

Climate Change & Mosquito Spread Diseases: Modeling Malaria and Dengue

Makena V. Grigsby, Sharvee Joshi

June 4, 2025

Table of Contents

- 1 Introduction and Motivation
- 2 Research Question and Objectives
- 3 Background
- 4 Exploratory Data Analysis
- 5 Methodology
- 6 Global Warming Simulations
- 7 Conclusions

Introduction

- Vector-borne diseases (VBDs) are infections transmitted to humans or animals through the bite or sting of infected arthropods, such as mosquitoes, ticks, fleas, and flies.
 - Some examples of VBDs are malaria and dengue.
 - Symptoms include, but are not limited to: fever, headache, rash, nausea, organ failure, and in some cases, death.
- Climate change is a rapidly growing environmental problem that can cause extreme weather conditions over time.
- Predicting how climate change may influence disease spread is critical; we can be proactive in health planning, especially in third world countries.
- The project aims to simulate climate change by creating realistic adjustments to existing climate variables to reflect projected future scenarios.

Motivation Behind Tracking Disease

- Health systems in newly affected areas may be unprepared for VBD outbreaks.
- Rising temperatures can expand the range of disease-carrying mosquitoes to these newly affected areas.
- Currently, we focus on stopping the spread of disease in particular areas; instead, we want to focus on what to do if extreme weather conditions force disease to be spread in unprepared areas.

Research Question and Objectives

Main Research Question:

- How might climate change and environmental conditions impact the number of cases of malaria and dengue across regions?

Project Objectives:

- 1 Simulate realistic global warming scenarios by changing climate variables in the dataset.
- 2 Build a predictive model to estimate malaria and dengue cases based on temperature, precipitation, UV index, and air quality.
- 3 Evaluate how projected changes in climate may influence disease outcomes under different warming scenarios.

Literature Review

- **Malaria** is caused by *Plasmodium* parasites which is transmitted by *Anopheles* mosquitoes. It thrives in **warm, humid** environments and is most prevalent in **Sub-Saharan Africa**.
- **Dengue** is caused by a flavivirus and spread by *Aedes aegypti* mosquitoes, which prefer urban areas and stagnant water. Typically, transmission increases with **high temperatures and more rainfall**.

Climate change is expected to change a multitude of these climate variables.

- Ryan et al. [4] project that dengue may affect an additional 1.5 billion people by 2080.
- Caminade et al. [1] show malaria transmission potential is shifting toward higher altitudes and latitudes.

Introduction to Data

- Monthly panel data from 2000 to 2020, across multiple countries and regions. [2]
- **Variables:**
 - Climate: Avg. Temp, Precipitation, UV Index, AQI
 - Outcome: Malaria Cases, Dengue Cases
 - Time Data: Year, Month
 - Location: Country, Region
 - Economic Factors: Population Density, Health Budget
- 34,560 observation.
 - 33,866 observations after removing data from Antarctica and Greenland.

Note: This dataset was generated to approximate real-world patterns in climate and disease but is not derived from official surveillance systems. As such, it serves as a tool for model development and hypothesis exploration, rather than for drawing precise empirical conclusions.

Table Summaries

Table: Summary Statistics for Climate and Disease Variables

Statistic	Malaria	Dengue	Precip (mm)	Temp (°C)	AQI	UV	Pop. Dens.	Health \$
Min.	0.00	0.00	0.00	-2.56	0.00	0.87	50.00	205.00
1st Qu.	32.00	48.00	124.64	14.40	15.32	8.92	136.25	1435.50
Median	69.00	76.00	195.32	20.35	40.69	10.76	277.50	2859.00
Mean	70.52	75.94	192.67	20.40	43.79	10.21	266.90	2764.65
3rd Qu.	107.00	104.00	260.15	26.34	68.40	12.00	384.75	3996.75
Max.	201.00	181.00	485.70	44.58	168.85	12.00	498.00	4969.00

Dataset: Monthly panel data (2000–2020), 34,560 rows, 12 variables across multiple countries and regions.

Malaria Cases Around the World

Total Malaria Cases by Country (2000–2020)

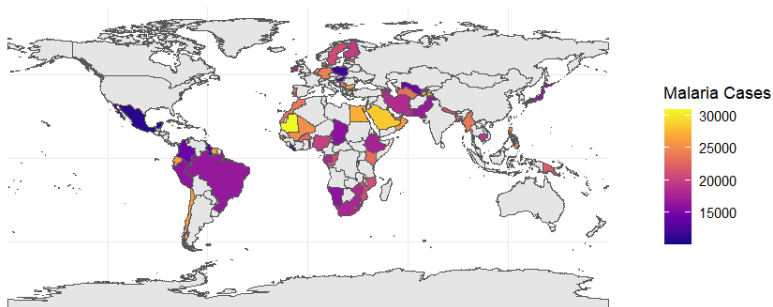


Figure: Malaria Cases

Dengue Cases Around the World

Total Dengue Cases by Country (2000–2020)

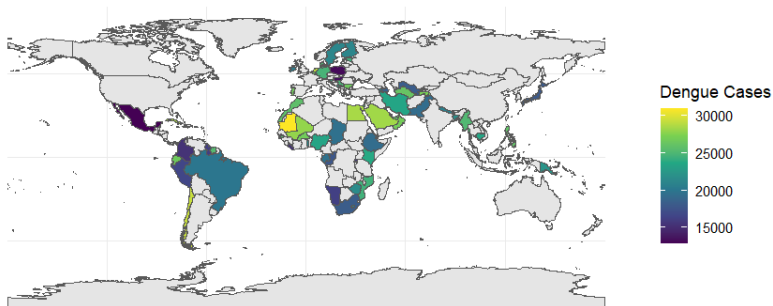
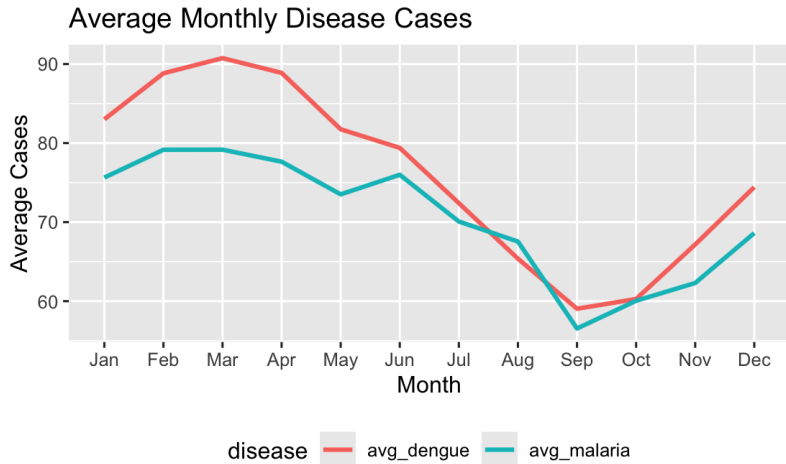


Figure: Dengue Cases

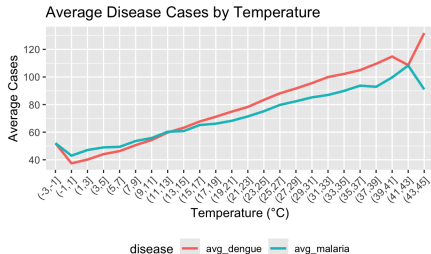
Monthly Seasonality in Malaria Cases

Plot: Average monthly dengue and malaria cases across all countries (2000–2020)



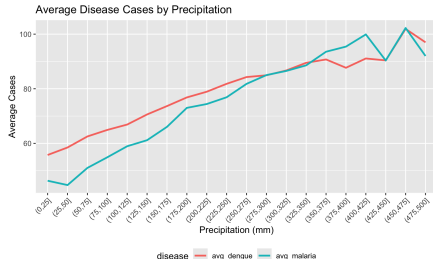
Climate Effects on Malaria Cases

Plots: Average disease cases across temperature and precipitation



Increase in cases with rising temperature supports findings in Liu et al. (2023).

Summary: Both climate factors exhibit strong, positive associations with disease prevalence.



Dengue and malaria rise with precipitation, likely due to favorable mosquito breeding conditions.

Random Forest: Baseline

- We used the **tidymodels** framework to predict malaria cases using climate variables:
 - `temp_lag1`, `precip_lag1`, `uv_index`, `air_quality_index`, `month`, `region`, `country`
- Data was cleaned and prepared using **dplyr** and **recipes**:
 - Created lagged temperature and precipitation variables via `robust` function.
 - Converted categorical variables with `step_dummy()`.
- A **Random Forest** model was fit using `rand_forest()` via the `ranger` engine.
- A 60/40 split was used for training and testing using `initial_split()` with stratification on malaria cases.
- **Why Random Forest?**
 - Flexible, non-parametric model.
 - Automatically captures nonlinear relationships and complex interactions.
 - No strict distributional assumptions.

Poisson Model: Accounting for Count Data

- We used the **tidymodels** framework to fit a Poisson regression model using the same predictors:
 - temp_lag1, precip_lag1, uv_index, air_quality_index, month, region, country
- Data was cleaned and prepared using **dplyr** and **recipes**:
 - Created lagged temperature and precipitation variables via robust function.
 - Applied `step_dummy()` to categorical variables.
 - No normalization applied due to Poisson model assumptions.
- The model was fit using `glm()` with the `poisson()` family.
- A 60/40 split was used for training and testing using `initial_split()` with stratification on malaria cases.
- **Why Poisson?**
 - Directly models count data using a log link.
 - Provides interpretable coefficients for understanding predictor effects.
 - Appropriate for count-based outcomes like malaria cases.
 - **Assumption:** Variance equals mean.

Why Use a GAM-NB Model for Disease Prediction?

- Used `mgcv::gam()` with `family = nb()` to model overdispersed disease count data.
- **Why not just Negative Binomial?** Because climate-disease relationships are nonlinear.
- Included **smooth terms** for climate predictors:
 - `s(avg_temp_c_lag1)`, `s(precipitation_mm_lag1)`, `s(uv_index)`, `s(air_quality_index)`
- Linear terms included: `population_density`, `healthcare_budget`
- A General Additive Model allows climate variables to affect cases in a flexible, non-monotonic way.
- Negative Binomial handles overdispersion in counts, which is typical in real-world disease data.
- Model fit assessed using R^2 Nakagawa and Schielzeth [3]

Modeling Results

Table: Malaria - Model Comparisons

Metric	Random Forest	Poisson	GAM-NB
RMSE	43.566	42.819	43.393
R-squared (R^2)	0.132	0.139	0.116

Table: Dengue - Models Comparisons

Metric	Random Forest	Poisson	GAM-NB
RMSE	33.598	32.423	33.14
R-squared (R^2)	0.230	0.243	0.213

Explanation of Four Scenarios

Planned Simulation Scenarios:

- Case 1: Warming Only - Increase temperature 10 degrees Celsius
- Case 2: Warming + Rain Variability – Increase temperature by 10 degrees Celsius and $\pm 20\%$ variability in precipitation
- Case 3: Extreme Climate Change – Mimic extreme weather conditions and increase temperature by 10 degrees Celsius, $\pm 20\%$ variability in precipitation, increase UV index by 3, and worsen air quality index by 10 points.
- Case 4 (Simulated Using the GAM-NB Model): Increased Health Investment – Simulate a 50% increase in healthcare budget to examine its potential to offset rising disease burden.

Comparison of Models Using Dengue Simulated Data

Random Forest			
Scenario	Mean Predicted	Total Cases	% Change
Baseline	75.62	1,024,420	0.00%
Warming Only	78.80	1,067,440	4.20%
Warming + Precipitation	78.72	1,066,430	4.10%
Full Scale	80.17	1,086,119	6.02%

Poisson			
Scenario	Mean Predicted	Total Cases	% Change
Baseline	75.48	1,022,462	0.00%
Warming Only	75.37	1,021,135	-0.144%
Warming + Precipitation	75.37	1,021,140	-0.129%
Full Scale	75.37	1,020,986	-0.129%

Distribution of Dengue Cases Across Scenarios

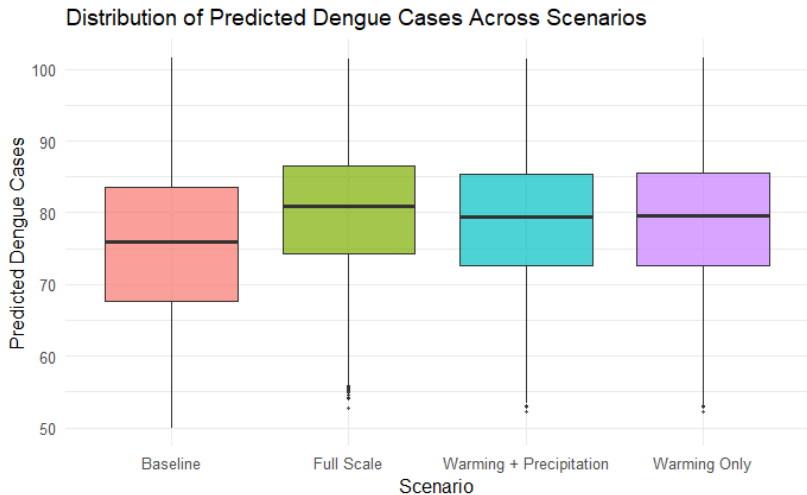


Figure: Distribution of Dengue Cases - Random Forest

Comparison of Models: GAM-NB

Malaria Model

Scenario	Mean Predicted	Total Cases	% Change
Baseline	76.20	1,028,367	0.00%
Case 1: Warming	92.70	1,250,706	+21.7%
Case 2: Warm + Precipitation	93.01	1,254,637	+22.1%
Case 3: Extreme Climate	95.18	1,283,594	+24.9%
Case 4: Budget Increase	70.90	956,880	−7.0%

Dengue Model

Scenario	Mean Predicted	Total Cases	% Change
Baseline	76.09	1,015,929	0.00%
Case 1: Warming	92.58	1,234,831	+21.6%
Case 2: Warm + Precipitation	92.73	1,236,835	+21.8%
Case 3: Extreme Climate	94.50	1,261,143	+24.1%
Case 4: Budget Increase	71.33	951,114	−6.4%

Why GAM-NB Predicts More Cases Despite Lower R^2

- **R^2 reflects fit to historical data** — not necessarily predictive power under new climate scenarios.
- **Poisson model assumes linearity and equal variance**, which may slightly inflate R^2 but limits responsiveness to extreme changes.
- **GAM-NB captures nonlinear effects** in temperature, precipitation, and air quality — better reflecting thresholds or nonlinear growth in disease risk.
- **Handling overdispersion** using a Negative Binomial family, making it more realistic for noisy disease count data.
- **Bottom Line:** GAM-NB may have a lower R^2 , but it better simulates case surges under extreme climate conditions — which is critical for scenario planning.

Future Work: Enhancing Climate-Disease Modeling

1. Biological Lag Effects:

- Use time-lagged climate variables to reflect incubation processes:
 - *Mosquito eggs*: 7–14 days (temperature-dependent)
 - *Human disease*: 10 days (dengue), 12–14 days (malaria)
- Align predictors with delayed disease response for improved fit

2. Advanced Mixed Models:

- Fit Generalized Linear Mixed Models (GLMMs) with:
 - **Random intercepts** to account for baseline differences across countries or regions
 - **Random slopes** to allow climate effects (e.g., temperature, rainfall) to vary by location

3. Move toward Early Warning Systems:

- Combine climate forecasts + disease models to anticipate outbreaks

Biological Motivation: Toward Mixed-Effects Models

- Disease spread is also influenced by biological and ecological processes:
 - **Mosquito egg incubation period**
 - **Human disease incubation period**
- These introduce lags between climate conditions and observed cases.
- Incorporating them would require:
 - Lagged predictors across multiple time scales
 - Modeling of mosquito and host dynamics
- This motivates future use of a **mixed-effects framework**, which can account for:
 - Region-specific effects
 - Mosquito and host dynamics
 - Nested biological timing processes (e.g. Mosquito egg → larva → adult)
- *Takeaway:* Our current model is strong, but the biological system is more complex—mixed models would get us even closer.

Team Contributions

Makena Grigsby:

- Data cleaning, transformation, and feature engineering
- Exploratory Data Analysis: Seasonal trends, temperature and precipitation patterns
- Researched how to conduct an analysis on ecological data
- Implemented and evaluated GAM-NB model for malaria and dengue
- Formatted and analyzed metrics (RMSE, R^2)

Sharvee Joshi:

- Researched malaria and dengue and found sources explaining climate change
- Data cleaning: removing data from random countries and adding lagged variables for temperature and precipitation
- Literature Review and Slides Set Up
- Initial Random Forest Model and Poisson Model
- Map Visualization
- Formatted and analyzed metrics (RMSE, R^2)

References I

- [1] Cyril Caminade et al. "Impact of climate change on global malaria distribution". In: *Proceedings of the National Academy of Sciences* 111.9 (2014), pp. 3286–3291.
- [2] hopeofchange. *Climate-driven Disease Spread*.
<https://www.kaggle.com/datasets/hopeofchange/climate-driven-disease-spread>. Accessed: 2025-05-21. 2023.
- [3] Shinichi Nakagawa and Holger Schielzeth. "A general and simple method for obtaining R^2 from generalized linear mixed-effects models". In: *Methods in Ecology and Evolution* 4.2 (2013), pp. 133–142. DOI: 10.1111/j.2041-210x.2012.00261.x.
- [4] Sadie J Ryan et al. "Global expansion and redistribution of Aedes-borne virus transmission risk with climate change". In: *Nature Microbiology* 8 (2023), pp. 71–80.

Any Questions?

Thank you!