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 Campos**[2], **Ángel G. Muñoz**[2*], **and Ana Riviére-Cinnamond**[3]

[1]Complutense University of Madrid, Department of Earth Science and Astrophysics, Madrid, 28040, Spain
[2]Barcelona Supercomputing Center, Earth Sciences Department, 08034, Spain
[3]Pan-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

Vector-borne diseases, particularly those transmitted by *Aedes* mosquitoes such as dengue, Zika, and chikungunya, pose significant public health challenges worldwide. These diseases are highly sensitive to environmental conditions, with climate factors playing a crucial role in determining their transmission potential and geographic distribution. As global temperatures continue to rise due to climate change, understanding the complex relationship between climate patterns and disease transmission becomes increasingly critical for public health planning and intervention strategies. However, the transmission dynamics of these diseases, while known to be partly driven by climate, are not so well understood.

Using data from a number of sources, we explore and analyze the behavior of the climate-component of *Aedes*-borne disease transmission, in order to understand its role on the dynamics of disease outbreaks in the context of a changing climate. Our analysis is composed of three different studies: 1) a timescale decomposition of disease transmissibility values, thereby guiding officials to understand the climatology and behaviour of outbreaks for budget and resource allocation; 2) a correlation analysis between transmissibility values and different climate variability indices, 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.

1 Introduction

2 Methods

In order to understand the climatological behavior of vector-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 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 completely susceptible population. It includes the effects of the vector's biology (whether the vector is present in the area), the human behavior (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 metric is more so understood 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 therefore serve as a first approximation of the role of the climate in the dynamics of vector-borne diseases, and how it can be used to forecast disease outbreaks.

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

2.1 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.

2.1.1 Data

The timescale decomposition analysis was undertaken using R_0 outputs from the *Aedes* Disease Environmental Suitability 2's (AeDES2) 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.

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.

2.1.2 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 a locally estimated scatterplot smoothing technique (LOESS). In order to obtain the best smoothing parameter for the analysis, a search was performed over a range of values between 1 and 504 months for the fit of the spatial median of the R_0 observational data. The best smoothing parameter for said regression was selected using 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 GCV value. The GCV value is prioritized over the other two metrics, as it is a more robust measure of the goodness of fit of the model to the data, penalizing overfitting.

Once the ideal smoothing parameter is found for the R_0 data, the R_0 time-series for each gridpoint 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.

Once variance maps obtained, Strongest Seasonal Signal Regions (SSSRs) are identified as regions with a significant percentage of variance explained by the seasonal component of the data, to be used in the following parts of the analysis. The boundaries for the selection of SSSR regions are defined by the current Intergovernmental Panel on Climate Change set of reference regions for subcontinental analysis of climate model data.

After the variance maps and SSSR regions 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 form the AeDES2's monitoring system datasets as the independent variable (a multi-reference observational ensemble containing 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.

2.2 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.

2.2.1 Data

Correlation studies are performed over both global and SSSR regions, using the temperature-based detrended R_0 data as in the previous analysis over the different seasons and over the whole time period. 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.

2.2.2 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, a Monte Carlo method was used, with a p-value threshold of 0.05.

2.3 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, as opposed to correlation, 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 SSSR regions and over the different seasons and whole time period. 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.

2.3.1 Data

The datasets that were used for the causality analysis are the same detrended datasets as those employed in Section 2.2.1.

2.3.2 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.

3 Results

3.1 Analysis 1: Multi-timescale climate decomposition of R_0

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

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

4 Discussion

Discussion text goes here...

5 Conclusion

Conclusion text goes here...

Figure references

Author contributions statement

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.

Code and data availability statement

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

Competing interests

The authors declare no competing interests.