

Can climate patterns predict dengue outbreaks? A causal-based analysis on the role of climate change in *Aedes*-borne disease transmission

Javier Corvillo^{1, 2*}, Verónica Torralba^{2*}, Diego Campos², Ana Riviére-Cinnamond³, and Ángel G. Muñoz¹

¹Complutense University of Madrid, Department of Earth Science and Astrophysics, Madrid, 28040, Spain

²Barcelona Supercomputing Center, Earth Sciences Department, 08034, Spain

³Pan-American Health Organization, Communicable Diseases and Health Analysis, Panama City, 0843-03441, Panama

*javier.corvillo@bsc.es / veronica.torralba@bsc.es / angel.g.munoz@bsc.es

ABSTRACT

Background: Vector-borne diseases transmitted by *Aedes* mosquitoes such as dengue, Zika, and chikungunya, pose significant public health challenges worldwide in the wake of human-driven climate change. However, while their transmission is known to be susceptible to some climate variables like temperature or the amount of rainfall, the overall role of climate patterns on the emergence of these diseases is not so well understood.

Methods: Using data from a number of sources, we explore and analyze the response of *Aedes*-borne disease transmission to climate patterns, in order to understand its influence on disease outbreaks. Our analysis is composed of three different studies: 1) a timescale decomposition of disease transmissibility values, thereby guiding officials to understand the behavior of outbreaks for budget and resource allocation; 2) a correlation analysis between transmissibility values and different climate patterns, such as El Niño Southern Oscillation, in order to understand the effects of natural climate patterns onto *Aedes*-borne outbreaks; and 3) a causality analysis to solidify findings obtained through correlation, identifying the most relevant predictors and their applicability in a climate-and-health service framework for forecasting the transmissibility of *Aedes*-borne diseases.

Results:

Conclusions:

Keywords: Public Health, Vector-borne Diseases, Epidemiology, Climate Change, Climate Services, Environmental Sciences

1 Multilingual abstracts

Please consult the [Additional files](#) section for abstract translations into the other five official languages of the United Nations (Arabic, Chinese, French, Russian and Spanish).

2 Background

2.1 The emergence of vector-borne diseases in the context of climate change

2.2 Known drivers on *Aedes*-borne disease transmission dynamics and their impact

2.3 Climate patterns as an aggregate of unknown transmissibility drivers

2.4 The applicability of climate services in vector-borne disease prevention

3 Methods

We utilize a select number of datasets and methodologies to undertake three distinct analyses, in order to characterise

the behavior of *Aedes*-borne disease transmissibility, and to understand the role of climate patterns in their potential predictability.

By using global climate products that transform climate variables into the climate-driven component of *Aedes*-borne transmissibility, we can explore the behavior of vector-borne diseases at multiple timescales (seasonal, inter-annual, decadal, and long-term trends). We later employ a series of correlation and causality analyses in order to understand the role of climate patterns on disease emergence across regions and seasons. By highlighting which climate patterns are dominant over the different areas and seasons, we can assess their impact on the transmissibility of *Aedes*-borne diseases, therefore serving as predictors for disease outbreaks.

3.1 Redefining R_0 as a bridge between climate and health

In order to understand the climatological behavior of *Aedes*-borne diseases, we first need to understand the behavior of the climate component of the disease transmission. Generally speaking, the basic reproduction number, or R_0 , is a commonly used epidemiological metric that quantifies the transmissibility of vector-borne diseases, and is defined as the average number of secondary cases generated by a single infected individual in a susceptible population. It includes the effects of the vector's behavior (e.g. absence or presence in the area), human behavior (e.g. whether an infected host can transmit the disease by traveling), and the climate (whether the conditions are favorable for the vector to transmit the disease). However, by only integrating the climate component of the disease transmission for the computation of R_0 , then the resulting R_0 metric is more so interpreted as the role of environmental conditions in the spread of the disease. Thus, this definition of R_0 , which we can understand as the diseases' environmental suitability, can be understood as a first approximation of the role of the climate in the dynamics of vector-borne diseases.

In this context, our analysis is then composed of three different studies under this definition of R_0 :

3.2 The *Aedes* Disease Environmental Suitability 2's (AeDES2) monitoring system

AeDES2's monitoring system was used throughout this study in order to obtain the R_0 values for the analysis. Improving over its predecessor, AeDES2 is a climate-and-health service that provides real-time monitoring of the environmental suitability for *Aedes*-borne diseases. The system uses a climate variables like temperature and precipitation, and through its integration with ento-epidemiological models, as well as with calibration with recorded DENV cases, R_0 values are computed for different regions and seasons. The monitoring system is designed to be used by public health officials and researchers to assess the risk of disease outbreaks and to inform prevention and control strategies.

3.3 Analysis 1: Multi-timescale climate decomposition of R_0

With the intent of isolating the human-driven signal from the natural variability R_0 data, a timescale decomposition methodology was used to obtain the total variance across different time-scales. This approach allows us to separate the complex R_0 signal into distinct temporal components, each providing insight into the underlying climatology of *Aedes*-borne disease transmission. This decomposition is particularly crucial in regions endemic to *Aedes*-borne diseases, such as tropical and subtropical areas of Latin America, Southeast Asia, and parts of Africa, where understanding how different climate processes operate at various temporal scales to influence disease transmissibility is essential for developing effective public health strategies. In these hotspot regions, the ability to

distinguish between predictable seasonal patterns, multi-year climate oscillations, and long-term warming trends enables health authorities to optimize resource allocation, implement targeted intervention strategies, and develop more accurate early warning systems for disease outbreak prevention.

3.3.1 Data

The timescale decomposition analysis was undertaken using R_0 outputs from the AeDES2's monitoring system. The 1980-2021 monthly-mean period of AeDES2's R_0 values was selected for the analysis, for a total of 504 months or 167 full seasons. Considering that vector borne diseases are extending to previously unaffected areas due to the effects of man-made climate change, AeDES2's coverage has been increased since its inception to contain global outputs, allowing for a comprehensive analysis of the relationship between climate variability indices and R_0 both in current *Aedes* hotspots and emerging regions.

3.3.2 Methodology

As R_0 doesn't follow a clearly defined probability distribution function, the temporal analysis filters a given R_0 time-series of any given grid-point by employing the non-parametric locally estimated scatterplot smoothing technique (LOESS). Sensitivity tests have been conducted in order to obtain the best LOESS smoothing parameter for the analysis, using three verification metrics for the goodness of fit of the model: the highest R squared value (R^2), the lowest Akaike Information Criterion (AIC) value, or the lowest Generalized Correlated Cross-Validation (GCV) value. Whenever these metrics yield conflicting results, the GCV value is prioritized as the primary selection criterion. Unlike R^2 , which can artificially inflate with increased model complexity, or AIC , which relies on asymptotic assumptions that may not hold for finite samples, GCV provides a more robust assessment of model generalizability by directly penalizing overfitting through its leave-one-out validation procedure.

Once the ideal LOESS smoothing parameter is found for the R_0 data, the R_0 time-series for each grid-point is separated into four components: a long-term trend signal (understood to be the trend caused by anthropogenic climate change), an inter-annual signal (year to year), a decadal signal (10-30 years), and lastly, a remainder signal which contains other signals of the data (i.e., inter-annual and inter-decadal variability, among others). Variance maps for each of these four components capture the overall direction of the data over time, as well as the climatological variability of R_0 in any given grid-point.

Variance maps, as well as any results from following analyses, are shown for both global outputs and the Panama region. The Panama region is selected as a case study for the analysis, as it is known to be a present hotspot for *Aedes*-borne diseases, with a long history of DENV outbreaks and a complex interplay of climatic factors that influence

disease transmission.

After the variance maps are obtained, R_0 values are then detrended for the following correlation and causality analyses. While detrending through the assumption that R_0 changes linearly over time could be a valid approach, it fails to capture the temperature dependency of the data, expressed extensively in the literature. In order to capture this temperature dependency, a similar timescale decomposition analysis is performed on the detrended R_0 data, but using temperature data from the AeDES2's monitoring system datasets as the independent variable (monthly-mean temperature data consisting of the GHCN-CAMS project, the CPC Unified Global Temperature dataset, the ERA5 reanalysis dataset, and the ERA5Land reanalysis dataset). In this way, the obtained trend serves as a functional relationship between temperature and R_0 : a temperature-based description of the R_0 signal, attributed to the warming of the planet.

3.4 Analysis 2: Correlation studies between R_0 and climate variability indices

After analyzing the R_0 signal and its variability through timescale decomposition, we assess the impact of several climate variability indices on global R_0 values over the chosen 1980-2021 monthly-mean period.

3.4.1 Data

Correlation studies are performed over both global and Panama regions, using the temperature-based detrended R_0 data as in the previous analysis over the different seasons. A total of 9 temperature-based climate variability indices have been used for the correlation analysis, which have been computed using the detrended temperature data utilized in the previous analysis. Their periodicity, as well as their main pattern type, are listed and summarized in Table 1.

3.4.2 Methodology

The correlation analysis was performed using the Pearson correlation coefficient, which quantifies the linear relationship between two variables. For computation of statistical significance in correlation, the non-parametric Monte Carlo method was used, with a p-value threshold of 0.05.

3.5 Analysis 3: Causality studies between R_0 and climate variability indices. Outlining of predictors for disease outbreaks

Causal-based patterns can be identified after this analysis, which allow for a more robust foundation for the understanding of the underlying mechanisms between climate variability and R_0 patterns. These causality studies are performed over both global and Panama regions and over the different seasons. In discarding potentially spurious results obtained through correlation, this causality analysis can be used to outline the most relevant predictors for disease outbreaks. These predictors, in turn, can be used for the refining and building of AeDES2's prediction system for improving the accuracy and skill of the ensemble forecasts compared to its predecessor.

3.5.1 Data

The datasets that were used for the causality analysis are the same detrended datasets as those employed in Section 3.4.1.

3.5.2 Methodology

Causality analysis between R_0 and climate variability indices was performed by using Liang-Kleeman's proposed methodology for computing information flow between two entities of a dynamical system, quantifying the amount of information that one time series (the climate variability indices) can provide about another time series (R_0 patterns). Once the transfer entropy is computed, it is then normalized in order to account for the different scales of the two time series. Statistical significance is computed using Fisher's information matrix, with a p-value threshold of 0.05.

4 Results

The most relevant results from each of the three analyses are summarized over the following sections of the manuscript, highlighting the most relevant findings and their implications. The complete results from the analyses described above, including the correlation and causality study maps for each individual climate pattern over the different seasons, can all be found in the supplementary material provided along with this manuscript (see [Additional files](#)).

4.1 Analysis 1: Multi-timescale climate decomposition of R_0

4.2 Analysis 2: Correlation studies between R_0 and climate variability indices

4.3 Analysis 3: Causality studies between R_0 and climate variability indices. Outlining of predictors for disease outbreaks

5 Discussion

5.1 The added value of climate patterns in seasonal forecasting of *Aedes*-borne diseases

5.2 Analysis 1: Multi-timescale climate decomposition of R_0

5.3 Analyses 2 and 3: Correlation and causality studies between R_0 and climate variability indices

5.4 Notable limitations and constraints

6 Conclusions

1. Historical and climatological analyses of R_0 values for *Aedes*-borne diseases provide insight in understanding the role of climate in disease emergence, and can be used to improve the accuracy of seasonal forecasts through the identification of climate predictors. However, while global climate models are suitable for a broad, general-purpose understanding, high-resolution data is preferred when more nuanced analysis are performed in endemic regions, in order to provide more accurate and actionable information for public health officials.

- 2.

Index Name	Abbreviation	Periodicity	Pattern Type
Atlantic 3 Index	ATL3	Several months to a few years	Oceanic
Indian Ocean Basin	IOB	Several months to a few years	Oceanic
Indian Ocean Dipole	IOD	Between 2-7 years	Oceanic
El Niño 3.4 Index	Niño 3.4	Between 2-7 years	Oceanic
North Pacific Meridional Mode	NPMM	Several months to a few years	Atmospheric
South Atlantic Subtropical Dipole	SASD	Several months to a few years	Oceanic
Southern Indian Ocean Dipole	SIOD	Several months to a few years	Oceanic
South Pacific Meridional Mode	SPMM	Several months to a few years	Atmospheric
Tropical North Atlantic	TNA	Several months to a few years	Oceanic

Table 1. Summary of the climate variability indices used in the analysis used for the correlation and causality studies.

7 Additional files

8 Abbreviations

- AeDES2: *Aedes Disease Environmental Suitability 2*
- AIC: Akaike Information Criterion
- ATL3: Atlantic 3 Index
- CHIKV: Chikungunya
- DENV: Dengue
- ENSO: El Niño Southern Oscillation
- GCV: Generalized cross-validation
- IOB: Indian Ocean Basin
- IOD: Indian Ocean Dipole
- LOESS: Locally estimated scatterplot smoothing
- NPMM: North Pacific Meridional Mode
- R^2 : Coefficient of determination
- SASD: South Atlantic Subtropical Dipole
- SIOD: Southern Indian Ocean Dipole
- SPMM: South Pacific Meridional Mode
- TNA: Tropical North Atlantic
- ZIKV: Zika

9 Acknowledgements

10 Funding

11 Availability of data and materials

Code for the generation of R_0 values employed for this study, computed using AeDES2's monitoring system, is available under request, and its values can be visualized in an operational, in-development Shiny App (link) for any region and grid-point. Additionally, the necessary datasets, functions and scripts to generate the maps and plots for this

manuscript and supplementary material are available under the following GitHub repository: https://github.com/jacorvillo/monitoring_system_analysis

12 Authors' contributions

Á.M., V.T. and D.C. conceived the methodology to be undertaken in this manuscript. Data sources, code and figures were obtained and developed from the ground up by J.C, who also analysed the results. All authors have reviewed the manuscript.

13 Ethics approval and consent to participate

Not applicable.

14 Consent for publication

Not applicable.

15 Competing interests

The authors declare no competing interests.

16 Author details

17 References