



2024 Infodengue Sprint: Dengue Fever Predictive Modeling in Brazil

Important Dates:

1. Release of the announcement - June 20, 2024
2. Forecast submission deadline - August 16th, 2024
3. Presentation of participant's models - August 16th, 2024
4. Presentation of the ensemble model- August 30th, 2024

Organization

This initiative is led by the Mosqlimate project, in collaboration with the Harmonize and IDExtremes projects.

Challenge

The year 2024 has seen an exceptional number of reported dengue fever cases in various parts of the world. In Brazil, the disease has spread to areas in the south and at altitudes where epidemics were not previously recorded, and the incidence rate has far exceeded that of previous years. The objective of this sprint is to promote, in a standardized way, the training of predictive models with the aim of developing an ensemble forecast for dengue in Brazil.

The challenge has two testing targets and one forecast target. The date range of interest is the period between epidemiological week (EW) 41 from one year and week 40 of the next year, corresponding to a typical dengue season in Brazil.

1. **Validation test 1.** Predict the weekly number of dengue cases by state (UF) in the 2022-2023 season [EW 41 2022- EW40 2023], using data covering the period from EW 01 2010 to EW 25 2022;



-
2. **Validation test 2.** Predict the weekly number of dengue cases by state (UF) in the 2023-2024 season [EW 41 2023- EW40 2024], using data covering the period from EW 01 2010 to EW 25 2023;
 3. **Forecast.** Predict the weekly number of dengue cases in Brazil, and by state (UF), in the 2024-2025 season [EW 41 2024- EW40 2025], using data covering the period from EW 01 2010 to EW 25 2024;

Afterwards, the best-performing models in the first two validation runs will be included in an ensemble forecast model for 2025 built by the organizers.

Outcome variables

Predictions should be made at the state (federation unit) level. Participants are encouraged to generate results for all 27 federal units but, if this is not possible, the minimum set of states is:

- * North: Amazonas (AM)
- * Northeast: Ceará (CE)
- * Midwest: Goiás (GO)
- * Southeast: Minas Gerais (MG)
- * South: Paraná (PR)

The following outcomes should be provided by the models:

1. Curve of dengue cases by week (EW 41 to 40) for the 2022-23 season and the 2023-2024 season: point estimates and 90% predictive intervals.
2. Total cumulative number of probable dengue cases by UF in the 2022-23 season and the 2023-2024 season: point estimate of total cases and 90% predictive intervals.¹

¹ This outcome is optional



Data:

The training datasets will be available for download as tar.gz files on the Mosqlimate platform², along with a variable dictionary. Participants are welcome to peruse the [Mosqlimate data API](#) for specific subsets of data. Participants can use other data sources as long as they share them with other participants through the organizers. The data used must be open-access and updatable. It also must be available for all the Brazilian states.

List of available datasets [these can be updated at any time]:

Type	Variables	Description
Epidemiological data	Number of probable ³ dengue cases per week per municipality ; Effective reproductive number	From SINAN (aggregated by Infodengue)
Demographic data	Population per municipality (2010-2022); Classification of municipalities in terms of their area of influence (REGIC); % urban population	Source: IBGE
Climate data	Weekly temperature, humidity, precipitation; ENSO	From ERA5, aggregated by week
Miscellaneous	Holiday Calendar	Holidays affect disease reporting
Environment data	Koppen climate classification ; Biome type	Municipality main climate and biome

² <http://mosqlimate.org/> is hosting this modeling initiative.

³ Probable dengue cases = reported cases - discarded cases. They include both confirmed cases and cases under investigation.



Forecast Evaluation

For the evaluation, the logarithmic score, CRPS⁴ and the interval score³ will be computed. The calculation of the interval predictions will be done on a per-week basis.

Expected Results:

- A scientific publication will be made with the results of the individual and ensemble models for targets 1 and 2. All groups that contributed models that meet the minimum quality criteria will be part of the publication. Main contributors will be invited as co-authors.
- A technical note will be published by Infodengue, in Portuguese, with the results of the ensemble model. All participants of the sprint will be acknowledged.
- In November 2025, there will be an award for the model(s) that made the most useful forecast(s), in a ceremony at the E-Vigilância Congress, in Rio de Janeiro.
- It is hoped that this sprint will be a pilot experiment that will be replicated next year.

Registration:

To [register for the event](#), you should complete the registration form and agree to the event rules (described in this document).



Sprint Rules:

1. **GitHub Repository Template:** A GitHub repository template will be available on the Mosqlimate project's GitHub, including links to data downloads, simple model examples in Python and R, and forecast upload scripts. The participant should read and follow all the instructions.
2. **Public GitHub Repository:** All participating models must maintain their code in a public GitHub repository created from this template permanently, even after the end of the event.
3. **Forecast Submission Format:** The forecast submission format will be documented in the template repository. The upload code will check the formatting before submission. Code and data must be available for independent analysis.
4. **Data Sharing:** Participants may use other data sources as long as they share them with other participants through the organizers. The data used must be open-access, updatable, and generalizable, i.e., available for states not included in the sprint.
5. **Discussion Forum:** A discussion forum for participants will be available on Discord.
6. **Participant Guidance:** Weekly virtual meetings will be organized to answer questions and explain the model/forecast submission processes.

References

1. Banholzer, N. *et al.* A comparison of short-term probabilistic forecasts for the incidence of COVID-19 using mechanistic and statistical time series models.
2. Gneiting, T. & Raftery, A. E. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Stat. Assoc.* **102**, 359–378 (2007).
3. Bracher, J., Ray, E. L., Gneiting, T. & Reich, N. G. Evaluating epidemic forecasts in an



interval format. *PLOS Comput. Biol.* **17**, e1008618 (2021).

4. Gneiting, T., Raftery, A. E., Westveld, A. H. & Goldman, T. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. (2005) doi:10.1175/MWR2904.1.

Contact:

alerta_dengue@fiocruz.br

Financial Support:

Project Mosqlimate (grant 226088/Z/22/Z)

