AeDES2's Monitoring System Analysis Paper

Javier Corvillo[1], Verónica Torralba[1], Ana-Riviére Cinnamond[2*]

[1]Barcelona Supercomputing Center
[2]Pan-American Health Organization

Abstract

This will be the abstract for the paper...

Introduction

Introduction text goes here...

Methods

Methodology text goes here...

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.

Data

The timescale decomposition analysis was undertaken using R_0 outputs from the *Aedes* Disease Environmental Suitability 2's (AeDES2) monitoring system, which is a global, next-generation operational dataset that detects historical outbreaks for *Aedes*-borne diseases using climate variables. AeDES2's monitoring system collects monthly-mean temperature and precipitation values from three different observational sources: NOAA's GHCN-CAMS, CPC Unified Global, and upscaled 0.5º resolution data from Era5Land's reanalysis. After transforming climate information into R_0 health information using four different ento-epidemiological models, outputs are then calibrated to real-life *Aedes*-borne data by using quantile mapping, before merging them together into a 12 member observational ensemble.

The 1980-2022 monthly-mean period was selected for the analysis. Considering that vector borne diseases are extending to previously unaffected areas due to the effects of man-made climate change, the selected R_0 contains global coverage, allowing for a comprehensive analysis of the relationship between climate variability indices and R_0 both in current *Aedes* hotspots and emerging regions.

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) with a 12 month frequency. This non-parametric regression method fits local polynomial regressions to the data, separating the R_0 time-series into three components: a long-term trend signal (understood to be the trend caused by anthropogenic climate change), an inter-annual signal (year to year), and a decadal signal (10-30 years).

Variance maps for each of these three 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 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 (Iturbide et al., 2020).

Time series, variance maps and AeDES2's R_0 data are freely available in an operational, in-development Shiny App (link) for any region and grid-point. On the other hand, AeDES2's code and data are available on demand.

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 1980-2022 period.

Data

Correlation studies are performed over both global and SSSR regions, using the same R_0 data as in the previous analysis. A total of 16 climate variability indices have been used for the correlation analysis. Their respective sources, as well as detrending methods for each, are listed and summarized in Table 1.

*Corresponding author: jane@smith.com
Received: October 20, 2023, **Published:** December 14, 2023

Index Name	Abbreviation	Periodicity	Pattern Type	Source	Detrending Method
Arctic Oscillation	AO	Several weeks to months	Atmospheric	NOAA's Climate Prediction Center (CPC)	Linear
Atlantic Multidecadal Oscillation	AMO	Between 60-80 years	Oceanic	NOAA's Kaplan Extended SST data	Linear
Atlantic 3 Index	ATL3	Several months to a few years	Oceanic	Detrended ORAS5 reanalysis data	-
Indian Ocean Basin	IOB	Several months to a few years	Oceanic	Detrended ORAS5 reanalysis data	-
Indian Ocean Dipole	IOD	Between 2-7 years	Oceanic	Detrended ORAS5 reanalysis data	-
North Atlantic Oscillation	NAO	Several days to decades	Atmospheric	NOAA's CPC	Linear
El Niño 3.4 Index	Niño 3.4	Between 2-7 years	Oceanic	NOAA's CPC	Linear
North Pacific Meridional Mode	NPMM	Several months to a few years	Atmospheric	Detrended ORAS5 reanalysis data	-
Pacific Decadal Oscillation	PDO	Between 20-30 years	Oceanic	NOAA's ERSSTv5	Linear
Pacific-North American Pattern	PNA	Several weeks to months	Atmospheric	NOAA's CPC	Linear
Quasi Biannual Oscillation	QBO	∼2 years	Atmospheric	NOAA's NCEP/NCAR Reanalysis	Linear
South Atlantic Subtropical Dipole 1	SASD1	Several months to a few years	Oceanic	Detrended ORAS5 reanalysis data	-
Southern Indian Ocean Dipole	SIOD	Several months to a few years	Oceanic	Detrended ORAS5 reanalysis data	-
Southern Oscillation Index	SOI	Between 2-7 years	Pressure-based	NOAA's CPC	Linear
South Pacific Meridional Mode	SPMM	Several months to a few years	Atmospheric	Detrended ORAS5 reanalysis data	-
Tropical North Atlantic	TNA	Several months to a few years	Oceanic	Detrended ORAS5 reanalysis data	-

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

Methodology

The correlation analysis was performed using the Pearson correlation coefficient, which quantifies the linear relationship between two variables. The Pearson correlation coefficient is defined as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

where x_i and y_i are the values of the two variables, $\bar{x}$ and $\bar{y}$ are their respective means, and n is the number of observations. The correlation coefficient ranges from -1 to 1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation. For computation of statistical significance in correlation, a Monte Carlo method was used, with a p-value threshold of 0.05.

In order to avoid spurious correlation outputs, anthropogenic signal from the natural variability of the data is isolated by using the long-term trend component obtained in the timescale decomposition analysis. For the climate variability indices, we apply the detrending methods listed in Table 1. The correlation analysis is performed over the different seasons over the 1980-2022 period.

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. 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 respect its predecessor.

Data

The datasets that were used for the causality analysis are the same as those employed in assessing the impact of climate variability indices on R_0 values across the globe (Section).

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 (Liang and Kleeman, 2005), quantifying the amount of information that one time series (the climate variability indices) can provide about another time series (R_0 patterns). This formalism is based on the concept of transfer entropy, which allows to compute causality as:

$$T_{2\to 1} = \frac{r}{1-r^2}(r'_{2,\partial 1} - r'_{1,\partial 1}) \tag{2}$$

where $T_{2\to 1}$ is the rate of entropy transfer from time series 2 to time series 1, r is the correlation coefficient between the two time series, and $r'_{2,\partial 1}$ and $r'_{1,\partial 1}$ are the partial correlation coefficients of time series 2 and 1 with respect to each other. While normalizing the transfer entropy can help to streamline the analysis, it is not advised for the purposes of this study, as higher correlation values in the denominator of the equation will naturally lead to very high values of transfer entropy that can influence in the normalization process.

Much like for the correlation analysis, the causality analysis is performed over the different seasons, regions, and time period, with the detrending methods listed in Table 1. For the causality analysis, a p-value threshold of 0.05 was used for statistical significance, which, following Liang-Kleeman's causality formalism, has been computed using Fisher's information matrix.

Results

Results text goes here...

Analysis 1: Multi-timescale climate decomposition of R_0

Analysis 2: Correlation studies between R_0 and climate variability indices

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

Discussion

Discussion text goes here...

Conclusion

Conclusion text goes here...

Code and Data Availability

References

Chen, Mingyue et al. (2008). "Assessing objective techniques for gauge-based analyses of global daily precipitation". en. In: *Journal of Geophysical Research: Atmospheres* 113.D4. ISSN: 2156-2202. DOI: 10.1029/2007JD009132. URL: https://onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009132 (visited on 11/26/2024).

Fan, Yun and Huug van den Dool (2008). "A global monthly land surface air temperature analysis for 1948–present". en. In: *Journal of Geophysical Research: Atmospheres* 113.D1. ISSN: 2156-2202. DOI: 10.1029/2007JD008470. URL: https://onlinelibrary.wiley.com/doi/abs/10.1029/2007JD008470 (visited on 11/26/2024).

Kelly-Hope, Louise and Madeleine C. Thomson (2008). "Climate and Infectious Diseases". en. In: *Seasonal Forecasts, Climatic Change and Human Health: Health and Climate*. Ed. by Madeleine C. Thomson, Ricardo Garcia-Herrera, and Martin Beniston. Dordrecht: Springer Netherlands, pp. 31–70. ISBN: 978-1-4020-6877-5. DOI: 10.1007/978-1-4020-6877-5_3. URL: https://doi.org/10.1007/978-1-4020-6877-5_3 (visited on 01/13/2025).

Muñoz-Sabater, Joaquín et al. (Sept. 2021). "ERA5-Land: a state-of-the-art global reanalysis dataset for land applications". English. In: *Earth System Science Data* 13.9. Publisher: Copernicus GmbH, pp. 4349–4383. ISSN: 1866-3508. DOI: 10.5194/essd-13-4349-2021. URL: https://essd.copernicus.org/articles/13/4349/2021/ (visited on 11/26/2024).

Xie, Pingping, Phillip A. Arkin, and John E. Janowiak (2007). "CMAP: The CPC Merged Analysis of Precipitation". en. In: *Measuring Precipitation From Space: EURAINSAT and the Future*. Ed. by Vincenzo Levizzani, Peter Bauer, and F. Joseph Turk. Dordrecht: Springer Netherlands, pp. 319–328. ISBN: 978-1-4020-5835-6. DOI: 10.1007/978-1-4020-5835-6_25. URL: https://doi.org/10.1007/978-1-4020-5835-6_25 (visited on 11/26/2024).