RESEARCH IN STATISTICS
2025, VOL. 3, NO. 1, 2479722
https://doi.org/10.1080/27684520.2025.2479722

Taylor & Francis
Taylor & Francis Group



# Spatio-temporal variation of malaria incidence in Sub-Saharan Africa from 2011 to 2020

Changkuoth Jock Chol[a,b], Denekew Bitew Belay[a,c], Haile Mekonnen Fenta[a,c] and Ding-Geng Chen[c,d]

[a]Department of Statistics, College of Science, Bahir Dar University, Bahir Dar, Ethiopia; [b]Department of Statistics, College of Natural and Computational Science, Gambella University, Gambella, Ethiopia; [c]Department of Statistics, University of Pretoria, Pretoria, South Africa; [d]College of Health Solutions, Arizona State University, Phoenix, AZ, USA

**ABSTRACT**

**Background:** Malaria continues to pose a severe danger to health, social, and economic development in sub-Saharan Africa (SSA). Malaria in SSA requires a tremendous effort to eradicate. This study aimed to explore clusters of malaria incidence both spatially and temporally in the regions of the SSA countries using spatial autocorrelation statistics and univariate functional spatial scan statistics.

**Methods:** The study was done in 634 administrative areas throughout SSA nations. We used secondary data from the Malaria Atlas Project between 2011 and 2020. To convert malaria cases into incidence, the population at risk has been projected. To measure spatial autocorrelation, Global Moran's *I* was selected to examine feature positions and values. The Local Moran statistic was applied to identify regions dissimilar from their neighborhood. The Getis-Ord Local G* was employed to identify hot spots and cold spots areas in the SSA region. Furthermore, the univariate functional spatial scan statistic was used to see if spatial data tended to aggregate.

**Results:** The results revealed that the spatial and temporal clusters were statistically significant during the study period. Spatio-temporal scanning identified three high-risk malaria incidence areas in SSA. The regions around West-central, Central, and Southeast of SSA have a high-risk malaria incidence throughout the study period.

**Conclusions:** The West-Central, Central, and Southeast regions of SSA had a high risk of malaria. Due to that, global end malaria councils should mobilize multi-sectional action, resources, advocacy, and accountability to support the national malaria program and national strategic plan to take and coordinate action across sectors to fight malaria.

**ARTICLE HISTORY**
Received 15 September 2024
Accepted 7 March 2025

**KEYWORDS**
Malaria incidence; spatial scan; spatio-temporal; sub-Saharan Africa

## 1. Introduction

Malaria is caused by the protozoan parasite *Plasmodium* in humans (WHO 2023). It is spread exclusively through bites of infected female Anopheles mosquitoes (Walter and John 2022). The mosquito bite presents the parasites from the mosquito's spit into a person's blood (WHO 2023). The parasites go to the liver, where they proliferate and grow, which cause the malaria disease. The disease is broad in the tropical and subtropical locales that exist in a wide band around the equator (Caraballo and King 2014), this includes much of sub-Saharan Africa (SSA), Asia, and Latin America (WHO 2023). Methods utilized to avoid intestinal sickness incorporate drugs, mosquito disposal, and the anticipation of chomps. High densities of populations of individuals and Anopheles mosquitoes, as well as high rates of transmission between humans and mosquitoes, are necessary for the occurrence of malaria in a given location (WHO 2015). Malaria has been successfully eliminated or significantly reduced in certain areas, but not globally. Most of Europe, North America, Australia, North Africa, the Caribbean, and parts of South America, Asia, and Southern Africa have moreover dispensed with jungle fever (Meade and Emch 2010). "Elimination" (also known as "malaria-free") is defined by the World Health Organization (WHO) as the absence of domestic transmission (indigenous cases) during the previous three years. They too characterize "pre-elimination" and "disposal" stages when a nation has less than 5 or 1, separately, cases per 1000 individuals at risk per year (WHO 2023).

For evidence-informed research and policy-making to determine malaria-free, the spatial cluster detection methods are valuable tools for objective detection and localization of statistically significant aggregation of events indexed in space. In the writing, a few cluster identification techniques have been put forth. Specifically, spatial scan statistics (Kulldorff 1997; Kulldorff and Nagarwalla 1995) for Poisson and Bernoulli models are effective techniques for identifying statistically significant geographical clusters that, when pre-

CONTACT Ding-Geng Chen dinchen@asu.edu Department of Statistics, University of Pretoria, Pretoria, South Africa.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/27684520.2025.2479722.

selection bias is absent, and with a changeable scanning window, may be characterized as a collection of locations exhibiting an anomalous concentration of an observable variable (objective detection of the cluster). Many scholars have modified spatial scan statistics to fit various spatial data distributions, such as exponential (Huang et al. 2007), ordinal (Jung et al. 2007), normal (Kulldorff et al. 2009), Weibull (Bhatt and Tiwari 2014), etc., after Kulldorff's original work. Nonparametric methods are employed by Jung and Cho (2015) and Cucala et al. (2019), who extend the Wilcoxon-Mann-Whitney test for temporal or spatial scan statistics, respectively. Because the two approaches broaden the method created by Jung and Cho (2015) to find either high or low clusters, it should be emphasized that they are equivalent when working with geographical data.

Technological advances resulted in the collection of data in large volumes over time, which led to the introduction of functional data analysis (FDA) (Ramsay and Silverman 2005). In spaces where information normally shows a spatial component, the development of useful information in this way is driven by the advancement of spatial efficient data (Delicado et al. 2010; Fernández de Castro and Manteiga 2008; Menafoglio et al. 2013). In this framework, data interpolation methods (such as *kriging* (Bohorquez et al. 2016; Giraldo et al. 2010) or *co-kriging* (Monestiez and Nerini 2008)), regression (Ahmed et al. 2022; Ternynck 2014), or clustering (Giraldo et al. 2012; Vandewalle et al. 2021), approaches have been proposed.

In spatial scan measurements, using a "naïve" univariate technique based on time-averaged information in each geographic region across the study period would result in significant data loss. High dimensionality and high correlation problems would be faced by a multivariate approach that treats every observation time point as a variable, which would result in a dramatic reduction in power and a noticeably higher false positive rate (Cucala et al. 2017; Smida et al. 2022). Although an analysis of variance (ANOVA) for functional data has been published by Cuevas et al. (2004) and Frévent, Ahmed, Marbac, et al. (2021) noted that a parametric scan statistic for functional data has not before been offered; nonetheless, it should be feasible to construct a scan method based on this test. Then, in the last few years, statistical tests for high dimensional data were created by resuming the information after the computation of the statistics for each basic detail (Lin et al. 2023). Thus, combining this method with (Cucala 2014) distribution-free scan statistic technique, (Frévent, Ahmed, Marbac, et al. 2021) developed a scan statistic. They put out concepts of spatial clusters with spatial functional data. The next two parametric spatial scan statistics they offer for functional data are the distribution-free functional scan statistic (DFFSS) and the functional ANOVA-based functional scan statistic (Frévent, Ahmed, Marbac, et al. 2021). Furthermore, they suggest an additional point-wise rank-based spatial scan statistic, called the univariate rank-based functional spatial scan statistic (URBFSS). When dealing with Gaussian data, the pointwise DFFSS is the method of choice for detecting spatial clusters; in other cases, the pointwise rank-based technique (URBFSS) is employed (Smida et al. 2022).

The field of epidemiology has a long history of investigating the variables that influence the variation in the incidence or mortality of chronic and infectious diseases. One of the most important criteria in assessing the performance and distribution of health care has been the regional (or spatial) variability of health outcomes (Greene et al. 2016). The degree of noise in the data and patterns of reliance have also been demonstrated by spatial variance in health outcomes. More recently, the way that health variables change over time has been investigated using time-series studies (Diggle et al. 2019). The advantage of spatio-temporal studies over solely spatial or time-series analyses is that they enable the scholar to simultaneously examine how patterns endure over time and shed light on any anomalous patterns. Adding space-time interaction variables can also help identify data clustering that might be a sign of recurring problems in the data recording process or new environmental dangers (Diggle et al. 2019).

Despite recent sub-national, national, and international efforts to combat it, malaria continues to be one of the most serious vector-borne infectious illnesses in SSA. It is evident from the data that the SSA has failed to meet the 2020 WHO strategy milestones for lowering the morbidity and mortality rates of malaria. A number of reasons, including high-impact health emergencies and rapid population increase, have significantly impeded progress. Some parts of regions (states) in SSA countries have continuously increased the burden of malaria. Few writers used spatial scan statistics to investigate the prevalence of malaria in certain nations. Nevertheless, no researchers researched malaria incidence in any country using a functional spatial scan statistic. This study aimed to fill this knowledge gap to explore the spatio-temporal clusters of malaria incidence in the regions of the SSA countries using spatial autocorrelation statistics and univariate functional spatial scan statistics.

The structure of this study is organized as follows: The first section which an introduction of the stud outlines the research problem, objectives, and the existing literature gap regarding the Spatio-Temporal Variation of Malaria incidence. The methodology section details the study design, data source, and the application of Spatio-temporal to our malaria data. The results section presents the key findings, including the spatial and temporal difference in malaria incidence across sub-Saharan African countries. The discussion interprets these findings in the context of prior research and highlights their policy and intervention implications. Finally, the conclusion summarizes the study's contributions, emphasizes its practical applications, and suggests directions for future research.

## 2. Methods

### 2.1. Settings

This study has been conducted in the SSA countries at an administration level 1 area (sub-national level). The SSA lies south of the Sahara, these include Central Africa,

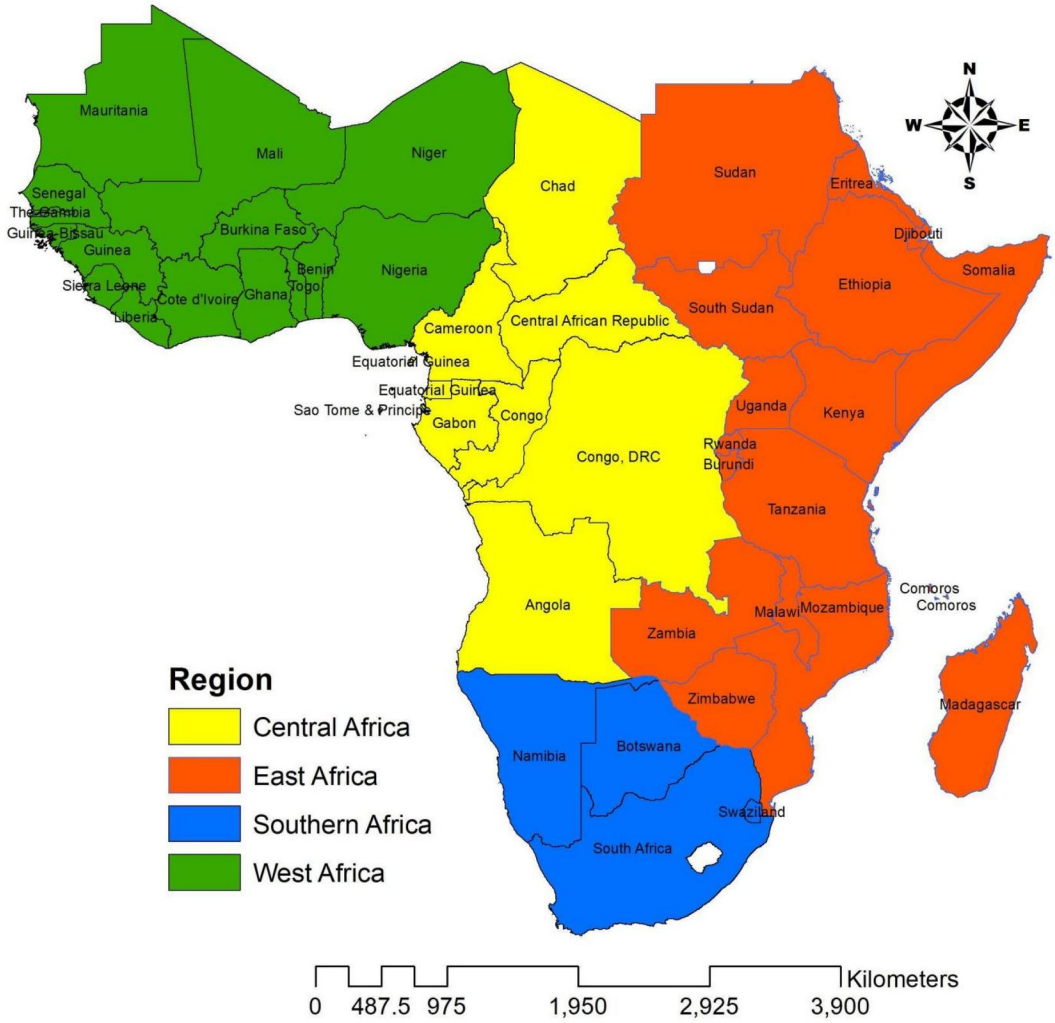


**Figure 1.** Map of the study area in the SSA countries.

East Africa, Southern Africa, and West Africa except Cape Verde (Figure 1). We excluded Lesotho, Mauritius, and Seychelles in this study due to inconsistent data at the sub-national level. We have conducted this study in 45 countries in SSA, within 45 countries there are 634 administration level 1 areas. Administration level 1 has different names in SSA countries; however, for this study, they all were referred to as regions. The map and name of the regions of the study area were provided in the additional file (Figure S1 and Table S1), respectively.

### 2.2. Data

Secondary data from the Malaria Atlas Project was used (MAP 2024). Different tools for learning about, examining, and utilizing malaria data that may be found on the Malaria Atlas website are offered via the Malaria Atlas Data Platform. We generate annual model estimates of the malaria risk landscape and burden of the disease at subnational levels in the SSA countries from 2011 to 2020. To transform malaria cases into incidence, an estimate of the population at risk was used. The proportion of the population at high, low, or no risk of malaria was provided by the National Malaria Programme (NMP). This was applied to United Nations (UN) population estimates (UN 2024), to compute the number of people at risk of malaria. This number sustained over time to ensure comparability of incidence estimates across years in the same cohort of countries endemic since 2011. The population at risk at the sub-national level was aggregated and included the population of all endemic countries since 2011.

### 2.3. Statistical methods

In this work, the data analysis techniques employed were spatial autocorrelation statistics and spatial scan statistics for functional data. Based on the locations and values of the features, we employed the Global Moran's $I$ to measure the malaria incidence spatial autocorrelation in SSA global between 2011 and 2020. It assesses if a set of highlights and an associated characteristic result in a pattern that is scattered, random, or clustered during the study period. The Local Moran statistics were employed to identify regions that are malaria incidence different from their neighborhood from 2011 to 2020 in the study area. The Getis-Ord Local $G^*$ was used to identify hot spots and cold spot areas of malaria incidence in the 634 regions at SSA between 2011 and 2020. To detect statistically significant spatial clusters in

the absence of any priori information about the location of the fixed focus, the univariate functional spatial scan statistic was utilized to ascertain whether some spatial observations tend to aggregate on malaria incidence in SSA, which screens for the presence of one or more clusters around the fixed focus during the study period. The findings from this study will play an essential role in malaria control and elimination in SSA countries.

### 2.3.1. Global Moran's I value

Moran's $I$ test is a frequently employed method for assessing spatial dependency in spatial lattice data (Waller and Gotway 2004). This test can be connected to the information straightforwardly, or the residuals from a few spatial relapses demonstrated. In this study, let $\{Z_{i(t)} : i = 1, \ldots, m = 634\}$ represent spatially referenced data (or residuals) for $m = 634$ spatial locations (regions) at time $t$, where $t = 2011, \ldots, 2020$. Then, we use the formula

$$I(t) = \frac{m\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}\left(Z_{i(t)} - \overline{Z(t)}\right)\left(Z_{j(t)} - \overline{Z(t)}\right)}{\left(\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}\left(\sum_{i=1}^{m}(Z_{i(t)} - \overline{Z(t)}\right)^2\right)}, \quad (1)$$

to Moran's $I$ statistic, where $\overline{Z(t)} = (1/m)\sum_{i=1}^{m} Z_{i(t)}$ is the spatial mean at each specific year and $w_{ij}$ are spatial adjacency "weights" between locations $i$ and $j$ (where we require $w_{ii} = 0$, for all $i = 1, \ldots, 634$). Therefore, $z_{I(t)} = \frac{I(t) - E(I(t))}{\sqrt{V[I(t)]}}$, where $E(I(t)) = {}^{-1}/_{(m-1)}$ and $V(I(t)) = E\left[I(t)^2\right] - (E[I(t)])^2$, is the formula for calculating the $z_{I(t)} - score$ for the Moran $I$ statistic, given that $E(I(t))$ is the expected value and $V(I(t))$ is the variance.

It can be seen from Equation (1) that the Moran's $I$ statistic is simply a weighted form of the usual Pearson correlation coefficient, where the weights are the spatial proximity weights, and it takes values between −1 and 1. If (1) is positive, then neighboring regions tend to have similar values, and if it is negative, then neighboring regions tend to have different values of malaria incidence in SSA as whole between 2011 and 2020.

### 2.3.2. Anselin's Local Moran

The first LISA statistic was created by Luc Anselin (Anselin 1995) and is known as Anselin's Local Moran statistic. The method applies Moran's "I" statistics to individual regions permitting them to be recognized as comparative or distinctive to their nearby pattern. The definition of "$I_{i(t)}$" is from (Getis and Ord 1996):

$$I_{i(t)} = \frac{\left(Z_{i(t)} - \overline{Z(t)}\right)}{S^2_{Z(t)}}\sum_{j=1}^{m-1}\left[w_{ij}(Z_{j(t)} - \overline{Z(t)})\right], \quad (2)$$

where $t = 2011, \ldots, 2020$, $Z_{i(t)}$ is the intensity of observation $i$ at time $t$, $\overline{Z}(t)$ is the mean intensity over all observations at time $t$, $Z_{j(t)}$ is intensity for all other observations, $j$ (where $j \neq i$) at time $t$, $S^2_{Z(t)}$ is the variance over all observations at time $t$, and $w_{ij}$ is a distance weight for the interaction between observations $i$ and $j$ during the study period. The first term in (2) refers only to observation $i$ although the second term is the sum of the weighted values for all other observations (but not including $i$ itself). The expected "$I_{i(t)}$" is formulated as $E\left(I_{i(t)}\right) = \frac{\sum_{i=1}^{m} w_{ij}}{m-1}$, where $w_{ij}$ is the distance weight for the interaction between observations $i$ and $j$. The variance, standard deviation, and an approximate (pseudo) standardized score of $I_{i(t)}$ is

$$Var\left(I_{i(t)}\right) = \frac{\left(\sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij}^2\right)(m - b_2)}{m-1} + \frac{2w_{i(kh)}(2b_2 - m)}{(m-1)(m-2)} + \frac{\left(\sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij}^2\right)}{(m-1)^2},$$

$S\left(I_{i(t)}\right) = \sqrt{Var\left(I_{i(t)}\right)}$, and $Z\left(I_{i(t)}\right) = \frac{I_{i(t)} - E(I_{i(t)})}{S(I_{i(t)})}$. We define $b_2 = \frac{\sum_{i=1}^{m}\frac{\left(Z_{i(t)} - \overline{Z(t)}\right)^4}{m}}{\left[\sum_{i=1}^{m}\frac{\left(Z_{i(t)} - \overline{Z(t)}\right)^4}{m}\right]^2}$ and $2w_{i(kh)} = \sum_{k=1}^{m-1}\sum_{h=1}^{m-1} W_{ik}W_{ih}$, where $k \neq i$ and $h \neq i$. This term is twice the sum of the cross-products of all weights for $i$ with themselves, using $k$ and $h$ to maintain a strategic distance from the use of indistinguishable subscripts.

Anselin's Local Moran has several uses. First, it can identify regions that are different (dissimilar) from its neighbors. This can be a good first step in finding areas that either have higher malaria incidence rates (a hot spot) or lower malaria incidence rates (a cold spot) than the neighboring areas. This can focus researchers' efforts on identifying the problems that cause the region to be higher in the case of a hot spot or in the event of a cold spot, to determine the elements that reduce the occurrence of malaria during the study period. Second, another use of Anselin's Local Moran statistic is to identify 'outliers', regions that are different from their neighbors. In this case, regions with a high negative I esteem (e.g., with an "$I_{i(t)}$" smaller than two standard deviations underneath the mean) are demonstrative of outliers. They either have a high incidence rate in comparison to their neighbors' low incidence rate, or they have the reverse pattern-low incidence rates among areas with high incidence rates.

### 2.3.3. Getis-Ord local G*

The Getis-Ord Local G* statistic determines whether specific areas are spatially connected to the nearby area by applying the Getis-Ord "G" statistic to individual regions. The Getis-Ord Local G* is applied to each region, unlike the global Getis-Ord G*, but similar to Anselin's Local Moran. The formulation presented here is taken from (Wong and Lee 2005). In this study, the G*(t) value at time $t = 2011, ., 2020$ is calculated concerning a specified search distance, namely:

$$G^*(t) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} w_{i,j}x_i x_j}{\sum_{i=1}^{m}\sum_{j=1}^{m} x_i x_j}, \forall_{j \neq i} \quad (3)$$

where $w_{i,j}$ is the spatial weight between features i and j, and $x_i$ and $x_j$ are the attribute values for features i and j, $m$ is the number of features in the dataset (634 regions) and $\forall_{j \neq i}$

indicates that features i and j cannot be the same feature. Therefore, $z_{G^*(t)} = \frac{G^*(t)-E[G^*(t)]}{\sqrt{V[G^*(t)]}}$ yields the $z_{G^*(t)} - score$ for the statistic, where $V[G^*(t)] = E[G^*(t)^2] - (E[G^*(t)])^2$ *and* $E[G^*(t)] = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}}{n(n-1)}, \forall_{j\neq i}$.

The Getis-Ord Local $G^*$ is very effective at identifying hot and cold spots. As seen earlier, Anselin's Local Moran can only identify positive or negative spatial autocorrelation, or how similar or unlike the regions are from one another. Those regions with positive spatial autocorrelation could occur because regions with high values are near another region with high values or they may happen because regions with low values are near another region with low values of malaria incidence in SSA between 2011 and 2020. The Getis-Ord Local $G^*$ can distinguish those two types.

### 2.3.4. A univariate functional spatial scan statistic

**2.3.4.1. General principle.** Let $s_1, \ldots, s_n$ be $n$ (for example $n = 634$ administration level 1 areas in this study) different locations of an observation domain $S \subset R^2$ and $X_1, \ldots, X_n$ be the observations of a variable $X$ in $s_1, \ldots, s_n$ (i.e., $s_1, \ldots, s_{634}$ in this study). From now on all perceptions are considered autonomous, which is a classical presumption in scan statistics. Spatial scan statistics recommend recognizing geographical groups and assessing their statistical significance. Hence, one tests a null hypothesis $H_0$ (the absence of a cluster) against a composite alternative hypothesis $H_1$ (the presence of at least one cluster $w \subset S$ presenting abnormal values of $X$). A spatial scan statistic has two phases for this purpose.

To begin with, there is a discovery stage utilizing a checking window of variable estimate and shape. We will center here on the approach of Kulldorff and Nagarwalla (1995), which use a circular filtering window of variable center and span, in any case, it ought to be apparent that other shapes can be considered (Cucala et al. 2013; Kulldorff et al. 2006). Limiting the maximum size to half of the researched region is an often recommended method because it avoids the detection of "negative clusters" in the areas outside the clusters that encompass almost the whole examined region (Kulldorff and Nagarwalla (1995). Next, using

$$W = \left\{ w_{i,j} / 1 \leq \left| w_{i,j} \right| \leq \frac{n}{2}, 1 \leq i,j \leq n \right\}, \tag{4}$$

the scanning window enables the definition of a collection of possible clusters $W$, where $w_{i,j}$ is the disc centered on $s_i$ that passes through $s_j$ and $\left| w_{i,j} \right|$ corresponds to the number of sites in $w_{i,j}$. The highest concentration index over the set of possible clusters $W$ may then be used to determine the spatial scan statistic. The moment step is the assurance of the factual noteworthiness of the spatial scan statistic. This is because the overlapping nature of $W$ makes the distribution of the scan statistic unmanageable under $H_o$. One popular solution, which was taken into consideration here, is to employ a Monte Carlo technique.

In the functional framework, (Frévent et al. 2023) state that the DFFSS and the URBFSS in the univariate context are the methods that get the greatest results. Despite detecting more false positives than the DFFSS, the URBFSS often exhibits greater powers and true positive rates. In this study, applying spatial scan statistics for univariate functional data will thus allow highlighting of areas where the malaria incidence curves are abnormally high or low. We select to utilize the URBFSS scan strategy given that it regularly presents higher powers and genuine positive rates than other univariate functional methods.

**2.3.4.2. Univariate functional data: a novel rank-based spatial scan statistic.** By modifying the method of (Frévent et al. 2023), a pointwise technique based on locations and the nonparametric scan statistic for univariate information may be presented in the univariate functional system. For a time $t$, (Jung and Cho 2015) proposed to test $H_0 : \forall w \in W, F_{w,t} = F_{W^c,t}$ where, $F_{w,t}$, and $F_{W^c,t}$ is the cumulative distribution functions of $X(t)$ in $w$ and outside $w$, by using the Wilcoxon rank-sum test statistic. The Wilcoxon rank-sum test statistic, for a period $t$ and a probable cluster $w$, is $W(t)^{(w)} = \sum_{i, S_i \in w} R_i(t)$, where $R_i(t)$ is the rank of $X_i(t)$ in $\{X_1(t), \ldots, X_n(t)\}$, calculated using the average rank in the event of tied observations. At that point the standardized version of this measurement is

$$T(t)^{(w)} = \frac{W(t)^{(w)} - \mathbb{E}\left[W(t)^{(w)}\right]}{\sqrt{\mathbb{V}\left[W(t)^{(w)}\right]}}, \tag{5}$$

where the variance and expected value of $W(t)^{(w)}$ under $H_0$ are, respectively, $\mathbb{V}\left[W(t)^{(w)}\right] = \frac{|w||w^c|(n+1)}{12}$ and $\mathbb{E}\left[W(t)^{(w)}\right] = \frac{|w|(n+1)}{2}$. The idea put out by Jung and Cho (2015) was to reduce the p-value linked to $T(t)^{(w)}$ on the set of possible clusters $W$. They propose to adapt their approach by simply using $\left|T(t)^{(w)}\right|$ as a point-wise statistic.

The null hypothesis in cluster discovery is defined as $H_0 : \forall w \in W, \forall t \in T, F_{w,t} = F_{w^c,t}$. $H_1^{(w)} : \exists t \in T, F_{w,t}(x) = F_{w^c,t}(x - \Delta_t), \Delta_t \neq 0$, is the alternative hypothesis $H_1^{(w)}$ connected to a possible cluster $w$.

As before, they propose to globalize the information over time with $T^{(w)} = \sup_{t\in T}\left|T(t)^{(w)}\right|$ and to use this quantity as a concentration index, yielding the following URBFSS: $_{URBFSS} = \max_{w\in W} T^{(w)}$.

**2.3.4.3. Computing the statistical significance of the MLC.** Detected is one of the most likely clusters (MLC), its measurable significance must be assessed. Because $W$ is overlapping, the distribution of the scan statistic ($S =$ $_{URBFSS}$) is unmanageable under $H_0$. Then we choose to obtain a large set of simulated datasets by randomly permuting the observations $X_i$ in the spatial locations. This technique is called "random labeling" and has already been used in spatial scan statistics (Cucala et al. 2019; Kulldorff et al. 2009).

Let $S^{(1)}, \ldots, S^{(M)}$ be the observed scan statistics on the simulated datasets, and let $M$ be the number of random permutations of the original dataset. According to Dwass

(1957), the $p$-value for $S$ observed in the real data is estimated by

$$\hat{p} = \frac{1 + \sum_{m=1}^{M} 1_{S^{(m)} \geq S}}{M + 1} \quad (6)$$

Lastly, the MLC is measurably significant if the associated $\hat{p}$ is less than the type I error. Because the computation time is multiplied by $M + 1$, this method can be extremely costly from a computational standpoint, particularly if the precision needed for the estimation of the cluster's p-value (and hence the number of Monte-Carlo simulations) is high.

#### 2.3.5. Software

For the spatial auto-correlation statistics, the researcher has been provided the *Arcmap* version 10.8 to analyze patterns and mapping clusters to investigate whether the pattern exposed is clustered, dispersed, or random and identify the dissimilar from their neighborhood, hot-spots, and cold-spots areas. The R package *HDSpatialScan* (Frevent, Ahmed, Soula, et al. 2021), provides a function *SpatialScan* to compute the univariate functional spatial scan statistic. You may find the package https://CRAN.R-project.org/package=HDSpatialScan on CRAN.

## 3. Results

### 3.1. Spatial autocorrelation statistics

The spatial autocorrelation was estimated using Global Moran's I statistic, and the result for each year during the study period is given in Table 1. All $p$-values (all $p < 0.05$) are less than 0.05, the $z$-score is positive, and all Moran I indexes are greater than 0.78. The null hypothesis was rejected. Compared to what would be expected if the underlying geographical processes were random, the dataset's high value spatial distribution is more spatially concentrated. This indicates that the pattern expressed is clustered in the study area during the study period.

In this study, the spatial and temporal distribution of malaria incidence of the regions (subnational levels) in the SSA region from 2011 to 2020 was presented in Figure 2. The result shows that the risk of malaria incidence varies spatially and temporally in the SSA region during the period of the study. The findings also revealed that the malaria incidence risk is high in the West-central, Central, and Southeast and the malaria risk is low in the Southern, Northwest, Northeast, and some parts of the Eastern SSA during the study period.

In this study, Figure 3 presented the outcome that identifies regions that either have higher malaria incidence or lower malaria incidence rates than the neighboring areas and also identify outliers regions that are very different from their neighbors, to evaluate which regions have a high number of malaria incidences at the same time their neighbors have a low number of malaria incidence or which regions have a low number of malaria incidence whereas their neighbors have a high number of malaria incidence. The result revealed that the West-central, Central, and some parts of regions in the Southeast were high cluster areas in the SSA between 2011 and 2020. Then, the Northwest, Northeast, and some parts of eastern and Southern regions in the SSA were low cluster areas in SSA from 2011 to 2020. The low-high and high-low outliers results of malaria incidence in the regions of the study area during the study period were provided in the additional file (Supplementary Text S1).

Figure 4 represents the result of Getis-Ord Local $G^*$ to identify hot spots and cold spots in the SSA region. The finding indicated that the West-central, Central, some parts of Eastern, and Southeast were hot spot regions in the SSA countries during the study period. On the other hand, the Northwest, Northeast, and some parts of the Eastern and Southern part regions of the SSA were cold spot areas for malaria incidence from 2011 to 2020.

Table 1. Global Moran's *I* autocorrelation value for annual malaria incidence cases in the SSA, 2011–2020.

| Year | Moran's *I* | Variance | *Z*-score | *p*-Value |
|---|---|---|---|---|
| 2011 | 0.875090 | 0.000583 | 36.318915 | 0.001 |
| 2012 | 0.864849 | 0.000582 | 35.898101 | 0.001 |
| 2013 | 0.835739 | 0.000582 | 34.696966 | 0.001 |
| 2014 | 0.813986 | 0.000582 | 33.798309 | 0.001 |
| 2015 | 0.799559 | 0.000582 | 33.200776 | 0.001 |
| 2016 | 0.801741 | 0.000582 | 33.289580 | 0.001 |
| 2017 | 0.801401 | 0.000582 | 33.274514 | 0.001 |
| 2018 | 0.803800 | 0.000582 | 33.371070 | 0.001 |
| 2019 | 0.799288 | 0.000582 | 33.182077 | 0.001 |
| 2020 | 0.788771 | 0.000582 | 32.745841 | 0.001 |

### 3.2. Univariate functional spatial scan statistic

We consider malaria data provided by the Malaria Atlas project platform, which is available on the website. This lattice data consists of the yearly concentrations (from 2011 to 2020) of malaria incidence for every 634 regions (subnational levels of SSA) characterized by spatial polygons and located by their center of gravity $s_1, \ldots, s_{634}$. We consider the pollutant of malaria incidence applying spatial scan statistics for univariate functional data will thus allow highlighting areas where the malaria incidence concentration curves are abnormally high or low. We selected the URBFSS scan procedure because it often presents higher powers and true positive rates than the other univariate functional methods. The results of the URBFSS scan procedure are presented in Tables 2 and 3 and Figures 5 and 6.

Table 2 indicates the summary statistics for malaria incidence for each year during the study period in the SSA, which summarizes the lowest, maximum, standard deviation, and mean. The findings show that the average malaria incidence in the SSA decreased from 2011 to 2019 and increased from 2019 to 2020. In addition, the highest malaria incidence rate was reported in 2013 which was 633. 1443 Cases per thousand people during the study period. Moreover, there were zero cases per thousand people within some 634 regions between 2011 and 2020 in the study area.

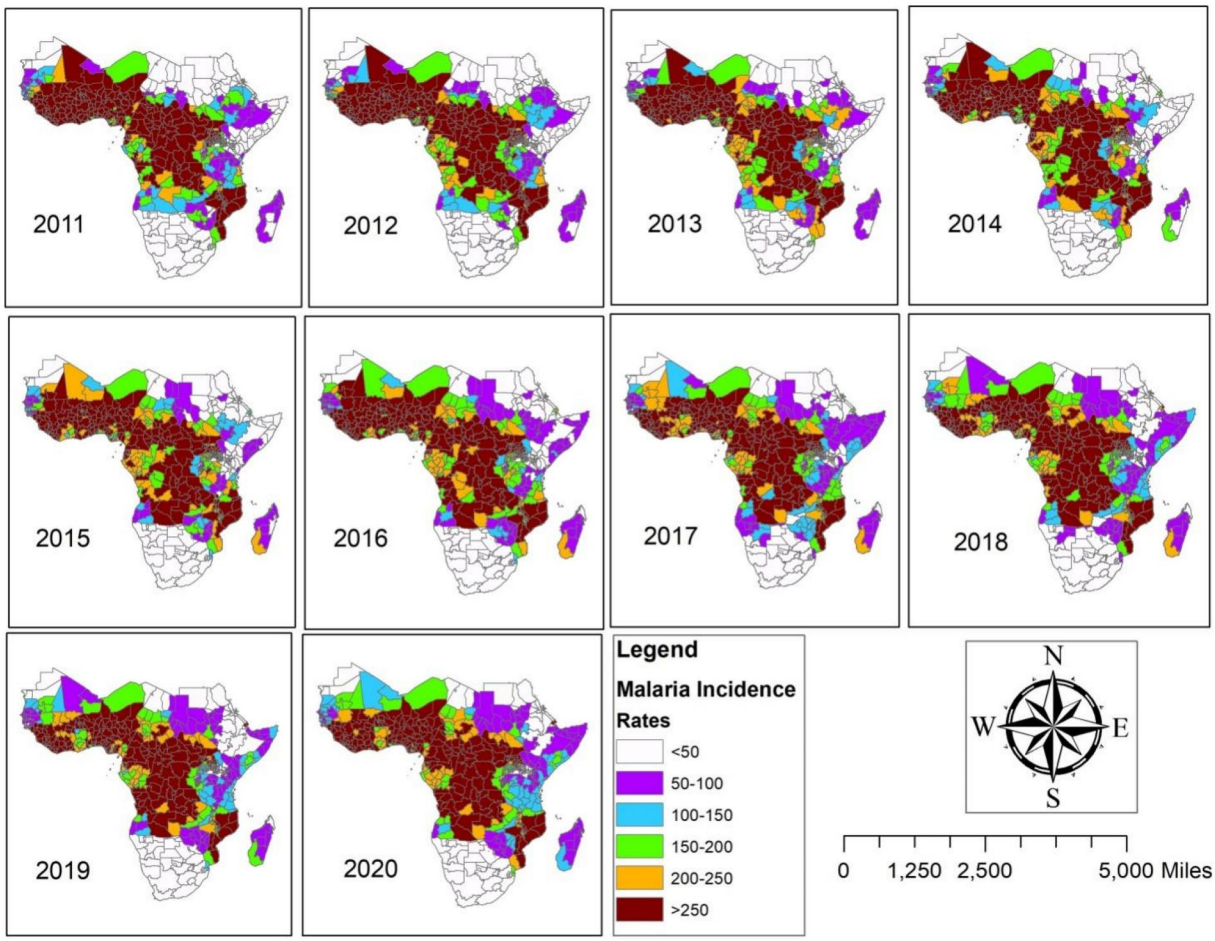


**Figure 2.** Spatial and temporal distribution of malaria incidence at the Sub-national level in the SSA region from 2011 to 2020.

Table 3 and Figure 5 show the statistically significant spatial clusters of high or low malaria incidence rates detected between 2011 and 2020. The findings show that 6 clusters are statistically significant spatial clusters. The spatial distribution of the average concentration of malaria incidence over the period between 2011 and 2020 overall (634 regions) is 207.698 in the SSA countries. The first cluster was detected as the third highest malaria incidence in SSA which has 317 regions in the SSA countries with an average malaria incidence of 283.135 and a radius of 2556.848 which is a larger significant cluster area than other significant clusters. The first cluster lies in the West-central, and Central in the SSA. In addition, the second cluster was detected as the second low cluster with an average malaria incidence of 72.773 throughout the study with a radius of 2003.290 and lies in the Horn of Africa in the study area.

Moreover, the third cluster was a statistically significant spatial cluster of the lowest malaria incidence rates detected between 2011 and 2020. The average malaria incidence of the third cluster is 21.532, which is the smallest average malaria incidence than other clusters and has a radius of 1379.744. The third cluster was located in the Southern part of the SSA and had 47 regions in the study area. Despite that, the fourth cluster is the highest malaria incidence area has only 26 regions in the study area detected over the period from 2011 to 2020, which is located around Uganda, Rwanda, and Burundi with a mean of 355.171 and a radius of 158.729, which is the smallest radius than all significant clusters.

However, the fifth cluster is the third low malaria incidence area in the SSA during the study period with an average malaria incidence of 119.458 and a radius of 434.535. The fifth cluster has 33 regions, located in Senegal and some parts in Mauritania, The Gambia, and Guinea-Bissau. Finally, the sixth cluster is a statistically significant spatial cluster as the second high cluster malaria incidence area detected between 2011 and 2020 in the SSA. The sixth cluster has 294.921 averages of malaria incidence and 415.319 radiuses. The sixth cluster has 28 regions and is located in Mozambique and Malawi.

The curves in Figure 6 show a marked temporal variability from 2011 to 2020 for the most likely clusters, which are statistically significant. The finding revealed that there is temporal variability of malaria incidence within six significant clusters in the SSA during the study period. The global mean of malaria incidence in the study area between 2011 and 2020 was represented by the black line in each cluster in Figure 6, the red line represented the temporal variation of malaria incidence of the regions inside each cluster from 2011 to 2020, and the gray line represented the temporal variation of malaria incidence of the regions outside each cluster during the study period.

In Figure 6 for each cluster's curve, the zero (0) is indicated at the time axis in each cluster showing the average

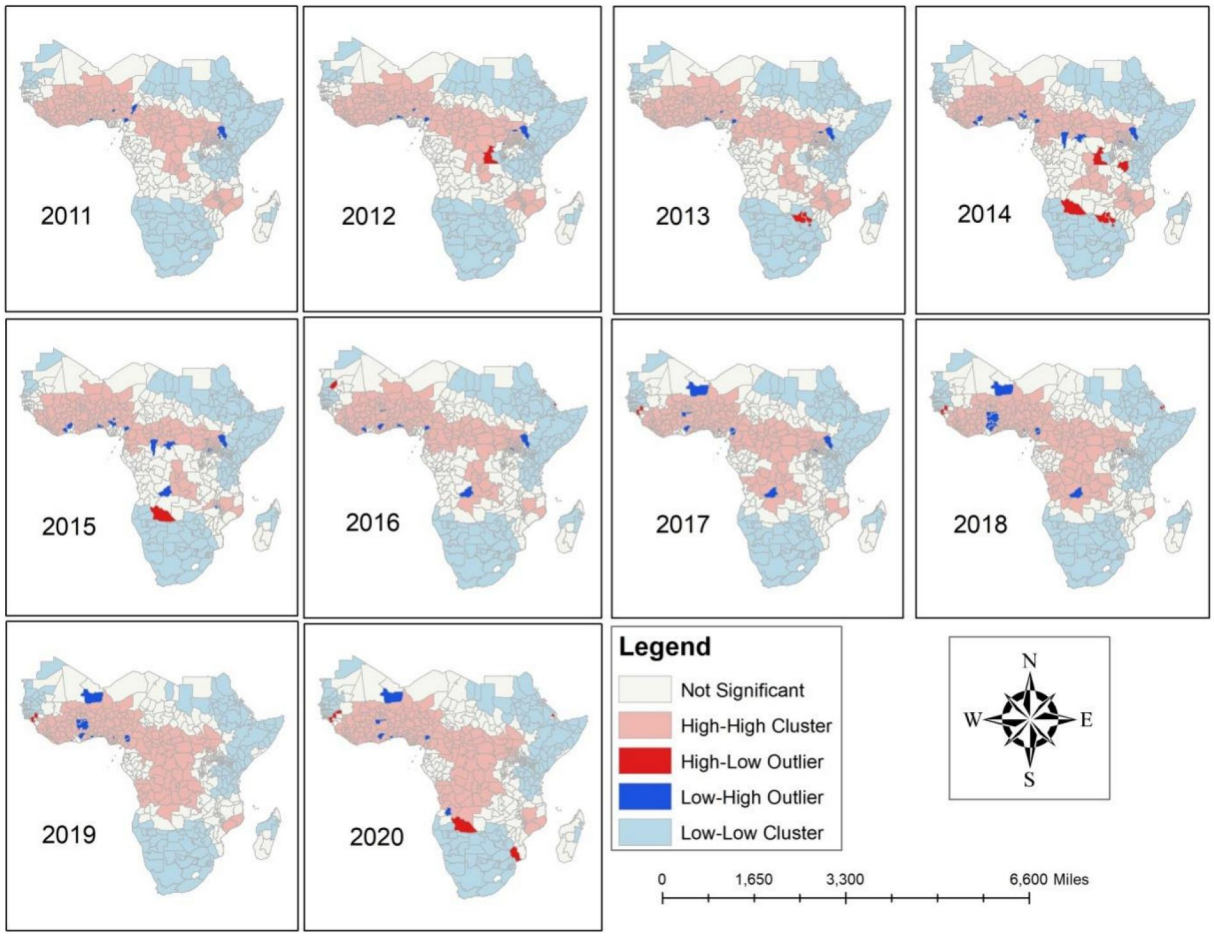


**Figure 3.** Cluster and outlier analysis of the Sub-national level in SSA from 2011 to 2020.

malaria incidence within each region between 2011 and 2012. The first cluster is the largest cluster area in the SSA which is a significant spatial cluster that has 317 (50%) regions (sub-national level) in the study area and the 317 regions that are inside cluster one show temporary variability during the study period. However, many regions inside cluster one has an average malaria incidence above the global average. The second cluster, which is the second low malaria incidence area in the study area shows a temporary variation between 2011 and 2020, and there are some regions in which their average malaria incidence exceeds then the global average malaria incidence. The third cluster has the lowest malaria incidence than other clusters; the average malaria incidence within regions during the study period is below the global average. However, the fifth cluster, which is the second low malaria incidence area in the SSA, shows a temporary variation between 2011 and 2020. In addition, from 2011 to 2015 the regions whose average malaria incidence exceeds the global mean inside cluster five show decreases from 2011 to 2016, and the mean malaria incidence was below the global mean between 2015 and 2020. Finally, the fourth and sixth cluster, which is the highest and second high cluster malaria incidence area in the SSA, respectively had a temporary variation during the study period. Our results revealed that the average malaria incidences in many regions inside the fourth and fifth clusters were above the global average between 2011 and 2020.

## 4. Discussion

In this study, the annual malaria incidence shows positive autocorrelation during the study period in the SSA, the Global Moran's *I* statistic ranging from 0.7888 to 0.8751, and all *p*-values are less than 0.05. This means the patterns explicitly were clustered in the SSA between 2011 and 2020. Our findings revealed that the malaria incidence risk varies spatially and temporarily in the study area over the study period between 2011 and 2020. The risk of malaria incidence is high in West-central, Central, and Southeast (Mozambique and Malawi) regions during the study period. However, the risk of malaria incidence is low in Southern and Northwest regions around Mauritania, Senegal, and The Gambia, and the Horn of Africa regions located in Djibouti, Somalia, Eritrea, Ethiopia, Sudan, and Kenya. Certain areas of Ghana in the West-Central SSA exhibit a decline from 2015 to 2020. In addition, the malaria incidence decreased around Ethiopia in the Horn of Africa during the study period; some regions around Djibouti, Sudan, and Somalia increased the malaria incidence between 2015 and 2020. The AU established a 2020 aim to reduce malaria incidence and death by 40% from 2015 levels, according to African Union (2020). Ethiopia, Mauritania, The Gambia, and Ghana reduced the incidence by at least 40%. In addition, the Eswati, Equatorial Guinea, Kenya, Rwanda, Senegal, and Togo reduced incidence by 25–40%. An increase in

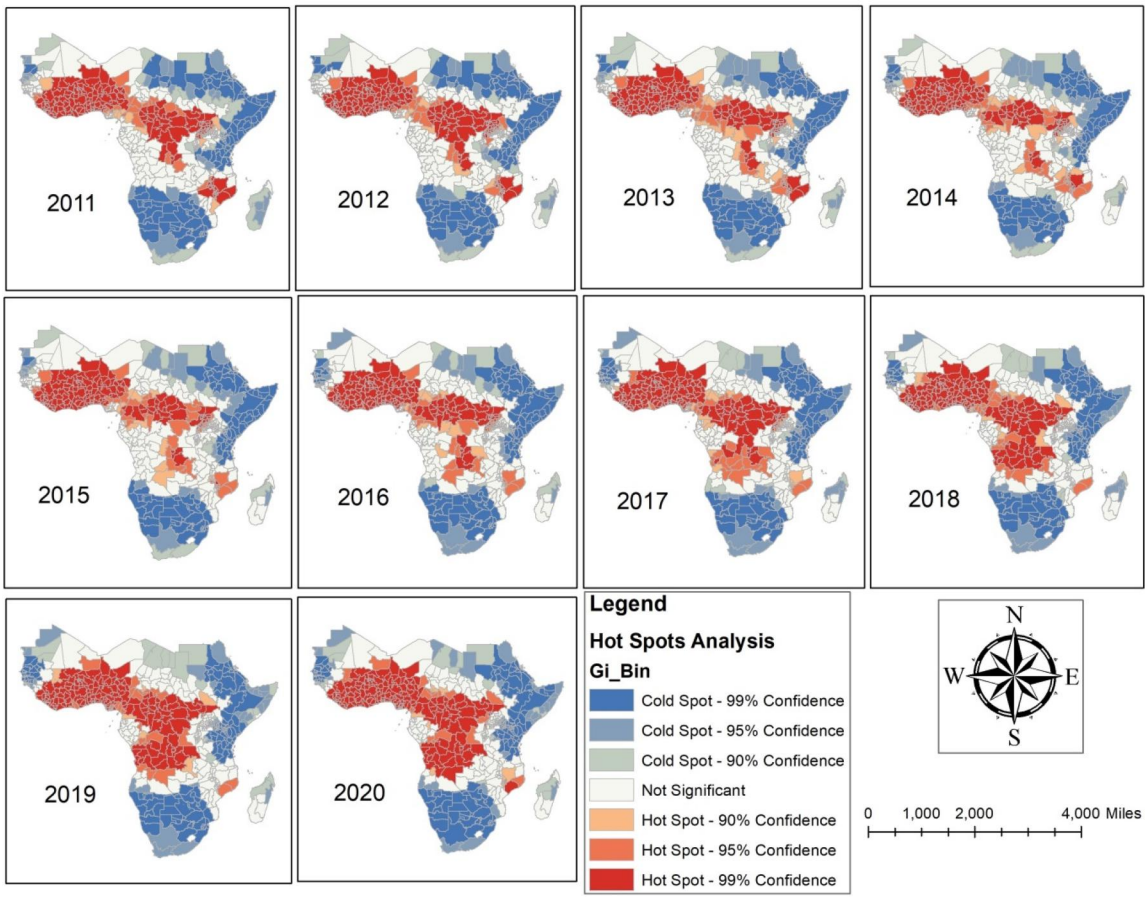


**Figure 4.** Hot spot analysis of the sub-national level in SSA from 2011 to 2020.

**Table 2.** Summary statistics.

| Year | Mean | Std. Dev. | Minimum | Maximum |
|---|---|---|---|---|
| 2011 | 233.4671 | 182.2421 | 0.0000 | 613.6464 |
| 2012 | 224.8597 | 173.3228 | 0.0000 | 632.2385 |
| 2013 | 215.8956 | 162.4275 | 0.0000 | 633.1443 |
| 2014 | 208.5888 | 156.4282 | 0.0000 | 621.3306 |
| 2015 | 204.9666 | 149.9497 | 0.0000 | 597.1162 |
| 2016 | 200.4698 | 144.2182 | 0.0000 | 588.9143 |
| 2017 | 199.2246 | 140.2140 | 0.0000 | 584.0099 |
| 2018 | 196.2044 | 136.9920 | 0.0000 | 557.3779 |
| 2019 | 191.3397 | 134.4873 | 0.0000 | 564.7912 |
| 2020 | 201.9617 | 135.5859 | 0.0000 | 577.9063 |

Std. Dev.: standard deviation.

estimated malaria cases was seen in Uganda, Mozambique, Malawi, Sudan, Somalia, and Djibouti in the Eastern SSA between 2015 and 2020. Our study revealed that the West-central, Central, and some parts of the Southeast were high-high clusters although, Southern, Northwest, Northeast, and some parts of Eastern regions in the study area were low-low clusters areas between 2011 and 2020. In addition, the findings indicated that the West-central, Central, and some parts of Eastern regions like South Sudan, Uganda, Rwanda, Burundi, Malawi, and Mozambique were the hot spots in the SSA countries between 2011 and 2020. Moreover, the Southern, Northwest like Mauritania and Senegal, the Horn of Africa, and some regions in Eastern SSA like Tanzania and Madagascar were cold spots for malaria incidence during the study period.

According to WHO (2023), between 2000 and 2019 malaria incidence cases in the WHO African region reduced from 368 to 222 per 1000 population at risk but increased to expand to 232 in 2020, basically because of disturbances to administrations amid the pandemic widespread. In this study, the average malaria incidence decreased from 2011 to 2019 and increased from 2019 to 2020 in the study area due to pandemic reasons. The highest malaria incidence rate was reported in 2013, which is 633.1443 cases per thousand people in the Zamfara region, Nigeria country. The spatial distribution of the average concentration of malaria incidence from 2011 to 2020 in 634 regions is 207.698 per thousand people in the SSA.

The statistically significant spatial clusters of high or low malaria incidence rates were detected from 2011 to 2020 in our study area. The six clusters are statistically significant, according to the findings. With mean values of 355.171 from 2011 to 2020, the fourth cluster-which includes just 26 areas and is situated in the Eastern SSA surrounding Uganda, Rwanda, and Burundi-was shown to have the greatest malaria incidence risk area. In addition, the sixth cluster is the second highest malaria incidence risk area with a mean of 294.921 and 28 regions inside, and the cluster was located in Mozambique and Malawi in the Southeast of the SSA. Moreover, the first cluster is the biggest cluster area and the third highest malaria incidence risk area with 317 regions inside the cluster with an average malaria incidence of 283.135 per thousand people. The first cluster was located

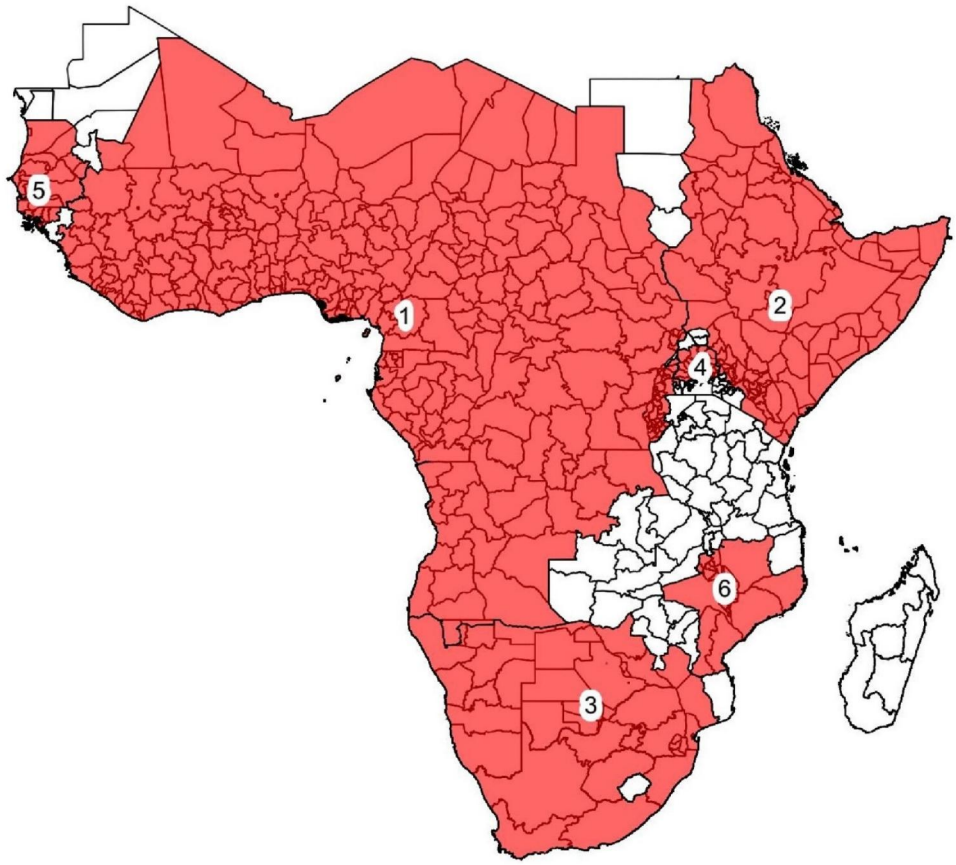


**Figure 5.** The spatial clusters of high or low malaria incidence rates detected over the period from 2011 to 2020.

**Table 3.** Summary of the clusters obtained with a URBFSS.

| | Inside | | | Outside | | | | |
|---|---|---|---|---|---|---|---|---|
| Cluster | *N* | Mean | Std. Dev. | *N* | Mean | Std. Dev. | Radius | *p*-Value |
| 1 | 317 | 283.135 | 129.660 | 317 | 132.260 | 136.357 | 2556.848 | 0.001 |
| 2 | 93 | 72.773 | 94.130 | 541 | 230.892 | 149.075 | 2003.290 | 0.001 |
| 3 | 47 | 21.532 | 46.351 | 587 | 222.604 | 148.645 | 1379.744 | 0.001 |
| 4 | 26 | 355.171 | 153.135 | 608 | 201.391 | 149.742 | 158.729 | 0.001 |
| 5 | 33 | 119.458 | 87.107 | 601 | 212.543 | 154.303 | 434.535 | 0.003 |
| 6 | 28 | 294.921 | 91.716 | 606 | 203.668 | 154.004 | 415.319 | 0.006 |
| Overall | 634 | 207.698 | 152.942 | | | | | |

*N*: number of the regions; Std. Dev.: standard deviation.

around West-central, Central, and some parts of the Eastern of the SSA. In the West-central, the countries inside cluster one was Guinea, Sierra Leone, Liberia, Mali, Cote d'Ivoire, Burkina Faso, Niger, Nigeria, Benin, Togo, and Ghana. In the Central of the SSA, the countries inside cluster one was Chad, Cameron, Equatorial Guinea, Gabon, Congo, Central African Republic, DR Congo, and Angola. Some of the regions in Sudan and South Sudan in the Eastern SSA are inside cluster one. According to WHO (2020), in the Western of the SSA, Burkina Faso, Guinea, Mali, Niger, Sierra Leone, and Togo were a reduction in malaria incidence of less than 40% in 2020 compared to 2015 and Benin, Cote d'Ivoire, Guinea-Bissau, and Liberia increased the malaria incidence in 2020 compared with 2015. The estimated incidence of malaria rose in the Central SSA, Burundi, Angola, Congo, Democratic Republic of Congo, and Chad between 2015 and 2020 but, did not change in Cameroon or the Central African Republic. Malaria is steady throughout much of Malawi, Mozambique, South Sudan, Uganda, Tanzania, and Zambia, with significant transmission in the Eastern SSA. In Mozambique and Uganda, the high burden to high impact (HBHI) project has been started.

According to African Union (2020), Namibia saw a large decrease in the number of malaria incidence estimated cases in 2019 compared with 2018, cases significantly increased again in 2020, and Botswana also had a significant increase in the estimated cases in 2020, which was seven times higher than estimated cases in 2019, Zimbabwe saw no reduction for the malaria incidence between 2015 and 2020. In addition, Eswatini and South Africa met the *Global Technical Strategy for Malaria 2016–2030* (GTS) target of a 40% reduction in malaria incidence by 2020 compared with the GTS baseline 2015. In this study, the first lowest malaria incidence area is the third cluster with an average incidence of 21.532 per thousand people and 47 regions inside. The third cluster is located around Southern parts of countries in the SSA. The nations with the lowest risk of malaria incidence in the SSA include South Africa, Namibia, Botswana, Zimbabwe, and Eswatini. In addition, the second low malaria incidence risk area in the SSA was detected as the second cluster with an average malaria incidence of 72.773 and 93 regions inside. These second clusters were located in Djibouti, Eritrea, Ethiopia, Somalia, and Kenya, and some regions were in Sudan and South Sudan. The fifth cluster was detected as the third low malaria incidence

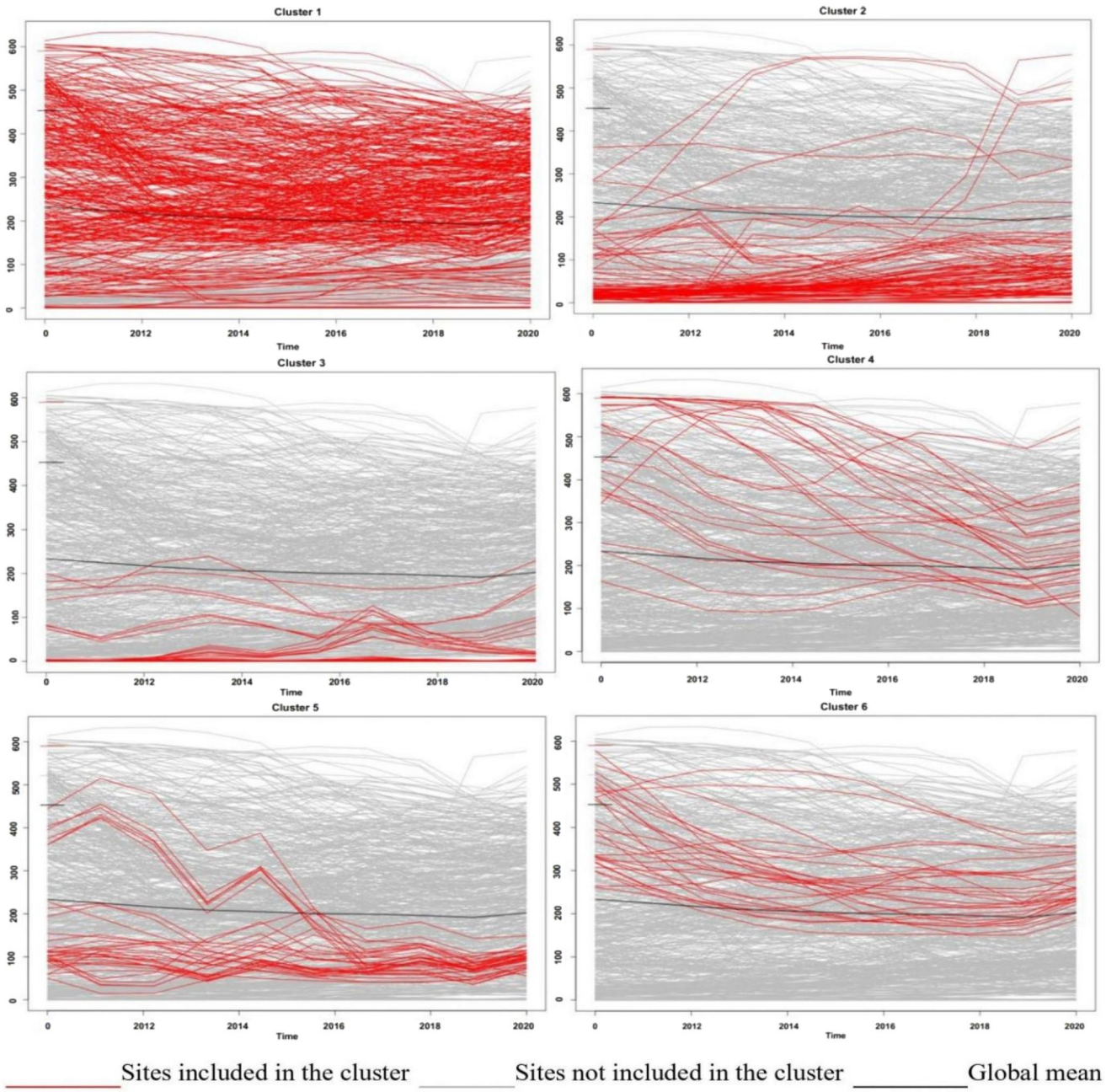


**Figure 6.** The most likely clusters are characterized by mean concentration curves of malaria incidence higher or lower from 2011 to 2020.

risk area in the SSA with a mean value of malaria incidence of 119.458. The fifth cluster has only 33 regions inside that were located in Senegal, and the Gambia, and some were in Mauritania, and Guinea-Bissau. According to WHO (2020), the Gambia, Mauritania, the Niger, and Senegal in the Western of the SSA have reoriented their program towards malaria subnational elimination. Kenya, Ethiopia, Djibouti, Eritrea, Somalia, and Sudan have decreased the malaria incidence from 2015 to 2019, but there was an increase between 2019 and 2020 in Sudan, Somalia, and Djibouti. Our findings revealed that there are temporary variations of malaria incidence within the six most likely clusters which are statistically significant in the SSA between 2011 and 2020.

## 5. Conclusions

Our study analyzed the spatial, temporal, and space-time clusters of malaria incidence at the regional level in the SSA countries between 2011 and 2020, using the Global Moran's I statistic, Local Moran statistic, Getis-Ord Local $G^*$ statistic, and the URBFSS methods. The spatial and temporal clusters were statistically significant every year during the study period, and the space-time scanning results indicated three high-risk areas and three low-risk areas for malaria incidence in the SSA. The three high-risk areas for malaria incidence were predominantly located in the West-central, Central, and Southeast of the SSA. On the other hand, the three low-risk areas for malaria incidence were located in

the Southern, Northwest, and Northeast of the SSA. These results indicate that it is urgent to establish preventive and control strategies to decrease the incidence in SSA countries through the global malaria program. We recommend that the global end malaria councils should mobilize multisectional action, resources, advocacy, and accountability to support the national malaria program and the national strategic plan in the SSA countries to take and coordinate action across sectors to fight malaria.

Furthermore, initiatives are needed to address the increasing incidence of malaria and eventually eliminate and control it in SSA. Additionally, to manage resources for successful implementation, it is imperative to continuously improve efficient data modeling and distribution of medications and other palliatives to eliminate and manage malaria. Furthermore, technological innovations and domestic information and communication technologies should be developed to improve malaria response in terms of epidemiological surveillance and statistical modeling tools. This will help geospatial mapping and vector parasite monitoring to address the increasing problems linked to malaria control.

Our study also demonstrated the usefulness of spatial and temporal clustering analysis using the ArcGIS and R software to identify the significant space-time clusters of malaria incidence in the SSA. This could be used to provide strategies plan for malaria elimination and prevention at regional levels in the study area. However, the study had limitations analysis. First, we didn't include the several factors that influence malaria incidence risks like climate and socio-economic factors, because the dataset did not contain these variables. Second, we excluded the coverage of malaria vector control intervention factors, which have a huge role in understanding which populations may be less well protected at subnational levels over the study period in the SSA, due to missing data at subnational levels in some of the malaria years. Malaria eradication and elimination efforts in SSA have not been particularly successful over the years, primarily due to a lack of a well-planned and coordinated synergy among the health care providers and the beneficiary (the community).

## Acknowledgments

The authors thank the Ethiopian Ministry of Education, Bahir Dar University, and Gambella University for their administrative support. The authors would also acknowledge the South Africa National Research Foundation (NRF) and South Africa Medical Research Council (SAMRC) (South Africa DSTNRF-SAMRC SARCHI Research Chair in Biostatistics) for their partial support in this research work. We gratefully acknowledge the Malaria Atlas Project particularly the Bill and Melinda Gates Foundation who are principally funded by this data platform.

## Author contributions

C.J.C. analyzed and drafted the manuscript. D.B.B. and H.M.F., supervisors for C.J.C., designed and supervised from the design to the write-up of the manuscript. D.G.C. reviewed and revised the manuscript. All authors read and approved the final manuscript.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work is partially supported by the South Africa National Research Foundation (NRF) and South Africa Medical Research Council (SAMRC) [South Africa DST-NRF-SAMRC SARCHI Research Chair in Biostatistics, Grant number 114613]. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF and SAMRC.

## ORCID

Ding-Geng Chen http://orcid.org/0000-0002-3199-8665

## Data availability statement

The malaria data used to support the outcome of this paper are available at the Malaria Atlas website at https://malariaatlas.org/.

## References

African Union. 2020. Decision on the report on malaria – Doc.Assembly/AU/13(XXXIII). In: Conference and Publications Management Department (DCPM). Vol. Assembly AU DEC 770 (XXXIII). Addis Ababa, Ethiopia.

Ahmed M-S, Broze L, Dabo-Niang S, Gharbi Z. 2022. Quasi-maximum likelihood estimators for functional linear spatial autoregressive models. In: Mateu J, Giraldo R, editors. Geostatistical functional data analysis. Hoboken, NJ: John Wiley & Sons Ltd.

Anselin L. 1995. Local indicators of spatial association-LISA. Geog Anal. 27:93–115.

Bhatt V, Tiwari N. 2014. A spatial scan statistic for survival data based on Weibull distribution. Stat Med. 33:1867–1876.

Bohorquez M, Giraldo R, Mateu J. 2016. Optimal sampling for spatial prediction of functional data. Stat Methods Appl. 25:39–54.

Caraballo H, King K. 2014. Emergency department management of mosquito-borne illness: malaria, dengue, and West Nile virus. Emerg Med Pract. 16:1–23.

Cucala L, Demattei C, Lopes P, Ribeiro A. 2013. A spatial scan statistic for case event data based on connected components. Comput Stat. 28:357–369.

Cucala L, Genin M, Lanier C, Occelli F. 2017. A multivariate Gaussian scan statistic for spatial data. Spatial Stat. 21:66–74.

Cucala L, Genin M, Occelli F, Soula J. 2019. A multivariate nonparametric scan statistic for spatial data. Spatial Stat. 29:1–14.

Cucala L. 2014. A distribution-free spatial scan statistic for marked point processes. Spatial Stat. 10:117–125.

Cuevas A, Febrero M, Fraiman R. 2004. An Anova test for functional data. Comput Statist Data Anal. 47:111–122.

Delicado P, Giraldo R, Comas C, Mateu J. 2010. Statistics for spatial functional data: some recent contributions. Environmetrics. 21:224–239.

Diggle P, Giorgi E, Chipeta M, Macfarlane SB. 2019. Tracking health outcomes in space and time: spatial and spatio-temporal methods. In: Macfarlane SB, AbouZahr C, editors. The Palgrave handbook of global health data methods for policy and practice. London: Palgrave Macmillan. p. 383–401.

Dwass M. 1957. Modified randomization tests for nonparametric hypotheses. Ann Math Statist. 28:181–187.

Fernández de Castro BM, Manteiga WG. 2008. Boosting for real and functional samples: an application to an environmental problem. Stoch Environ Res Risk Assess. 22:27–37.

Frévent C, Ahmed M-S, Dabo-Niang S, Genin M. 2023. Investigating spatial scan statistics for multivariate functional data. J R Stat Soc Series C: Appl Stat. 72(2):450–475.
Frévent C, Ahmed M-S, Marbac M, Genin M. 2021a. Detecting spatial clusters in functional data: New scan statistic approaches. Spatial Stat. 46:100550.
Frevent C, Ahmed M-S, Soula J, Smida Z, Cucala L, Dabo-Niang S, Genin M. 2021b. Multivariate and functional spatial scan statistics [R package HDSpatialScan version 1.0.2].
Getis A, Ord J. 1996. Local spatial statistics: an overview. In: Longley P, Batty M, editors. Spatial analysis: modelling in a GIS environment. New York: John Wiley & Sons. p. 261–277.
Giraldo R, Delicado P, Mateu J. 2010. Continuous time-varying kriging for spatial prediction of functional data: an environmental application. J Agricult Biol Environ Statist 15(1):66–82.
Giraldo R, Delicado P, Mateu J. 2012. Hierarchical clustering of spatially correlated functional data. Stat Neerl. 66:403–421.
Greene SK, Peterson ER, Kapell D, Fine AD, Kulldorff M. 2016. Daily reportable disease spatiotemporal cluster detection, New York City, New York, USA, 2014–2015. Emerg Infect Dis. 22:1808–1812.
Huang L, Kulldorff M, Gregorio D. 2007. A spatial scan statistic for survival data. Biometrics. 63:109–118.
Jung I, Cho HJ. 2015. A nonparametric spatial scan statistic for continuous data. Int J Health Geogr. 14:
Jung I, Kulldorff M, Klassen AC. 2007. A spatial scan statistic for ordinal data. Stat Med. 26:1594–1607.
Kulldorff M, Huang L, Konty KJ. 2009. A scan statistic for continuous data based on the normal probability model. Int J Health Geogr. 8: 58–58.
Kulldorff M, Huang L, Pickle L, Duczmal L. 2006. An elliptic spatial scan statistic. Stat Med. 25:3929–3943.
Kulldorff M, Nagarwalla N. 1995. Spatial disease clusters: detection and inference. Stat Med. 14:799–810.
Kulldorff M. 1997. A spatial scan statistic. Commun Statist Theory Meth. 26:1481–1496.
Lin Z, Lopes ME, Müller H-G. 2023. High-dimensional MANOVA via bootstrapping and its application to functional and sparse count data. J Am Stat Assoc. 118:177–191.
MAP. 2024. Malaria atlas project | home. MAP [Internet]. https://malariaatlas.org.
Meade MS, Emch M. 2010. Medical geography. New York: Guilford Press.
Menafoglio A, Secchi P, Dalla Rosa M. 2013. A Universal Kriging predictor for spatially dependent functional data of a Hilbert space. Electron J Statist. 7:2209–2240.
Monestiez P, Nerini D. 2008. A Cokriging method for spatial functional data with applications in oceanology. Contrib statist. 237–242.
Ramsay JO, Silverman BW, Springerlink (Online Service). 2005. Functional data analysis. New York, NY: Springer New York.
Smida Z, Cucala L, Gannoun A, Durif G. 2022. A Wilcoxon-Mann-Whitney spatial scan statistic for functional data. Computat Statist Data Anal. 167:107378.
Ternynck C. 2014. Spatial regression estimation for functional data with spatial dependency. J de la Société Française de Statistique. 155: 138–160. http://www.numdam.org/item/JSFS_2014__155_2_138_0/.
UN. 2024. World population prospects. United Nations [Internet]. https://population.un.org/wpp/.
Vandewalle V, Preda C, Dabo-Niang S. 2021. Clustering spatial functional data. In: Mateu J, Giraldo R, editors. Wiley series in probability and statistics. p. 155–174.
Waller LA, Gotway CA. 2004. Applied spatial statistics for public health data. Hoboken, NJ, USA: John Wiley & Sons, Inc.
Walter K, John CC. 2022. Malaria. JAMA. 327:597.
WHO. 2015. Guidelines for the treatment of malaria. 3rd ed. Geneva, Switzerland: World Health Organization. https://iris.who.int/handle/10665/162441.
WHO. 2020. The potential impact of health service disruptions on the burden of malaria: a modelling analysis for countries in Sub-Saharan Africa. Geneva: World Health Organization.
WHO. 2023. Malaria. World Health Organization [Internet]. https://www.who.int/news-room/fact-sheets/detail/malaria.
Wong DWS, Lee J. 2005. Statistical analysis of geographic information with ArcView GIS and ArcGIS. Hoboken, N.J.: John Wiley & Sons.