, non-linear, and DLNMs posterior estimates





Evaluate effects: random effects posterior estimates





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

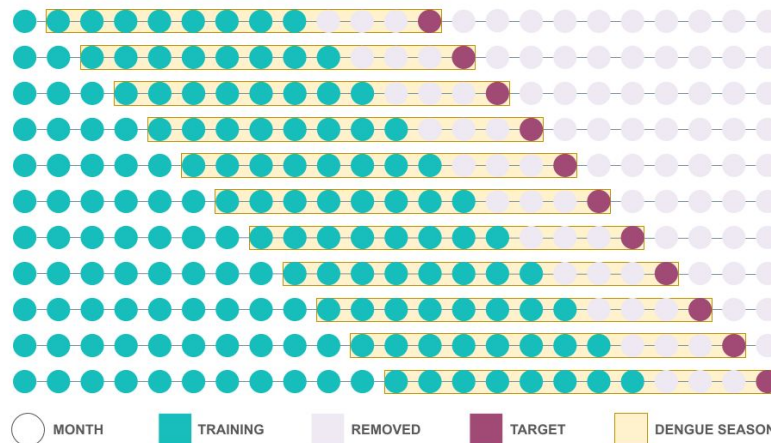


**PROF RACHEL LOWE**  
LEADING RESEARCHER

Expected release date:  
December 2026



## Expanding window cross-validation



Source: Fletcher et al. (2025)

Cross-validation

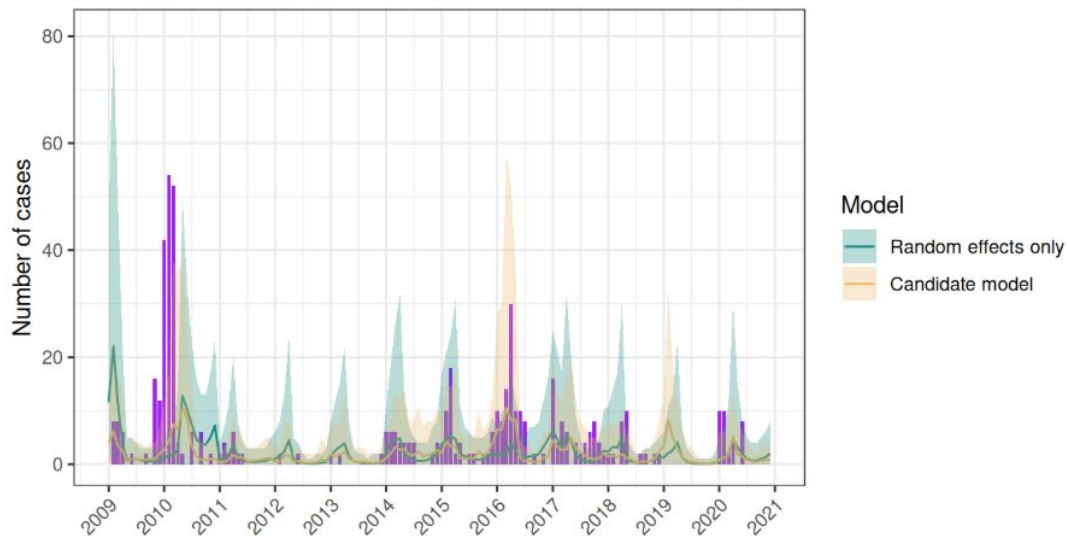


Predictive  
performance

Computes forecast accuracy metrics  
for case counts prediction:

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

Observed and predicted cases



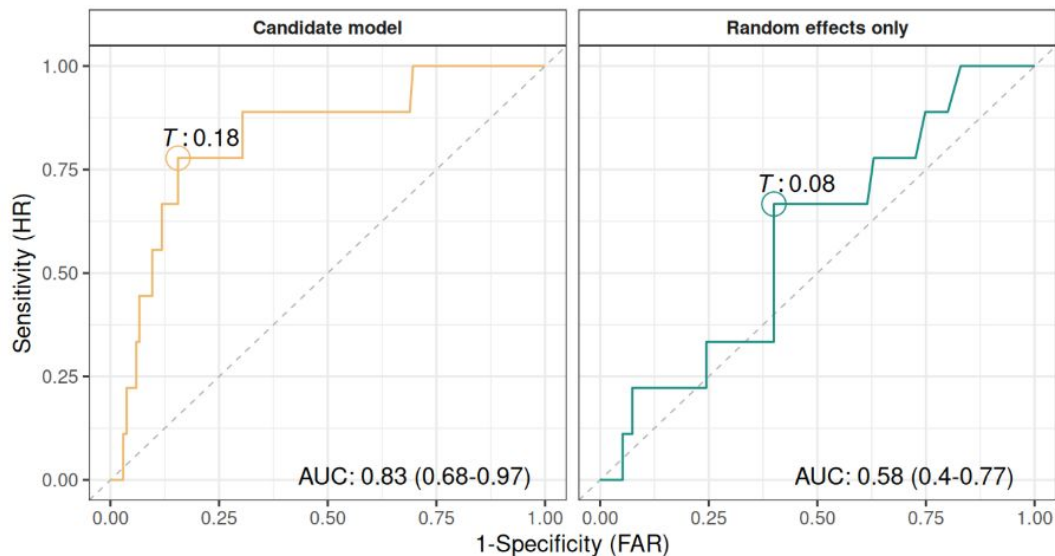
Cross-validation



Predictive  
performance

Computes forecast accuracy  
metrics for outbreak detection:

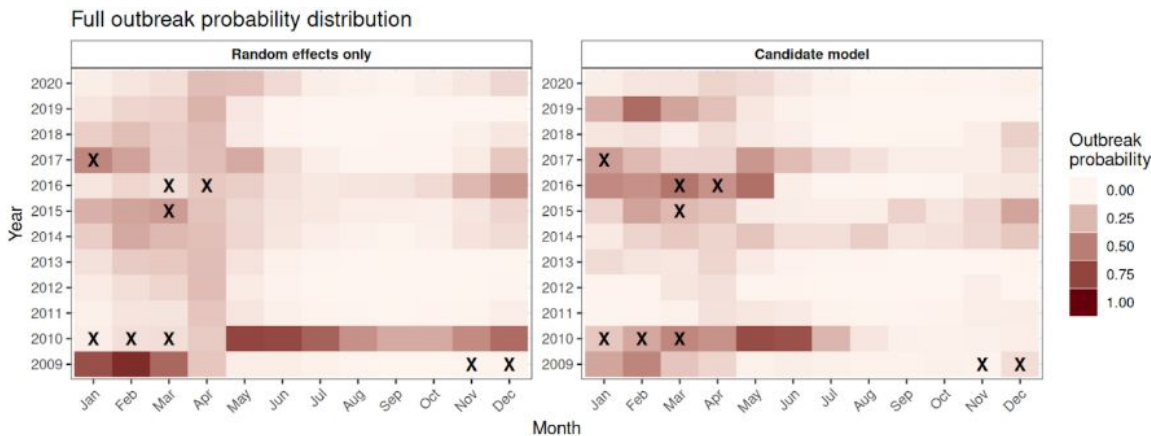
- Multiple thresholds to define outbreaks.
- Metrics: ROC, Brier, Confusion matrix
- Optimal trigger outbreak probability based on ROC curves.

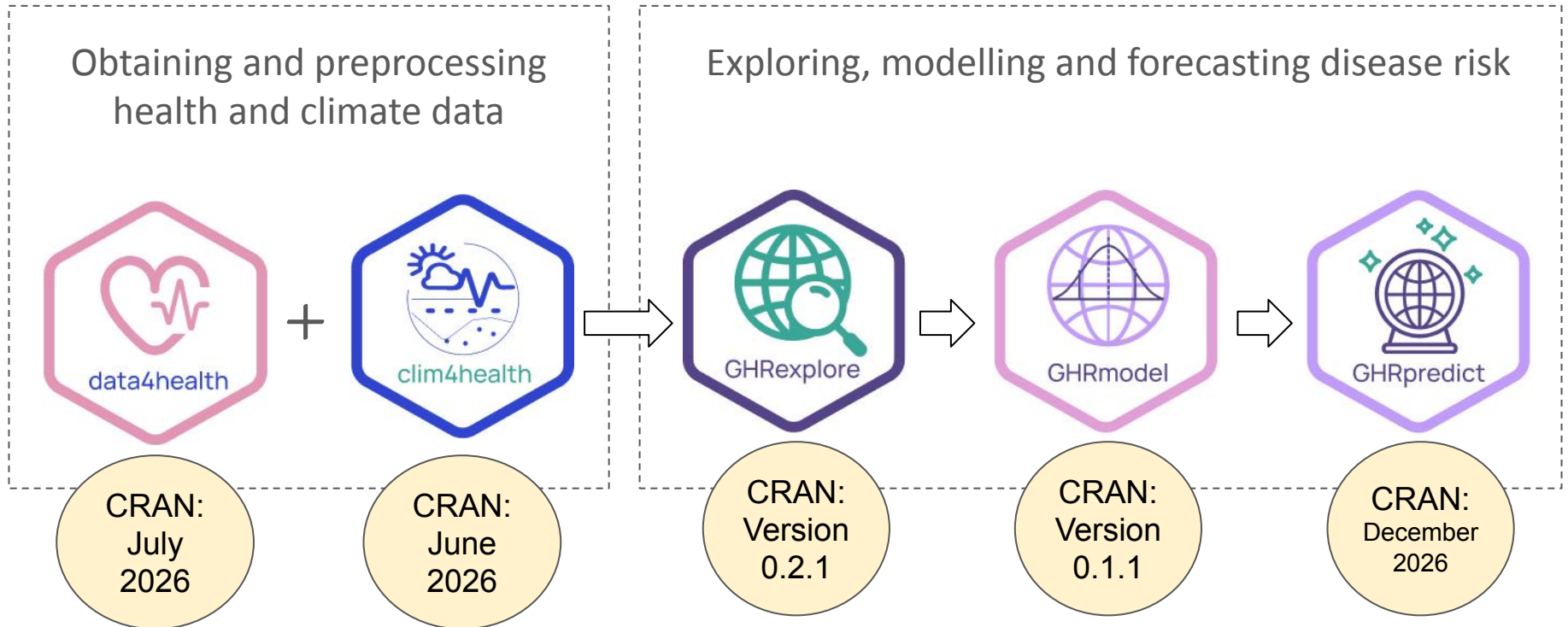




Computes forecast accuracy metrics for outbreak detection:

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



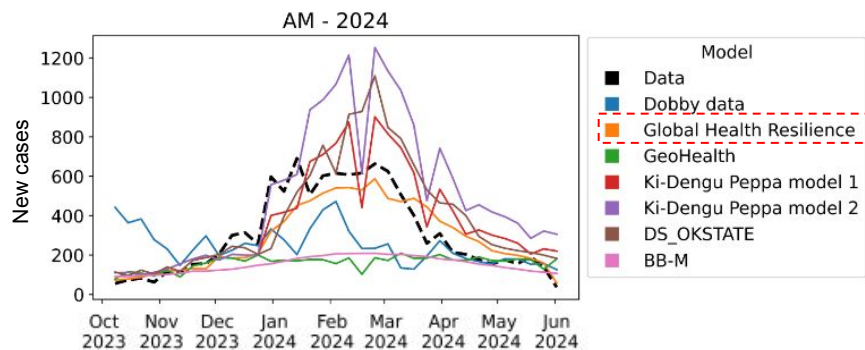


# Applications and use cases of the GHR tools



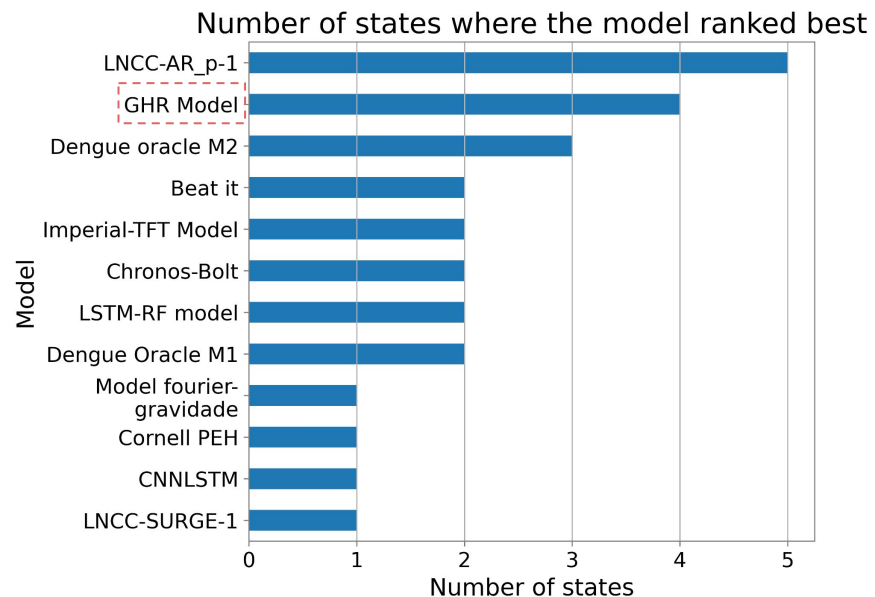
## Brazil - Dengue forecast

### InfoDengue Forecast Sprint 2024



Source: Codeço Coelho (2024)

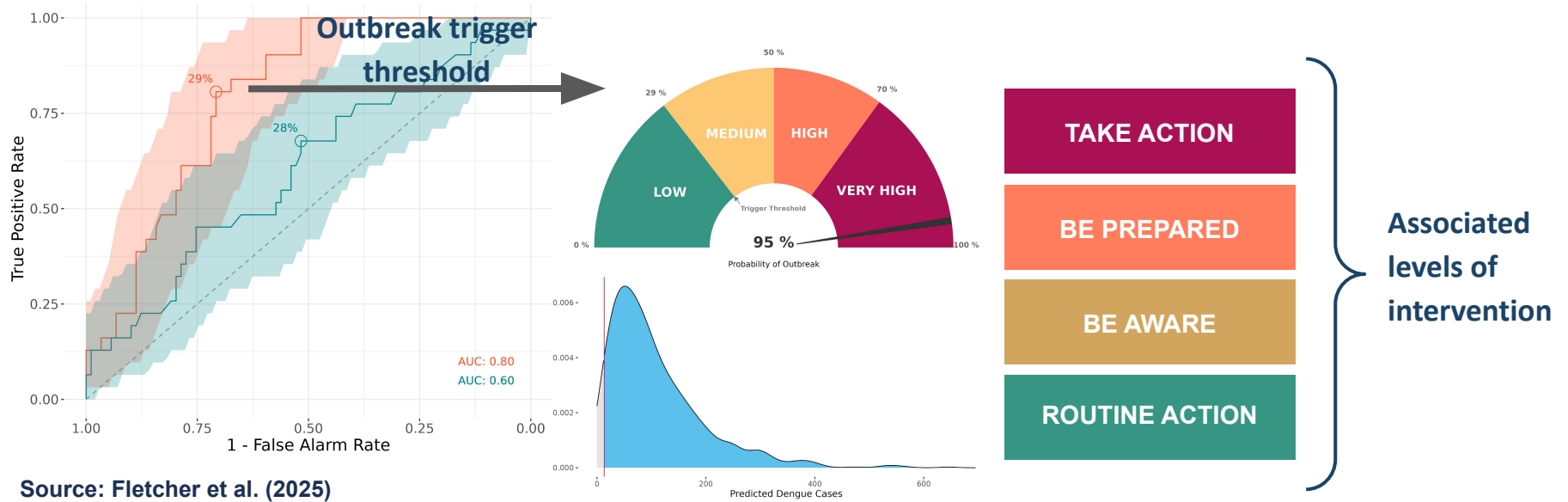
### InfoDengue Forecast Sprint 2025



Source: Milà & Fletcher (2025) *in prep*



## Barbados - Dengue forecast



Source: Fletcher et al. (2025)



# GHR tools use cases: ENDCast platform in South America

## Welcome!

This R Shiny platform hosts climate-informed forecasts for dengue, malaria and leptospirosis outbreaks across the Latin American and Caribbean (LAC) region, from 1 to 6 months ahead. These forecasts will be made operational during the 2023/24 El Niño event in a selection of LAC hotspots that are sensitive to El Niño.

Select region from the map...



Select disease:

Dengue

Threshold:

- 75th percentile
- 95th percentile

Target month:

- Aug 2023
- Sep 2023
- Oct 2023
- Nov 2023
- Dec 2023
- Jan 2024

Select color palette:

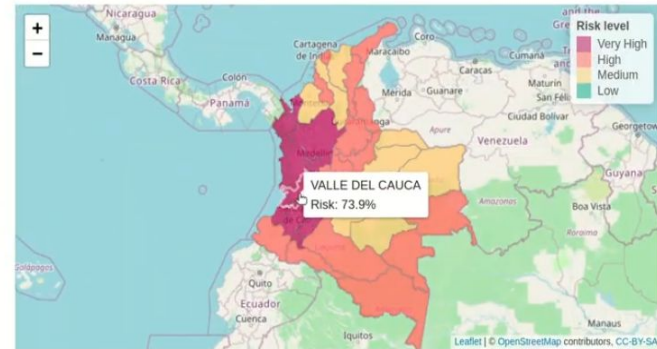
- Traffic light
- Colorblind friendly
- Greyscale

Select forecast period:

Jul 2023 (valid for Aug 2023 - Jan 2024)

Display absolute threshold values

National Local



This spatial map shows the outbreak risk level for the forecast month over Colombia.

You can manually change the forecast month of interest in the options on the left. This enables users to track how the outbreak risk level changes over time and space.

The risk levels are assigned as followed:

Color	Category	Outbreak probability
Green	Low	[0, 23)%
Yellow	Medium	[23, 40)%
Orange	High	[40, 70)%
Red	Very High	[70, 100)%

# GHR tools use cases: health impact attribution

Warming



Counterfactual climate

Hazard



Exposure-response relationship

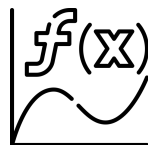
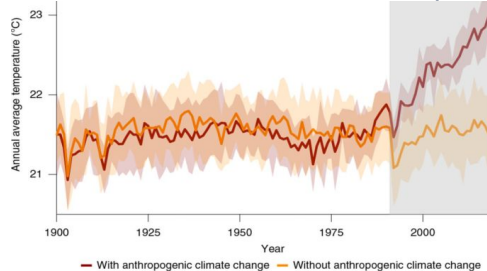
Disease



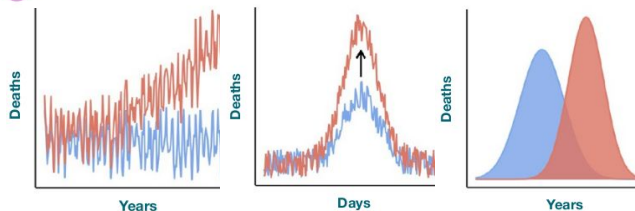
Health impact attribution



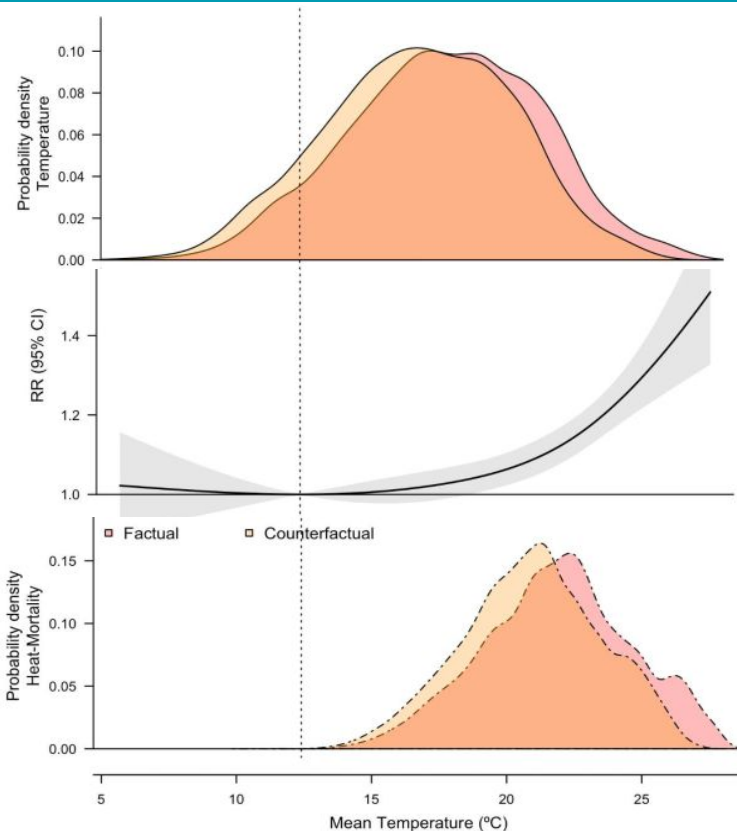
Source: Vicedo-Cabrera et al. (2021)



Model



Source: Carlson et al. (2024)




Source: Stuart-Smith et al. (2024)

GHRtools Home data4health clim4health GHRexplore GHRmodel GHRpredict About GHR Q

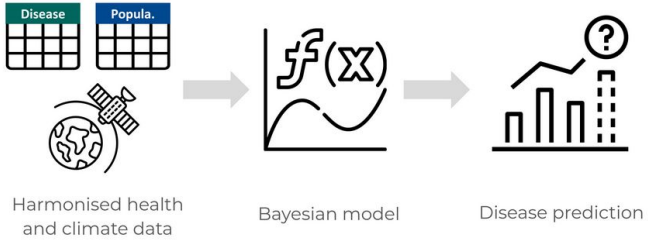
## GHRtools

A suite of R packages for climate and health workflows created by the Global Health Resilience (GHR) group at the Barcelona Supercomputing Center.



**From health and climate data to disease forecasts**

GHRtools is a suite of R packages that support **climate and health workflows** by streamlining health and climate data collection and harmonization, exploratory analysis, Bayesian modelling and forecasting.



Harmonised health and climate data      Bayesian model      Disease prediction

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



- Documentation
- Vignettes and tutorials
- Software manuals

# Questions?

ania.kawiecki@bsc.es

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

