

Spatio-temporal dynamics of malaria vector niche overlaps in Africa

Eric Ali Ibrahim ^{a,b,*}, John Odindi ^b, Mark Wamalwa ^a, Henri E.Z. Tonnang ^{a,b,c}

^a International Centre of Insect Physiology and Ecology (ICIPE), PO Box 30772, Nairobi, Kenya

^b School of Agricultural, Earth, and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg 3209, South Africa

^c International Institute of Tropical Agriculture (IITA), PMB, 5320, Oyo Road, Idi-Oshe, Ibadan, Nigeria

ARTICLE INFO

Keywords:
 Vector ecology
 Cellular automata (CA)
 Vector control
 Environmental factors
 Primary vector
 Secondary vectors

ABSTRACT

Malaria remains a significant public health challenge in sub-Saharan Africa, with transmission heightened by the dynamics of primary and secondary mosquitoes infected with *Plasmodium* parasites. Regions where both vector types co-exist face heightened likelihood of intensified malaria transmission. Hence, understanding vectors' ecological interactions, especially their niche overlaps in geographic or environmental space, is crucial for targeted malaria control and elimination strategies. We employed a dynamic cellular automata (CA) model to map niche overlaps among primary (*Anopheles gambiae* complex, *An. fuscus* group) and secondary (*An. pharaoensis*, *An. coustani*) malaria vectors across Africa, using open-access environmental and vector occurrence datasets sourced from open-access geospatial portals, and spanning 1985 to 2021. Prior to modeling, we conducted exploratory data analysis (EDA) involving descriptive statistics, correlation and cluster analysis to glean insights into the relationships between the variables. Spearman correlation analysis revealed weak significant correlations ($|r| < 0.3$, $p\text{-value} < 0.001$) between environmental variables and vectors occurrence, while environmental variables exhibited strong intercorrelations. Furthermore, *An. gambiae* complex prevailed at higher elevations with a minimum relative humidity of 22 %, while secondary vectors prevailed at lower elevations with humidity $>38\%$ and temperatures above 20 °C. Our model, with accuracy exceeding 0.9 following validation, revealed expanding malaria vector niche overlaps across Africa, attributed to vectors expansion beyond their native regions. Such expanding vector niche overlaps predisposes numerous areas at risk of sustained and prolonged malaria transmission, underscoring the need for targeted malaria vector control interventions. Furthermore, dynamic modeling approaches, incorporating continuous data updates, captured ecological interactions accurately.

1. Introduction

Malaria remains a significant global health concern, leading to deaths, particularly in sub-Saharan Africa, which accounts for 95 % of all cases and 96 % of all deaths (WHO, 2022; WHO, 2023). Commonly, the disease acutely affects young children, pregnant women, travelers, and persons with compromised immune systems (CDC, 2021; Packard, 2007; WHO, 2022). Furthermore, malaria causes havoc on households, healthcare systems, education and economic development in endemic countries (World Bank, 2024). Additionally, malaria transmission and prevalence in humans is heightened by dynamics of mosquitoes infected with *Plasmodium* parasites (CDC, 2021; Lee et al., 2016; Mahdizadeh et al., 2019; Packard, 2007).

Malaria control has remained elusive despite targeted pharmaceutical and non-pharmaceutical public health interventions. For example,

socio-cultural and economic barriers prevent the effective use of proven intervention strategies such as long-lasting insecticide-treated nets (LLINs), and indoor residual spraying of insecticides (IRS) (Ibrahim et al., 2024a; Packard, 2007). Furthermore, the interplay between the biological, social, environmental and economic forces play a role in malaria incidence (Packard, 2007). Additionally, the occurrence of malaria vectors and the reproduction of malaria parasites are influenced by environmental variables and climatic changes (Diamond and Wilson, 1999). Hence, understanding the occurrence and spread of these vectors is crucial for mitigating malaria outbreaks (Joshi and Miller, 2021; Ndenga et al., 2011).

Malaria is transmitted by female *Anopheles* mosquitoes classified as primary and secondary vectors based on their transmission roles. Primary and secondary vectors are responsible for over 95 % and 5 % of malaria transmission respectively (Afrane et al., 2016). In Africa,

* Corresponding author.

E-mail address: eibrahim@icipe.org (E.A. Ibrahim).

primary vectors include *Anopheles gambiae* s.s, *An. coluzzii*, *An. arabiensis*, *An. funestus*, *An. moucheti* and *An. Nili*, while secondary vectors include *An. pharoensis*, *An. coustani*, *An. rivolorum*, *An. ziemanni* and *An. squamosus* (Afrane et al., 2016). However, recent trends in malaria transmission indicate that secondary vectors play significant role in sustaining malaria transmission, particularly in areas where primary vectors have been largely controlled (Mustapha et al., 2021). Moreover, there is a discernible trend of secondary vectors such *An. arabiensis* as assuming primary roles in transmission (Afrane et al., 2016; Mukama and Mwangi, 1989).

The growing role of secondary malaria vectors in transmission is linked to their exophagic and exophilic tendencies, reducing the efficacy of indoor-focused control strategies, thereby preserving their populations and enhancing their role in malaria spread (Antonio-Nkondjio et al., 2006b; Gross et al., 2022; Mustapha et al., 2021). Anthropogenic environmental changes and extensive insecticide use, which often suppress primary vectors, inadvertently allow secondary vectors to dominate (Afrane et al., 2016; Olabimi et al., 2021). Moreover, the shared habitats between primary and secondary vectors, alongside the latter's geographical expansion, sustain malaria transmission (Afrane et al., 2016; Olabimi et al., 2021). These niche overlaps occur when co-occurring species share part of the niche space (Tsafack et al., 2021). Therefore, a species niche is defined by its geographical or environmental space, including factors such as climate, terrain, and interactions with other organisms (Kroes, 1977; Polechová and Storch, 2008). To estimate species' niche, their structures and dynamics, researchers use geo-referenced presence/absence data and environmental variables as predictors. In particular, under climate change scenarios, the distributions of malaria vector species may shrink or expand, highlighting the dynamic nature of ecological niches and the potential impact of environmental changes. Recently, increased indoor detection of secondary malaria vectors, suggest a potential shift towards endophagic and endophilic habits (Kamau et al., 2006; Saily et al., 2023). In regions where both primary and secondary vectors co-exist, the likelihood of intensified malaria transmission increases, necessitating a thorough assessment of niche overlaps using metrics such as Schoener's D statistic, Warren's I values, Pianka's niche overlap index and Jaccard similarity index (Broennimann et al., 2012). Areas with high niche overlaps potentially intensify malaria transmission (Moffett et al., 2007), with minimal competition among co-existing vectors (Collins et al., 2019; Kirby and Lindsay, 2009; Koenraadt et al., 2004; Pascual-Rico et al., 2020).

To define the ecological niches of malaria vectors, various studies have utilized niche models that incorporate environmental factors, predominantly using the maximum entropy algorithm (Adeogun et al., 2023; Ayala et al., 2009; Kulkarni et al., 2010; Moffett et al., 2007; Olabimi et al., 2021). However, these studies have paid less attention to the spatio-temporal dynamics of overlaps between primary and secondary vectors. Additionally, most of these models are static, failing to reflect the dynamic nature of vector distributions. The initial distribution of these species, influenced by niche characteristics and the ability to disperse over distances, plays a critical role in determining their future spatial patterns (Plotnick and Gardner, 2002). Given the inherent uncertainties and non-linearities in species distribution data, this study utilizes a cellular automata (CA) approach to effectively model niche overlaps. Prior, a database search on PubMed and web of science using key words "niche overlap", "malaria" and "cellular automata", found no prior studies applying CA in mosquito ecological niche modeling.

While machine learning (ML) algorithms such as random forest (RF) and artificial neural networks (ANN) result in strong predictive performance, they often struggle to capture the dynamic spread of phenomena over time and space (Yu and Wang, 2024). This capability is inherent to the CA model, which models the spread of a phenomenon from an initial location to surrounding neighborhoods, reporting its presence at each timestep. Furthermore, predictions of ML models are based on the training scope (Povalej et al., 2005), and are often considered

"blackboxes" (Rahman et al., 2021; Visser et al., 2022). Similarly, while ensemble techniques have been explored, they too can be complex and challenging to interpret. Consequently, it is essential to consider the trade-offs between predictive power and interpretability while using ensemble techniques (Ramampiandra et al., 2023). Subsequently, similar studies have used ML models to define transition rules for the CA based model (Charif et al., 2012; Xu et al., 2022; Zambrano-Asanza et al., 2023).

The CA approach, is effective in modeling the spread of phenomena influenced by environmental factors, particularly due to its capacity to accurately integrate initial distribution points (Guimapi et al., 2016). However, widely dispersed or poorly reported spatio-temporal data records during the initial stages of a study, can limit CA model's ability to capture the underlying spatio-temporal dynamics. Additionally, the increasing spread of other vector species, such as *Anopheles stephensi*, often facilitated by transportation and not necessarily dependent on initial distribution points, poses a challenge. To address these issues, our study enhances the CA model by incorporating data collected at every time-stamp beyond the initial distributions and environmental variables. This methodological advancement aims to improve accuracy by integrating temporal data points, thereby effectively capturing the evolving dynamics of vector occurrence and niche overlaps. Our study is guided by three core objectives: firstly, to establish the extent of niche overlaps among primary and secondary malaria vectors within Africa, secondly, to investigate the environmental variables that affect the vectors' occurrences and lastly, refine our modeling approach to account for continuous reporting of vector occurrences, to ensure our model remains relevant and accurate over different time frames. Ultimately, the study seeks to provide a more comprehensive understanding of the vector's spatio-temporal distribution and changes.

2. Material and methods

2.1. Study area

The focus of our study was the African continent because it bears significant malaria burden (WHO, 2023). With over forty diverse vector species known to facilitate the transmission of *Plasmodium* to humans (Antonio-Nkondjio et al., 2006a; Ntabi et al., 2024), Africa remains a key region for studying residual malaria transmissions.

2.2. Datasets and sources

We sourced data from the Malaria Atlas Project (MAP) (<https://malariaatlas.org/>). The data includes extensive spatial and temporal records on vector occurrence, ecological traits, and insecticide resistance patterns from 1974 to 2021. Our analysis focused on two primary (*An. gambiae* complex and *An. funestus* group) and two secondary (*An. pharoeensis* and *An. coustani*) vectors, due to their roles in malaria transmission (Ayala et al., 2009; Mukama and Mwangi, 1989). The entomological data contained a total of 26,331 records for targeted vector species, 25,826 from 1974 to 2017 and 505 from 2018 to 2021. The temporal distribution is presented in Appendix C while spatio-temporal distributions for selected years are presented in the results section (see maps).

The modeling of niche overlaps was based on various environmental and land use land cover variables that impact the distribution and ecology of the vectors (Agyekum et al., 2021a; Kibret et al., 2019; Minakawa et al., 2005). The variables are presented in Table 1.

The selection of these variables was informed by literature (Appendix D). Prior to undertaking the analysis, we pre-processed the vector occurrence data, excluding entries lacking spatial and temporal information. Next, using geo-referenced coordinates of the vector occurrences, we extracted corresponding environmental variables from various sources (refer to Appendix A). This process linked the specific locations of the vector occurrence with corresponding environmental

Table 1
Environmental and land use land cover related variables used.

Variable	Unit of measurement
1 Temperature	°C
2 Relative humidity	%
3 Irrigation proximity	m
4 Elevation	m
5 Build-ups proximity	m
6 NDVI	—
7 Shrubs proximity	M
8 Climate moisture index	Kg/m ²
9 Cloud area fraction	%
10 Wind speed	m/s
11 Potential evapotranspiration	Kg/m ²
12 Precipitation	Kg/m ²
13 Solar radiation	w/m ²
14 LULC classes	A factor, hence no measurement unit

Kg – kilogram, m = meters, C = celcius, w = energy transferred, – no units, LULC - land use land cover, NDVI - Normalized difference vegetation index.

data. After integrating these datasets, we proceeded with the analysis. A detailed discussion of the data processing is presented in the section below.

2.3. Data processing

We utilized various open-access geospatial portals to download different covariates (Table 1 and Appendix A) in raster formats. These covariates were sourced monthly or yearly, some with different spatial and temporal resolutions between 1985 and 2021 covering Africa. Due to availability of several Geographic Information System (GIS) layers on the same covariates, we selected the ones for our use based on the following data requirements; high spatial resolution of less or equal to 5×5 km (25 km^2), monthly temporal resolution, spatial coverage (at least Africa), temporal range covering most of the study period, and open access. Further, the covariates were reprojected into World Geodetic System 1984 (WGS84) and resampled to 5×5 km (25 km^2) resolution using the ‘resampling’ function embedded in the ‘rasterio’ package in Python (Gillies, 2013) to harmonize the different covariates. We harmonized all variables using mean resampling approach, except for irrigation, shrubs, built-ups, and LULC, which were harmonized using mode, and sum for precipitation. Furthermore, we segregated and computed relative distances using the euclidean distance function of the spatial analyst tool for the build-up, irrigation, and shrubland land use/cover classes in ArcGIS Pro version 3.3.0 (ESRI, 2024). This resulted in a continuous raster surface with pixels closest to respective land use/cover classes assigned a minimum distance of 0 m and a distance for pixels away from the class area increasing incrementally. Using the extraction function of the ‘rasterio’ (Gillies, 2013) and ‘rioarray’ packages (Snow, 2019) in Python, we overlaid the raster files with the vector occurrence records and extracted respective information, creating a unified record of each covariate at a location for exploratory analysis. Furthermore, using the ‘geopandas’ package (Jordahl, 2014) in Python, we created a 5×5 km (25 km^2) grid for the entire continent of Africa and computed centroid (central points) for each grid, which we used to extract information for the subset of covariates retained after the exploratory data analysis. Following this extraction of variables in monthly time steps, we merged each variable sequentially for the entire study period using ‘pandas’ (The pandas development team, 2020) and ‘numpy’ packages (Harris et al., 2020) in python. This resulted in spatio-temporal data-frame, which was then used to model the spatio-temporal dynamics of malaria overlaps across the continent over time in raster format projected into WGS84 in R using the ‘raster’ package (Hijmans, 2024). The data with varying frequency of timesteps and temporal periods were standardized, that is, for covariates with a starting year after 1985, the value from the earliest reported year was used for all previous years. For covariates ending before 2021, the value from the latest reported year

was used for all subsequent years. Similarly, for data with yearly records, such recorded data were used for each month within such year. Data processing steps are presented in Fig. 1.

2.4. Data analysis

Data analysis involved exploratory data analysis (EDA) and modeling the niche overlaps among *An. gambiae* complex, *An. funestus* group, *An. pharoensis* and *An. coustani*. EDA played a critical role in establishing collinearities among the environmental variables, underlying trends and relationships, ultimately informing the vector niche overlap modeling.

2.4.1. Exploratory data analysis (EDA)

Exploratory data analysis (EDA) involved descriptive statistics, correlation tests, and cluster analysis to reveal relationships between environmental variables and niche overlaps in malaria vector species, providing valuable insights for developing transition rules for the CA model. Correlation tests and cluster analysis, visualized through correlation plots using ‘metan’ package, and dendograms respectively, provided clarity on the relationships, collinearities and clustering of predictor variables. Exclusion or inclusion of predictor variables was based on correlation coefficient (r) ($|r| < 0.7$), clustering outcome, and insights from literature. In Dormann et al. (2013), a $|r| > 0.7$ distorts model prediction and estimation, and the use of ecological understanding was critical in variable selection (Dormann et al., 2013).

To address multicollinearity, we eliminated one variable from each cluster, retaining the one with stronger correlation to vector occurrence. Highly correlated predictor variables were dropped while retaining only one of the variables, with the highest correlation coefficient with the vector species occurrence. However, predictor variables that were weakly or moderately correlated (correlation coefficient $|r| < 0.7$) were retained. Descriptive statistics, on the other hand, provided insights into the vectors’ niche requirements, including their occurrence ranges and distribution patterns around mean values, informing the formulation of transition rules. The EDA was implemented using R version 4.0.5 (CoreTeam, 2020).

2.5. Key assumptions

Following selection of the variables, and prior to modeling malaria vector niche overlaps, we made several assumptions:

1. We set 1985 as the baseline year, treating vector occurrence records prior to this year as initial data points. While there are reports of malaria vectors dating back before the 20th century (Afrane et al., 2016), spatial records are not robustly documented before 1985.
2. We estimated the average dispersal rate of vectors at 10 km per month. Although there is no universally accepted rate for *Anopheles* vector dispersal, this assumption considers resource availability, species-specific environmental adaptability, transportation influences, wind speed, and topography (Manga et al., 1993; Ntabaliba et al., 2023).
3. Populating for months within a year with data values for that particular year, for the environmental variables which had only yearly temporal resolution, was assumed to effectively capture the spatio-temporal dynamics of the vectors’ niche overlaps.
4. We postulated that there is no competition for resources among the vectors at the adult stage (Collins et al., 2019), aligning with existing literature evidence.

2.5.1. Cellular automata (CA) approach

The CA approach simulates the distribution of phenomena over time, based on the concept of neighborhood interactions. This methodology involves segmenting space into a grid of cells, each characterized by a

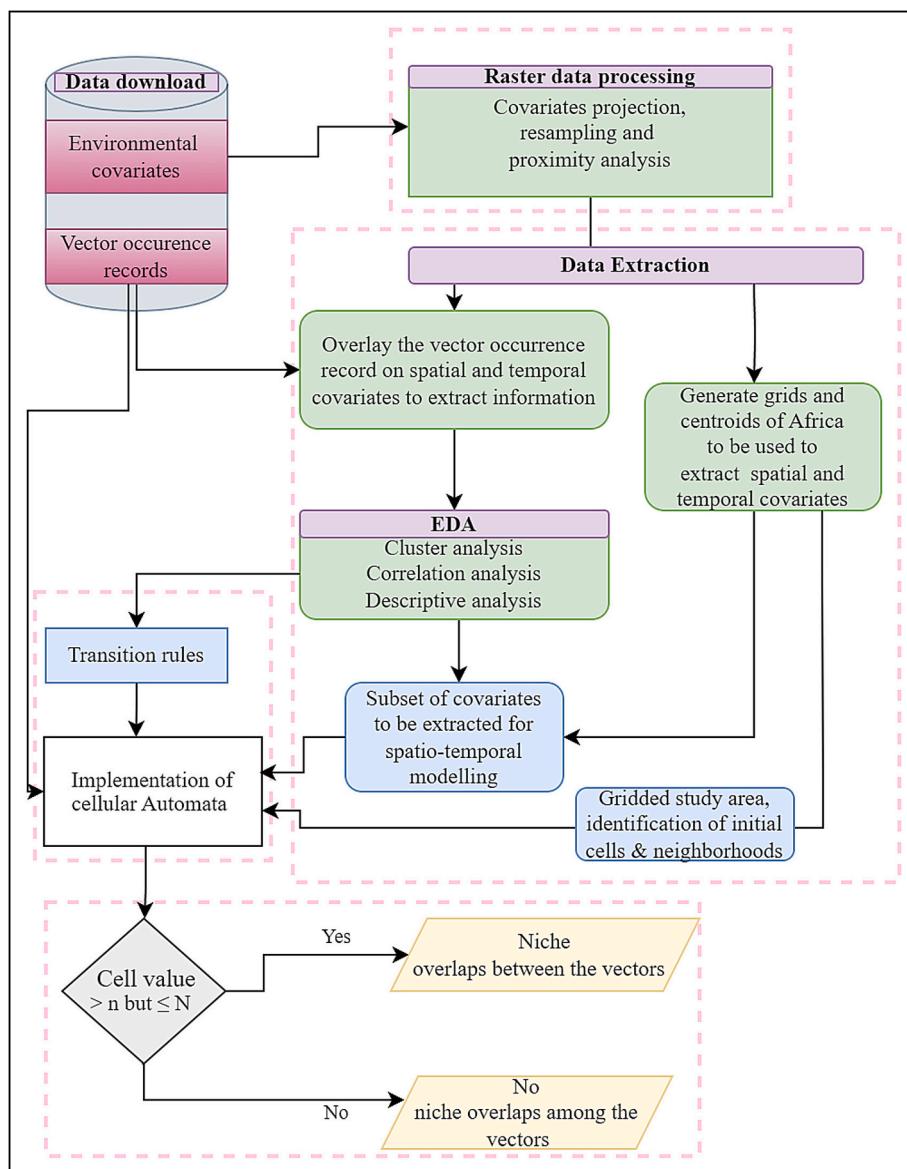


Fig. 1. Summary flowchart of the data processing and cellular automata model implementation. n is the highest value assigned to the species, N denotes the highest value assigned to the niche overlaps.

finite state that evolves over time based on the states of its neighboring cells and predefined transition rules. Suppose the cell with initial state, at time $t = 0$ is defined as $S_{ij}^{t=0}$, then the state of cell (i,j) at $t + \Delta t$, (ie $S_{ij}^{t+\Delta t}$) can be expressed as follow:

$$S_{ij}^{t+\Delta t} = f(S_{ij}^t, N_{ij}^t, T) \quad (1)$$

where;

S_{ij}^t is the state of cell(s) at defined time t ,

N_{ij}^t is the state of cells the neighborhood of cell S_{ij}^t ;

T is the defined transition rules; and.

Δt is the time-step.

The CA approach has been used successfully to model invasion and spread of *Tuta absoluta* in Africa (Guimapi et al., 2016) and to characterize spatio-temporal dynamics of phenotypic insecticide resistance in malaria vectors (Ibrahim et al., 2024b), demonstrating robust performance in both cases. In this study, the CA approach specifically sought to characterize niche overlaps among malaria vector species. This step was realized by identifying grid cells concurrently colonized by multiple

vector species, effectively capturing the spatio-temporal dynamics of their distribution. A significant challenge addressed by the CA model is the initial sparsity and poor quality of reporting in the early stages of our temporal data. To overcome this, we enhanced the CA model to facilitate continuous updates at each time increment $t + \Delta t$. This enhancement extends beyond the reliance on predefined transition rules by incorporating occurrences of vector species reported at different times throughout the duration of the study. Such integration of real-time data updates is designed to enhance the model's accuracy and predictive capabilities in delineating vector niche overlaps. This methodological improvement is anticipated to provide a more dynamic and detailed perspective of the ecology and interaction patterns of malaria vectors. For novelty, we extended the model's transition rule to be a function of two distinct terms;

- "V" this term represents the traditional transition rules resulting from the relationship between the environmental variables and occurrence of the species, and
- $W_{ij}^{t+\Delta t}$ indicates the newly acquired information about the vectors' occurrence at a specific cell (i,j) at updated time $t + \Delta t$.

The integration of these two terms lead to an updated state of a cell in CA model, with improved prediction capability. The revised equation, reflecting this advancement is expressed as follows:

$$S_{ij}^{t+\Delta t} = f(S_{ij}^t, N_{ij}^t, T(V, W_{ij}^{t+\Delta t})) \quad (2)$$

where;

S_{ij}^t is the state of cell(s) at defined time t ,

N_{ij}^t is the state of cells the neighborhood of cell S_{ij}^t ;

T is the transition rules which is a function of rules formulated based on relationship between variables (V), and $W_{ij}^{t+\Delta t}$ is the term for incorporating into the model new records on occurrence of the malaria vectors.

Δt is the time-step;

With the adjustments made to our CA model, our next step was to define the niche overlap, as the state of each cell. In this context, the niche overlap, indicating either presence or absence, is quantitatively assessed using several statistical measures including Schoener's D statistic, Warren's I values, Pianka's niche overlap index, and the Jaccard similarity index, all yield values ranging from 0 (no overlap) to 1 (complete overlap).

Schoener's D statistic, a widely recognized method for quantifying niche overlap in SDMs (Collins et al., 2017) is defined as (Schoener, 1970):

$$D = 1 - \frac{1}{2} \sum_{k=1}^n |p_{ik} - p_{jk}| \quad (3)$$

where, D represents Schoener's overlap index, n is the number of niche dimensions, and p_{ik} and p_{jk} are the proportions of the i^{th} and j^{th} species in the k^{th} niche dimension, respectively. The Schoener's D statistic values range from 0 to 1, where:

- $D = 0$: Indicates no overlap, suggesting that the species occupy completely distinct ecological niches.
- $0 < D < 1$: Signifies partial overlap of niches. The closer the D to 1, the more significant the overlap between the species' niches.
- $D = 1$: Denotes complete niche overlap, implying that the species share identical ecological niches.

A higher Schoener's D statistic value implies a greater degree of niche overlap between species, while a lower value signifies less overlap. This statistic was instrumental in assessing the extent to which malaria vector species share ecological niches. In our study, D was zero where no species niche overlap occurred, and 1 where there was niche overlap.

2.5.2. Model implementation and validation

Based on EDA, we formulated transition rules for governing the dynamics of each niche overlap. The rules were employed in an IF-THEN format, that is, IF ($i_1 < X_1 < j_1 \& \dots \& i_n < X_n < j_n$), THEN (*niche overlap*). Here, i and j represented the lower and upper bounds of respective environmental variable X_m , $m = 1, 2, \dots, n$. n signifies the total number of environmental variables considered. The rules were formulated for all the possible vector niche overlaps and are presented alongside the shared code.

After formulating the transition rules, we imported the gridded African raster, the covariates, and vector occurrence data into R. We converted the covariates and vector occurrence data to spatial point dataframes, projected them to WGS84 and arranged them sequentially from 1985 (y_j) to 2021 (y_m), where y_j and y_m denote the initial and end years of the study period. This setup allowed for the sequential implementation of the model over time. The process involved rasterizing the covariates and vector occurrence dataframes at each timestep using 'raster' package in RStudio, then determining the state of neighborhood grids, at every timestep, based on the transition rules.

During model implementation, each cell was assigned an initial

value $i = 1, 2, \dots, n$ corresponding to specific vector species, with additional values $n + 1, n + 2, \dots, N$ indicating all possible niche overlaps and $i = 0$ for cells without vectors. Here, n and N were 4 and 15, respectively. This categorization was vital for differentiating states of cells at every time-step (details in Appendix B). Finally, the CA model was implemented, considering the state of neighboring cells using an extended Moore neighborhood, and vector occurrences at monthly intervals, as depicted in Fig. 1. Following model implementation, we validated its performance using data from 2018 to 2021. This involved assessing how well the model's predicted niche overlaps corresponded with the actual shared niche. The model was implemented and validated using R version 4.0.5 (CoreTeam, 2020).

Following model implementation, we visualized the resulting raster outputs using ArcGIS Pro version 3.3.0 (ESRI, 2024). The outputs identified niche-overlaps for the four-vector species in Africa where 1 = *An. gambiae* complex; 2 = *An. funestus* group; 3 = *An. pharoensis* and 4 = *An. coustani*, while values between 5 and 15 indicate the regions with malaria vector niche overlaps (Appendix B).

3. Results

3.1. Exploratory data analysis

3.1.1. Correlation analysis

Spearman correlation, demonstrated significant correlations between all environmental variables considered in this study and the occurrence of all the four malaria vector species (p -value <0.001), except relative humidity (Fig. 2). However, the correlation coefficients ($|r| < \pm 0.3$), indicated that these correlations are relatively weak (as shown in Fig. 2). Additionally, the analysis revealed correlations among the environmental variables, notably strong positive correlation between precipitation and potential evapotranspiration ($r = 0.71$, p -value <0.001), and a moderately strong positive correlation between elevation and temperature range ($r = -0.54$, p -value <0.05) (Fig. 2).

3.1.2. Cluster analysis

We observed significant clustering among various environmental variables. These include near-surface wind speed and solar radiation, shrubs and irrigation, potential evapotranspiration and precipitation, built-up areas and the NDVI, as well as the climate moisture index and cloud area fraction (Fig. 3). These clusters informed the selection of environmental variables for modeling malaria vector niche overlaps.

A subset of relevant variables incorporated into the CA model, based on a comprehensive review of relevant literature and findings from correlation and cluster analyses, included temperature range, relative humidity, precipitation, elevation, NDVI, land use land cover (LULC), built-up areas, and irrigation. For highly correlated and clustered pairs of predictor variables, we retained only one of the variables. For example, relative humidity was selected over cloud fraction area and climate moisture index (Fig. 3). These variables were chosen for their known influence on malaria vector habitats and behaviors, ensuring our model comprehensively reflects the factors impacting vector distribution and niche overlaps.

3.1.3. Descriptive statistics

The descriptive analysis of data yielded notable insights into the habitat preferences of different malaria vector species. Our findings revealed that *An. gambiae* complex and *An. funestus* group were predominant at higher elevations compared to *An. pharoensis* and *An. coustani*. Specifically, *An. gambiae* complex was identified in areas with relatively higher elevations and the lowest recorded monthly average relative humidity (22%). Conversely, the other vector species, including the secondary vectors, were more prevalent at lower elevations, with monthly average relative humidity exceeding 38%. Additionally, the monthly average temperature recorded at sites with secondary vectors averaged above 20 °C, suggesting their preference for warmer

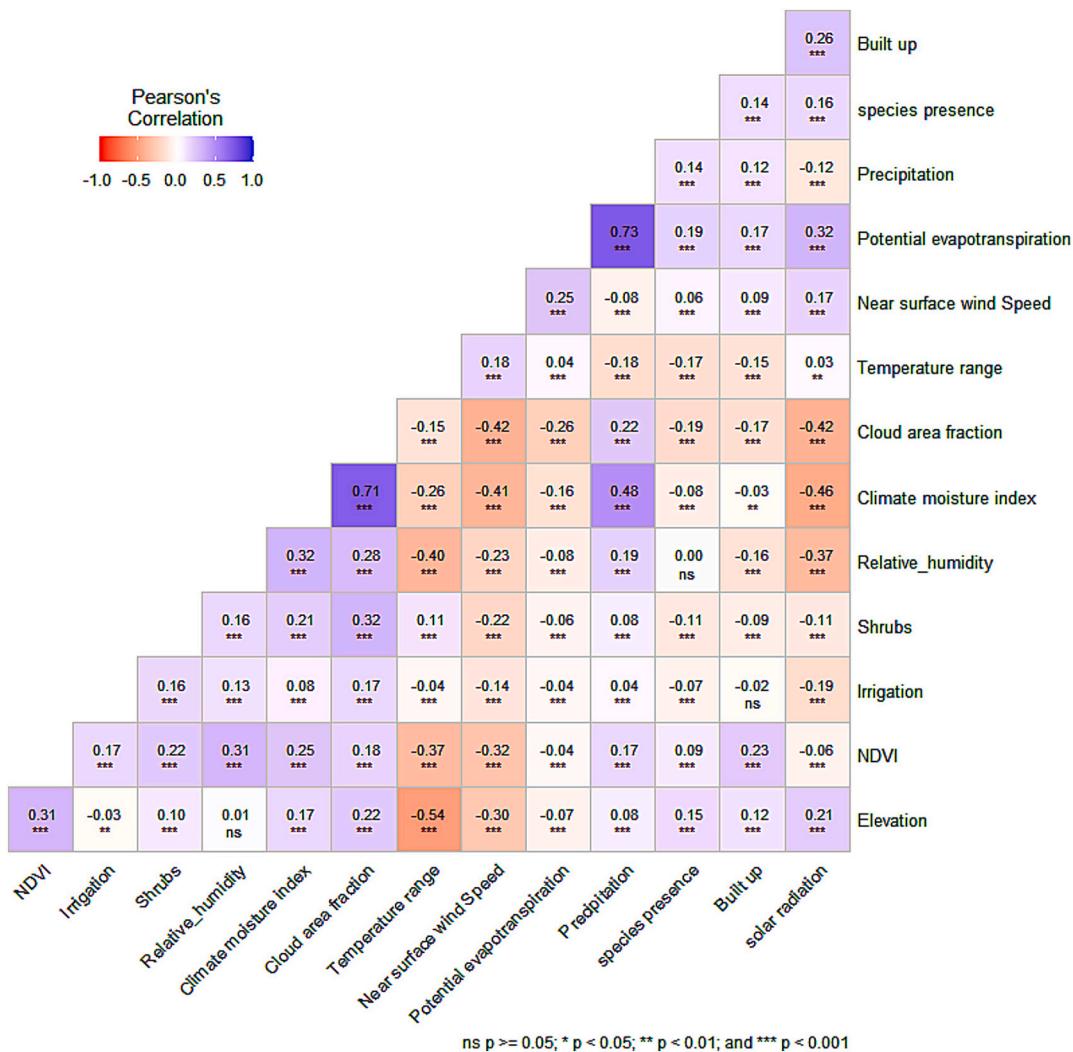


Fig. 2. Results of correlation between the study variables.

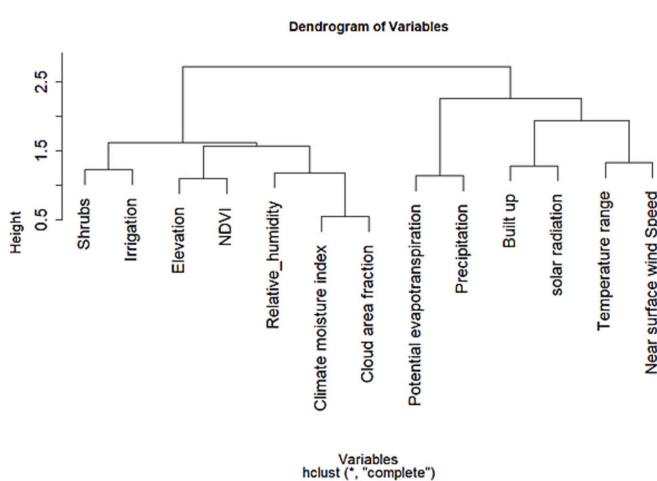


Fig. 3. Results of cluster analysis using dendrogram.

conditions. Our results also highlighted differences in proximity to human settlements between primary and secondary vectors. Secondary vectors were typically found at approximately 140 m from built-up areas, whereas primary vectors were generally located further. These spatial patterns, detailed in Table 2, underscore the differing ecological

niches and behavior between primary and secondary malaria vectors.

Additionally, descriptive analysis of the ranges of environmental conditions at which niche overlaps between vectors occurred is presented in Table 3. These were key in formulating the transition rules.

3.2. Validation of the niche overlap model results

We observed a high degree of alignment between the predicted niche overlaps and the actual occurrences of malaria vectors when model-generated maps were overlayed onto vector occurrence data (Fig. 4). Notably, the model effectively captured vector distribution trends, even in isolated regions like Madagascar, with an accuracy exceeding 0.9. This successful validation underscores the importance of regularly updating the model with new information, such as recent vector occurrence data, to overcome challenges associated with inadequate reporting and evolving vector dynamics. The model's ability to reflect current trends and distributions enhances its reliability as a tool for understanding and predicting malaria vector behavior.

The accuracy metrics derived from model validation indicated that the model effectively captured niche overlaps between primary, secondary, or both primary and secondary vectors with high precision.

3.3. Niche overlaps between primary and secondary vector species

The visualization of niche overlaps between primary and secondary

Table 2
Summary statistics.

Malaria vectors		Built up	Temperature range	Elevation	Irrigation	NDVI	Relative humidity	solar radiation	cloud fraction area	precipitation
<i>Anopheles gambiae</i> complex	Minimum	0.000	18.233	0.000	0.000	-0.090	22.470	10.966	8.4	0.000
	1st quantile	3.606	29.200	55.000	0.000	0.366	53.310	17.286	282.3	6690
	Median	12.000	30.820	301.000	9.000	0.488	60.720	19.258	388.9	19,235
	Mean	18.076	30.940	496.500	44.200	0.473	59.640	19.223	389.7	60,276
	3rd quantile	24.042	33.600	832.000	35.900	0.588	64.270	21.175	487.8	57,458
	Maximum	206.940	37.740	2804.000	1467.200	0.851	89.730	27.341	799.9	2,138,871
<i>Anopheles funestus</i> group	Minimum	0.000	21.290	0.000	0.000	-0.079	40.500	10.966	10.3	0.000
	1st quantile	7.071	28.320	214.000	0.000	0.449	57.640	18.709	260.9	12,350
	Median	18.385	29.630	592.000	7.000	0.541	61.090	20.221	316.9	40,526
	Mean	26.201	29.710	686.500	35.610	0.516	61.360	19.927	351.0	87,956
	3rd quantile	35.341	31.100	1159.000	30.270	0.609	64.810	21.289	434.2	130,544
	Maximum	202.800	37.120	2527.000	1467.210	0.851	88.610	27.023	811.5	2,138,871
<i>Anopheles pharoensis</i>	Minimum	0.000	24.030	0.000	0.000	-0.134	38.140	11.899	2.6	0.000
	1st quantile	1.000	29.110	7.000	0.000	0.270	49.270	19.315	124.0	155
	Median	9.899	30.960	299.000	0.000	0.438	51.530	21.184	246.1	7480
	Mean	15.439	31.850	421.200	17.870	0.394	54.360	21.093	253.6	42,139
	3rd quantile	20.591	34.810	855.000	5.000	0.507	57.870	22.797	372.3	46,387
	Maximum	137.535	38.060	2031.000	485.370	0.793	85.340	29.284	707.6	876,076
<i>Anopheles coustani</i>	Minimum	0.000	20.300	0.000	0.000	-0.079	44.850	10.966	35.2	61
	1st quantile	3.606	28.830	208.200	0.000	0.408	54.050	17.286	286.8	9388
	Median	15.554	29.940	410.000	4.500	0.510	59.690	19.258	360.8	53,724
	Mean	18.279	30.290	745.600	49.160	0.488	60.010	19.223	360.9	126,665
	3rd quantile	27.221	31.150	1177.000	65.000	0.578	64.620	21.175	428.8	150,633
	Maximum	84.345	37.290	2056.000	468.110	0.784	84.610	27.341	792.0	2,138,871

NDVI- normalized difference vegetation index.

Table 3
The upper and lower bounds of the environmental conditions required for niche overlaps between vectors to occur.

Niche overlaps between vectors		Built up	Temperature range	Elevation	Irrigation	NDVI	Relative humidity	Solar radiation	Cloud fraction area	Precipitation
1 & 2	Minimum	0.000	21.290	0.000	0.000	-0.079	40.500	10.966	8.4	0.000
	Maximum	202.800	37.120	2527.000	1467.200	0.851	88.610	27.023	799.9	213,887.1
1 & 3	Minimum	0.000	24.030	0.000	0.000	-0.090	38.140	11.899	8.4	0.000
	Maximum	137.535	37.740	2031.000	485.370	0.793	85.340	27.341	707.6	87,607.6
1 & 4	Minimum	0.000	20.300	0.000	0.000	-0.079	44.850	10.966	35.2	61
	Maximum	84.345	37.290	2056.000	468.110	0.784	84.610	27.341	792.0	213,887.1
2 & 3	Minimum	0.000	24.030	0.000	0.000	-0.079	40.500	11.899	10.3	0.000
	Maximum	137.535	37.120	2031.000	485.370	0.793	85.340	27.023	707.6	87,607.6
2 & 4	Minimum	0.000	21.290	0.000	0.000	-0.079	44.850	10.966	35.2	6.1
	Maximum	84.345	37.120	2056.000	468.110	0.784	84.610	27.023	792.0	213,887.1
3 & 4	Minimum	0.000	24.030	0.000	0.000	-0.079	44.850	11.899	35.2	6.1
	Maximum	84.345	37.290	2031.000	468.110	0.784	84.610	27.341	707.6	87,607.6
1, 2 & 3	Minimum	0.000	24.030	0.000	0.000	-0.079	40.500	11.899	10.3	0.000
	Maximum	137.535	37.120	2031.000	485.370	0.793	85.340	27.023	707.6	87,607.6
1, 2 & 4	Minimum	0.000	21.290	0.000	0.000	-0.079	44.850	10.966	35.2	6.1
	Maximum	84.345	37.120	2056.000	468.110	0.784	84.610	27.023	792.0	213,887.1
1, 3 & 4	Minimum	0.000	37.290	0.000	0.000	-0.079	44.850	11.899	35.2	6.1
	Maximum	84.345	37.290	2031.000	468.110	0.784	84.610	27.341	707.6	213,887.1
2, 3 & 4	Minimum	0.000	24.030	0.000	0.000	-0.079	44.850	11.899	35.2	6.1
	Maximum	84.345	37.120	2031.000	468.110	0.784	84.610	27.023	707.6	87,607.6
1,2,3 & 4	Minimum	0.000	24.030	0.000	0.000	-0.079	44.850	11.899	35.2	6.1
	Maximum	84.345	37.120	2031.000	468.110	0.784	84.610	27.023	707.6	213,887.1

Vectors; 1 = *An. gambiae* complex, 2 = *An. funestus* group, 3 = *An. pharoensis*, 4 = *An. coustani*.

malaria vectors in sub-Saharan Africa, as depicted in Fig. 5, demonstrates a notable spatial-temporal increase in overlaps. We observed that the niche overlaps among the three vector species were particularly pronounced in the lower western and eastern regions of Africa, as well as parts of central Africa. Significantly, countries like Nigeria, Uganda, and Mozambique, which account for a major share of malaria cases and deaths in sub-Saharan Africa, exhibited overlaps involving more than two vector species, encompassing both primary and secondary types. In the Democratic Republic of Congo (DRC), persistent niche overlaps between *Anopheles gambiae* complex and *Anopheles funestus* group were consistently observed across most regions during the study period. These

findings underscore the importance of vector niche overlaps in contributing to the malaria burden in these areas. Additionally, our analysis identified the coastal regions of the Horn of Africa as areas with significant niche overlaps between *Anopheles gambiae* complex and *Anopheles pharoensis*. This highlights the dynamic and evolving nature of vector habitats and interactions, with implications for malaria transmission patterns and control strategies.

4. Discussion

Understanding the niche overlaps between primary and secondary

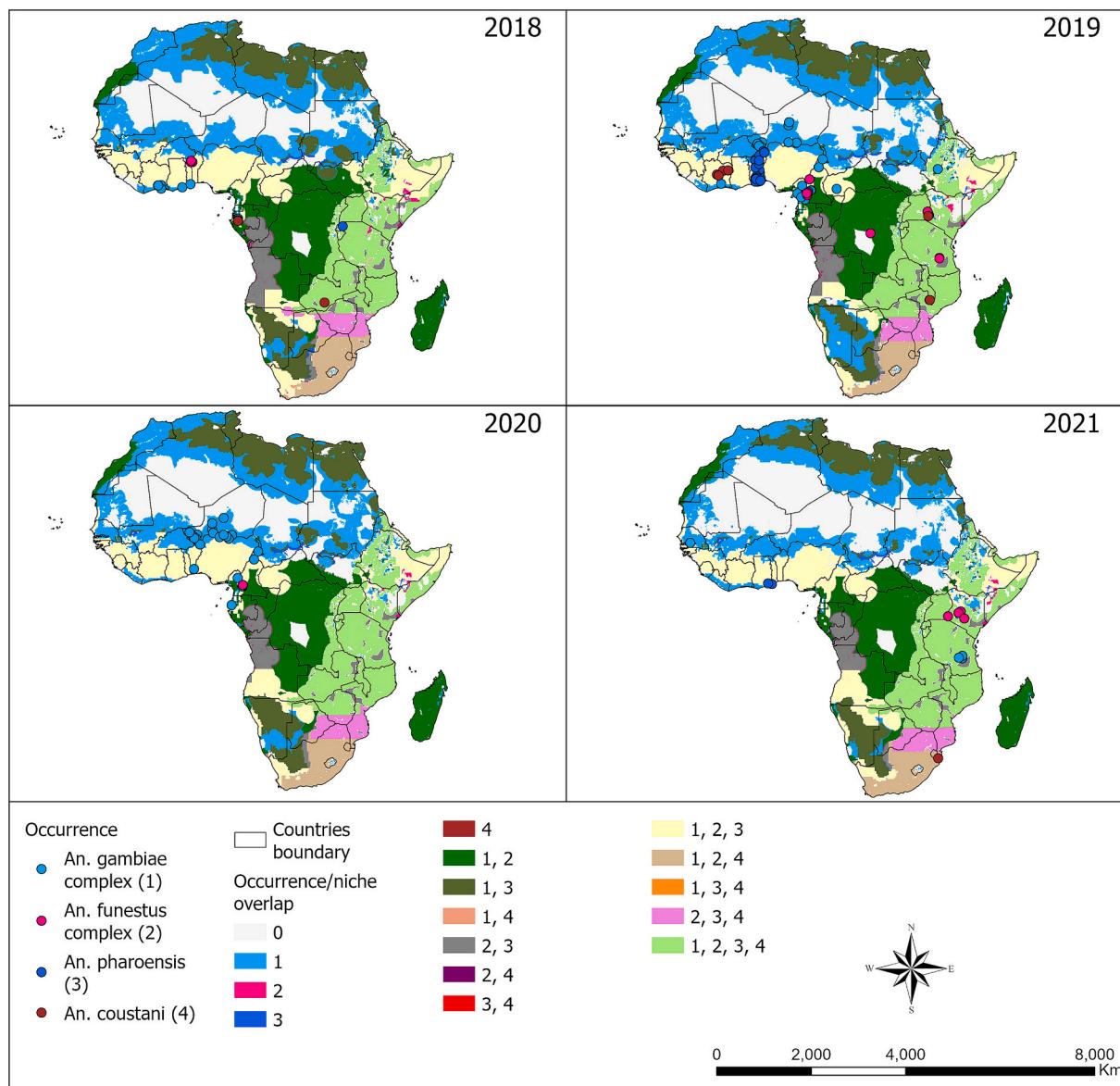


Fig. 4. Spatio-temporal dynamics of the niche overlaps between primary and secondary malaria vectors in Africa. Zero (0) denotes absence of vectors while 1, 2, 3 and 4 denote *An. gambiae* complex, *An. funestus* group, *An. pharoensis* and *An. coustani* respectively. The niche overlaps between vectors are denoted by the conjoint values of individual vectors.

malaria vectors, as well as their occurrences, is critical in pinpointing regions at elevated risk of sustained and prolonged malaria transmission. These regions often serve as hotspots where malaria can persist or resurface, thus becoming key focal points for mitigation and elimination efforts. Knowledge of these overlaps is instrumental for strategic planning and implementing integrated vector control measures, which can significantly mitigate the disease's impact. Tailored interventions, such as indoor residual spraying, and larval source management, are more effective when aligned with the specific behaviors and habitats of the vectors involved. To achieve this, our study was structured into two distinct phases of modeling that ensured a thorough and nuanced understanding of vector dynamics, crucial for effective malaria vector control.

The initial phase involved a comprehensive EDA, which revealed weak correlations between various environmental factors and the occurrence of malaria vector species. Despite these weak correlations, significant positive or negative correlations were observed among different environmental factors. A likely reason for the weak correlations with vector occurrence is the existence of non-linear relationships

between the environmental factors and the vector species. For example, temperature's impact on ectotherms is known to be non-linear, where their performance improves to an optimal temperature, but declines beyond that threshold (Agyekum et al., 2021b). This phenomenon might mirror the behavior of malaria vectors, where certain environmental conditions are favorable only to a certain extent, after which they could hinder vector development or survival. These findings underscore the complexities and uncertainties inherent in the relationships between species and environmental variables. The observed weak correlations might be a result of these complexities, suggesting that simple, linear associations might not fully capture the intricate ecological interactions at play. This understanding is essential, as it indicates that straightforward correlations may not effectively represent the true dynamics of interactions between malaria vectors and their environments, highlighting the need for more complex and adaptive modeling approaches.

The second phase highlighted specific environmental conditions under which niche overlaps between primary (*An. gambiae* complex and *An. funestus* group) and secondary (*An. pharoensis* and *An. coustani*) malaria vectors are likely to occur. Understanding these conditions is

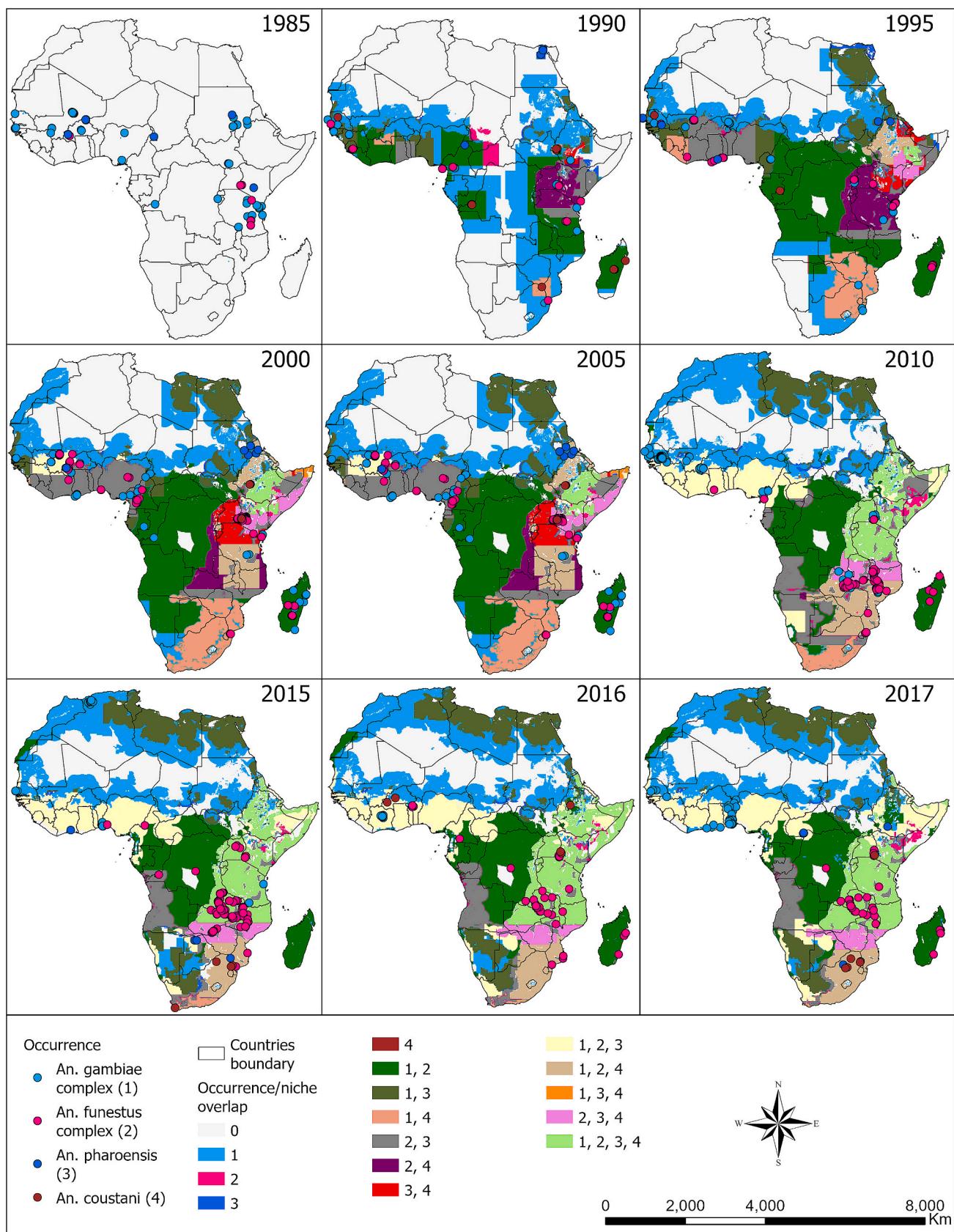


Fig. 5. Spatio-temporal dynamics of the niche overlaps between primary and secondary malaria vectors in Africa. Zero (0) denotes absence of vectors while 1, 2, 3 and 4 denote *An. gambiae* complex, *An. funestus* group, *An. pharoensis* and *An. coustani* respectively. The niche overlaps between vectors are denoted by the conjoint values of the vectors.

crucial for identifying areas at risk of sustained and prolonged malaria transmission. We established that niche overlaps between *An. gambiae* complex and *An. funestus* group often occur across a range of elevations, including higher altitudes, and in cooler temperature zones compared to *An. pharoensis* and *An. coustani*. Remarkably, the *An. gambiae* complex shows adaptability to a wide range of extreme environmental conditions. Prior research has indicated the ability of the *An. gambiae* complex to thrive at higher elevations, a phenomenon possibly exacerbated by deforestation, which raises local temperatures and creates conducive environments for mosquito development and survival (Tuno et al., 2005). Additionally, global warming may render higher altitudes, previously inhospitable to mosquitoes, as suitable habitats (Afrane et al., 2012). The observation that primary vectors occur in cooler temperatures suggests a higher tolerance to cooler conditions compared to their secondary counterparts (Afrane et al., 2012; Lindblade et al., 2000). This adaptability of primary vectors to diverse environmental conditions, including cooler climates, emphasizes the need for adaptive and localized vector control strategies, especially in regions undergoing particularly high environmental change.

The spatio-temporal distribution and vector niche overlap maps indicate a progressive increase in niche overlap among malaria vector species. This increase involves primary vectors, a combination of primary and secondary vectors, as well as secondary vectors alone. Such trends can be attributed to the ongoing expansion of these vector species into new territories, leading to the development of shared ecological niches. This observation aligns with previous research findings that have documented the continued expansion of various species beyond their native ranges, a phenomenon largely influenced by environmental change (Ferguson et al., 2010). The presence of these niche overlaps suggests that multiple areas within Africa are ecologically suitable for both primary and secondary vectors. Specifically, the high degree of overlap between *Anopheles gambiae* complex and other vectors can be linked to the well-established and widespread presence of *An. gambiae* across numerous parts of Africa. *Anopheles gambiae* is prevalent throughout the continent and is possibly a native African species (Ayala and Coluzzi, 2005). Consequently, the increased niche overlaps identified in our study highlight a heightened risk for malaria in many regions in Africa. These areas are potentially more susceptible to prolonged and sustained malaria transmissions due to the coexistence of multiple vector species, including both primary and secondary vectors. The findings underscore the necessity for comprehensive vector surveillance and control strategies that address the diverse and expanding ecological niches of malaria vectors across Africa.

The dynamic spatial-temporal modeling approach employed in our study demonstrate exceptional accuracy in delineating niche overlaps among malaria vectors. The model's adeptness in accurately predicting the spatio-temporal dynamics of niche overlaps, along with the occurrences of vector species, can be attributed to several key factors. Foremost is the model's ability to dynamically adapt to the evolving nature of each species niche. Additionally, the model's design allows for the incorporation of new vector occurrence data as it becomes available, ensuring that the model remains current and reflective of the latest ecological scenarios. This approach highlights the effectiveness of models that are designed to leverage continuous data updates, showcasing their superior performance in capturing the dynamic and complex nature of ecological phenomena. Additionally, our results highlight a significant advancement over traditional ecological niche modeling approaches (Adeogun et al., 2023; Ayala et al., 2009; Kulkarni et al., 2010; Moffett et al., 2007; Olabimi et al., 2021), which predominantly employ static methodologies. Such static models often fail to capture the dynamic aspects of ecological niche and species adaptations. In contrast, our CA model effectively reflects the current state of species niche by integrating updates from previous time spans, thereby accommodating the continuous evolution of species and their habitats. Additionally, a comparison of our CA model output with past studies reveals agreement between our results and maps of vector spatial distribution, including

An. gambiae complex, *An. funestus* group (Adeogun et al., 2023; Koffi et al., 2023; Wangrawa et al., 2024), and secondary malaria vectors (Afrane et al., 2016).

While our CA model has shown strong capabilities in identifying niche overlaps among malaria vectors, it is important to consider its limitations and areas for potential enhancement. One primary limitation lies in the use of environmental data with a yearly temporal resolution such as irrigation and built-ups. Given the short life cycle of mosquitoes, shifting to a more granular temporal resolution, such as monthly, weekly, or even daily, could uncover more intricate dynamics of niche overlap. This adjustment could capture subtle variations and trends that yearly data might miss, leading to a deeper understanding of how environmental changes throughout the year influence vector behaviors and interactions. Moreover, the selection of the initial study period is crucial. Historical data indicates that populations of *An. funestus* and *An. gambiae* significantly increased during the 1970s and 1980s, a period noted for high sporozoite-positive rates in both species (Afrane et al., 2016). However, most current databases commence their records around the year 2000, potentially missing vital information from earlier decades. This gap might limit our understanding of the vectors' historical distributions and the early stages of their niche development. Incorporating data from these earlier periods, if available, could provide a more comprehensive picture of the distribution and evolution of vector niches over time. Additionally, our study recognizes that the CA model struggles to accurately capture the dynamics of phenomena in the initial temporal stages, particularly when data reporting is sparse or unreliable. This underscores the need for thorough and precise data collection in the early phases of modeling to ensure the model's accuracy, reliability and insensitivity at initial stages. In summary, our model offers significant insights into the niche overlaps of malaria vectors. However, refining the temporal resolution of environmental data and extending the historical range of our analysis could greatly enhance our understanding of these complex ecological interactions. Also, as malaria vectors continue to adapt to changing environmental chance, future research could benefit from incorporating additional environmental variables, while exercising caution because this increases the risk of collinearity among the variables. Secondly, future research could explore integration of ML and artificial intelligence (AI) into the CA model, leveraging the increasing availability of vector occurrence data. Additionally, we recommend extending the model to other malaria and non-malaria vectors. To facilitate the implementation of this study's recommendations, application of the model to other phenomena, or reproduction of the findings, all utilized data are open-access, and the codes are made available.

5. Conclusion

The observed niche overlaps between primary and secondary malaria vectors span extensive areas across Africa and are expanding due to the vectors moving beyond their native regions. This expansion places numerous areas at risk of sustained and prolonged malaria transmission and raises concerns about the potential spread of malaria if reintroduced into currently malaria-free zones. Moreover, the occurrence of these vectors and their niche overlaps is intricately linked to environmental conditions. On numerical results, environmental variables exhibit weak but significant correlations with the occurrence of malaria vector species. The *An. gambiae* complex tends to occur in areas with higher elevations and the lowest monthly average relative humidity, whereas *An. funestus* group and the secondary vectors (*An. pharoensis* and *An. coustani*) show a preference for warmer conditions. Therefore, as niche overlaps increase, pinpointing their locations and closely monitoring transmission rates in these areas become crucial for effective malaria control. This knowledge is vital for implementing swift and targeted vector control interventions when necessary. In light of these findings, our study emphasizes the need for ongoing surveillance and adaptive strategies in vector control, particularly in regions where niche overlaps are prominent.

CRediT authorship contribution statement

Eric Ali Ibrahim: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **John Odindi:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Mark Wamalwa:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Henri E.Z. Ton-nang:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Funding

The present study was executed by the International Centre of Insect Physiology and Ecology. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the donor.

Appendix A. Appendix

Appendix A

Data sources.

Variable	Spatial resolution	Source
1 Temperature	30 arcsec	https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav2%2F&prefix=%2F
2 Relative humidity	30 arcsec 0.0833 decimal	https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav2%2F&prefix=%2F
3 Irrigation	degree	https://data.apps.fao.org/catalog/iso/f79213a0-88fd-11da-a88f-000d939bc5d8
4 Elevation	30 m	https://www.worldclim.org/data/worldclim21.html
5 Built-ups	500 m	https://e4ftl01.cr.usgs.gov/MOTA/MCD12Q1.061/
Normalized difference vegetation index (NDVI)	1 km	https://e4ftl01.cr.usgs.gov/MOLT/MOD13A2.061/
7 Shrubs	500 m	https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav2%2F&prefix=%2F https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav2%2F&prefix=%2F
8 Climate moisture index	1 km	https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav2%2F&prefix=%2F
9 Cloud area fraction	1 km	https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav2%2F&prefix=%2F
10 Wind speed	1 km	https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav2%2F&prefix=%2F
11 Potential evapotranspiration'	1 km	https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav2%2F&prefix=%2F
12 Precipitation	1 km	https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav2%2F&prefix=%2F
13 Solar radiation	1 km	https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav2%2F&prefix=%2F
14 Land use land cover	1 km	https://e4ftl01.cr.usgs.gov/MOTA/MCD12Q1.061/
15 Vector occurrence		Malaria Atlas Project vector occurrence layers under the explorer icon (https://data.malariaatlas.org/maps?layers=)

Appendix B

Differentiation of cells in the study area during model implementation.

Vectors	Value assigned to cell containing the vector(s) and their niche overlaps	Value assigned to neighborhood cells and cells where neighborhood intersections occurred
No vector present	0	-
1	1	0.5
2	2	1.5
3	3	2.5
4	4	3.5
1 & 2	5	4.5
1 & 3	6	5.5
1 & 4	7	6.5
2 & 3	8	7.5
2 & 4	9	8.5

(continued on next page)

Appendix B (continued)

Vectors	Value assigned to cell containing the vector(s) and their niche overlaps	Value assigned to neighborhood cells and cells where neighborhood intersections occurred
3 & 4	10	9.5
1,2 & 3	11	10.5
1, 2 & 4	12	11.5
1, 3 & 4	13	12.5
2, 3 & 4	14	13.5
1, 2, 3, 4	15	14.5

1 = *An. gambiae* complex, 2 = *An. funestus* group, 3 = *An. pharoensis*, 4 = *An. coustani*.

Appendix C

Temporal distribution of the vector data.

Year	<i>An. coustani</i>	<i>An. funestus</i> group	<i>An. gambiae</i> complex	<i>An. pharoensis</i>
1974			4	
1975			12	
1980			1	
1981			6	
1982			40	
1983			20	
1984		11	29	11
1985		50	190	36
1986	11	71	180	23
1987	1	56	261	15
1988	1	99	298	13
1989	5	127	357	7
1990	14	109	229	14
1991	8	74	286	4
1992	4	185	484	23
1993	5	67	178	94
1994	2	92	182	11
1995	2	64	236	9
1996	2	60	210	3
1997	1	88	226	3
1998	4	125	211	3
1999	7	107	375	6
2000	2	182	476	10
2001	3	207	522	8
2002	7	201	442	18
2003	8	147	331	27
2004	1	142	311	29
2005	8	89	424	17
2006	6	73	427	9
2007		36	239	
2008	1	58	447	6
2009		45	1136	
2010		183	1174	
2011		155	1195	
2012		244	1333	
2013	3	101	1368	1
2014	5	279	1525	
2015	4	406	1679	5
2016	6	231	1765	5
2017	6	118	2187	11
2018	3	3	46	1
2019	7	49	198	31
2020		1	134	
2021	1	4	24	3

Appendix D. Association of the predictor variables with vector occurrence

Precipitation, as a measure of rainfall, is crucial in forming breeding habitats, thereby directly influencing vector populations (Bayoh and Lindsay, 2003; Kibret et al., 2019; Kirby and Lindsay, 2009), while temperature plays a vital role in the development, reproduction, and survival of vectors. Relative humidity is of considerable importance, impacting vectors' longevity and fecundity, which in turn influences their survival and ability to reproduce (Cohen et al., 2008). Varying elevations result in different climatic conditions, thereby influencing the distribution and activity of malaria vectors (Minakawa et al., 2006).

Agricultural practices, notably those involving water-intensive crops like rice and sugarcane, significantly influence the ecology of malaria vectors. These farming methods often result in the formation of extensive water bodies, providing ideal breeding sites for *Anopheles* mosquitoes. Furthermore, irrigation schemes, integral to such agricultural activities, create favorable conditions for these vectors to breed, thrive, and potentially increase in density and biting frequency (Haileselassie et al., 2021; Hawaria and Kibret, 2023; Minakawa et al., 2005). The formation of stagnant water bodies due to farming activities, is conducive to larval development and the subsequent emergence of adult mosquitoes. Additionally, the proximity to human

settlements is a crucial predictor of vector habitat suitability. Urban and peri-urban areas, characterized by human-built environments, often attract malaria vectors, increasing the risk of malaria transmission in densely populated areas (Ayala et al., 2009).

The Normalized Difference Vegetation Index (NDVI) is a key metric in understanding vector ecology by measuring photosynthetic activity, which reflects vegetation's spatial and temporal dynamics (Lourenço et al., 2011). The variations in vegetation have an indirect but significant impact on the development and reproduction of mosquito vectors (Altamiranda-Saavedra et al., 2017; Lourenço et al., 2011). Additionally, changes in land use and land cover, such as deforestation and alterations in land coverage, profoundly affect vector habitats (Srivastava et al., 2013). As vegetation dynamics and land-use changes influence microclimate, they indirectly dictate the viability of breeding and resting sites for mosquito vectors, thereby influencing their population dynamics and potential for disease transmission.

Appendix E. ISO-Compliant metadata

The ISO-compliant metadata have been deposited in the DMMG Unit repository (https://dmmg.icipe.org/data/malaria_vectors_niche_overlaps/) under the 'metadata files'.

Data availability

The data used in this study has been deposited in the Data Management, Modeling, and Geo-Information (DMMG) Unit repository (https://dmmg.icipe.org/data/malaria_vectors_niche_overlaps/) including the source codes (<https://github.com/icipe-oficial/Niche-Overlap-Model>). The links to other data sources are provided in Appendix A.

References

- Adeogun, A., Babalola, A.S., Okoko, O.O., Oyeniyi, T., Omotayo, A., Izekor, R.T., Adetunji, O., Olakiigbe, A., Olagundoye, O., Adeleke, M., Ojianwuna, C., Adamu, D., Daskum, A., Musa, J., Sambro, O., Adedayo, O., Inyama, P.U., Samdi, L., Obembe, A., Dogara, M., Kennedy, P., Mohammed, S., Samuel, R., Amajoh, C., Adesola, M., Bala, M., Esema, M., Omo-Eboh, M., Sinka, M., Idowu, O.A., Ande, A., Olayemi, I., Yayo, A., Uhomobibhi, P., Awolola, S., Salako, B., 2023. Spatial distribution and ecological niche modeling of geographical spread of *Anopheles gambiae* complex in Nigeria using real time data. Sci. Rep. 13, 1–18. <https://doi.org/10.1038/s41598-023-40929-5>.
- Afrane, Y.A., Githeko, A.K., Yan, G., 2012. The ecology of Anopheles mosquitoes under climate change: case studies from the effects of environmental changes in East Africa highlands. Ann. N. Y. Acad. Sci. 9, 1249. <https://doi.org/10.1111/j.1749-6632.2011.06432.x.The>.
- Afrane, Y.A., Mariangela, B., Guiyun, Y., 2016. Secondary malaria vectors of Sub-Saharan Africa: threat to malaria elimination on the continent? Curr. Top. Malar. <https://doi.org/10.5772/65359>.
- Gyekum, T.P., Botwe, P.K., Arko-Mensah, J., Issah, I., Acuah, A.A., Hoggarh, J.N., Dwomoh, D., Robins, T.G., Fobil, J.N., 2021. A systematic review of the effects of temperature on Anopheles mosquito development and survival: implications for malaria control in a future warmer climate. Int. J. Environ. Res. Public Health 18, <https://doi.org/10.3390/ijerph18147255>.
- Altamiranda-Saavedra, M., Arboleda, S., Parra, J.L., Peterson, A.T., Correa, M.M., 2017. Potential distribution of mosquito vector species in a primary malaria endemic region of Colombia. PLoS One 12, 1–14. <https://doi.org/10.1371/journal.pone.0179093>.
- Antonio-Nkondjio, C., Kerah, C.H., Simard, F., Awono-Ambene, P., Chouaibou, M., Tchuinkam, T., Fontenille, D., 2006. Complexity of the malaria vectorial system in Cameroon: contribution of secondary vectors to malaria transmission. J. Med. Entomol. 43, 1215–1221. [https://doi.org/10.1603/0022-2585\(2006\)43\[1215:COTMVS\]2.0.CO;2](https://doi.org/10.1603/0022-2585(2006)43[1215:COTMVS]2.0.CO;2).
- Ayala, F.J., Coluzzi, M., 2005. Chromosome speciation: humans, *Drosophila*, and mosquitoes. Proc. Natl. Acad. Sci. USA 102 (Suppl), 6535–6542. <https://doi.org/10.1073/pnas.0501847102>.
- Ayala, D., Costantini, C., Ose, K., Kamdem, G.C., Antonio-Nkondjio, C., Agbor, J.-P., Awono-Ambene, P., Fontenille, D., Simard, F., 2009. Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. Malar. J. 8, 307. <https://doi.org/10.1186/1475-2875-8-307>.
- Bayoh, M.N., Lindsay, S.W., 2003. Effect of temperature on the development of the aquatic stages of *Anopheles gambiae* sensu stricto (Diptera: Culicidae). Bull. Entomol. Res. 93, 375–381. <https://doi.org/10.1079/ber2003259>.
- Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., Petitpierre, B., Pellissier, L., Yoccoz, N.G., Thuiller, W., Fortin, M.J., Randin, C., Zimmermann, N.E., Graham, C. H., Guisan, A., 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. Glob. Ecol. Biogeogr. 21, 481–497. <https://doi.org/10.1111/j.1466-8238.2011.00698.x>.
- CDC, 2021. Malaria's Impact Worldwide.
- Charif, O., Omrani, H., Basse, R.-M., Trigano, P., 2012. Cellular automata model based on machine learning methods for simulating land use change. In: Proceedings of the 2012 Winter Simulation Conference, pp. 1–23.
- Cohen, J.M., Ernst, K.C., Lindblade, K.A., Vulule, J.M., John, C.C., Wilson, M.L., 2008. Topography-derived wetness indices are associated with household-level malaria risk in two communities in the western Kenyan highlands. Malar. J. 7, 40. <https://doi.org/10.1186/1475-2875-7-40>.
- Collins, S.D., Abbott, J.C., McIntyre, N.E., 2017. Quantifying the degree of bias from using county-scale data in species distribution modeling: can increasing sample size or using county-averaged environmental data reduce distributional overprediction? Ecol. Evol. 7, 6012–6022. <https://doi.org/10.1002/ee3.3115>.
- Collins, C.M., Bonds, J.A.S., Quinlan, M.M., Mumford, J.D., 2019. Effects of the removal or reduction in density of the malaria mosquito, *Anopheles gambiae* s.l., on interacting predators and competitors in local ecosystems. Med. Vet. Entomol. 33, 1–15. <https://doi.org/10.1111/mve.12327>.
- CoreTeam, R, 2020. R: A Language and Environment for Statistical Computing [WWW Document]. URL <https://www.r-project.org/> (accessed 8.16.21).
- Cross, D.E., Healey, A.J.E., McKeown, N.J., Thomas, C.J., Macarie, N.A., Siaziyu, V., Singini, D., Liywali, F., Sakala, J., Silumesii, A., Shaw, P.W., 2022. Temporally consistent predominance and distribution of secondary malaria vectors in the Anopheles community of the upper Zambezi floodplain. Sci. Rep. 12, 240. <https://doi.org/10.1038/s41598-021-04314-4>.
- Diamond, J., Wilson, E.O., 1999. More Praise for Guns, Germs, and Steel: The Fates of Human Societies, New York.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münckmüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography (Cop.) 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- ESRI, 2024. ArcGIS Pro: Version 3.3.0. Environmental Systems Research Institute, Redlands, CA.
- Ferguson, H.M., Dornhaus, A., Beeche, A., Borgemeister, C., Gottlieb, M., Mulla, M.S., Gimnig, J.E., Fish, D., Killeen, G.F., 2010. Ecology: a prerequisite for malaria elimination and eradication. PLoS Med. 7, 1–7. <https://doi.org/10.1371/journal.pmed.1000303>.
- Gillies, S., 2013. Rasterio: Geospatial Raster I/O for {Python} Programmers.
- Guimapi, R.Y.A., Mohamed, S.A., Okeyo, G.O., Ndjomatchoua, F.T., Ekesi, S., Tonnang, H.E.Z., 2016. Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. Ecol. Complex. 28, 77–93. <https://doi.org/10.1016/j.ecocom.2016.08.001>.
- Haileselassie, W., Zemene, E., Lee, M.-C., Zhong, D., Zhou, G., Taye, B., Dagne, A., Deressa, W., Kazura, J.W., Yan, G., Yewhalaw, D., 2021. The effect of irrigation on malaria vector bionomics and transmission intensity in western Ethiopia. Parasit. Vectors 14, 516. <https://doi.org/10.1186/s13071-021-04993-y>.
- Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, S., Wieser, E., Taylor, J., Berg, P., Smith, N.J., Kern, R., 2020. Array programming with NumPy. Nature 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
- Hawaria, D., Kibret, S., 2023. Increased malaria incidence following irrigation practices in the endorheic Rift Valley Basin of Sidama Region, Ethiopia. PLoS One 18, 1–12. <https://doi.org/10.1371/journal.pone.0284247>.
- Hijmans, R.J., 2024. Raster: Geographic Data Analysis and Modeling.
- Ibrahim, E.A., Wamalwa, M., Odindi, J., Tonnang, H.E.Z., 2024a. Insights and challenges of insecticide resistance modelling in malaria vectors: a review. Parasit. Vectors 17, 1–12. <https://doi.org/10.1186/s13071-024-06237-1>.
- Ibrahim, E.A., Wamalwa, M., Odindi, J., Tonnang, H.E.Z., 2024b. Spatio - temporal characterization of phenotypic resistance in malaria vector species. BMC Biol. 1–19. <https://doi.org/10.1186/s12915-024-01915-z>.
- Jordahl, K., 2014. GeoPandas: Python Tools for Geographic Data.
- Joshi, A., Miller, C., 2021. Review of machine learning techniques for mosquito control in urban environments. Ecol. Inform. 61, 101241. <https://doi.org/10.1016/j.ecoinf.2021.101241>.
- Kamau, L., Mulaya, N., Vulule, J.M., 2006. Evaluation of potential role of *Anopheles ziemanni* in malaria transmission in western Kenya. J. Med. Entomol. 43, 774–776. [https://doi.org/10.1603/0022-2585\(2006\)43\[774:EOPOA\]2.0.co;2](https://doi.org/10.1603/0022-2585(2006)43[774:EOPOA]2.0.co;2).
- Kibret, S., Glenn Wilson, G., Ryder, D., Tekie, H., Petros, B., 2019. Environmental and meteorological factors linked to malaria transmission around large dams at three ecological settings in Ethiopia. Malar. J. 18, 1–16. <https://doi.org/10.1186/s12936-019-2689-y>.
- Kirby, M.J., Lindsay, S.W., 2009. Effect of temperature and inter-specific competition on the development and survival of *Anopheles gambiae* sensu stricto and an. *Arabiensis*

- larvae. Acta Trop. 109, 118–123. <https://doi.org/10.1016/j.actatropica.2008.09.025>.
- Koenraadt, C.J.M., Majambere, S., Hemerik, L., Takken, W., 2004. The effects of food and space on the occurrence of cannibalism and predation among larvae of *Anopheles gambiae* s.l. Entomol. Exp. Appl. 112, 125–134. <https://doi.org/10.1111/j.0013-8703.2004.00186.x>.
- Koffi, A.A., Camara, S., Ahoua Alou, L.P., Oumbouke, W.A., Wolie, R.Z., Tia, I.Z., Sternberg, E.D., Yapo, F.H.A., Koffi, F.M., Assi, S.B., Cook, J., Thomas, M.B., N'Guessan, R., 2023. Anopheles vector distribution and malaria transmission dynamics in Gbéké region, Central Côte D'Ivoire. Malar. J. 22, 1–12. <https://doi.org/10.1186/s12936-023-04623-1>.
- Kroes, H.W., 1977. The niche structure of ecosystems. J. Theor. Biol. 65, 317–326. [https://doi.org/10.1016/0022-5193\(77\)90327-7](https://doi.org/10.1016/0022-5193(77)90327-7).
- Kulkarni, M.A., Desrochers, R.E., Kerr, J.T., 2010. High resolution niche models of malaria vectors in northern Tanzania: a new capacity to predict malaria risk? PLoS One 5, e9396. <https://doi.org/10.1371/journal.pone.0009396>.
- Lee, K.Y., Chung, N., Hwang, S., 2016. Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. Ecol. Inform. 36, 172–180. <https://doi.org/10.1016/j.ecoinf.2015.08.011>.
- Lindblade, K.A., Walker, E.D., Onapa, A.W., Katungu, J., Wilson, M.L., 2000. Land use change alters malaria transmission parameters by modifying temperature in a highland area of Uganda. Trop. Med. Int. Health 5, 263–274. <https://doi.org/10.1046/j.1365-3156.2000.00551.x>.
- Lourenço, P.M., Sousa, C.A., Seixas, J., Lopes, P., Novo, M.T., Almeida, A.P.G., 2011. Anopheles atroparvus density modeling using MODIS NDVI in a former malarious area in Portugal. J. Vector Ecol. 36, 279–291. <https://doi.org/10.1111/j.1948-7134.2011.00168.x>.
- Mahdizadeh, G.N., Mesgari, M.S., Hooshangi, N., 2019. Developing an agent-based model for simulating the dynamic spread of plasmodium vivax malaria: a case study of Sarbaz, Iran. Ecol. Inform. 54, 101006. <https://doi.org/10.1016/j.ecoinf.2019.101006>.
- Manga, L., Fondjo, E., Carnevale, P., Robert, V., 1993. Importance of low dispersion of *Anopheles gambiae* (Diptera: Culicidae) on malaria transmission in hilly towns in South Cameroonian. J. Med. Entomol. 30, 936–938. <https://doi.org/10.1093/jmedent/30.5.936>.
- Minakawa, N., Sonye, G., Yan, G., 2005. Relationships between occurrence of *Anopheles gambiae* s.l. (Diptera: Culicidae) and size and stability of larval habitats. J. Med. Entomol. 42, 295–300. <https://doi.org/10.1093/jmedent/42.3.295>.
- Minakawa, N., Omukunda, E., Zhou, G., Githeko, A., Yan, G., 2006. Malaria vector productivity in relation to the highland environment in Kenya. Am. J. Trop. Med. Hyg. 75, 448–453. <https://doi.org/10.4269/ajtmh.2006.75.448>.
- Moffett, A., Shackelford, N., Sarkar, S., 2007. Malaria in Africa: vector species' niche models and relative risk maps. PLoS One 2, e824. <https://doi.org/10.1371/journal.pone.0000824>.
- Mukiama, T.K., Mwangi, R.W., 1989. Seasonal population changes and malaria transmission potential of Anopheles pharoensis and the minor anophelines in Mwea Irrigation Scheme, Kenya. Acta Trop. 46, 181–189. [https://doi.org/10.1016/0001-706X\(89\)90035-1](https://doi.org/10.1016/0001-706X(89)90035-1).
- Mustapha, A.M., Musembi, S., Nyamache, A.K., Machani, M.G., Kosgei, J., Wamuyu, L., Ochomo, E., Lobo, N.F., 2021. Secondary malaria vectors in western Kenya include novel species with unexpectedly high densities and parasite infection rates. Parasit. Vectors 14, 252. <https://doi.org/10.1186/s13071-021-04748-9>.
- Ndenga, B.A., Simbauni, J.A., Mbugi, J.P., Githeko, A.K., Fillinger, U., 2011. Productivity of malaria vectors from different habitat types in the western Kenya highlands. PLoS One 6, e19473. <https://doi.org/10.1371/journal.pone.0019473>.
- Ntabaliba, W., Vavassori, L., Stica, C., Makungwa, N., Oduluwa, O.G., Swai, J.K., Lekundayo, R., Moore, S., 2023. Life expectancy of Anopheles funestus is double that of Anopheles arabiensis in Southeast Tanzania based on mark-release-recapture method. Sci. Rep. 13, 1–9. <https://doi.org/10.1038/s41598-023-42761-3>.
- Ntabi, J.D.M., Lissom, A., Djontu, J.C., Nkemngu, F.N., Diafouka-Kietela, S., Mayela, J., Missonsta, G., Djogbenou, L., Ndo, C., Wondji, C., Adegnika, A.A., Lenga, A., Borrman, S., Ntoumi, F., 2024. Entomological indicators of plasmodium species transmission in Goma Tsé-Tsé and Madibou districts, in the Republic of Congo. Malar. J. 23, 1–13. <https://doi.org/10.1186/s12936-023-04823-9>.
- Olabimi, I.O., Ileke, K.D., Adu, B.W., Arotolu, T.E., 2021. Potential distribution of the primary malaria vector *Anopheles gambiae* Giles [Diptera: Culicidae] in Southwest Nigeria under current and future climatic conditions. J. Basic Appl. Zool. 82. <https://doi.org/10.1186/s41936-021-00261-8>.
- Packard, R.M., 2007. The Making of a Tropical Disease: A Short History of Malaria. The Johns Hopkins University Press.
- Pascual-Rico, R., Sánchez-Zapata, J.A., Navarro, J., Eguía, S., Anadón, J.D., Botella, F., 2020. Ecological niche overlap between co-occurring native and exotic ungulates: insights for a conservation conflict. Biol. Invasions 22, 2497–2508. <https://doi.org/10.1007/s10530-020-02265-x>.
- Plotnick, R.E., Gardner, R.H., 2002. A general model for simulating the effects of landscape heterogeneity and disturbance on community patterns. Ecol. Model. 147, 171–197. [https://doi.org/10.1016/S0304-3800\(01\)00418-5](https://doi.org/10.1016/S0304-3800(01)00418-5).
- Polechová, J., Storch, D., 2008. Ecological Niche. Encycl. Ecol. Five-Volume Set 1–5, pp. 1088–1097. <https://doi.org/10.1016/B978-008045405-4.00811-9>.
- Povalej, P., Kokol, P., Družovec, T.W., Stiglic, B., 2005. Machine-learning with cellular automata. In: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 3646 LNCS, pp. 305–315. https://doi.org/10.1007/11552253_28.
- Rahman, M.S., Pientong, C., Zafar, S., Ekalaksananan, T., Paul, R.E., Haque, U., Rocklöv, J., Overgaard, H.J., 2021. Mapping the spatial distribution of the dengue vector *Aedes aegypti* and predicting its abundance in northeastern Thailand using machine-learning approach. One Health 13, 100358. <https://doi.org/10.1016/j.onehlt.2021.100358>.
- Ramampiandra, E.C., Scheidegger, A., Wydler, J., Schuwirth, N., 2023. A comparison of machine learning and statistical species distribution models: quantifying overfitting supports model interpretation. Ecol. Model. 481, 110353. <https://doi.org/10.1016/j.ecolmodel.2023.110353>.
- Saili, K., de Jager, C., Sangoro, O.P., Nkya, T.E., Masaninga, F., Mwenya, M., Sinyolo, A., Hamainza, B., Chanda, E., Fillinger, U., Mutero, C.M., 2023. Anopheles rufipes implicated in malaria transmission both indoors and outdoors alongside *Anopheles funestus* and *Anopheles arabiensis* in rural South-East Zambia. Malar. J. 22, 1–12. <https://doi.org/10.1186/s12936-023-04489-3>.
- Schoener, T.W., 1970. Nonsynchronous Spatial Overlap of Lizards in Patchy Habitats. Author (s): Thomas W. Schoener Reviewed work (s). Ecological Society of America Stable. <http://www.jstor.org/stable/1935376>. Ecology 51, 408–418.
- Snow, A.D., 2019. Rioxarray · PyPI.
- Srivastava, A.K., Kharbuli, B., Shira, D.S., Sood, A., 2013. Effect of land use and land cover modification on distribution of anopheline larval habitats in Meghalaya, India. J. Vector Borne Dis. 50, 121–126.
- The Pandas Ddevelopment Team, 2020. Pandas-Dev/Pandas: Pandas.
- Tsafack, N., Wang, X., Xie, Y., Fattorini, S., 2021. Niche overlap and species co-occurrence patterns in carabid communities of the northern Chinese steppes. Zookeys 2021, 929–949. <https://doi.org/10.3897/zookeys.1044.62478>.
- Tuno, N., Okeka, W., Minakawa, N., Takagi, M., Yan, G., 2005. Survivorship of *Anopheles gambiae* sensu stricto (Diptera: Culicidae) larvae in western Kenya highland forest. J. Med. Entomol. 42, 270–277. <https://doi.org/10.1093/jmedent/42.3.270>.
- Visser, H., Evers, N., Bontsema, A., Rost, J., Niet, A., Vethman, P., Mylius, S., Linden, A., Roovaart, J., Gaalen, F., Knoben, R., Lange, H., 2022. What drives the ecological quality of surface waters? A review of 11 predictive modeling tools. Water Research 208 (2022), 117851. <https://doi.org/10.1016/j.watres.2021.117851>.
- Wangrawan, D.W., Odero, J.O., Baldini, F., Okumu, F., Badolo, A., 2024. Distribution and insecticide resistance profile of the major malaria vector *Anopheles funestus* group across the African continent. Med. Vet. Entomol. 38, 119–137. <https://doi.org/10.1111/mve.12706>.
- WHO, 2022. World Malaria Report 2022. World Health Organization, Geneva.
- WHO, 2023. World Malaria Report 2023.
- World Bank, 2024. Malaria [WWW Document]. URL <https://www.worldbank.org/en/topic/health/brief/malaria>.
- Xu, T., Zhou, D., Li, Y., 2022. Integrating ANNs and cellular automata-Markov chain to simulate urban expansion with annual land use data. Land 11, 1–15. <https://doi.org/10.3390/land11071074>.
- Yu, R., Wang, R., 2024. Learning dynamical systems from data: an introduction to physics-guided deep learning. Proc. Natl. Acad. Sci. USA 121, 1–10. <https://doi.org/10.1073/pnas.2311808121>.
- Zambrano-Asanza, S., Morales, E., Montalvan, J.A., Franco, J.F., 2023. Integrating artificial neural networks and cellular automata model for spatial-temporal load forecasting. Int. J. Electr. Power Energy Syst. 148, <https://doi.org/10.1016/j.ijepes.2022.108906>.