

RESEARCH ARTICLE

Malaria intensity in Colombia by regions and populations

Alejandro Feged-Rivadeneira^{1,2}*, Andrés Ángel^{3,4}, Felipe González-Casabianca³, Camilo Rivera⁵

1 Department of Anthropology, Stanford University, Stanford, CA, United States of America, **2** Department of Urban Management and Design, Universidad del Rosario, Bogotá, Colombia, **3** Department of Mathematics, Universidad de los Andes, Bogotá, Colombia, **4** Department of Mathematics and Statistics, Universidad del Norte, Barranquilla, Colombia, **5** Walmartlabs, Sunnyvale, CA, United States of America

* These authors contributed equally to this work.

* alejandro.feged@urosario.edu.co



OPEN ACCESS

Citation: Feged-Rivadeneira A, Ángel A, González-Casabianca F, Rivera C (2018) Malaria intensity in Colombia by regions and populations. PLoS ONE 13(9): e0203673. <https://doi.org/10.1371/journal.pone.0203673>

Editor: Luzia Helena Carvalho, Instituto Rene Rachou, BRAZIL

Received: June 26, 2018

Accepted: August 26, 2018

Published: September 12, 2018

Copyright: © 2018 Feged-Rivadeneira et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data set used in this study can be found on the repository <https://figshare.com/> (link: <https://figshare.com/s/f6d49359c4f536ed4488> DOI: 10.6084/m9.figshare.6863780, DOI: 10.6084/m9.figshare.6863780).

Funding: The funder Walmartlabs provided support in the form of salaries for author CR, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author

Abstract

Determining the distribution of disease prevalence among heterogeneous populations at the national scale is fundamental for epidemiology and public health. Here, we use a combination of methods (spatial scan statistic, topological data analysis and epidemic profile) to study measurable differences in malaria intensity by regions and populations of Colombia. This study explores three main questions: What are the regions of Colombia where malaria is epidemic? What are the regions and populations in Colombia where malaria is endemic? What associations exist between epidemic outbreaks between regions in Colombia? *Plasmodium falciparum* is most prevalent in the Pacific Coast, some regions of the Amazon Basin, and some regions of the Magdalena Basin. *Plasmodium vivax* is the most prevalent parasite in Colombia, particularly in the Northern Amazon Basin, the Caribbean, and municipalities of Sucre, Antioquia and Cordoba. We find an acute peak of malarial infection at 25 years of age. Indigenous and Afrocolombian populations experience endemic malaria (with household transmission). We find that *Plasmodium vivax* decreased in the most important hotspots, often with moderate urbanization rate, and was re-introduced to locations with moderate but sustained deforestation. Infection by *Plasmodium falciparum*, on the other hand, steadily increased in incidence in locations where it was introduced in the 2009-2010 generalized epidemic. Our findings suggest that Colombia is entering an unstable transmission state, where rapid decreases in one location of the country are interconnected with rapid increases in other parts of the country.

1 Introduction

Malaria in Colombia has been studied from a variety of disciplines that describe disease patterns with dimensions such as the diversity of the vector [1, 2], characteristics of the parasite [3], social phenomena affecting disease transmission [4, 5], and physical phenomena such as climate, weather and land use, [6–8], just to name some of the most important. Mainly, national and local contexts are well understood for a country that presents unusual diversity of

contributions' section. The first author would like to thank Stanford University, the Zaffaroni family, and the Morrison Institute for their financial support. The second author acknowledges and thanks the financial support of the grant P12.160422.004/01-FAPA ANDRES ANGEL from Vicedecanatura de Investigaciones de la Facultad de Ciencias de la Universidad de los Andes, Colombia. The third author acknowledges and thanks the financial support of the grant Proyecto Semilla 2017-1 from Vicedecanatura de Investigaciones de la Facultad de Ciencias de la Universidad de los Andes, Colombia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The funder Walmartlabs provided support in the form of salaries for the author Camilo Rivera, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

environments and social backgrounds (including vast cultural diversity), which, in turn, represent different characteristics of malaria transmission. In contrast with Sub-Saharan Africa, where malaria is commonly a deadly disease affecting primarily children, Colombia is not considered particularly relevant in malarial disease studies given the relatively low mortality when compared with Sub Saharan Africa. However, malaria in Colombia presents certain characteristics that resemble those observed in Southeast Asia. Colombia was one of the first countries where resistance to chloroquine-based treatment was reported. Varied malaria intensity among segregated and diverse populations inhabiting different and unique environments make Colombia one of the few cases where malaria is endemic and where disease patterns are regionally inconsistent, in contrast to several countries that follow a consistent pattern of infection, or whose segregated vulnerable populations do not differ in their epidemic patterns [9, 10]. This does not mean that other countries have a homogeneous experience of malaria intensity across subpopulations or regions.

Malaria is a complex disease, and factors associated to disease severity and resistance have been reported, yet genetic resistance to malaria is better understood than to any other human infectious disease [11]. However, the strong geographical association between resistance to the pathogen and disease severity remains a major challenge to assess the causality of human genetic resistance [11]. We know from evolutionary theory that two critical factors for selection must occur: 1) a population with genetic diversity has to exist for selection to operate; 2) the trait in question must confer a differential in reproductive value for adaptation to evolve. Because African populations exhibit both high genetic diversity and experience severe malaria, genetic resistance to the pathogen appears to have emerged independently in different foci [12]. However, unlike Africa, Colombia has no record of *de novo* emergence of human genetic resistance to malaria. On the contrary, the parasite appears to have developed resistance to treatment. Until recently, it was unknown whether observed pathogen resistance was the result of selection of adaptive mutant strains under drug pressure, or the spread of resistant strains [13]. Genetic evidence suggests that resistant malaria emerged in at least 4 different geographical foci, consistent with the history of reports of resistant pathogens for *Plasmodium falciparum* in the Thailand-Cambodia border and Colombia in the 1950s, then spreading for two decades to South America, Asia and India, and then to Africa in Kenya and Tanzania in the late 1970s [14–16]. Resistant *Plasmodium vivax* was first reported in Papua-New Guinea in 1989, it is currently present in South East Asia, and South America [16].

Studies have found resistant *Plasmodium vivax* at a rate of 11% in representative samples of all blood smears collected in two endemic geographical regions in Colombia: *Llanos orientales* (Eastern Plains) and Urabá [17], while others have found no evidence of resistant *Plasmodium vivax* forms in the Pacific Coast and the Amazon Basin [18]. However, therapeutic failure rates of *Plasmodium falciparum* (for representative samples of all blood smears collected) have been reported as high as 78% for these same regions [18], and 67% in Antioquia [19]. More recent assessments of malaria prevalence in endemic areas also suggest that uncomplicated malaria due to low parasitemia is one of the biggest challenges for malaria control strategies [5, 20, 21], and studies indicate that the observed differences are not attributable to human genetic traits that confer resistance [22].

Few studies have addressed the epidemiology of malarial infection by regions and populations to explore the role of intensity in the emergence of resistant forms of the parasite. However, the role of Colombia in the global epidemiological context of malaria indicates that the country may present unique characteristics for disease transmission. Mainly, the presence, absence, and most importantly, emergence of resistant forms of the parasite in different regions suggests that isolated and distinct epidemic regions exist within the national boundaries, and such characteristics may play a distinctive role in the evolution of the parasite. Here

we address the intensity of malaria by regions and human populations in Colombia, and the degree that the epidemic characteristics between regions affect each other.

One key aspect remains poorly understood about malaria dynamics in Colombia: how many different epidemic regions exist, and how do subpopulations in these regions experience malaria. During our research, we interacted with local health officials who conducted malaria prevention programs at both local and national levels. Each public health official had knowledge and expertise about epidemic dynamics in their specific territorial assignment. However, a lack of systematic approaches hamper the ability to formalize such knowledge. Malaria intensity and the social aspects that condition the transmission of the parasite drive public health interventions. However, the regional designations are yet to be formalized based on analysis of malaria dynamics. Decisions about prevention strategies, and how to target the most vulnerable populations are made based primarily on the expertise of local health officials, as indicated by the different authors [6, 23, 24].

Here we generate a systematic classification of the malarious regions and subpopulations of Colombia, to characterize locations and subpopulations with epidemiological aspects of the parasite. Here we address three basic questions surrounding the intensity of malarial infection by ethnicity and region:

1. Is this population experiencing higher intensity of malarial infection than other regions of the country?
2. Is the parasite persisting endemically within this population?
3. Are the epidemic characteristics of this subpopulation affecting other subpopulations?

Specifically, we analyze eight years of malarial case reports, they are examined for both malaria intensity, synchrony and segregation by ethnicity. First, we employ an outbreak detection algorithm [25] widely used [26–28] to identify clusters in space with outbreaks of malaria. Second, we apply methods from Topological Data Analysis (TDA) [29, 30] to visualize synchronous outbreaks: areas that present similar time patterns of malarial epidemics. Finally, regional case reports are explored with descriptive statistics to analyze the intensity of malaria exposure by ethnicity.

2 Background

From John Snow's seminal study of cholera in London, epidemiology has been a spatial discipline [31]. Geographical disease patterns have been widely described for numerous pathogens and regions. We use three methods to analyze malaria in Colombia: disease clustering, disease visualization, and ecological analysis.

2.1 Importance of detecting endemicity

The production of good quality maps to understand and visualize risk of disease transmission is recognized as one of the fundamental tools for malaria control strategies, specifically, understanding the relationship between malaria endemicity and the health impact of malaria [32]. Studies suggest that annual entomological inoculation rates (commonly computed as the product of the daily human biting rate, the sporozoite rates from the caught mosquitoes, and the days per year [33]) in Ghana (100–1000), Kenya (10–60) and Gambia (less than 10) are associated to prevention of all cause childhood mortality rates by insecticide treated bed nets, with efficacy of 17%, 33%, and 63%, respectively [32]. These results suggest that public health policies should vary according to malaria endemicity, since bed nets have been the linchpin of malaria prevention strategies since DDT was discontinued as a viable alternative. Yet, evidence

suggests that there are several contexts in which bed nets are not efficient [32]. In locations where malaria is intense, the use of bed nets is less efficient to prevent the burden of the disease.

Due to the scarcity of multi-sited studies across different countries, variation of the relationship between endemicity and overall health remains unknown [32]. However, within country variation of malaria has been subject of numerous studies. One study that addresses such relationship is produced by Omumbo et al. [34], using GIS and malaria case reports to map malaria intensity in Kenya. Their results also question the use of treated bed nets in regions where malaria is intense, because in these communities, bednets are the most inefficient [34].

2.2 Analyzing malaria from a spatial point of view

Spatial descriptions of variation in malarial infection within countries has been addressed producing maps of risk of contracting the disease using a variety of methods. For example, [35] produced a more accurate visualization of risk of contracting malaria in Mali, by combining regression analysis with “kriging” (i.e., an interpolation method similar to smoothing fitted values) to account for local responses to environmental conditions such as weather, population and other topographic and sociological features. Using those methods, they are able to identify regions where the risk is higher than represented in traditional maps [35]. A similar approach, but based on entomological and demographic geo-coded records, is implemented with a GIS analysis to describe local risk of infection based upon proximity to breeding sites and human populations [36]. Beck [37] implemented a variation of these risk maps by integrating remote sensing data to identify locations of high transmission based on human-vector interaction for a region in Mexico, including variation by season.

Estimating the effect of migration on pathogen loads has been a growing interest of epidemiologists in the past years, and multiple methods have been implemented to address such interaction. A data-driven approach to examine the effect of human mobility on epidemics has been the gravity model, used to evaluate measles outbreaks, both by age-classes and by urban and rural settings [38–40]. The main finding of this approach was that population densities were the main driver of outbreak seasonality across different environments [38–40]. Furthermore, the same group has used nighttime light imagery to estimate the effect of changing patterns of population densities on disease outbreaks [41]. Unfortunately, few comparative studies exist to determine which method is more effective under which conditions and for which diseases. However, the method of nighttime light imagery provides good estimates of mobility of populations without access to phone services, often the most vulnerable populations in terms of disease prevalence.

Although the gravity model has been mostly used for directly transmitted diseases, understanding the effect of human mobility on disease epidemics, and more generally how disease disperses over space and time, has been one of the fundamental questions in contemporary spatial epidemiology.

2.3 Malaria in the Americas

Contributing 42% of the cases of malaria in the region, Brazil is one of the most important players of the disease in the Americas [8]. Contrary to the spectrum of infection in Colombia, most cases of malaria in Brazil are both Amazonian and caused by *Plasmodium vivax*, with challenges such as asymptomatic and submicroscopic infections, emergence of drug resistance, and social and environmental factors driving risk of disease over a consistently decreasing area of land [8, 42]. Brazil stands both as an example in terms of innovative and effective measures for malaria control, on the one hand, and of policies that promoted malarial infection

epidemics on the other [8, 42, 43]. A recent geospatial assessment in the Brazilian Amazon found that malarial infection is more frequently observed in areas with greater forest cover and close to gold mining, and the observed spatial heterogeneity of disease risk suggests that the Acre region should be prioritized by public health interventions, highlighting the relevance of spatial analysis to inform policy [44].

2.4 Synchrony in the disease outbreaks

Two approaches have been used to analyze synchrony of disease outbreaks over space and time, controlling for seasonal and environmental variables. Spectral analysis has been used to describe the association between aggravation of asthma symptoms and temperature or atmospheric contamination levels [45, 46], and the association between air pollution and mortality [46]. This technique has been used to study the effect of climatic variation on cholera [47, 48], malarial epidemics [49], and the seasonality of sexually transmitted diseases (STDs) [50].

Some authors suggest that these methods have limitations, because they can only be used for time-series data in which statistical proprieties do not change over time, yet, epidemic data are inherently complex and non-stationary [46]. Furthermore, evidence suggests that epidemic data characteristics do change over time [46, 51, 52].

The limitations of the spectral decomposition methods led to the implementation of the second technique that is most widespread in understanding disease dynamics over space and time, coupled with climatic and environmental conditions from a non-stationary perspective: wavelets, a method used to show how time-series vary as a function of time and space [46, 53].

Wavelet analysis has been used to study geographical hierarchies of measles epidemics, and the observed effect of vaccination policies over time [54]. Associations between dengue epidemics and El Niño Southern Oscillation (ENSO) have also been documented using this method [55]. Kreppel [56] found an association between ENSO, Indian Ocean Dipole (IOD) and plague dynamics in Madagascar, and [57] found similar effect of those two climatic phenomena on infectious gastroenteritis in Japan. [58] documented that Buruli ulcer is affected by short and long rainfall patterns in French Guiana, as well as stochastic events such as ENSO. The relationship between ENSO and cutaneous leishmaniasis has also been documented for Costa Rica [59]. Jose [60] have studied the changing patterns and seasonality of Australian rotavirus epidemics comparing a multiplicity of methods including wavelet analysis, and detected seasonal biannual and quinquennial periods, yet, a three year epidemic period was also found to be dominant. Spectral analysis confirms that serotype harmonics interact in a complex, non-linear fashion, yielding an observable overall pattern beyond the isolated dynamics of each separate serotype, that is more than the sum of the parts, and inherent dynamics remain unchanged but the amplitude of disease infection is modified [60].

2.5 Disease cluster identification

Spatial analysis methods have been applied in disease cluster identification. The main approaches used are: K-cluster analysis, detects global clusters based on each case point [61]; the geographical machine [62] and the scan statistic [25], work by aggregating cases in different areas and performing a hypothesis test based on a Bernoulli null model, with the advantageous difference for the scan statistic that it can perform multiple tests simultaneously. We present an implementation of the scan statistic in this case study. Small, isolated outbreaks of malaria in specific communities have been identified as “discrete mini epidemics”, which represent disease severity by using space-time cluster identification [63]. Disease risk by geographical location has also been implemented as simple logistic regressions that include altitude, and physical coordinates of each individual within a case-control study [64]. The scan

statistic method has been used by [65] to identify disease clusters over space and time in a South African region. [66] have also implemented the scan statistic method in China to identify clusters and suggest public health resource optimization. Faires [67] used this method to identify clusters of *Clostridium difficile* over time in Ontario, Canada. Duczma [68] implemented the scan statistic to study Chagas' disease in Brazil, while [69] do the same for end-stage renal disease (ESRD) in northern France. In Virginia, the increasing burden of Lyme disease was documented using spatiotemporal scan statistics [70]. In all cases, studies were able to identify areas with more cases than expected, highlighting in many cases the relevance of regions that did not present a comparatively higher incidence.

Globally, [71] have used maximum likelihood methods (i.e. a similar approach to the scan statistic) to predict areas where malaria is likely to expand as a result of climate change.

A branch that can also be used for cluster identification is Topological Data Analysis (TDA), which applies ideas from the mathematics area of topology to study high dimensional data by obtaining invariants and useful representations of the shape of the data.

In recent years, TDA has been used to find subgroups of individuals with unique genetic and prognostic profiles of breast cancer [30, 72], political alliances in the congress [30], profiles of basketball players [30], pathogen persistence in soil [73], novel patterns in spinal and brain injury [74], subgroups of individuals with different complications from type 2 diabetes [75], among others. We use TDA to find synchronic outbreak clusters among the data.

3 Methods

The analysis for this study was generated from case reports based on active and passive detection methods compiled by the Colombian government. Reports of malaria cases are mandatory. Treatment is provided for free to each case, and a positive test is required to disburse the medication. Data was accessed by requesting a user to query the Sispro database of reported cases by the Ministry of Health.

All cases are laboratory confirmed and geocoded to municipality level. We included data for 1,122 municipalities, that range in area from 1 to 65,786 km²; total area sampled was 1'144,385 km². For each municipality, we also analyzed sex and ethnic membership of each case, comprising 3,369 different populations. The ethnic denominations used by the Ministry of Health are:

1. Indigenous.
2. Romani.
3. Raizal.
4. Palenquero.
5. Afro, which includes: black, mulatto, african-colombian and afrodescendant.
6. Other, which consists of predominantly mestizos.

Since the combined total cases with ethnic denomination Romani, Raizal or Palenquero are less than the 0.3% of the database, they are excluded from our analysis.

The distribution of malaria cases and segmentation of the data summarized [Table 1](#). See also the section of additional figures.

3.1 Clustering

The two main objectives are to determine if malaria outbreaks exist in Colombia, and, if so, to determine their location. To address these objectives, we apply scan statistics to perform a

Table 1. Summary and segmentation of registered malaria cases in Colombia between the years 2007 and 2015. Although distribution of sexes among the Colombian population is almost equal, registered malaria cases are predominantly men. *Plasmodium vivax* has a similar ethnic distribution to the Colombian population and as for *Plasmodium falciparum*, the predominant cases are of afro denomination. [76]

Attribute		Total Cases	Percentage (%)
Sex	Male	228075	63.33
	Female	132075	36.67
Ethnicity	Afro	110333	30.63
	Indigenous	38199	10.60
	Other	211618	58.75
Type of Malaria	<i>Plasmodium falciparum</i>	115260	32.00
	<i>Plasmodium vivax</i>	244890	67.99
Total		360150	

<https://doi.org/10.1371/journal.pone.0203673.t001>

hypothesis test in each municipality, examining whether it presents an outbreak. These approaches have been widely used in epidemiological studies [25], [77], and [78].

The approach to test hypothesis is mainly based on the Bernoulli model of the Kulldorff Scan Statistic developed in [25] using the r-package Spatial-Epi [79].

Given the data aggregated by municipality for 2007-2015, each record is assigned to the centroid of the municipality. Because the set of possible outbreaks (all possible aggregations of neighboring municipalities) is almost unlimited in terms of shape and size, a solution is to approximate this set. In this case, a grid G of N by N was overlaid onto Colombia's jurisdictional boundaries and then the set of possible outbreaks is limited to all the possible sub rectangles within the grid.

Now, under the Bernoulli model we consider a measurement m for each rectangle R contained in G , where $m(R)$ corresponds to an integer and in our specific case, the number of individuals inside the given rectangle. This leads us to assume that there is a rectangle Z contained in G such that each individual inside Z has a probability p of being infected, while the individuals outside Z have a probability q . Let n_R be the number of observed malaria cases inside R , so by assuming a Bernoulli model and the following hypothesis for our unknown variables p and q :

$$H_0 : p = q \tag{1}$$

$$H_1 : p > q \tag{2}$$

we have these possible distributions:

- Assuming H_0 :

$$n_R \sim Bin(m(R), p) \quad \forall R \subseteq G \tag{3}$$

- Assuming H_1 :

$$n_R \sim Bin(m(R), p) \quad \forall R \subseteq Z \quad \text{and} \quad n_R \sim Bin(m(R), q) \quad \forall R \subseteq Z^C \tag{4}$$

And hence, under H_1 , we have that Z is a region with potential malaria outbreak.

Lastly, the third and final step is to establish a measure of density for each subrectangle, in this case the likelihood ratio. This measurement of density has desirable properties to compare different sized rectangles [78]. So, [25] derives the formula for likelihood ratio of a generic

region. The scan statistic (ss) is defined as the highest density measurements for all subrectangles:

$$ss^* = \max_R ss(R) \tag{5}$$

$$ss(R) = p^{n_R}(1 - p)^{m(R) - n_R} q^{n_G - n_R}(1 - q)^{(m(G) - m(R)) - (n_G - n_R)} \tag{6}$$

The local measurement $ss(R)$ can be interpreted as the likelihood that subrectangle R is an outbreak.

To test the hypothesis represented in Eq 1 a Monte Carlo simulation was used to obtain the histogram of the statistic ss^* under the null hypothesis. Finally, it assesses the value of ss^* with the observed data. If p is not greater than 0.05 under the null model, H_0 is rejected and we assume an outbreak.

3.2 Synchronous epidemic visualization

The main objectives are to determine whether abnormal behaviors are related across municipalities and finding, if a relationship exists, groups of them that have a similar temporal patterns, independent of their geographical layout. To address these matters we turn to Topological Data Analysis (as mentioned in section 2.5).

We apply the method *Mapper*, introduced in [29], that constructs a representation of the data in the form a graph. This graph allows both the analysis and visualization of the data. The vertices of the graph correspond to local clusters and the interactions between these clusters are encoded on the edges of the graph.

Mapper detects phenomena that appear both at large and small scale better than other methods, such as principal component analysis (PCA) and cluster algorithms. *Mapper* can be considered a hybrid method that is doing partial clustering, the regions where the clustering is done is guided by the filter. It is a refinement of clustering and scatterplot methods like PCA.

The input of the method *Mapper* is a collection of points with a notion of similarity and a filter, a function defined on the collection of points. The filter is used to define pieces that cover the collection of points. We apply a clustering algorithm to each piece to obtain a set of local clusters. These are the vertices of the graph. Edges are added to the graph in the following way: two local clusters are connected if they have points in common. Since each vertex and edge correspond to subcollection of points, we can consider the graph to have weights. Figs 1–3 give a visual road map through the *Mapper* algorithm applied to a set sampled from the unitary circle.

In this study, we implement TDA on disease data in Colombia to find topological characteristics that describe spatiotemporal patterns for both *Plasmodium vivax* and *falciparum* separately. We constructed the three *mapper* algorithm’s inputs (collection of points, similarity notion and filter function) as follows, for each incidence type independently:

- **Collection of Points:** For each one of the 1,122 municipalities we calculated an epidemic occurrence vector consisting of binary values for each week from 2007 to 2015. The idea is that each binary value indicates if there was a malaria epidemic on that municipality for the given week. To construct each weekly entry, we executed a Kulldorf Scan Statistic procedure (as explained in the previous section) among all municipalities, but only with data from the given week. So a municipality k will have 1 on a certain entry if in the corresponding week it suffered a malaria epidemic (i.e the obtained p value from the Kulldorf procedure is not greater than 0.15) or 0 otherwise. We discard municipalities that never showed an outbreak, leaving us with a total of 449 for *Plasmodium vivax* and 269 for *Plasmodium falciparum*.

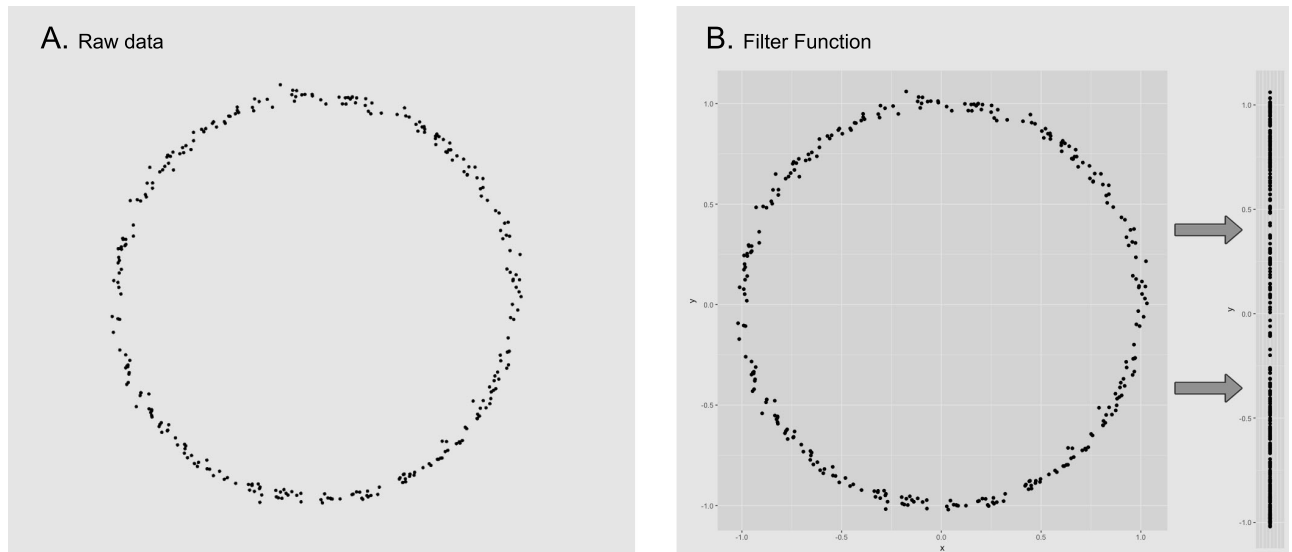


Fig 1. We start with a given data set (image A), for this example the points correspond to a sample of the unitary circle with a small amount of noise. For convenience we will use the euclidean distance to calculate the distance between each pair of points. In the next step, we select the projection onto the Y coordinate as our filter function and apply it to the data set (image B).

<https://doi.org/10.1371/journal.pone.0203673.g001>

- Similarity notion:** Since our collection of points is now a set of binary vectors of dimension 477 (the number of weeks between 2007 and 2015) that represent outbreaks, we computed the similarity between records using the Wasserstein distance. In short, this distance measures the amount of effort required to transform a certain distribution X into another Y. So, if we think of the binary vectors as distributions, this metric captures our intuitive similarity notion among the epidemic records.

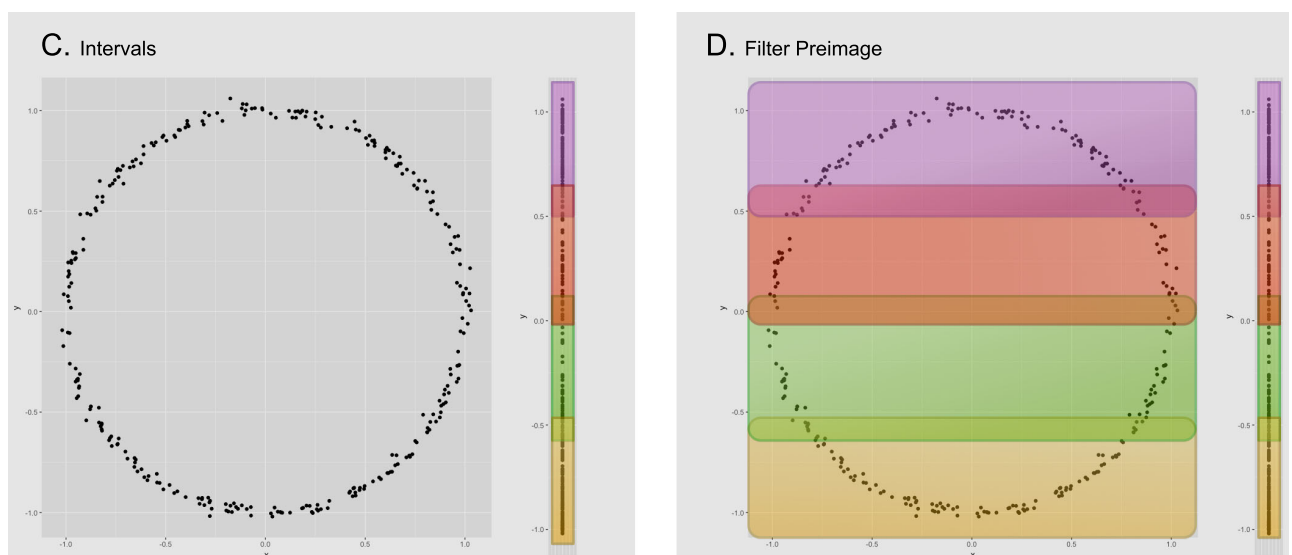


Fig 2. We now divide the image of the data set (under the filter function) into evenly distributed overlapping intervals (image C) and compute the corresponding points in their pre-image (image D). Notice how each pair of overlapping intervals, defines two different subsets of data that can have elements in common.

<https://doi.org/10.1371/journal.pone.0203673.g002>

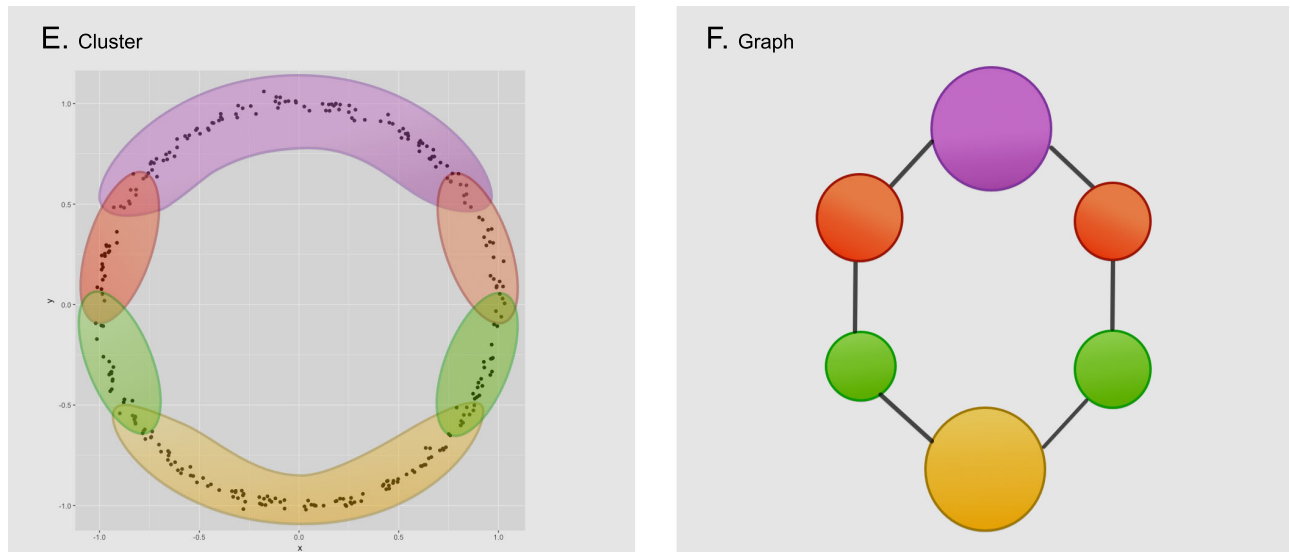


Fig 3. Inside every defined subset of data, we execute a clustering algorithm to detect isolated groups of points (image E). Each of the resulting groups will correspond to a node on the output graph (image F). Notice how nodes are joined together by edges when their corresponding groups have points of the data set in common. Also, the size of the node in the cluster corresponds to the amount of points in its corresponding group.

<https://doi.org/10.1371/journal.pone.0203673.g003>

Nevertheless, there is a small shortcoming of using only the Wasserstein metric. When comparing two binary vectors, the metric assumes that they distribute the same amount of mass along all the dimensions, regardless of the amount of positive entries. This means that a binary vector with a single positive entry holds a total m mass in that entry, whereas a binary vector with n positive entries will hold m/n mass in each entry. This is a problem because intuitively a binary vector with a single positive entry is very different from a vector with numerous ones, and the mentioned metric does not capture this fact. We solved this problem by adding a factor of the difference of total positive entries among the vectors. Thus, if X and Y are binary vectors, the similarity notion used in our experiments is:

$$d(X, Y) = w(X, Y) + \frac{|\sum X - \sum Y|}{\eta} \tag{7}$$

where w is the Wasserstein metric and

$$\eta = \max_{X, Y} \sum X - \sum Y$$

- **Filter Function:** We used a dimension reduction technique called t-SNE [80]. This technique focuses on preserving the probability that points are chosen as neighbors (meaning that they are similar to one another under some notion), while sending them into an euclidean space of lower dimension. For our particular case, we used t-SNE to represent our binary vectors in \mathbb{R}^3 , scattered to preserve the Wasserstein metric.

The output of the *Mapper* algorithm is a graph where we select significant subgraphs by disease intensity and number of epidemic weeks. To further visualize the municipalities that appear in the selected subgraphs, we make another graph where the nodes are now the

municipalities that appear in nodes of the *Mapper* output and draw connections between those municipalities to visualize them over the Colombian territory.

3.3 Ethnicity

Case reports, collected by the national surveillance system and confirmed by laboratory, include components of age and ethnicity in the notification form. This filled form is required by law for every case reported in the country, and treatment to cure the disease is provided by the government for each case. Ethnicity and age are self-reported, and should be interpreted with caution (e.g. no genetic resistance can be inferred, for example). As mentioned before, the three main ethnic groups that were used in this paper are: *Indigenous*, *Afro* and *Other*.

Data for age and ethnicity were displayed with descriptive statistics to develop patterns of malaria intensity by population for the clusters identified with TDA analysis, and to identify two patterns of intensity. First, the occupational hazard risk profile is characterized by a peak within a particular age-group and contains a pronounced sex difference [42, 81]. Second, an endemic risk profile is characterized by intense malaria exposure at young ages and reduced malaria at later ages, due to overexposure [42, 81, 82].

3.4 Social and environmental change

To examine whether epidemic characteristics are associated with land use patterns and development, we visualized deforestation and expansion of the anthropogenic transformation print. To discuss the association between the patterns of malarial infection and deforestation, we plot deforestation alerts published by the government office SIAC [83] for the years available (2013 and 2014). Anthropogenic change, measured using nighttime lights imagery, was visualized using Google Earth Engine to calculate the difference (using map algebra, the difference between two rasters) in stable lights between 1999 and 2013 using an average of 5 years for each pixel for each layer. Mining activities were plotted using the data published by the United Nations Office on Drugs and Crime [84].

4 Results

4.1 Clustering

All significant outbreaks are highlighted in Fig 4. When considering all parasites, the method detects significant outbreaks of malarial infection along the Pacific Coast, the Magdalena river Basin, and the Amazon river Basin. For *Plasmodium falciparum*, significant clusters were observed in municipalities along the Pacific Coast, the border with Panama, and Northern Antioquia (the tertiary Cauca river Basin). The municipalities: Policarpa and Cumbitirá in Nariño, El Retén in Magdalena and Calima in Valle del Cauca appear to be a hidden cluster for this parasite, since they were not marked as epidemic when considering all malarial parasites. And as for *Plasmodium vivax*, significant clusters were observed in municipalities of departments: Cordoba, Vichada and Antioquia. There are also hidden clusters, that appear when only considering this parasite, including: Orito in Putumayo, Piamonte in Cauca and Achi in Bolivar.

4.2 Synchronous epidemic visualization

TDA enabled us to find at least 11 groups of municipalities with similar temporal behaviors (6 for *Plasmodium falciparum* and 5 for *Plasmodium vivax*) Figs 5 and 6. These groups were selected by high overall disease intensity, defined as the number of cases with infants (age

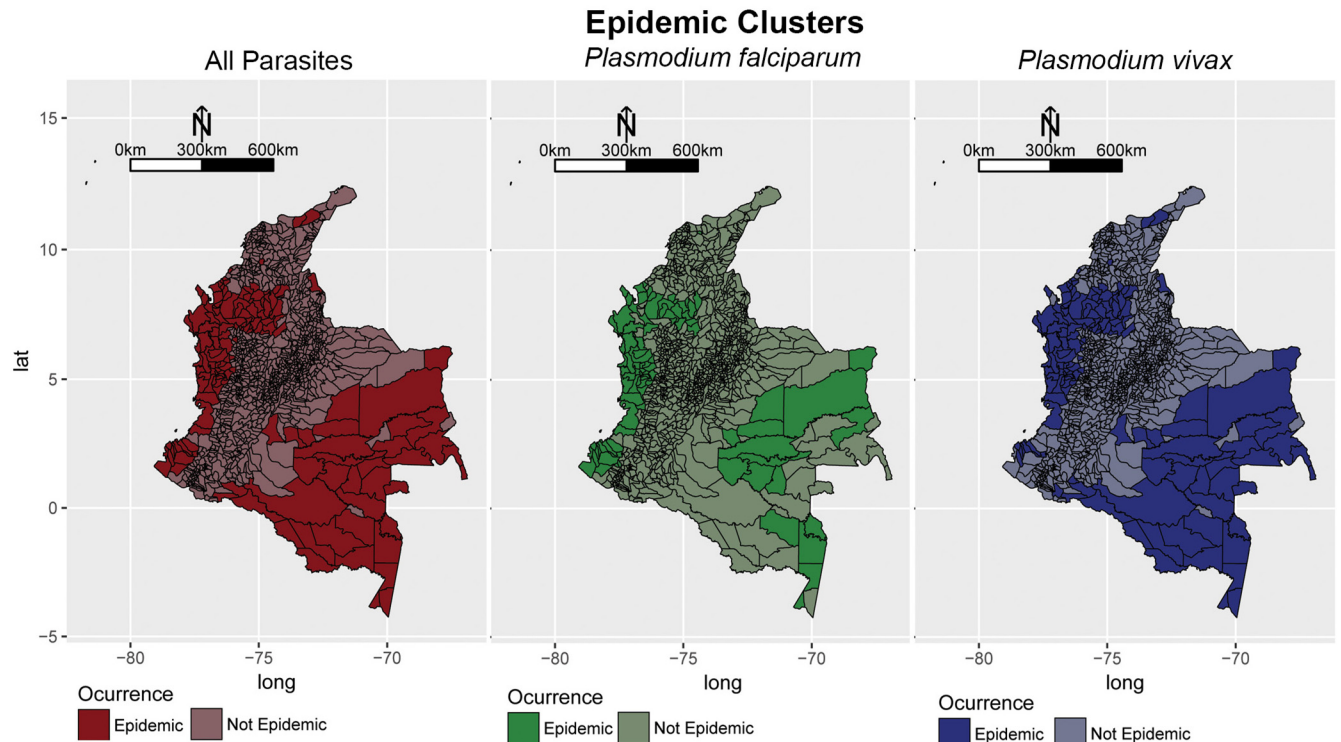


Fig 4. Significant outbreaks of malaria in Colombia from 2007-2015, calculated using the scan statistic developed by [25] based on a likelihood ratio. The significance threshold parameter was calculated using a Bernoulli model where cases were simulated for each municipality, and taking the maximum value. The process was iterated many times and the distribution of the maximum values was calculated to determine the 95% confidence interval.

<https://doi.org/10.1371/journal.pone.0203673.g004>

below 5 years) over the total number of cases, or high epidemic rate, understood as the average amount of epidemic weeks among the municipalities in the group.

In the TDA graphs each node represents a group of municipalities. The size of each node will be proportional to the number of municipalities in the group, and two nodes will have an arch between them if they have at least one municipality in common.

Now, it is possible to visualize the municipalities appearing in the TDA graph geographically, to get a notion of the part of Colombia the corresponding cluster represents, as seen in Fig 7.

Each of the colored small dots in the maps corresponds to a municipality, contained in some node of the corresponding colored cluster. Notice there are arrows between some points in the map, they represent connections between the municipalities (these connections are not the same as the edges between nodes in the graph). Before we mention how these connections are constructed, let us explain what centrality means for a municipality:

Given a certain municipality, its centrality corresponds to the number of nodes in the graph in which it appears. This means that if we remove a municipality with high centrality, it is very possible that the resulting graph will have less arches and in turn be disconnected.

Now, the connection scheme is as follows: Only municipalities that appear together in several nodes in the TDA graph can have a connection. No municipality will be connected to itself. Municipalities will be connected towards the municipalities in its node with the highest centrality. Note that there could be municipalities with multiple outgoing connections.

TDA Graph and t-SNE Components (*Plasmodium falciparum*)

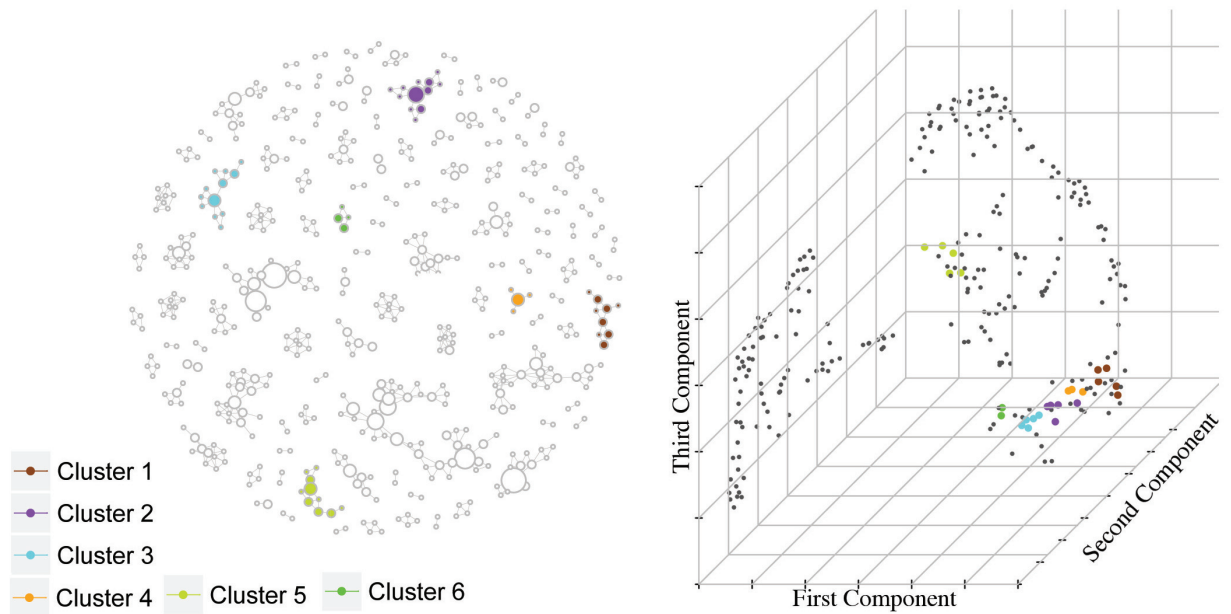


Fig 5. Graph constructed using TDA and t-SNE component plot using the epidemic occurrence vectors, where selected groups have been highlighted. These groups were selected by high overall disease intensity and high epidemic rate. Each cluster can be interpreted as a group of municipalities with *Plasmodium falciparum* incidence that have similar temporal behavior. Notice how the colored dots in the component plot are somewhat grouped together and since these represent municipalities with high epidemic rate, we have highlighted locations with several positive entries in the occurrence vector distributing differently across time.

<https://doi.org/10.1371/journal.pone.0203673.g005>

TDA Graph and t-SNE Components (*Plasmodium vivax*)

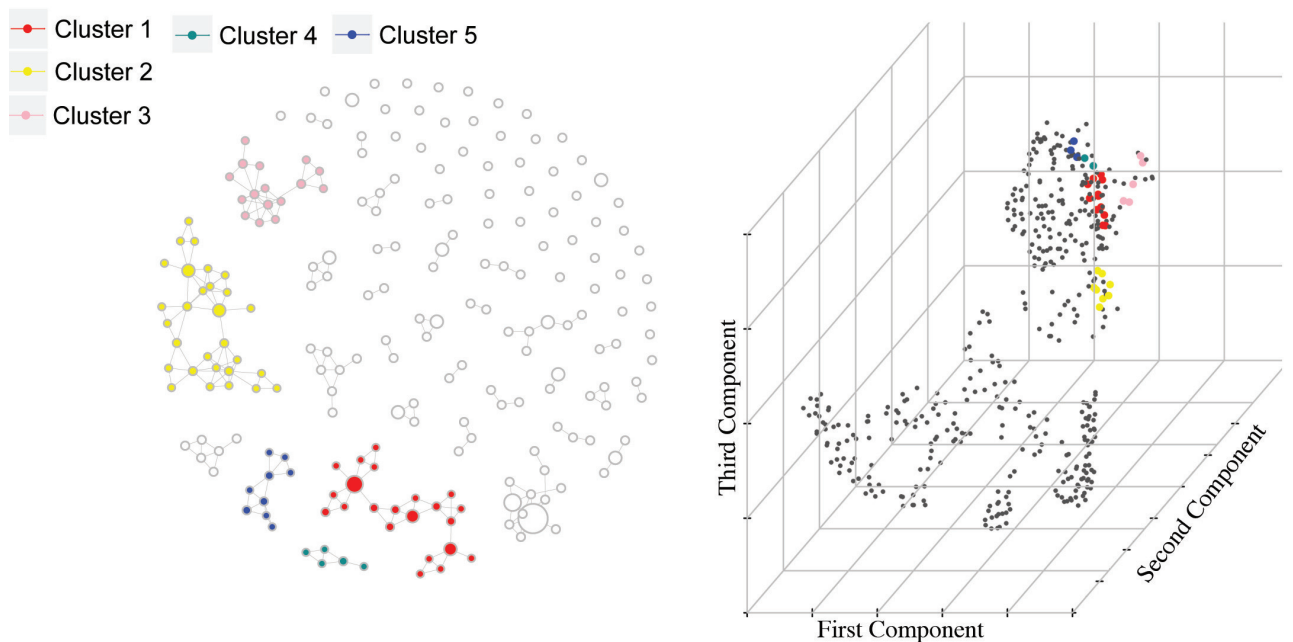


Fig 6. Similar to Fig 5, highlighted groups where selected by high overall disease intensity and high epidemic rate for *Plasmodium vivax*.

<https://doi.org/10.1371/journal.pone.0203673.g006>

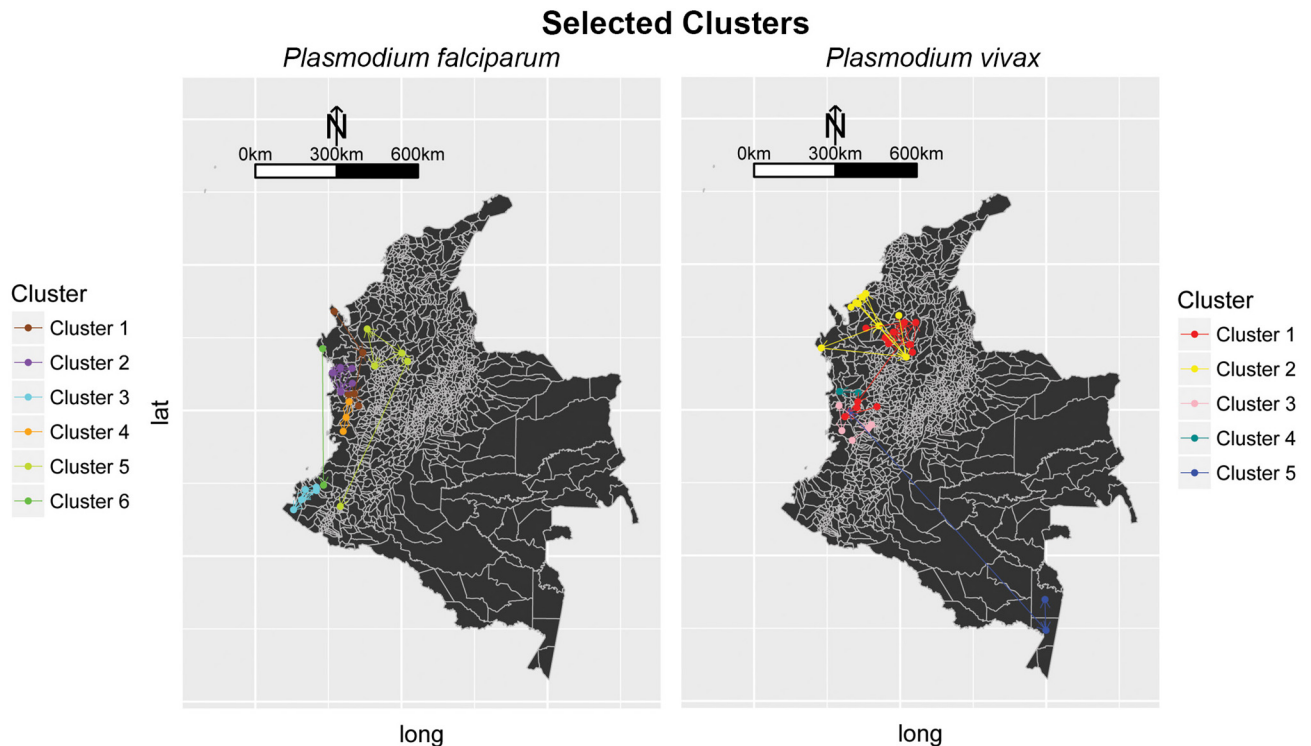


Fig 7. Selected municipalities by TDA over the Colombian territory for both parasites. As expected, the clusters follow some geographic pattern, since the time series were constructed using a Kullback procedure that detects clusters geographically. For *Plasmodium falciparum* all clusters are concentrated near the Pacific coast and northern Antioquia. Unexpected results happen in cluster 5 for *Plasmodium vivax*, where the grouped municipalities belong to two different geographic regions of the country. This means that the municipalities in this cluster from Chocó and Amazonas have similar time patterns, regardless of their geographical distance.

<https://doi.org/10.1371/journal.pone.0203673.g007>

We also identified significant municipalities with high centrality in the TDA graphs (Figs 5 and 6) and are reported in Tables 2 and 3. These municipalities are responsible for the connection among several nodes in their corresponding subgraphs appearing in overlapping zones of the selected TDA filter.

4.3 Ethnicity

Fig 8 shows the histograms of age reports of malaria by ethnicity and cluster. Two distinctive patterns consistently appearing throughout different regions of Colombia. First, an endemic profile risk (higher density of cases at young ages) was observed for the indigenous populations of clusters 3, 4 and 5 for *Plasmodium vivax*. Second, malarial infection suggesting occupational hazard and intensity of infection among the Afrocolombian populations of clusters 1, 2 and 3 for *Plasmodium falciparum*. Occupational hazard is consistently described for the population with no ethnic denomination across all clusters except cluster 4 for both parasites respectively. In all cases, within the same cluster, we find both endemic and occupational hazard infection patterns across populations segregated by ethnicity. In a histogram of case reports by age and sex, an occupational risk hazard has a unique and characteristic signature: one age class, typically for only one sex, presents an outstanding number of cases compared to any other age class. In the case of malaria in Colombia, we observed that men with no ethnic denomination of ages 20-25 were contracting malaria far more often than any other class. From this simple

Table 2. *Plasmodium falciparum*: Selected central municipalities after executing TDA over the epidemic occurrence vectors. These are the municipalities responsible for the connectivity among their respective groups and subgraphs. Quibdó, El Cantón Del San Pablo, Istmina and Roberto Payán appear among the top 10 municipalities with highest epidemic rate and Alto Baudó, Dabeiba and Atrato among the top 10 municipalities with highest disease intensity.

Cluster	Department	Municipality	Rural Pop.	Urban Pop.	Total Pop.
1	Antioquia	Dabeiba	15480	24084	39564
1	Chocó	Atrato	5073	2488	7561
2	Chocó	Quibdó	11752	101134	112886
2	Chocó	Alto Baudó	22739	6222	28961
2	Chocó	Bahia Solano	4864	4230	9094
2	Chocó	Bojayá	5369	4572	9941
3	Nariño	La Tola	2752	5656	8408
3	Nariño	Roberto Payán	16029	863	16892
3	Nariño	San Andres De Tumaco	75366	84668	160034
4	Chocó	El Cantón Del San Pablo	3722	2491	6213
4	Chocó	El Litoral Del San Juan	11186	1058	12244
4	Chocó	Istmina	5319	18320	23639
5	Antioquia	Amalfi	9589	10936	20525
5	Antioquia	Santafe De Antioquia	9267	13636	22903
5	Cordoba	Tierralta	45895	32875	78770
6	Cauca	Guapi	12390	16273	28663

<https://doi.org/10.1371/journal.pone.0203673.t002>

observation, we inferred the following: first, men of this age class were engaging in activities that posed a risk of contracting the disease. Second, women were not engaging in this activity, nor were men in other age classes. Third, there was no household transmission once they ceased to engage in the risky activity.

Most interestingly, our analysis identifies, without explicitly addressing it, locations where the proportion of cases of *Plasmodium falciparum* to *Plasmodium vivax* changed over the past few years Figs 9 and 10. This is one of the open questions in tropical malaria, and our implementation has findings: most of the reduction in *Plasmodium vivax* cases was seen in clusters 1 and 2, while in clusters 3-5 this type of malaria increased recently. Most of the increase in

Table 3. *Plasmodium vivax*: Selected central municipalities after executing TDA over the epidemic occurrence vectors. These are the municipalities responsible for the connectivity among their respective groups and subgraphs. Cáceres, Nechí and Tadó appear among the top 10 municipalities with highest epidemic rate and Alto Baudó and Medio San Juan among the top 10 municipalities with highest disease intensity.

Cluster	Department	Municipality	Rural Pop.	Urban Pop.	Total Pop.
1	Antioquia	Cáceres	22736	6209	28945
1	Antioquia	Nechí	10228	10440	20668
1	Antioquia	Valdivia	12442	4848	17290
1	Bolívar	Montecristo	9893	7080	16973
1	Chocó	Istmina	5319	18320	23639
1	Chocó	Tadó	6795	11246	18041
2	Antioquia	Vegachi	4824	6469	11293
2	Córdoba	Puerto Escondido	18252	3534	21786
3	Valle del Cauca	Zarzal	12184	28799	40983
4	Chocó	Alto Baudó	22739	6222	28961
5	Chocó	Medio San Juan	8794	4233	13027
5	Amazonas	La Pedrera	3711	0	3711
5	Amazonas	Tarapacá	3775	0	3775

<https://doi.org/10.1371/journal.pone.0203673.t003>

Malaria by species, age, sex, ethnicity and clusters

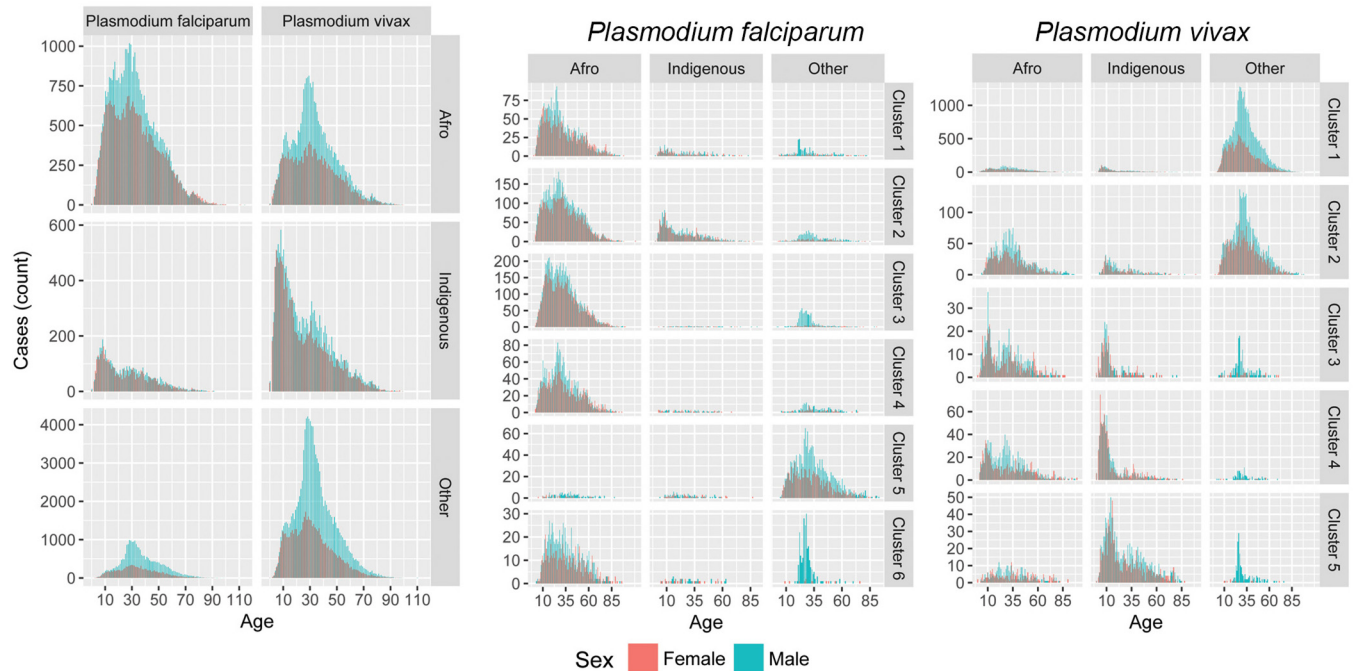


Fig 8. Malaria by parasite species, age, sex, ethnicity and cluster groups of human cases in Colombia. For all parasites, the indigenous ethnic group shows a pattern of endemicity, with most cases being reported for the youngest ages, while people with no ethnic denomination and the Afrocolombian population show a pattern consistent with occupational hazard risk. For *Plasmodium falciparum*, the indigenous and Afrocolombian populations in clusters 1, 2, 3 and 4 suggest that these populations experience intense exposure to malarial infection, with the Afrocolombian population showing occupational hazard transmission. Histograms for the population with no ethnic denomination in all clusters except 4 suggest malarial infection is associated with occupational hazard. And for *Plasmodium vivax*, the indigenous population in clusters 3, 4, and 5 suggest intense exposure to malarial infection among these populations. The population with no ethnic denomination experiences malarial infection as an occupational hazard in all clusters, except 4.

<https://doi.org/10.1371/journal.pone.0203673.g008>

Plasmodium falciparum happened in clusters 1-4, in the Pacific Coast (although not explicitly on the coastal areas, but rather in the lower-altitude Andes on the Pacific). Our findings also suggest that the change in ratio of *Plasmodium falciparum* to *Plasmodium vivax* observed at the whole country is comprised of different phenomena, and not just one single change in epidemic characteristics. First, infection by *Plasmodium falciparum* increased in locations where it was not previously present or not persistent (clusters 1-4), while at the same time receding in locations where it was previously a bigger burden (cluster 5, in the lower Cauca Basin). We also find evidence that infection by *Plasmodium vivax* also experienced different epidemic changes across Colombia. First, it receded in locations where it was consistently present (clusters 1-2, lower Cauca Basin and the Caribbean), and increased in locations that did not report cases before the 2009 epidemic (clusters 3-5, Pacific Coast and Amazon Basin). Thus, malarial infection epidemiology in Colombia over the past years has seen the rate of *Plasmodium falciparum* change, we suggest the following stages:

1. 2008-2012: An outbreak of both parasites that peaked in 2010, where *Plasmodium vivax* became the most prevalent parasite, particularly in the lower Cauca Basin and the Caribbean.
2. 2013-2014: A reduction in the total cases of malarial infection in the whole country, with acute reductions of *Plasmodium vivax* in the Caribbean and the lower Cauca Basin, and an

Malaria totals by weeks in Colombia (by clusters of *Plasmodium falciparum*)

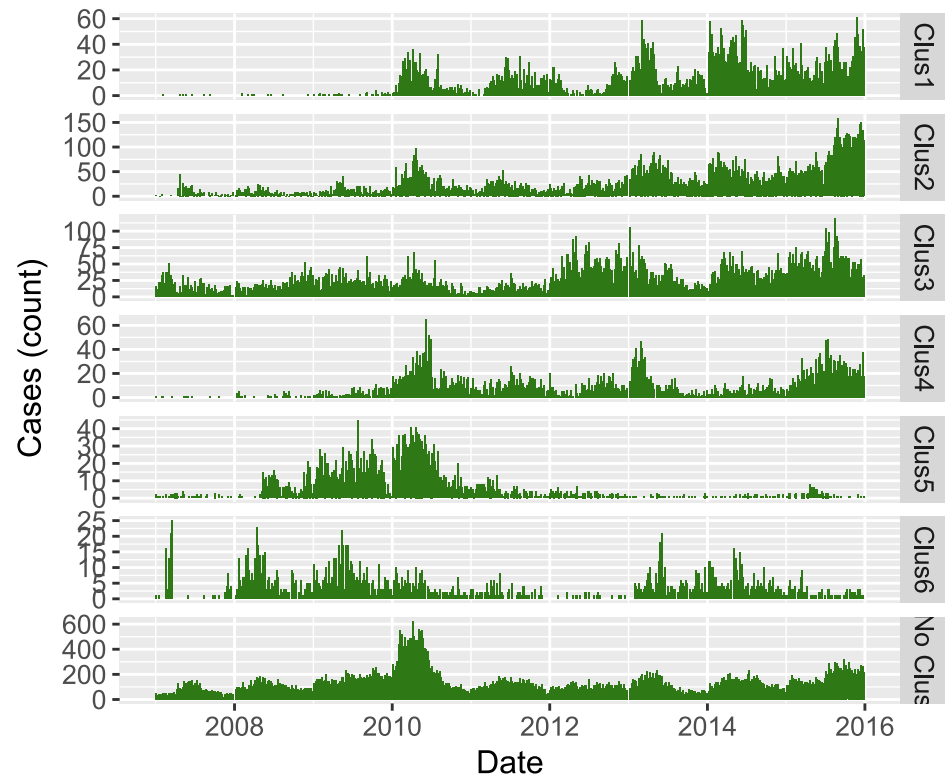


Fig 9. Total cases of *Plasmodium falciparum* by weeks, between the years 2007 and 2015.

<https://doi.org/10.1371/journal.pone.0203673.g009>

increase of *Plasmodium falciparum* cases in the corridor that connects Acandí and Capurganá with the central Pacific region.

- 2015-2016: The re-emergence of malarial infection by *Plasmodium vivax* in the central Pacific region and the Amazon.

In other words, the generalized outbreak of malarial infection in 2010 enabled *Plasmodium vivax* to persist in the lower Cauca Basin and the Caribbean, and also introduced the same parasite at lower prevalences in the central Pacific and the Amazon. When the outbreak of *Plasmodium vivax* infection was controlled where it was a bigger burden (lower Cauca Basin and Caribbean), it re-emerged in locations that had not experienced intense malarial infection since the 2010 epidemic (Amazon Basin and central Pacific).

5 Further exploration of the results

5.1 Deforestation and anthropogenic change

All the clusters identified for *Plasmodium falciparum* are located along the Pacific or the lower Cauca Basin, where the human occupation pattern is observed to change little, in comparison

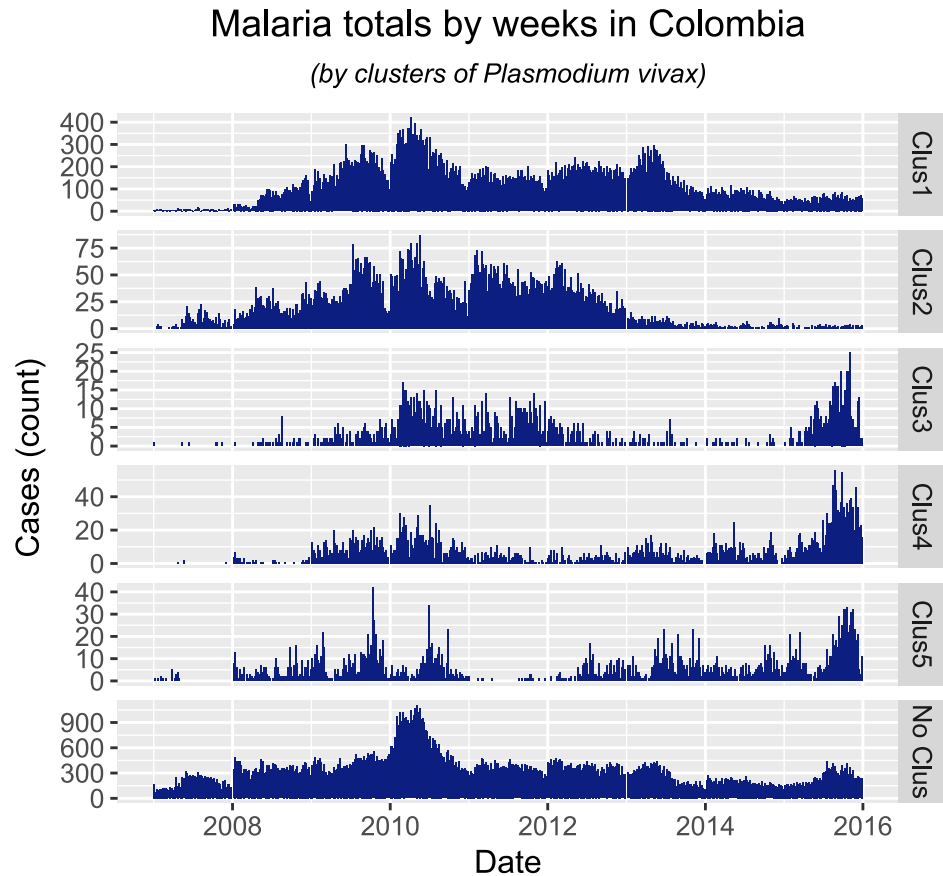


Fig 10. Total cases of *Plasmodium vivax* by weeks, between the years 2007 and 2015.

<https://doi.org/10.1371/journal.pone.0203673.g010>

to other parts of the country. We find evidence of a different pattern for *Plasmodium vivax*, where clusters 1, 2, and 3 hold both municipalities with low or no change in anthropogenic occupation, but also municipalities with rapid change (Fig 11).

Deforestation, on the other hand, happened at moderate and medium rates along the Pacific (Fig 12), where clusters of both parasites were observed. Interestingly, the pattern of deforestation along the Pacific is not observed to be the most intense in Colombia, but it is widely spread at lower rates. This suggests that habitat transformation does not need to be intense to harbor malaria transmission, and that a wide-spread moderate deforestation favors the transmission of malaria, perhaps to a greater extent, than rapid deforestation in a single location.

5.2 Gold exploitation

Our findings suggest a high correlation between gold exploitation and malaria occurrence. Specifically, clusters 2, 3 and 4 for *Plasmodium falciparum* and 1 and 4 for *Plasmodium vivax*, are located around medium and high gold exploitation areas in the country (Fig 13). Furthermore, all the central municipalities highlighted in Tables 3 and 2 (except for La Pedrera and Tarapacá in Amazonas) are located in the departments: Antioquia, Bolívar, Chocó, Córdoba,

Deforestation alerts in Colombia, 2013 – 2014

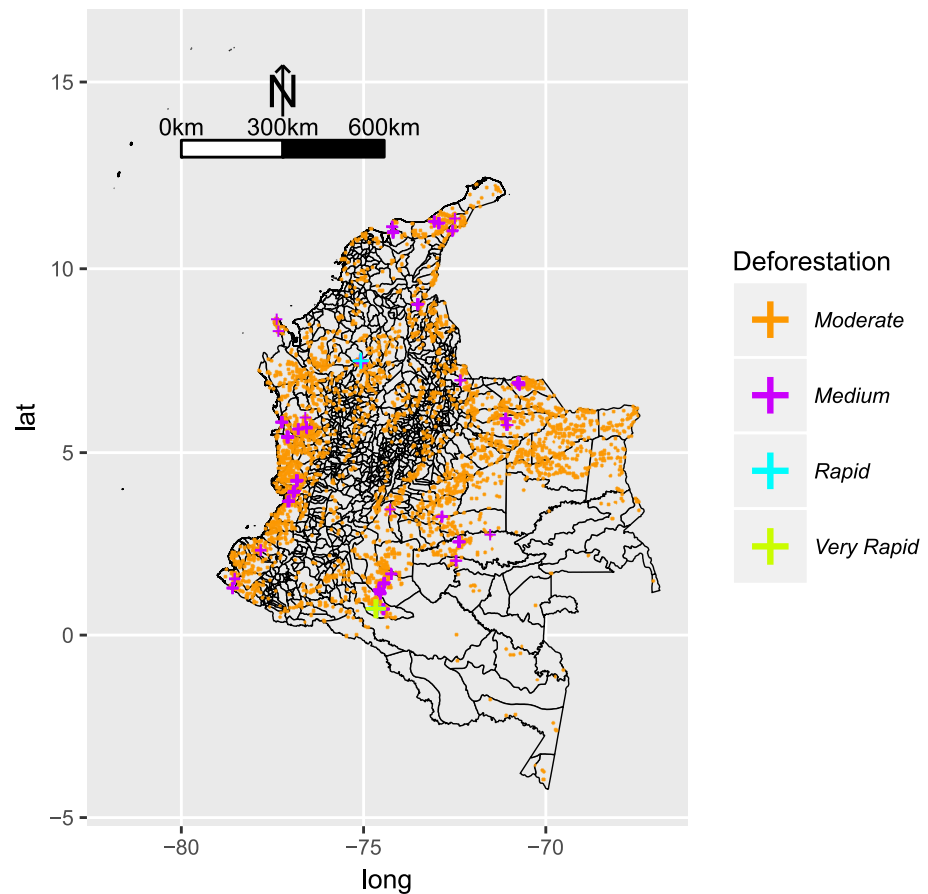


Fig 11. Anthropogenic change in Colombia, 1999-2013, using the nighttime lights dataset NOAA-DMSP-OLS. A mean for 5-year periods was computed for each pixel, and then map algebra was used to calculate the difference between the two periods. Very Rapid anthropogenic change was observed in the region of Bogotá and the Eastern Plains. Rapid change was observed in proximity of the main urban areas along the Andes (Bogotá, Cali, Medellín, and the Coffee Region, the urban areas of the Caribbean, and the Eastern Plains. Moderate and Medium anthropogenic change was observed throughout the Andes and the Caribbean, and the lower Cauca Basin.

<https://doi.org/10.1371/journal.pone.0203673.g011>

Valle del Cauca, Cauca and Nariño, which are precisely where gold-mining is most prevalent [85].

6 Discussion and conclusions

Malaria in Colombia was characterized by a different intensity, connectivity and segregation in each region. While there was a general pattern of risk throughout the country associated with occupational hazard, some populations experienced intense malaria exposure in endemic pockets. Understanding the interaction of such pockets is fundamental for designing appropriate malarial control strategies. Here we have produced a systematic approach that analyzes malaria under three dimensions: Clustering, Synchrony and Ethnicity.

Our findings suggest that controlling the epidemic in the lower Cauca Basin and the Caribbean had significant effects throughout the country. The reduction of *Plasmodium vivax* was followed by an increase of the same parasite, in connected clusters, in the Amazon Basin and

Anthropogenic Change in Colombia, 1999 – 2013

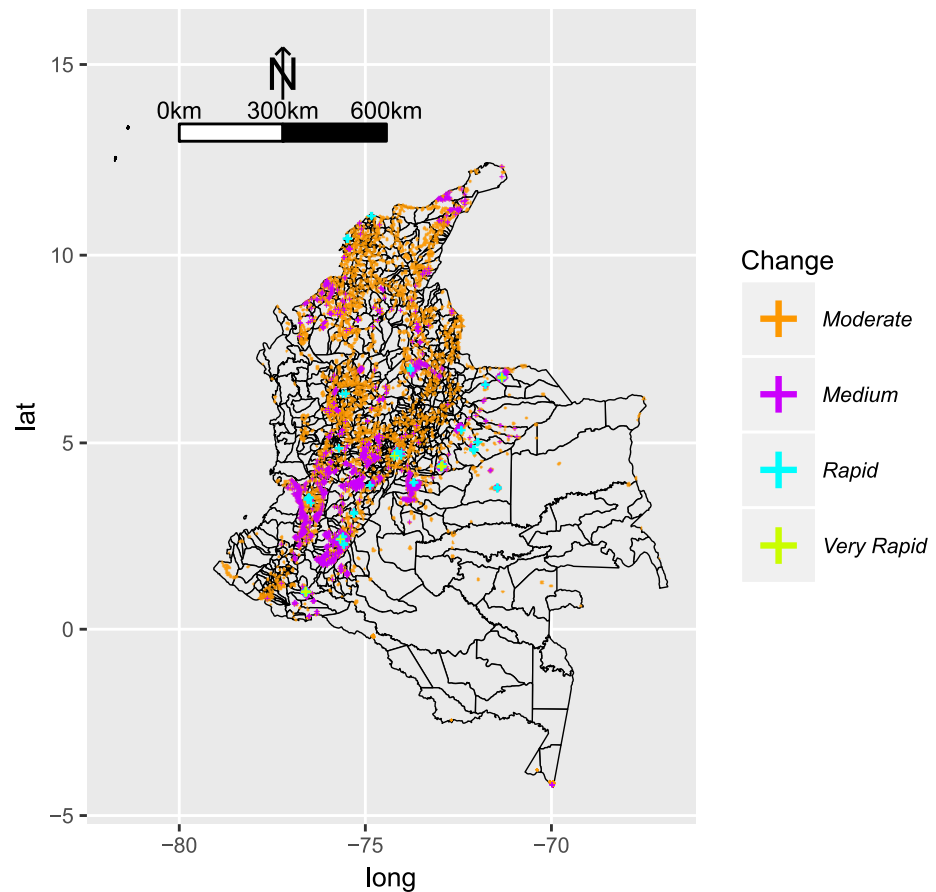


Fig 12. Deforestation alerts in Colombia for years 2013-14, as published by SIAC [83]. Moderate and medium rates of deforestation were observed during the study period along the Pacific Coast, and throughout other parts of the country but more scattered. Rapid deforestation rates were observed in the lower Cauca Basin. Very Rapid deforestation rates were observed in Caquetá, the North Eastern Region of the Amazon Basin.

<https://doi.org/10.1371/journal.pone.0203673.g012>

the Central Pacific region. The increase of *Plasmodium vivax* was observed in locations that also experienced sustained and moderate deforestation throughout the territory. We find that epidemic and endemic characteristics coexist in infection by *Plasmodium vivax* and *Plasmodium falciparum*. While we did not study the effect of one population on the other, further studies can explore if the reduction of cases in the Caribbean and lower Cauca Basin were first experienced in the epidemic (“other”) or endemic (“indigenous”) populations. Similarly, it would be interesting to explore if the increase in *Plasmodium vivax* and *Plasmodium falciparum* of the Pacific region was first experienced by the indigenous, Afrocolombian, or population with no ethnic denomination. It is possible that the endemic indigenous population has an asymptomatic parasite carriage which, as previous studies suggest [42], might contribute strongly to malaria transmission and progression in the mentioned areas. Further studies, including genetic samples, will confirm if these populations in fact show asymptomatic behavior.

We also find that habitat transformation intensity harbors or reduces intensity of malarial infection. In line with reports from other locations in Latin America [43], we find that

Gold Exploitation Density in Colombia, 2016

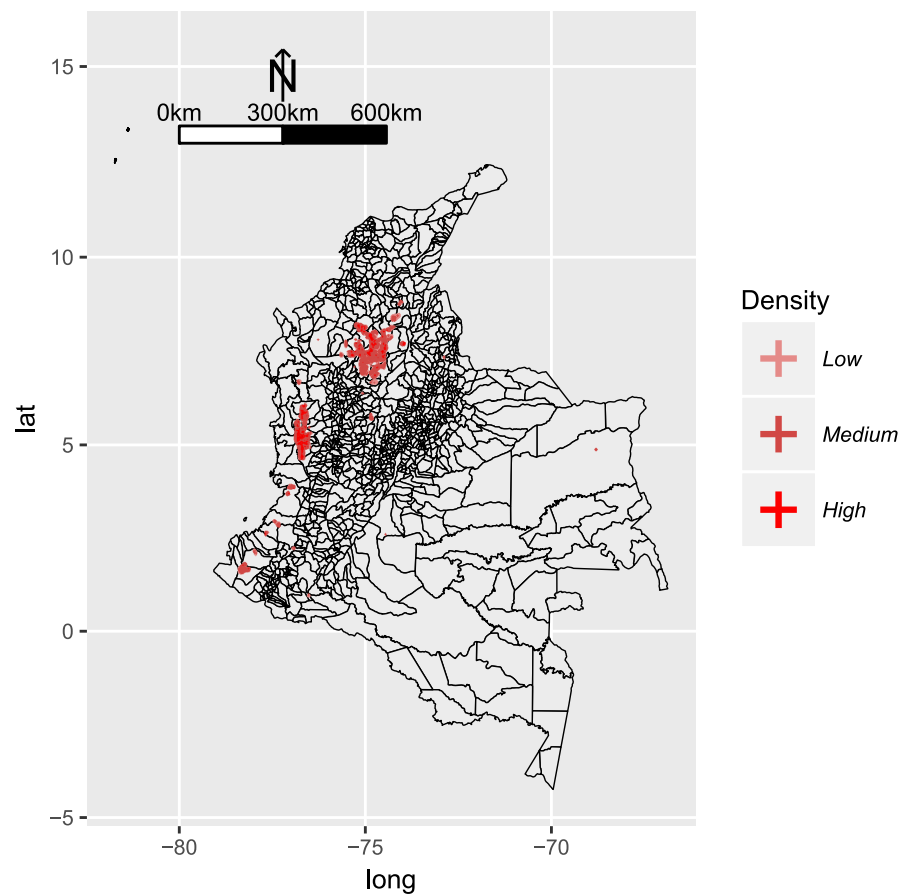


Fig 13. Density for evidence of alluvial gold exploitation in Colombia in 2016, as published by [84]. Three categories are included: Low (less 1 habitant per square kilometer), Medium (between 1.1 and 5 habitants per square kilometer) and High (more than 5 habitants per square kilometer). Evidence of intense mining activities was observed in the lower Cauca and Magdalena Basins, and the the Central Pacific region. Scattered mining activities were observed in the Southern Pacific, some parts of the Eastern Plains.

<https://doi.org/10.1371/journal.pone.0203673.g013>

moderate deforestation fosters malarial infection of both parasites, and our findings suggest that urbanization was associated with some of the most important reductions of *Plasmodium vivax* infection in the Caribbean. Thus, malaria control strategies should pay particular attention to habitat transformation in the sense that slow and steady habitat transformation of large territories may be conducive to increases in the burden of malarial infection, while a moderate process of urbanization in hotspots, can have drastic effects not only within the region, but it may also have both positive and negative effects elsewhere. These strategies could also contemplate gold exploitation as an incidence factor, since both our findings and several studies show a correlation between gold-mining areas and malaria. [85] [44]

Our findings also suggest that the epidemic changes of malarial infection in Colombia are complex, comprised of a multiplicity of interconnected trends by parasite and region. This is consistent with the unstable transmission pattern exhibited in countries with rapid reductions of malaria. We observe a 5 year period between peaks of transmission for *Plasmodium vivax*, while epidemics of *Plasmodium falciparum* have been less intense, but more frequent (3 years). Control strategies should not only aim to adapt and mitigate to current burdens of malarial

infection, but also be prepared for the re-introduction of malaria in locations where it was previously not endemic or epidemic. Such increases are to be expected in areas with sustained and moderate deforestation and without dense human occupation.

Our findings have potential implications for malarial infection control. First, we found that malaria in Colombia did present different, isolated pockets with distinctive epidemic characteristics. These hotspots play an important role in malaria transmission and when targeted, constitute an efficient way to reduce malaria intensity [86]. Furthermore, the magnitude of such differences in epidemic characteristics is relevant in studying the pressure of anti-malarials upon the parasite, since the emergence of resistance has been reported in the country. We found that *Plasmodium falciparum* was particularly acute among the Afrocolombian population of the Pacific region, while in the Cauca Basin, it constituted an isolated outbreak of the population with no ethnic denomination, while in the Pacific, the outbreak was dispersed among both the Afrocolombian and the indigenous populations.

Different parasite loads among ethnically and culturally distinct populations constitute the quintessential mechanism of selective pressures that are ideal for the evolution of parasites. The diversity of epidemic characteristics of malarial infection among the subpopulations of Colombia account for an ideal environment for parasite evolution, where plasmodia persist under different pressures of asymptomatic individuals, susceptible classes of ethnically distinct populations, and public health interventions using different anti-malarial strategies. Such diversity provides the necessary conditions, acting as isolated experiments, and then sharing “successful” results, for the emergence of resistant parasites.

Second, the patterns of endemicity observed in these populations suggested that prevention efforts should be population specific, and vary according to the epidemic characteristics exhibited by the parasite in the targeted population. Demographic, social and population factors confound studies limited to environmental factors and are also excluded from many studies [87] and thus could hold important insight on malaria control. We are currently experiencing an epidemic transition in malarial infection in Colombia, where parasite loads in populations are being transformed (potentially as populations are also subject of change). To understand such transitions, a systematic method to classify malarial infection in terms of vulnerability of populations over space, and the variation of those epidemic characteristics over time is crucial to implement public health strategies.

Therapeutic failures have been suggested to be correlated with high intestinal parasite loads [19]. The effectiveness of bed nets has been reported to be low among populations that experience intense malaria exposure [32]. We have identified populations that experienced malaria endemicity, where prevention efforts focused on the distribution of bed nets. Our findings, combined with previous knowledge suggest that public health interventions should integrate two aspects: 1) Diagnostic and treatment of asymptomatic malaria; and 2) Prevention strategies for occupational hazard.

Third, prevention strategies focusing on populations with endemic malaria would yield a reduction of occupational hazard malaria, since both types of epidemic characteristics exist in the same clusters of malaria, and interact in ways that determine their spatio-temporal epidemic characteristics.

6.1 Limitations

Our study was constrained by data availability at the spatial and temporal resolutions we observed for the epidemiological data. We are aware that it would have been ideal to include variables that describe vector ecology, distribution and behavior, pathogen genetic information, climatic variation during the study period, and other sociodemographic information.

Malaria Incidence in Colombia

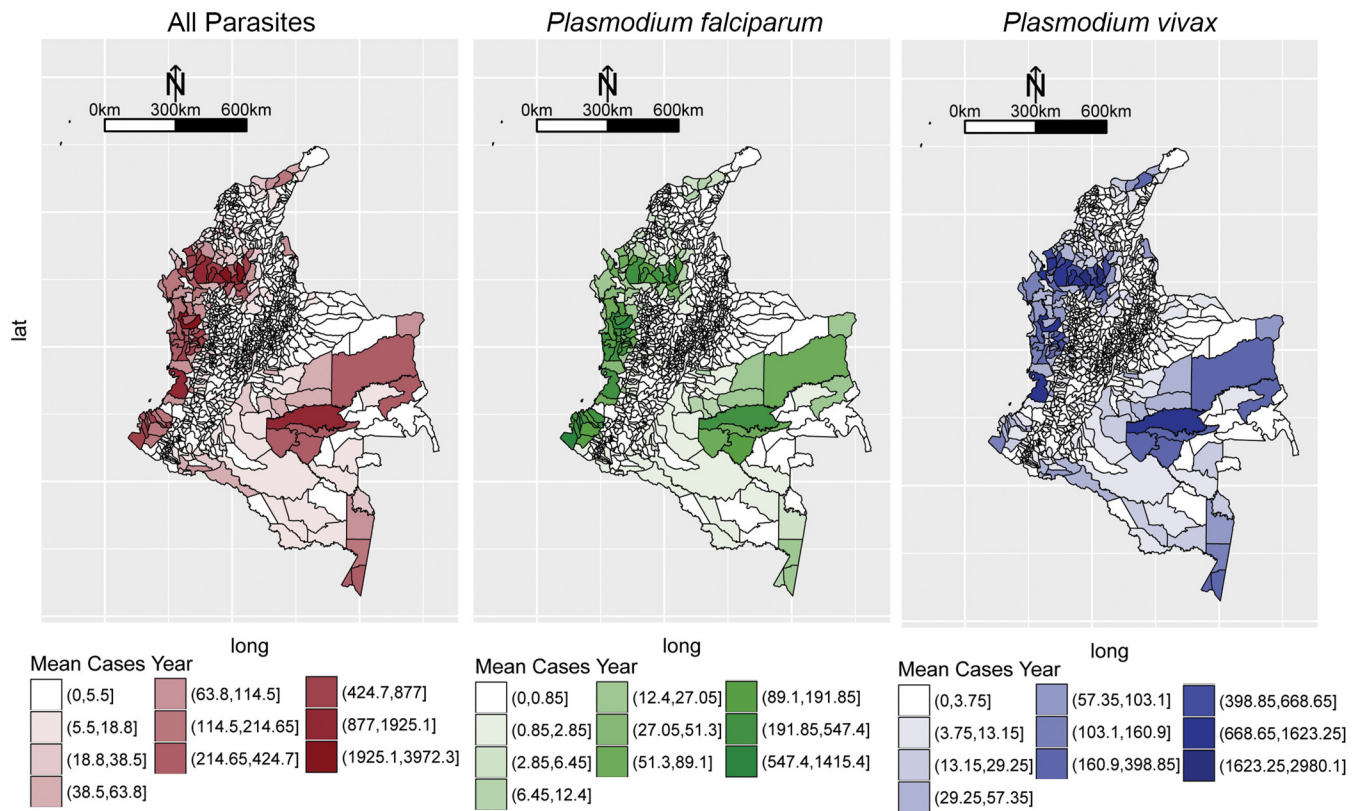


Fig 14. Malarial incidence for both species in Colombia from 2007-2015. Intervals were constructed using the Jenks procedure [88].

<https://doi.org/10.1371/journal.pone.0203673.g014>

Further research will show whether including these variables alters our findings. It is possible that such information will improve the ability of the model to correctly identify clusters, or more interestingly, the overlapping of clusters and variables may shed light into ongoing debates, such as vector displacement, for example. Our findings should be cautiously interpreted, but we are confident that they inform policy in a way that few other epidemiological studies currently advise, and furthermore, our findings are complementary to vector, pathogen and epidemic literature, to improve our understanding of malarial infection dynamics in Colombia at a national level, and the way in which such dynamics interact at a regional level.

7 Additional figures

For completeness, we include the malaria incidence extracted from the data in Fig 14.

Acknowledgments

The first author would like to thank Stanford University, the Zaffaroni family, and the Morrison Institute for their financial support. The second author acknowledges and thanks the financial support of the grant *P12.160422.004/01-FAPA ANDRES ANGEL* from Vicedecanatura de Investigaciones de la Facultad de Ciencias de la Universidad de los Andes, Colombia. The third author acknowledges and thanks the financial support of the grant *Proyecto Semilla*

2017-1 from Vicedecanatura de Investigaciones de la Facultad de Ciencias de la Universidad de los Andes, Colombia. We also want to thank Vladimir Corredor for comments and suggestions.

Author Contributions

Conceptualization: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca, Camilo Rivera.

Data curation: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

Formal analysis: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca, Camilo Rivera.

Funding acquisition: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

Investigation: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

Methodology: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca, Camilo Rivera.

Project administration: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

Resources: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

Software: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

Supervision: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

Validation: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

Visualization: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

Writing – original draft: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

Writing – review & editing: Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca.

References

1. Rubio-Palis Y, Zimmerman RH. Ecoregional classification of malaria vectors in the neotropics. *Journal of Medical Entomology*. 1997; 34(5):499–510. <https://doi.org/10.1093/jmedent/34.5.499> PMID: 9379453
2. Sáenz R, Molina JA, Quiñones ML, Brochero HL, Olano VA. Mapas preliminares de la distribución de especies de Anopheles vectores de malaria en Colombia. *Biomédica*. 2001; 21(4):402–408.
3. Carmona-Fonseca J. La malaria en Colombia, Antioquia y las zonas de Urabá y Bajo Cauca: panorama para interpretar la falla terapéutica antimalárica. Parte 2. *Iatreia*. 2004; 17(1):34–53.
4. Alexander N, Rodríguez M, Pérez L, Caicedo J, Cruz C, Prieto G, et al. Case-control study of mosquito nets against malaria in the Amazon Region of Colombia. *American Journal of Tropical Medicine and Hygiene*. 2005; 73(1):140–148. PMID: 16014849
5. Arévalo-Herrera M, Lopez-Perez M, Medina L, Moreno A, Gutierrez JB, Herrera S. Clinical profile of *Plasmodium falciparum* and *Plasmodium vivax* infections in low and unstable malaria transmission settings of Colombia. *Malaria journal*. 2015; 14(1):154. <https://doi.org/10.1186/s12936-015-0678-3> PMID: 25889074
6. Poveda G, Rojas W, Quiñones ML, Vélez ID, Mantilla RI, Ruiz D, et al. Coupling between annual and ENSO timescales in the malaria-climate association in Colombia. *Environmental health perspectives*. 2001; 109(5):489. <https://doi.org/10.1289/ehp.01109489> PMID: 11401760

7. Gagnon AS, Smoyer-Tomic KE, Bush AB. The El Niño Southern Oscillation and malaria epidemics in South America. *International Journal of Biometeorology*. 2002; 46(2):81–89. <https://doi.org/10.1007/s00484-001-0119-6> PMID: 12135203
8. Ferreira MU, Castro MC. Challenges for malaria elimination in Brazil. *Malaria journal*. 2016; 15(1):284. <https://doi.org/10.1186/s12936-016-1335-1> PMID: 27206924
9. Valero-Bernal MV. Malaria in Colombia: retrospective glance during the past 40 years. *Revista de Salud Pública*. 2006; 8(3):141–149. <https://doi.org/10.1590/S0124-00642006000300001> PMID: 17269215
10. WHO. Malaria entomology and vector control. World Health Organization; 2013.
11. Hill AV. Malaria resistance genes: a natural selection. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 1992; 86(3):225–232. [https://doi.org/10.1016/0035-9203\(92\)90282-H](https://doi.org/10.1016/0035-9203(92)90282-H) PMID: 1412635
12. Greenwood B, Marsh K, Snow R. Why do some African children develop severe malaria? *Parasitology Today*. 1991; 7(10):277–281. PMID: 15463389
13. Rosario V. Genetics of chloroquine resistance in malaria parasites. *Nature Publishing Group*. 1976; 261(5561):585–586.
14. Mita T, Tanabe K. Evolution of *Plasmodium falciparum* drug resistance: implications for the development and containment of artemisinin resistance. *Japanese journal of infectious diseases*. 2012; 65(6):465–475. <https://doi.org/10.7883/yoken.65.465> PMID: 23183197
15. Roper C, Pearce R, Bredenkamp B, Gumede J, Drakeley C, Moshia F, et al. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *The Lancet*. 2003; 361(9364):1174–1181. [https://doi.org/10.1016/S0140-6736\(03\)12951-0](https://doi.org/10.1016/S0140-6736(03)12951-0)
16. Wellems TE, Plowe CV. Chloroquine-resistant malaria. *Journal of Infectious Diseases*. 2001; 184(6):770–776. <https://doi.org/10.1086/322858> PMID: 11517439
17. Soto J, Toledo J, Gutierrez P, Luzz M, Llinas N, Cedeño N, et al. *Plasmodium vivax* clinically resistant to chloroquine in Colombia. *The American Journal of Tropical Medicine and Hygiene*. 2001; 65(2):90–93. <https://doi.org/10.4269/ajtmh.2001.65.90> PMID: 11508397
18. Castillo CM, Osorio LE, Palma GI. Assessment of therapeutic response of *Plasmodium vivax* and *Plasmodium falciparum* to chloroquine in a malaria transmission free area in Colombia. *Memórias do Instituto Oswaldo Cruz*. 2002; 97(4):559–562. <https://doi.org/10.1590/S0074-02762002000400020> PMID: 12118291
19. Blair-Trujillo S, Lacharme-Lora L, Carmona-Fonseca J. Resistance of *Plasmodium falciparum* to anti-malarial drugs in Zaragoza (Antioquia, Colombia), 1998. *Memórias do Instituto Oswaldo Cruz*. 2002; 97(3):401–406. <https://doi.org/10.1590/S0074-02762002000300022> PMID: 12048572
20. Vallejo AF, Chaparro PE, Benavides Y, Álvarez Á, Quintero JP, Padilla J, et al. High prevalence of sub-microscopic infections in Colombia. *Malaria journal*. 2015; 14(1):201. <https://doi.org/10.1186/s12936-015-0711-6> PMID: 25971594
21. Quintero JP, Siqueira AM, Tobón A, Blair S, Moreno A, Arévalo-Herrera M, et al. Malaria-related anaemia: a Latin American perspective. *Memorias do Instituto Oswaldo Cruz*. 2011; 106:91–104. <https://doi.org/10.1590/S0074-02762011000900012> PMID: 21881762
22. Ortega DC, Fong C, Cardenas H, Barreto G. Evidence of over-dominance for sickle cell trait in a population sample from Buenaventura, Colombia. *International Journal of Genetics and Molecular Biology*. 2015; 7(1):1–7. <https://doi.org/10.5897/IJGMB2014.0106>
23. Ruiz D, Poveda G, Vélez ID, Quiñones ML, Rúa GL, Velásquez LE, et al. Modelling entomological-climatic interactions of *Plasmodium falciparum* malaria transmission in two Colombian endemic-regions: contributions to a National Malaria Early Warning System. *Malaria Journal*. 2006; 5(1):66. <https://doi.org/10.1186/1475-2875-5-66> PMID: 16882349
24. Bouma M, Poveda G, Rojas W, Chavasse D, Quinones M, Cox J, et al. Predicting high-risk years for malaria in Colombia using parameters of El Niño Southern Oscillation. *Tropical Medicine & International Health*. 1997; 2(12):1122–1127. <https://doi.org/10.1046/j.1365-3156.1997.d01-210.x>
25. Kulldorf M. A Spatial Scan Statistic. *Communications in Statistics: Theory and Methods*. 1997; 6(26):1481–1496. <https://doi.org/10.1080/03610929708831995>
26. Kulldorf M, Athas W, Feurer E, Miller B, Key C. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health*. 1998; 88(9):1377–1380. <https://doi.org/10.2105/AJPH.88.9.1377> PMID: 9736881
27. Hjalmars U, Kulldorf M, Gustafsson G, Nagarwalla N. Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine*. 1996; 15(7-9):707–715. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960415\)15:7/9%3C707::AID-SIM242%3E3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960415)15:7/9%3C707::AID-SIM242%3E3.0.CO;2-4) PMID: 9132898

28. Burkom HS. Biosurveillance applying scan statistics with multiple, disparate data sources. *Journal of Urban Health*. 2003; 80(1):i57–i65.
29. Singh G, Memoli F, Carlsson G. *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*. Prague, Czech Republic: Eurographics Association; 2007. p. 91–100.
30. Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, et al. Extracting insights from the shape of complex data using topology. 2013; 3:1236 EP –.
31. Cameron D, Jones IG. John Snow, the Broad Street pump and modern epidemiology. *International Journal of Epidemiology*. 1983; 12(4):393–396. <https://doi.org/10.1093/ije/12.4.393> PMID: 6360920
32. Snow R, Marsh K, Le Sueur D. The need for maps of transmission intensity to guide malaria control in Africa. *Parasitology Today*. 1996; 12(12):455–457. [https://doi.org/10.1016/S0169-4758\(96\)30032-X](https://doi.org/10.1016/S0169-4758(96)30032-X)
33. Kilama M, Smith DL, Hutchinson R, Kigozi R, Yeka A, Lavoy G, et al. Estimating the annual entomological inoculation rate for *Plasmodium falciparum* transmitted by *Anopheles gambiae s.l.* using three sampling methods in three sites in Uganda. *Malaria journal*. 2014; 13(1):111. <https://doi.org/10.1186/1475-2875-13-111> PMID: 24656206
34. Omumbo J, Ouma J, Rapuoda B, Craig M, Lesueur D, Snow R. Mapping malaria transmission intensity using geographical information systems GIS: an example from Kenya. *Annals of Tropical Medicine and Parasitology*. 1998; 92(1):7–21. <https://doi.org/10.1080/00034989860120> PMID: 9614449
35. Kleinschmidt I, Bagayoko M, Clarke G, Craig M, Le Sueur D. A spatial statistical approach to malaria mapping. *International Journal of Epidemiology*. 2000; 29(2):355–361. <https://doi.org/10.1093/ije/29.2.355> PMID: 10817136
36. Kitron U, Pener H, Costin C, Orshan L, Greenberg Z, Shalom U. Geographic information system in malaria surveillance: mosquito breeding and imported cases in Israel, 1992. *The American Journal of Tropical Medicine and Hygiene*. 1994; 50(5):550–556. <https://doi.org/10.4269/ajtmh.1994.50.550> PMID: 8203702
37. Beck LR, Rodríguez MH, Dister SW, Rodríguez AD, Rejmankova E, Ulloa A, et al. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *The American Journal of Tropical Medicine and Hygiene*. 1994; 51(3):271–280. <https://doi.org/10.4269/ajtmh.1994.51.271> PMID: 7943544
38. Bharti N, Xia Y, Bjornstad ON, Grenfell BT. Measles on the edge: coastal heterogeneities and infection dynamics. *PloS One*. 2008; 3(4):e1941. <https://doi.org/10.1371/journal.pone.0001941> PMID: 18398467
39. Ferrari M, Djibo A, Grais R, Grenfell B, Bjørnstad O. Episodic outbreak bias estimates of age-specific force of infection: a corrected method using measles as an example. *Epidemiology and Infection*. 2010; 138(01):108–116. <https://doi.org/10.1017/S0950268809990173> PMID: 19538818
40. Ferrari MJ, Djibo A, Grais RF, Bharti N, Grenfell BT, Bjornstad ON. Rural–urban gradient in seasonal forcing of measles transmission in Niger. *Proceedings of the Royal Society B: Biological Sciences*. 2010; 277(1695):2775–2782. <https://doi.org/10.1098/rspb.2010.0536> PMID: 20427338
41. Bharti N, Tatem AJ, Ferrari MJ, Grais RF, Djibo A, Grenfell BT. Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science*. 2011; 334(6061):1424–1427. <https://doi.org/10.1126/science.1210554> PMID: 22158822
42. da Silva-Nunes M, Moreno M, Conn JE, Gamboa D, Abeles S, Vinetz JM, et al. Amazonian malaria: asymptomatic human reservoirs, diagnostic challenges, environmentally driven changes in mosquito vector populations, and the mandate for sustainable control strategies. *Acta Tropica*. 2012; 121(3):281–291. <https://doi.org/10.1016/j.actatropica.2011.10.001> PMID: 22015425
43. Castro M, Monte-Mor R, Sawyer D, Singer B. Malaria Risk on the Amazon Frontier. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103(7):2452–2457. <https://doi.org/10.1073/pnas.0510576103> PMID: 16461902
44. Valle D, Lima JMT. Large-scale drivers of malaria and priority areas for prevention and control in the Brazilian Amazon region using a novel multi-pathogen geospatial model. *Malaria journal*. 2014; 13(1):443. <https://doi.org/10.1186/1475-2875-13-443> PMID: 25412882
45. Bishop YM. Statistical methods for hazards and health. *Environmental Health Perspectives*. 1977; 20:149. <https://doi.org/10.1289/ehp.7720149> PMID: 598347
46. Cazelles B, Chavez M, de Magny GC, Guégan JF, Hales S. Time-dependent spectral analysis of epidemiological time-series with wavelets. *Journal of the Royal Society Interface*. 2007; 4(15):625–636. <https://doi.org/10.1098/rsif.2007.0212>
47. Pascual M, Rodó X, Ellner SP, Colwell R, Bouma MJ. Cholera dynamics and El Niño-southern oscillation. *Science*. 2000; 289(5485):1766–1769. <https://doi.org/10.1126/science.289.5485.1766> PMID: 10976073

48. Pascual M, Bouma MJ, Dobson AP. Cholera and climate: revisiting the quantitative evidence. *Microbes and Infection*. 2002; 4(2):237–245. [https://doi.org/10.1016/S1286-4579\(01\)01533-7](https://doi.org/10.1016/S1286-4579(01)01533-7) PMID: 11880057
49. Pascual M, Ahumada J, Chaves L, Rodo X, Bouma M. Malaria resurgence in the East African highlands: temperature trends revisited. *Proceedings of the National Academy of Sciences*. 2006; 103(15):5829–5834. <https://doi.org/10.1073/pnas.0508929103>
50. Grassly NC, Fraser C, Garnett GP. Host immunity and synchronized epidemics of syphilis across the United States. *Nature*. 2005; 433(7024):417–421. <https://doi.org/10.1038/nature03072> PMID: 15674292
51. Duncan C, Duncan S, Scott S. Whooping cough epidemics in London, 1701–1812: infection dynamics, seasonal forcing and the effects of malnutrition. *Proceedings of the Royal Society of London Series B: Biological Sciences*. 1996; 263(1369):445–450. <https://doi.org/10.1098/rspb.1996.0067> PMID: 8637925
52. Rohani P, Green C, Mantilla-Beniers N, Grenfell B. Ecological interference between fatal diseases. *Nature*. 2003; 422(6934):885–888. <https://doi.org/10.1038/nature01542> PMID: 12712203
53. Cazelles B, Cazelles K, Chavez M. Wavelet analysis in ecology and epidemiology: impact of statistical tests. *Journal of The Royal Society Interface*. 2014; 11(91):20130585. <https://doi.org/10.1098/rsif.2013.0585>
54. Grenfell B, Bjørnstad O, Kappey J. Travelling waves and spatial hierarchies in measles epidemics. *Nature*. 2001; 414(6865):716–723. <https://doi.org/10.1038/414716a> PMID: 11742391
55. Cazelles B, Chavez M, McMichael AJ, Hales S. Nonstationary influence of El Niño on the synchronous dengue epidemics in Thailand. *PLoS Medicine*. 2005; 2(4):e106. <https://doi.org/10.1371/journal.pmed.0020106> PMID: 15839751
56. Kreppel KS, Caminade C, Telfer S, Rajerison M, Rahalison L, Morse A, et al. A Non-Stationary Relationship between Global Climate Phenomena and Human Plague Incidence in Madagascar. *PLoS Neglected Tropical Diseases*. 2014; 8(10).
57. Onozuka D. Effect of non-stationary climate on infectious gastroenteritis transmission in Japan. *Scientific reports*. 2014; 4. <https://doi.org/10.1038/srep05157> PMID: 24889802
58. Morris A, Gozlian RE, Hassani H, Andreou D, Couppié P, Guégan JF. Complex temporal climate signals drive the emergence of human water-borne disease. *Emerging Microbes & Infections*. 2014; 3(8):e56. <https://doi.org/10.1038/emi.2014.56>
59. Chaves LF, Pascual M. Climate cycles and forecasts of cutaneous leishmaniasis, a nonstationary vector-borne disease. *PLoS medicine*. 2006; 3(8):e295. <https://doi.org/10.1371/journal.pmed.0030295> PMID: 16903778
60. José MV, Bishop RF. Scaling properties and symmetrical patterns in the epidemiology of rotavirus infection. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*. 2003; 358(1438):1625–1641. <https://doi.org/10.1098/rstb.2003.1291> PMID: 14561323
61. Cuzick J, Edwards R. Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society Series B (Methodological)*. 1990; p. 73–104.
62. Openshaw S, Charlton M, Wymer C, Craft A. A mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information System*. 1987; 1(4):335–358. <https://doi.org/10.1080/02693798708927821>
63. Snow R, Schellenberg JA, Peshu N, Forster D, Newton C, Winstanley P, et al. Periodicity and space-time clustering of severe childhood malaria on the coast of Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 1993; 87(4):386–390. [https://doi.org/10.1016/0035-9203\(93\)90007-D](https://doi.org/10.1016/0035-9203(93)90007-D) PMID: 8249058
64. Brooker S, Clarke S, Njagi JK, Polack S, Mugo B, Estambale B, et al. Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya. *Tropical Medicine and International Health*. 2004; 9(7):757–766. <https://doi.org/10.1111/j.1365-3156.2004.01272.x> PMID: 15228485
65. Coleman M, Coleman M, Mabuza AM, Kok G, Coetzee M, Durrheim DN. Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. *Malar J*. 2009; 8(68):10–1186.
66. Zhang W, Wang L, Fang L, Ma J, Xu Y, Jiang J, et al. Spatial analysis of malaria in Anhui province, China. *Malaria Journal*. 2008; 7(206):19.
67. Faires MC, Pearl DL, Ciccotelli WA, Berke O, Reid-Smith RJ, Weese JS. Detection of *Clostridium difficile* infection clusters, using the temporal scan statistic, in a community hospital in southern Ontario, Canada, 2006–2011. *BMC infectious diseases*. 2014; 14(1):254. <https://doi.org/10.1186/1471-2334-14-254> PMID: 24885351
68. Duczmal LH, Moreira GJ, Paquete L, Menotti D, Takahashi R, Burgarelli D. Dry Climate as a Predictor of Chagas' Disease Irregular Clusters: A Covariate Study. *Online Journal of Public Health Informatics*. 2015; 7(1).

69. Occelli F, Deram A, Génin M, Noël C, Cuny D, Glowacki F, et al. Mapping End-Stage Renal Disease (ESRD): Spatial Variations on Small Area Level in Northern France, and Association with Deprivation. *PLoS ONE*. 2014; 9(11). <https://doi.org/10.1371/journal.pone.0110132> PMID: 25365039
70. Li J, Kolivras KN, Hong Y, Duan Y, Seukep SE, Prisley SP, et al. Spatial and Temporal Emergence Pattern of Lyme Disease in Virginia. *The American Journal of Tropical Medicine and Hygiene*. 2014; 91(6):1166–1172. <https://doi.org/10.4269/ajtmh.13-0733> PMID: 25331806
71. Rogers DJ, Randolph SE. The global spread of malaria in a future, warmer world. *Science*. 2000; 289(5485):1763–1766. <https://doi.org/10.1126/science.289.5485.1763> PMID: 10976072
72. Nicolau M, Levine AJ, Carlsson G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108(17):7265–7270. <https://doi.org/10.1073/pnas.1102826108> PMID: 21482760
73. Ibekwe AM, Ma J, Crowley DE, Yang CH, Johnson AM, Petrossian TC, et al. Topological data analysis of *Escherichia coli* O157:H7 and non-O157 survival in soils. *Frontiers in Cellular and Infection Microbiology*. 2014; 4:122. <https://doi.org/10.3389/fcimb.2014.00122> PMID: 25250242
74. Nielson JL, Paquette J, Liu AW, Guandique CF, Tovar CA, Inoue T, et al. Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. *Nature Communications*. 2015; 6:8581. <https://doi.org/10.1038/ncomms9581> PMID: 26466022
75. Sardiù ME, Gilmore JM, Groppe B, Florens L, Washburn MP. Identification of Topological Network Modules in Perturbed Protein Interaction Networks. *Scientific Reports*. 2017; 7:43845. <https://doi.org/10.1038/srep43845> PMID: 28272416
76. Yearbook D. United Nations. New York. 2016;(67).
77. Neill DB. Expectation-based scan statistics for monitoring spatial time series data. *International Journal of Forecasting*. 2009; 25(3):498–517. <https://doi.org/10.1016/j.ijforecast.2008.12.002>
78. Neill D, Wong W. A tutorial on event detection. Presented at the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2009;.
79. Kim AY, Wakefield J. *SpatialEpi: Methods and Data for Spatial Epidemiology*; 2016. Available from: <https://CRAN.R-project.org/package=SpatialEpi>.
80. Maaten Lvd, Hinton G. Visualizing data using t-SNE. *Journal of machine learning research*. 2008; 9 (Nov):2579–2605.
81. Alves FP, Gil LHS, Marrelli MT, Ribolla PE, Camargo EP, Pereira Da Silva LH. Asymptomatic carriers of *Plasmodium* spp. as infection source for malaria vector mosquitoes in the Brazilian Amazon. *Journal of medical entomology*. 2005; 42(5):777–779.
82. Feged-Rivadeneira A, Evans S. Ethnography of a parasite: A quantitative ethnographic observation of forest malaria in the Amazon basin. *Scandinavian journal of public health*. 2018; p. 1403494818756561. <https://doi.org/10.1177/1403494818756561> PMID: 29783880
83. SIAC. Alertas tempranas por deforestación: 2013-1,2013-2,2014-1,2014-2. 2014;.
84. UNODC. Explotación de oro de aluvi6n. Evidencias a partir de percepci6n remota. 2016;.
85. Castellanos A, Chaparro-Narv6ez P, Morales-Plaza CD, Alzate A, Padilla J, Ar6valo M, et al. Malaria in gold-mining areas in Colombia. *Mem6rias do Instituto Oswaldo Cruz*. 2016; 111(1):59–66. <https://doi.org/10.1590/0074-02760150382> PMID: 26814645
86. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS medicine*. 2012; 9(1):e1001165. <https://doi.org/10.1371/journal.pmed.1001165> PMID: 22303287
87. Bannister-Tyrrell M, Verdonck K, Hausmann-Muela S, Gryseels C, Ribera JM, Grietens KP. Defining micro-epidemiology for malaria elimination: systematic review and meta-analysis. *Malaria journal*. 2017; 16(1):164. <https://doi.org/10.1186/s12936-017-1792-1> PMID: 28427389
88. Jenks GF, Caspall FC. Error on choroplethic maps: definition, measurement, reduction. *Annals of the Association of American Geographers*. 1971; 61(2):217–244. <https://doi.org/10.1111/j.1467-8306.1971.tb00779.x>