Prediction of mosquito vector abundance for three species in the *Anopheles gambiae* complex

Geoffrey R. Hosack*[1], Maud El-Hachem[1], Adrien Ickowicz[1], Nicholas J. Beeton[1], Andrew Wilkins[2], Keith R. Hayes[1], Samuel S. C. Rund[3], Sarah A. Kelly[4], and Mary Ann McDowell[3]

[1]CSIRO Data61, 3 Castray Esplanade, Hobart, 7001, Tasmania, Australia
[2]CSIRO Mineral Resources, 1 Technology Court, Pullenvale, 4069, Queensland, Australia
[3]Galvin Life Science Center, University of Notre Dame, Notre Dame, 46556, Indiana, USA
[4]Department of Life Sciences, Imperial College London, London, SW7 2AZ, UK

*Corresponding author: Geoffrey R. Hosack. Email: Geoff.Hosack@csiro.au

Abstract

Background: The dynamics of vector borne disease transmission depend on the abundances of vectors. The dominant malaria vector species complex of *Anopheles gambiae sensu lato* is a target of vector control strategies designed to reduce and eliminate malaria transmission. The three most widely distributed dominant malaria vectors within the species complex are *An. arabiensis*, *An. coluzzii* and *An. gambiae sensu stricto*.

Methods: Previous studies across the extent of the species complex range have been restricted to annual predictions of species occurrence or relative abundance. This study incorporated relative abundance data at the species level and abundance data at the species complex level to estimate and predict daily equilibrium vector abundances of each species. Statistical models with interpretable parameters were used to gain insight into how factors such as meteorological conditions, insecticide treated net use, and human density affect the spatial and temporal predictions. The statistical models were used to predict daily equilibrium vector abundance that is an important factor in indices of malaria transmission such as the basic reproduction number and vectorial capacity.

Results: Meteorological factors, such as precipitation and relative humidity, and human factors, such as population density and insecticide treated nets, were important predictors of these three malaria vector species in Africa. Insecticide treated net use was negatively associated with per capita vector abundance of the *An. gambiae* species complex and interacted with year to suggest an a negative effect on the ratio of *An. gambiae s.s.* to *An. arabiensis* at the end of the prediction period that extended from 2002 to 2020. A predicted increasing interannual trend for *An. coluzzii* was potentially caused by changing species identification methods.

Conclusion: The predicted equilibrium abundances of the three species showed potentially high levels of geographical overlap, niche overlap, and evidence for stable coexistence despite substantial interspecific competition. Improving collection of longitudinal species abundance data across the spatial range of the *An. gambiae* species complex will facilitate future explorations of causal hypotheses that relate vector abundance to control measures, malaria interventions and meteorological conditions.

Keywords: interspecific competition, generalised linear model, maturation delay, Leslie-Gower model, stability

Background

The reduction of mosquito vector abundance and the number of infectious bites on humans through vector control measures is an effective means to lower or eliminate human malaria [1–3]. Shifts in species abundances of mosquito vectors induce concomitant changes in behaviour and susceptibility to vector control that require adjustment of available intervention strategies such as insecticide treated nets, indoor residual spraying, larval source management, house screening, baited traps or novel techniques such as genetically modified vector suppression and replacement strategies [4–6]. The evaluation of conventional and potential novel vector control strategies on disease transmission dynamics thus benefits from prediction and estimation of vector abundances [7, 8]. Mathematical models of long term

or equilibrium vector abundance, for example, have long served as useful guides for vector control and malaria interventions [9]. Equilibrium vector abundance is the theoretical population size that would be attained if birth rates matched death rates at a given location. Equilibrium abundance is a useful concept allowing for the possibility that the true or actual abundance of a species may diverge from carrying capacity because of local environmental or anthropogenic perturbations, dispersal, source-sink meta-population dynamics or through interspecific competition [10–12]. For example, vector populations fluctuate around the equilibrium in response to seasonal changes in birth and death rates, and a range of other factors.

Applications with a time-varying basic reproduction number ($\mathcal{R}_0$) that depends on the ratio of vector abundance to human hosts, and derived indices such as time-varying vectorial capacity, are examples of an equilibrium based analysis. $\mathcal{R}_0$ is defined at the disease free equilibrium composed of completely susceptible vector and human populations in the standard Ross-Macdonald model [1, 2, 13] and related mathematical models of malaria transmission by anophelines [14]. The canonical basic reproduction number $\mathcal{R}_0$ corresponds to the number of malaria infections that arise from a single infectious case introduced into a completely susceptible population at the disease free equilibrium [15]. Macdonald acknowledges that the changing vector abundance and transmission parameters that define $\mathcal{R}_0$ suggest that "... the degrees of equilibrium seen in nature should make a continuous series ..." (p. 34 in [2]). The equilibrium index $\mathcal{R}_0$, for example, can be modelled as a continuous function of temperature and precipitation mediated by the average vector population at a climatically determined equilibrium [16]. The influential Garki model [9, 17, 18] accounts for seasonal malaria transmission in Nigeria via an interpolated function of time-varying vector population size that determines the vector population's contribution to $\mathcal{R}_0$ through the vectorial capacity [19]. Recent statistical analysis of the Garki data from the 1970s suggests that the equilibrium vector population size of *An. gambiae s.l.* defined by an environmentally driven carrying capacity approximately tracks weekly precipitation [20]. An analysis of seasonal malaria transmission in Burkina Faso suggests that the average vector population size of *An. gambiae* is proportional to weekly precipitation [21].

Despite the important contribution of the vector-to-human host ratio to the standard indices of malaria transmission risk such as the basic reproduction number $\mathcal{R}_0$ and vectorial capacity, abundance data for mosquito vector species are rarely collected in Africa [22]. In particular, the species complex *Anopheles gambiae sensu lato* (*s.l.*) is an important vector of malaria in Africa [23]. *An. gambiae s.l.* is estimated to have caused at least half of the 241 million malaria deaths of children under five years old in sub-Saharan Africa between 3000 BCE and 2000 CE [6]. The three primary malaria vector member or sibling species of the *An. gambiae* complex [6] that have the broadest geographical distribution [24] are *An. arabiensis*, *An. coluzzii* and *An. gambiae sensu stricto* (*s.s.*). However, longitudinal abundance data for species in the *An. gambiae* complex are rare. Species identification in field collections is often limited because the member species are not morphologically distinguishable, and so molecular identification from subsamples of field collections is required. Longitudinal count data for *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* are therefore uncommon, and most predictive studies at the continental scale rely only on species occurrence data [24–27].

Predictions of the probability of species occurrence from species distribution models link the distributions of species in the *An. gambiae* complex to environmental covariates such

as annual indices of meteorological conditions and climate [24–26]. Alternatively, climatological covariates may be used to predict relative abundance or species composition [28] based on species occurrence data [27, 29]. Sinka et al. [27] suggest that the consideration of relative abundance may defray some of the difficulties associated with the estimation of 'true' abundance that is spatially and temporally variable, and use species occurrence data to predict that the relative abundance of *An. arabiensis* versus the aggregate of the two species *An. coluzzii* and *An. gambiae s.s.* increases with insecticide treated net intervention. Tene Fossog et al. [29] use relative species frequencies of *An. coluzzii* and *An. gambiae s.s.* to predict relative probabilities of occurrence in western Africa based on climatic and geographical covariates with species specific responses determined by elevation and distance from coastline. Pombi et al. [30] find high niche overlap among *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* in certain regions based on an index that is a function of relative abundance data [31]. Niche overlap is sometimes suggested as an index of the strength of interspecific competition [32], although this correlative index based on the relative abundance of species may diverge from the underlying competitive relationships among species [31]. Competition among different species of mosquito larvae can lead to the competitive displacement or reduction of one species by another and so affect strategies for control of disease vectors [33, 34].

However, malaria transmission dynamics are driven by the abundance of mosquito species and not just relative abundance or the probability of species occurrence. This study therefore sought to estimate spatiotemporal abundance of *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* by collating data both on relative abundance at the species level (identified genotypically) and longitudinal population counts at the species complex level (identified morphologically) across the geographic range of *An. gambiae s.l.* The difficulty in acquiring such data should not be underestimated or understated. Regular vector surveillance is typically conducted to support operational activities, which in the African entomological context is overwhelmingly in support of malaria intervention through insecticide treated bednet (ITN) deployment or indoor residual spray (IRS) activity. Indeed, much of the regular surveillance is focused on determining local species present and their localised behaviours (e.g., measures of anthropophily, endophily, and biting time) to determine appropriate vector interventions [35]. Unlike some regions that employ weekly or daily surveillance to identify mosquito control action thresholds for vector suppression operations such as insecticidal fogging, the operational timescales for African ITN distributions or IRS operations are not based on the weekly or monthly status of mosquito species abundance. Hence, such data are less likely to be collected. This fact is highlighted in a recent global meta-analysis of mosquito control action thresholds [36], where none are listed from Africa. Consequently, although longitudinal surveillance of *Anopheles* mosquito populations is performed in Africa particularly for research purposes, most of the longitudinal studies investigated in this study either did not use genetic methods (e.g., polymerase chain reaction, PCR) to identify field collections to the species level or instead did not resolve changes in species diversity and relative abundance within a year. The ability to detect seasonal changes in species composition was therefore negated. Nonetheless, the species identified data collated by this study expand on previous efforts [24] by including finer scale temporal information on the sample collection time that thereby allows incorporation of seasonally varying environmental conditions into the analysis. Importantly, this data acquisition effort also currently represents

the largest collation of publicly available longitudinal abundance data for the *An. gambiae* species complex. These data investigation and collation efforts both highlight and help bridge a significant gap in available data for African malaria mosquito biology.

Using these assembled data, the equilibrium abundances of *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* were predicted across sub-Saharan Africa from 2002 to 2020. The species equilibrium abundances are not directly observable and were inferred from empirical observations as responding to known spatial and spatiotemporal covariates. Statistical relationships were estimated for established environmental and geographical factors, such as precipitation, temperature and human population density [16, 20, 21, 37], to enable spatiotemporal predictions of *An. gambiae s.l.* abundance. Previous relative abundance analyses for pairs of species in *An. gambiae s.l.* [27–29] were extended in this study to account for daily varying meteorological predictors to allow seasonally varying predictions of relative abundance for *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* The analysis of relative abundance data available at the sibling species level was combined with the longitudinal abundance analysis at the species complex level to predict potential sibling species abundance across sub-Saharan Africa. The predicted equilibrium abundances of *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* were then connected to the equilibrium of an underlying dynamic model that suggests the potential for strong interspecific competition but also stable coexistence among the three vector species.

Methods

The study area is defined as the geographic range of *An. gambiae s.l.* [24, 38] in sub-Saharan Africa. The genotype data, abundance data, spatial covariates and spatiotemporal covariates are described below with further details provided in the supplementary R package **AgslPredict** for this paper [39]. The code for analyses and graphical results of relative abundance predictions, abundance predictions and competition estimates are also available in **AgslPredict**.

Genotype and Abundance Datasets

The two sets of data, species level genotype data and species complex level abundance data, have been made publicly available at VectorBase.org and are each discussed in turn. The genotype data of the three species *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* were downloaded from VectorBase.org [40, 41] and spatially subsetted to the study area. Some samples of the *An. gambiae s.l.* species complex were identified by chromosomal form to species. Species recorded as "Anopheles gambiae chromosomal form Bamako" and "Anopheles gambiae chromosomal form Savanna" were assigned to "Anopheles gambiae sensu stricto", and "Anopheles gambiae chromosomal form Mopti" was assigned to "Anopheles coluzzii" [42]. The three species retained for analysis were "Anopheles arabiensis", "Anopheles coluzzii" and "Anopheles gambiae sensu stricto". Samples for which information on date of collection was missing were filtered out of the dataset. Information on time of collection varied across the remaining samples but could include durations of a single day, month, year or multiple years. The start and end dates of each collection were

extracted. Samples with collection dates reported only to the month were assigned start and end dates that corresponded to the start and end dates of the reported month. Samples with reported durations greater than 31 days were filtered out of the dataset. *An. coluzzii* and *An. gambiae s.s.* were not distinguished prior to 2001 [30] and samples with earlier initial dates were omitted. Samples were then identified by project ID, latitude and longitude, collection date and collection protocol to define the sample unit of analysis. The frequency counts of the three species identified for each unit sample were then recorded for use in analyses of species relative abundance (species composition) described in the next section. Sample units with shared collection dates and collection methods were aggregated to 5 km × 5 km grid cells across the spatial extent, which resulted in 1,740 records. The 5 km × 5 km grid cell resolution was chosen to capture the home range of a female mosquito since most mosquitoes disperse less than 5 km based on mark release recapture experiments [43] and for comparison with previous spatial analyses of genotyped data for *An. gambiae* species that also used 5 km × 5 km grid cell resolution [23, 24, 27, 29]. The collated records extended from 2001 to 2021 and are summarised in Table S1 and Figure S1.

The abundance data for *An. gambiae s.l.* were downloaded from VectorBase.org [40, 44] and spatially subsetted to the study area. Whereas the relative abundance data are typically derived from genotyped subsamples of field collections, the abundance data recorded from a field collection are most often counts of individuals morphologically identified to the level of the *An. gambiae* species complex. The observed counts of *An. gambiae s.l.* were treated as the response variable for the abundance analysis. Observation methods that could be categorised to specific collection methods were retained for analysis. The level of sampling effort was defined as the daily number of field collections by observation method in each grid cell. Records were then subsetted to samples identified as adult female mosquitoes. The available records extended from 1997 to 2021. Sites with long term longitudinal observations often also included long periods without observations. About 2% of the remaining samples had indeterminate sampling dates and were excluded from the analysis. Observations were aggregated to each grid cell by day and observation method, which resulted in 9,105 samples of abundance after excluding collection methods that targeted larvae or had low information content due to very few samples collected (less than ten records). The collated records are summarised in Table S2. These *An. gambiae s.l.* abundance records were sparsely distributed across sub-Saharan Africa (Figure S2).

Covariate Datasets

Precipitation, temperature and humidity are meteorological predictors of *An. gambiae* distribution [24, 26, 45]. Spatiotemporal meteorological data for daily precipitation totals, daily average temperature and daily average relative humidity were obtained from the National Aeronautics and Space Administration (NASA) Langley Research Center (LaRC) Prediction of Worldwide Energy Resource (POWER) Project funded through the NASA Earth Science/Applied Science Program [46] using the R package **nasapower** [47, 48]. These data are provided at a resolution of 0.625° longitude by 0.5° latitude. The meteorological covariates of a meteorological grid cell were assigned to all those 5 km × 5 km grid cells with centroids that occurred within the corresponding meteorological grid cell. In the Supplementary Information, the meteorological data are visualised by annual average (Figures S3,

S4, S5), quarterly average (Figures S6, S7, S8) and anomalies with respect to the average over the timespan 2002–2020 (Figures S9, S10, S11) used for the prediction period.

Distances to nearest coast, presence of freshwater, elevation, human population, insecticide treated net use, longitude and latitude are proposed spatial predictors of *An. gambiae* distribution [7, 24, 29, 37]. Spatial covariates were constructed using distances from 5 km × 5 km grid cell centroids to the nearest coast (Figure S12) using data provided by Natural Earth and obtained with the R packages **rnaturalearth** [49] and **rnaturalearthdata** [50]. Elevation data at 15 arcseconds resolution were obtained from the National Centers for Environmental Information, National Oceanic and Atmospheric Administration [51], and averaged to obtain mean elevation for each 5 km × 5km grid cell (Figure S13).

The presence of freshwater habitat was quantified by the probability of permanent freshwater occurring in each 5 km × 5km grid cell, which was defined as follows. The minimum probability of intermittence for streams and rivers in each cell, which is derived from the predicted probability of flow cessation for at least 30 days per year on average [52], was first calculated for all stream and river reaches that occurred in a cell. The probability of permanent stream or river freshwater was then defined by the complementary probability, that is, one minus the minimum probability of intermittence. If a cell intersected with a lake or reservoir [LakeATLAS; 53], then the probability of freshwater was assigned a value of one. The probability of permanent freshwater is depicted in Figure S14.

Human population is often proposed as a predictor that is positively associated with the areal density of *An. gambiae* s.l. in rural environments [7, 24, 29, 37, 54]. However, this relationship breaks down in urban environments where malaria transmission is generally lower compared to rural and peri-urban environments [55, 56] and also more variable [57], noting that localised areas of high vector abundance near larval habitats can still occur in urban settings [55]. Total human population in each 5 km × 5km grid cell was therefore censored above 50,000 human inhabitants, which is the threshold commonly used to demarcate urban areas formed by contiguous areas with greater than 1,500 inhabitants per square kilometre [58]. The corresponding threshold for a censored human density covariate in each 5 km × 5km grid cell was censored at 2,000 inhabitants per square kilometre for consistency with the censored total human population covariate. The censored human population covariates therefore vary across rural and peri-urban 5 km by 5 km grid cells, and are assigned their threshold values in urban cells. Annual human population data at 30 arcseconds resolution were obtained from Landscan (Oak Ridge National Laboratory, `https://landscan.ornl.gov/`), and averaged to obtain total human population and mean human population density per square kilometere for each 5 km by 5 km grid cell (Figure S15).

Annual vector intervention data for ITN percent use by individuals and IRS percent coverage of households is available for the study area across the years 2000–2022 [59–61]. The intervention data were averaged to the 5 km × 5 km grid cells to derive spatiotemporal covariates of intervention effort. Caution is required for assessment of intervention efficacy on the basis of observational studies [62]. For example, intervention effort may increase for locations and times with high vector densities. A spurious positive association could then be induced between vector populations and intervention effort that thereby masks an expected negative causal impact of the intervention. Such an outcome could even be likely where interventions are selectively applied, but possibly less likely or severe for broadly

applied interventions with deployment decisions that are less sensitive to vector population abundance. Universal coverage is recommended for ITNs [63], whereas effective vector control intervention by IRS should exceed 30% coverage of households [64] with recommended levels greater than 80% [65]. The ITN covariate shows a geographically broad increasing trend across much of sub-Saharan Africa with values that often exceed 10% in 2020 (Figure S16), whereas non-negligible IRS covariate values are comparatively limited in recent years (Figure S17). Across all samples, the maximum ITN covariate value by grid cell exceeded 10% in 18% of cells with observed genotype data and 86% of cells with observed abundance data. In contrast, the maximum IRS covariate value exceeded 10% in only 3% of grid cells with observed genotype or abundance data. The observational analyses of intervention was therefore restricted to ITNs that exhibited broad geographic deployment of non-negligible interventions. A single abundance study that occurred prior to 2000 was assigned ITN covariate values derived from contemporaneous ITN use data for the study location [66].

Analysis of Relative Species Abundance

Sinka et al. [27] use multinomial generalised additive models of genotype data to predict relative abundance (or composition) of *An. funestus*, *An. arabiensis* and the combination of the two species *An. coluzzii* + *An. gambiae s.s.* based on a habitat suitability index [23] and vector intervention data. Here, in contrast to [27], the VectorBase.org data genotyped to species were used to i) model each of the three species *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.*, which is suggested as a potential avenue for exploration by Sinka et al. [27], and ii) include intra-annually varying predictors for the relative abundances of these species. The analysis included covariates of temperature, precipitation and relative humidity that varied daily, in contrast to previous studies [24, 27–29]. The dates listed in the filtered VectorBase.org genotype data typically included a single day for each record, but some records were assigned a duration of up to a month (see previous section). The meteorological covariates for spatially and temporally varying precipitation, temperature and relative humidity used for the relative abundance analyses were constructed as follows. Thirty day averages were constructed from the daily meteorological observations that preceded both the first and last dates of the collection period for a sample. These two thirty day averages, which coincide for collections with a single day duration, were then averaged together.

ITN use is linked to selective pressure against *An. gambiae s.s.* and in favour of *An. arabiensis* [67, 68] and was also included as a covariate. Collection method was not explicitly included as a covariate. Genotype data for the three species of interest often do not distinguish between outdoor and indoor collection methods [24] that can affect interpretation of intervention efficacy [27]. For example, 27% of genotype samples in this study reported the collection method as "catch of live specimens" (collection method was incorporated into the analysis for the abundance data, see next section). The collected samples represent a mix of indoor and outdoor locations, which may help mitigate biased estimates [27]. Geographical covariates included distances to coasts, elevation, probability of permanent freshwater, human population density censored above 2,000 inhabitant per square kilometre, longitude and the absolute value of latitude as a measure of insolation. The year of collection was included to account for changes in species composition change due to increasing capacity for taxonomic discrimination at the species level for genotype data [69].

Multinomial logit generalised linear models (GLMs), see [70], were analysed using R package **nnet** [71] with the covariates scaled between zero and one based on the observed ranges. Let $m \in \{Aa, Ac, Ag\}$ indicate the three species of *An. arabiensis* (Aa), *An. coluzzii* (Ac) and *An. gambiae s.s.* (Ag) with probability $p_m(s,t)$ at grid cell s and time t. Let $\boldsymbol{\nu}(s,t)$ denote the set of covariates at grid cell s and time t. *An. arabiensis* was chosen as the baseline or reference category of the multinomial logit GLM. The logarithm of the odds $p_{m'}(s,t)/p_{Aa}(s,t)$ in favour of *An. coluzzii* ($m' = Ac$) or *An. gambiae s.s.* ($m' = Ag$) and against *An. arabiensis* were related to the covariates by the unknown parameter vector $\boldsymbol{\gamma}$. The statistical model is

$$\boldsymbol{z}(s,t) \sim \text{Multinomial}(\boldsymbol{p}(s,t), n(s,t))$$
$$\log \frac{p_{m'}(s,t)}{p_{Aa}(s,t)} = \boldsymbol{\nu}(s,t)^\top \boldsymbol{\gamma}, \quad m' \in \{Ac, Ag\},$$

where $\boldsymbol{z}(s,t)$ is the vector of counts of each species, given by $z_m(s,t)$ for species m, which sums to total observed abundances $n(s,t)$, and has corresponding probabilities $\boldsymbol{p}(s,t)$. The parameter vector $\boldsymbol{\gamma}$ was estimated by maximum likelihood. The following covariates were assessed by model selection. Trends in species composition were assessed by including the annually-varying covariates ITN use and year, which were included as linear terms together with their interaction. The interaction term was included to consider approximate changes in ITN efficacy over time, as could occur during pervasive trends in insecticide resistance that have been observed for *An. gambiae s.l.* in some regions [72]. A quadratic surface model was used as a second order approximation for the remaining meteorological and geographical covariates, that is, each of these covariates appeared as a squared term and a linear term, along with all pairwise interactions of the latter. Backwards stepwise model selection used the R package **MASS** based on the Bayesian information criterion (BIC) [71]. The model with the lowest BIC value was chosen as the final model.

Analysis of Abundance of *An. gambiae s.l.*

The observed abundance $y(s,t)$ of adult female *An. gambiae s.l.* per person at "site" (grid cell) s and time t was adjusted to account for different observation methods. The recorded collection methods from the VectorBase.org abundance dataset were mapped to covariates that assigned indoor human landing catch as the baseline method. Binary factors were scored for collection methods of light traps, exit traps, pyrethrum spray catches, animal baited traps and pit traps (Table S4), and whether or not a collection method was documented to occur outdoors. Projects with collection methods that were uncertain with respect to indoor or outdoor sampling were assigned a value of 0.5 for outdoors. The observation method according to Table S4 was encoded by the covariate vector $\boldsymbol{w}(s,t,k)$ with unknown coefficients $\boldsymbol{\zeta}$ to adjust the mean abundance per indoor human $f(s,t)$ by the multiplicative factor $\rho(s,t,k)$ to produce the observed mean abundance per indoor human $\mu(s,t,k)$ for collection method k. If $\boldsymbol{w}(s,t,k)$ is a zero vector then the corresponding collection method is "man biting catch - indoors" (MBCI, Table S4) with $\rho(s,t,k = \text{MBCI}) = 1$ and $\mu(s,t,k = \text{MBCI}) = f(s,t)$.

The equilibrium vector population abundance per indoor human $f(s,t)$ was assumed to be described by the spatially and temporally varying covariates $\boldsymbol{v}(s,t)$ with unknown coeffi-

cients $\boldsymbol{\beta}$. Human population density is a proxy index for anthropogenic habitat modification and urbanisation. For example, the density of houses is negatively associated with larval habitat area for *Anopheles* mosquitoes [73]. High human population density not only reduces larval habitat but also reduces per capita malaria exposure, suggesting lower vector areal density in urban areas [55, 56]. However, in rural areas, human population is positively associated with *An. gambiae s.l.* areal density [7, 37]. A quadratic relationship on the censored human population density covariate was therefore included in the model. The linear and squared terms allow the potential per capita *An. gambiae s.l.* abundance to increase at low population densities and decrease at high population densities approaching the urban population density threshold. The covariate vector $\boldsymbol{v}(s,t)$ also incorporated the lagged effect of precipitation on larval habitat. For example, the carrying capacity for *An. gambiae* is related to weekly total precipitation by [21]. Here, the immature phase of the life cycle was assumed to be about 10 days [37]. Thus, for a given day and location, the precipitation covariate was constructed by first calculating the weekly average precipitation that preceded each day of the corresponding 10-day larval period at that location. The weekly running means were then averaged over these 10 days to define the contribution of precipitation to the carrying capacity experienced by *An. gambiae s.l.* mosquitoes during the immature stage of development. The precipitation covariate was included in the model as a quadratic relationship to allow for increased per capita *An. gambiae s.l.* abundance at low precipitation levels and decreasing per capita *An. gambiae s.l.* abundance with high precipitation. Relative humidity, which varies with temperature and like precipitation is also a function of water vapour in the air, has been suggested to be an important factor for *An. gambiae* abundance [45] and so was additionally included as a covariate; the effect of temperature is further considered in the next section. Linear interaction terms between the censored human population density covariate and the meteorological covariates of precipitation and relative humidity were also included in the model.

The statistical model for vector abundance per indoor human given collection method k was described by a negative binomial GLM,

$$\begin{aligned} y(s,t,k) &\sim \text{NegativeBinomial}\left(\mu(s,t,k)u(s,t,k),\theta\right) \\ \log\mu(s,t,k) &= \log\rho(s,t,k) + \log f(s,t) \\ \log\rho(s,t,k) &= \boldsymbol{w}(s,t,k)^\top\boldsymbol{\zeta} \\ \log f(s,t) &= \boldsymbol{v}(s,t)^\top\boldsymbol{\beta}, \end{aligned}$$

with mean and variance

$$\begin{aligned} &\mathbb{E}[y(s,t,k) \mid \mu(s,t,k)u(s,t,k),\theta] = \mu(s,t,k)u(s,t,k), \\ &\mathbb{V}[y(s,t,k) \mid \mu(s,t,k)u(s,t,k),\theta] = \mu(s,t,k)u(s,t,k) + [\mu(s,t,k)u(s,t,k)]^2/\theta, \end{aligned}$$

where θ is the overdispersion parameter and $u(s,t,k)$ is an offset for the sampling effort for observation method k at cell s on day t as described above (Methods–Genotype and Abundance Datasets). Unknown parameters were estimated by maximisation of the likelihood with the covariates scaled between zero and one based on their ranges. The negative binomial GLM was fit in R v4.4.0 [74] using the R package **glm2** [75] and function `glm.nb` in package **MASS** [71].

Predicted Species Equilibrium Abundance

The equilibrium vector abundances of the three species *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* were predicted for the time period from 1 January 2002 to 31 December 2020 across the potential spatial range of *An. gambiae s.l.* [24, 38]. The predictions of relative abundance for the three species were based on the multinomial logit GLM with the year covariate set to the last year of the prediction period, 2020. The predictions for mean abundance of adult female *An. gambiae s.l.* per human at site (grid cell) s and time (day) t were based on the negative binomial GLM with the collection method set to indoor human landing catch (MBCI), such that $f(s,t) = \mu(s,t,k = \text{MBCI})$. Abundance predictions of *An. gambiae s.l.* are best related to precipitation in regions where temperature is not a limiting factor [76]. A temperature threshold rule was therefore applied. Laboratory experiments suggest that larval emergence of *An. gambiae s.l.* does not occur if water temperatures are less than 18°C or more than 34°C [77]. Air temperatures tend to be lower than water temperatures in *Anopheles gambiae s.l.* larval habitats [78], although the exact relationship between water and air temperatures can vary depending on local environmental factors [79]. About 1% of data observations included average air temperatures that were outside the 18°C – 34°C range. Seasonal reproduction could potentially still occur in such areas during favourable months, but such regions were not represented well enough in the abundance sample data to capture such detail in the predictions. Sites where air temperatures were outside these temperature tolerances for at least 30 continuous days in each year were filtered out to avoid overextrapolation and were provisionally assigned predictions of zero *An. gambiae s.l.* abundance, such that $\mu(s,t) = 0$ for all t at any site s that exceeded the temperature tolerance for larval emergence.

The negative binomial GLM provided per capita estimates of adult females for the *An. gambiae* species complex based on counts of *An. gambiae s.l.* Species level predictions therefore require incorporation of the multinomial GLM that provides relative abundance estimates for the three species of interest. A complication is that the predicted mean abundance $f(s,t) = \mu(s,t,k = \text{MBCI})$ in the negative binomial GLM is the average daily abundance of adult female *An. gambiae s.l.* per human based on indoor feeding mosquitoes, but the propensity for human feeding among the three species of interest is not the same and adult female mosquitoes do not feed every day. Human feeding rates across the species range of *An. gambiae s.l.* have been elicited from domain experts [80] accounting for species differences among *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* using a prior elicitation method for GLMs [81]. Estimates of the average daily probabilities of an adult female mosquito biting a human are $\boldsymbol{r} = [0.37, 0.49, 0.47]^\top$ for *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.*, respectively [80]. These average estimates reflect the greater degree of zoophilic feeding habit of *An. arabiensis* [23], although the human feeding rate will in actuality vary with covariates such as livestock density. The feeding rate adjustment for total adult females of the *An. gambiae* complex specifies that there are more adult female mosquitoes than those that may be observed feeding on humans, and this effect is greater on average for *An. arabiensis* compared to the other two species. Let $\xi(s,t)$ denote the average abundance of adult female *An. gambiae s.l.* per human after this adjustment for the mean daily probability of human feeding. Using the predicted per capita abundance $f(s,t)$ from the negative binomial GLM

and the predicted relative abundances $\boldsymbol{p}(s,t)$ from the multinomial logit GLM obtains

$$f(s,t) = \sum_{m \in \{Aa,Ac,Ag\}} \xi(s,t) p_m(s,t) r_m,$$

so that *An. gambiae s.l.* vector abundance per human is

$$\xi(s,t) = f(s,t) \left[\sum_{m \in \{Aa,Ac,Ag\}} p_m(s,t) r_m \right]^{-1}.$$

The predicted species vector abundances $\boldsymbol{x}(s,t)$ were then obtained by again using the multinomial GLM predictions,

$$x_m(s,t) = h(s) p_m(s,t) \xi(s,t), \quad m \in \{Aa, Ac, Ag\},$$

where $h(s,t)$ is the human population at site s and time t censored at the urban threshold of 50,000 inhabitants in a 5 km $\times$ 5 km grid cell. As described above, the total human population is censored because the relationship between vector abundance and human population is weaker in urban environments [55, 56] compared to non-urban areas [7, 37]. Note that the relative abundance of each species based on the total vector abundances is again $p_m(s,t) = x_m(s,t)/\sum_m x_m(s,t)$ as originally predicted by the multinomial GLM.

The multinomial and negative binomial GLMs each accounted for various environmental and human factors that affect the average carrying capacities of the three species. For prediction, the models were interpreted to track potential equilibrium vector abundances determined by the running averages of the meteorological covariates, rather than actual abundances, subject to local conditions not accounted for in the continent-wide spatial and temporal covariates used in this analysis. The predicted equilibrium vector abundance at a given site may not be realised for many important ecological factors not considered in this model, for example, immigration or emigration due to metacommunity dynamics (see [80] for an elaborated model to address these considerations including wind-assisted advection, sensu [82]). The predicted species abundances $\boldsymbol{x}(s,t)$ were therefore considered predictions of equilibrium species vector abundances across the various sites and times. An underlying dynamic model that allows for stable coexistence and competing species at equilibrium is presented in the next section.

Estimated Niche Overlap and Interspecific Competition

Pombi et al. [30] estimate the niche overlap among the three species *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* based on field samples collected at three sites in central Burkina Faso. The pairwise estimates of niche overlap α_{ij} between species i and species j use the method of Pianka [31]. For comparison, this method was applied to the equilibrium predictions that are available for all days and grid cells in Burkina Faso for years 2002–2020. Pianka's niche overlap [31] was estimated from the predicted mean proportion $\bar{p}_m(s) = \bar{x}_m(s)/\sum_m \bar{x}_m(s)$ of each species m with average vector abundance $\bar{x}_m(s)$ averaged over

seasons and years at each site s in Burkina Faso to obtain

$$\alpha_{ij} = \sum_s \bar{p}_i(s)\bar{p}_j(s) \left(\sum_s \bar{p}_i(s)^2 \sum_s \bar{p}_j(s)^2 \right)^{-1/2}, \quad i,j \in \{Aa, Ac, Ag\}.$$

The resulting niche overlap matrix is positive definite and symmetric with unit entries on the main diagonal [32]. The niche overlap index has been suggested as a measure of interspecific competition [32], although Pianka [31] lists limitations with this interpretation.

A contrasting approach estimates interspecific competition by identifying an underlying dynamic model [83] that is a multispecies version of the Beverton-Holt population model [84] or two-species Leslie-Gower model [85] with the added inclusion of a maturation delay. Envision a model for adult female mosquitoes given by

$$x_i(s,t+1) = \sigma_i x_i(s,t) + \frac{\lambda_i x_i(s,t)}{1 + q_i(s,t)\sum_j A_{ij} x_j(s,t-\delta)}, \quad i \in \{Aa, Ac, Ag\}.$$

The dynamics of the set of competing species are described at each day t with immature life history stages of duration $\delta = 10$ days. The dynamic model allows for competitive effects among immature individuals of species j on species i as mediated by $A_{ij} \geq 0$ that emerge as reproductive adults after the conclusion of the immature life history stage. It is assumed that $A_{ii} = 1$ so that strength of intraspecific competition is defined by $q_i(s,t)$. The density independent per-capita growth rate is λ_i. The probability of daily adult survival is σ_i. The positive parameter $q_i(s,t) > 0$ varies through space and time to set the equilibrium level of species i.

If $\boldsymbol{q}(s,t)$ is held constant, then from [83] the resulting equilibrium is

$$\boldsymbol{x}^*(s,t) = \boldsymbol{A}^{-1}\text{diag}\left[\boldsymbol{\lambda} - \mathbf{1} + \boldsymbol{\sigma}\right]\text{diag}[\mathbf{1} - \boldsymbol{\sigma}]^{-1}\boldsymbol{q}(s,t).$$

Let $\boldsymbol{q}(s,t)$ have spatial and temporal average $\mathbb{E}[\boldsymbol{q}(s,t)] = \bar{\boldsymbol{q}}$. The predicted average equilibrium is then defined by the spatial and temporal average,

$$\hat{\boldsymbol{x}}^* = \mathbb{E}\left[\boldsymbol{x}^*(s,t)\right] = \boldsymbol{A}^{-1}\text{diag}\left[\boldsymbol{\lambda} - \mathbf{1} + \boldsymbol{\sigma}\right]\text{diag}[\mathbf{1} - \boldsymbol{\sigma}]^{-1}\bar{\boldsymbol{q}}.$$

Given the above model, if it is assumed that the average vector abundance $\bar{x}_m$ of species m over all sites and times is an approximation of its predicted equilibrium vector abundance, such that $\bar{\boldsymbol{x}} \approx \hat{\boldsymbol{x}}^*$, then following [83] a minimisation problem can be constructed that seeks to shift the bottom of the "bowl" formed by the basin of attraction towards the estimated equilibrium $\bar{\boldsymbol{x}}$ while increasing the magnitude of interspecific competition described by the off-diagonal entries in $\boldsymbol{A}$. The underlying premise is that species "seek" to maximise available resources, thus increasing competition, and are more likely to be observed near an equilibrium that is stable and more resilient [83]. Stability is measured by the spectral radius, or largest eigenvalue by magnitude, of the linearised system around the equilibrium (see box). If the spectral radius is less than one, then the system is locally stable such that species abundances will eventually return to equilibrium after a small perturbation. As the spectral radius decreases below one, the return time to the equilibrium following a perturbation decreases and the system is said to be more resilient [86]. The symmetric positive definite structure of the niche overlap method is retained for $\boldsymbol{A}$ in this approach.

Methods Box: Understanding Stability and Resilience via the Spectral Radius

The potential for stable coexistence of competing species is assessed by examining how small disturbances behave around the equilibrium of a dynamic system. This mathematical approach uses a linearised approximation of a often nonlinear system to predict how species respond to small changes in abundances near equilibrium. For the nonlinear delayed model considered here, the linear approximation depends on the competition matrix $\boldsymbol{A}$, equilibrium abundances $\boldsymbol{x}^*$, and known fecundity and survival parameters [83]. The diagonal terms in the competition matrix $\boldsymbol{A}$ determine the strength of intraspecific competition, while off-diagonal terms determine the strength of interspecific competition. The largest eigenvalue of a linearised discrete time system is called the spectral radius that is denoted by ρ. The spectral radius determines whether stable coexistence is possible:

$\rho < 1$**:** the system is locally stable and will return to the equilibrium after a small disturbance, which means the species will persist and coexist.

$\rho > 1$**:** the system is unstable and may not return to the equilibrium after a small disturbance, which could lead to local species extinctions.

Moreover, the magnitude of the spectral radius determines the resilience of the system that is defined by the asymptotic return time to a stable equilibrium: a smaller spectral radius indicates greater resilience and faster rate of return after a perturbation that increases the chances of persistence and reduces the risk of extinction [86]. In this study, the concepts of stability and resilience defined by the spectral radius are used to evaluate the plausibility of competitive coexistence among *An. arabiensis*, *An. coluzzii*, and *An. gambiae s.s.*. The strength of competition among species is estimated by selecting parameter combinations that maximise the product of system resilience and interspecific competition given the observed species relative abundances [83]. This approach finds an optimised system where competition for resources and the likelihood to persist under natural environmental variability are both maximised.

The estimate used the ten day immature stage duration ($\delta_m = 10$ days) with fecundity rate of $\lambda_m = 9/2$ female adults estimated in [37] for *An. gambiae s.l.* and adjusted for equal sex ratio. See also [83] for an example using the empirical results of Pombi et al. [30] for central Burkina Faso only. Adult daily survival probabilities across the species range of *An. gambiae s.l.* were elicited from domain experts [80] accounting for species differences among *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* using a prior elicitation method for GLMs [81]. The average predicted survival probabilities of $\boldsymbol{\sigma} = [0.85, 0.84, 0.84]^\top$ for the species *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.*, respectively, were obtained from [80]. The estimate of competition does not depend on $\boldsymbol{q}$ in the method of [83], so that the above assumptions allowed estimation of the interspecific density dependent coefficients A_{ij} with $i \neq j$ over all sites in Burkina Faso for the period 2002–2020.

Results

Relative Abundance and Per Capita Abundance GLMs

The coefficient estimates and standard errors of the best fitting model for the log odds of *An. coluzzii* and *An. gambiae s.s.* against *An. arabiensis* as the baseline or reference category are presented in Table S3. The backwards stepwise selection resulted in a best fitting final model having a BIC value of 49,832 compared to 49,899 for the initial full model. The meteorological covariates of precipitation, temperature and relative humidity were retained as quadratic relationships, with both linear and quadratic terms retained as useful predictors of relative abundance, in the best fitting model. The relative abundances of both *An. coluzzii* and *An. gambiae s.s.* showed a decreasing trend with higher latitudes, though this effect was modified by interactions with temperature, human population density, elevation, distance to coasts and the probability of permanent water. Human population density was evidently an important predictor included as both linear and quadratic terms as well as interactions, where it showed a negative interaction with longitude for the log odds of *An. coluzzii* and *An. gambiae s.s.* The estimates indicated an increasing annual trend in the relative abundance of *An. coluzzii* and *An. gambiae s.s.*, but the temporal trends also interacted with ITN use. By the end of the observation period in 2021, the estimated effect of ITN use suggested a suppression of *An. gambiae s.s.* in favour of *An. arabiensis*.

Predictions from the multinomial GLM are averaged through time and presented across the spatial scope in Figure S21. The associated predicted variance on the linear predictor scale is presented in Figure S22. Higher relative abundance was predicted for *An. coluzzii* in areas located along the northwestern and eastern coasts of the Gulf of Guinea, and inland central West Africa and central Africa. The latter region lacked relative abundance data sources (Figure S1) and predictions of higher relative abundance for *An. coluzzii* versus *An. gambiae s.s.* were accompanied by high uncertainty (Figure S22). Greater relative abundance was predicted for *An. gambiae s.s.* versus *An. arabiensis* south of the Sahel in West Africa, eastern coastal Madagascar, and some regions near Lake Victoria. High relative abundance for *An. arabiensis* was predicted in southern Africa, the Sahel, and much of East Africa including the Horn of Africa, Ethiopian highlands and some regions near Lake Victoria.

The parameter estimates for the negative binomial GLM are provided in Table S5. Observed counts were strongly affected by observation method. Alternative methods to indoor human landing catch (MBCI) were generally estimated to produce higher counts per deployment. Both the precipitation and the censored human population density covariates indicated concave relationships with per capita vector abundance (Figure S23). The concave relationship between per capita vector abundance and human population density induced a unimodal relationship between total vector abundance and human population in a 5 km $\times$ 5 km grid cell. Total adult vector abundance peaked at intermediate human populations and declined towards zero as the human population approached the urban threshold of 50,000 people in a 5 km $\times$ 5 km grid cell. A positive interaction was estimated between human population density and precipitation, and a negative interaction between human population density and relative humidity (Table S5). ITN use was negatively associated with per capita abundance. The highest overall per capita abundance was predicted near the equator, along

the Gulf of Guinea, the Ethiopian Highlands and Madagascar (Figure S24a) with some areas of high uncertainty (Figure S24b) located in regions with comparatively few abundance data sources (Figure S2). These predictions from the negative binomial GLM do not account for temperature tolerances that are addressed in the next section below.

Predicted Equilibrium Species Abundances

The multinomial and negative binomial GLMs were used to predict the daily equilibrium abundances of adult female *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* in each 5 km × 5 km grid cell across the range of *An. gambiae s.l.* in Africa after accounting for temperature tolerances, daily probabilities of human feeding, and the human population. The average equilibrium vector abundances of each species are presented in Figure 1. The predicted equilibrium vector abundance of *An. arabiensis* was highest in eastern Africa and the Sahelian region and Madagascar with low areal densities in central Africa. The predicted equilibrium vector abundance of *An. coluzzii* was highest in western Africa and absent in eastern Africa. The highest predicted equilibrium vector abundance for *An. gambiae s.s.* occurred in western Africa, the Lake Victoria basin and Madagascar.

Seasonal predicted equilibrium abundances are visualised in Figures 2, 3 and 4, where seasonally high abundances for all three species occurred during the third quarter (Q3, July–September) in West Africa when precipitation was high (Figure S6). High abundance during this season was also predicted for *An. arabiensis* in regions of the Ethiopian highlands near the limits of the temperature tolerance for *An. gambiae s.l.* The opposite seasonal pattern was predicted in the southern parts of the ranges of *An. arabiensis* and *An. gambiae s.s.*, where precipitation is generally higher in the first quarter (Q1, January–March) compared to the third quarter (Figure S6). In these southern regions, high equilibrium abundances were predicted for both *An. arabiensis* and *An. gambiae s.s.* in the first quarter of year.

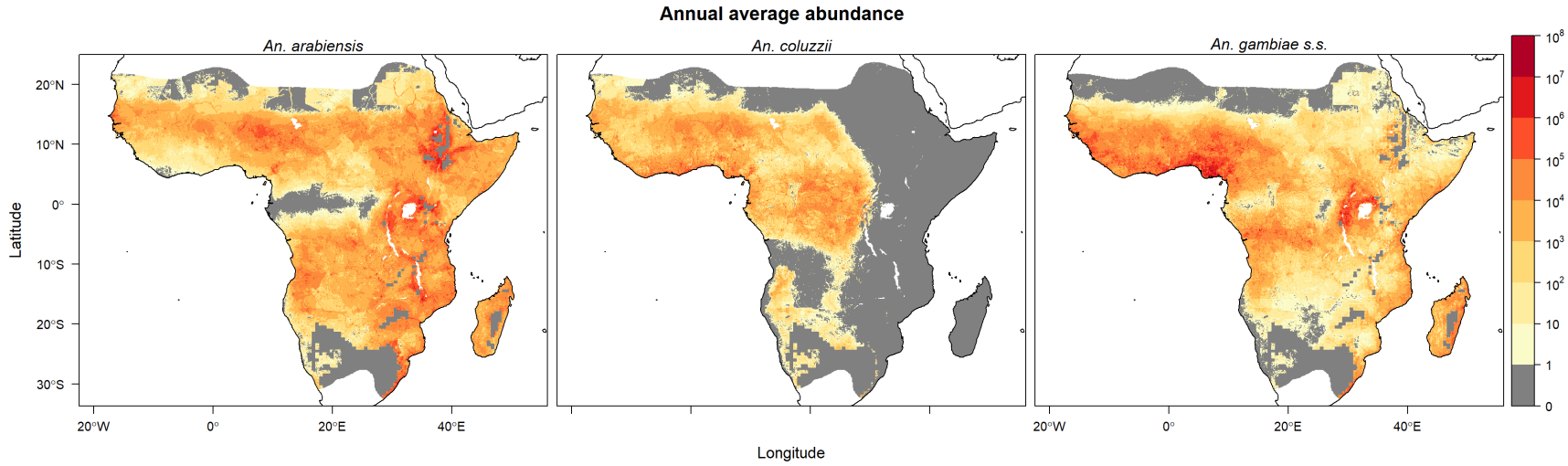


Figure 1: Annual average of predicted equilibrium vector abundances of dominant malaria vector species of *An. gambiae s.l.* per 25 square kilometres over years 2002–2020. Predictions are masked to the potential range of *An. gambiae s.l.* [24] provided by the Malaria Atlas Project (`https://malariaatlas.org/`) under the terms of a Creative Commons Attribution 4.0 International License available at `https://creativecommons.org/licenses/by/4.0/`. Note logarithmic scale.

Predicted equilibrium abundances for the three species also varied across years as illustrated by continent wide predictions separated by six year intervals between 2002–2020 (Figures S18, S19, S20). The annual average abundance of *An. arabiensis* was generally predicted to be higher in 2002 than in 2020 for most regions outside of the Ethiopian highlands. The annual average abundance of *An. coluzzii* was predicted to be higher in central Africa in 2002 than in 2020, and lower in West Africa in 2002 compared to 2020. Predictions of annual average anomalies for *An. gambiae s.s.* were more similar to *An. arabiensis* than *An. coluzzii*.

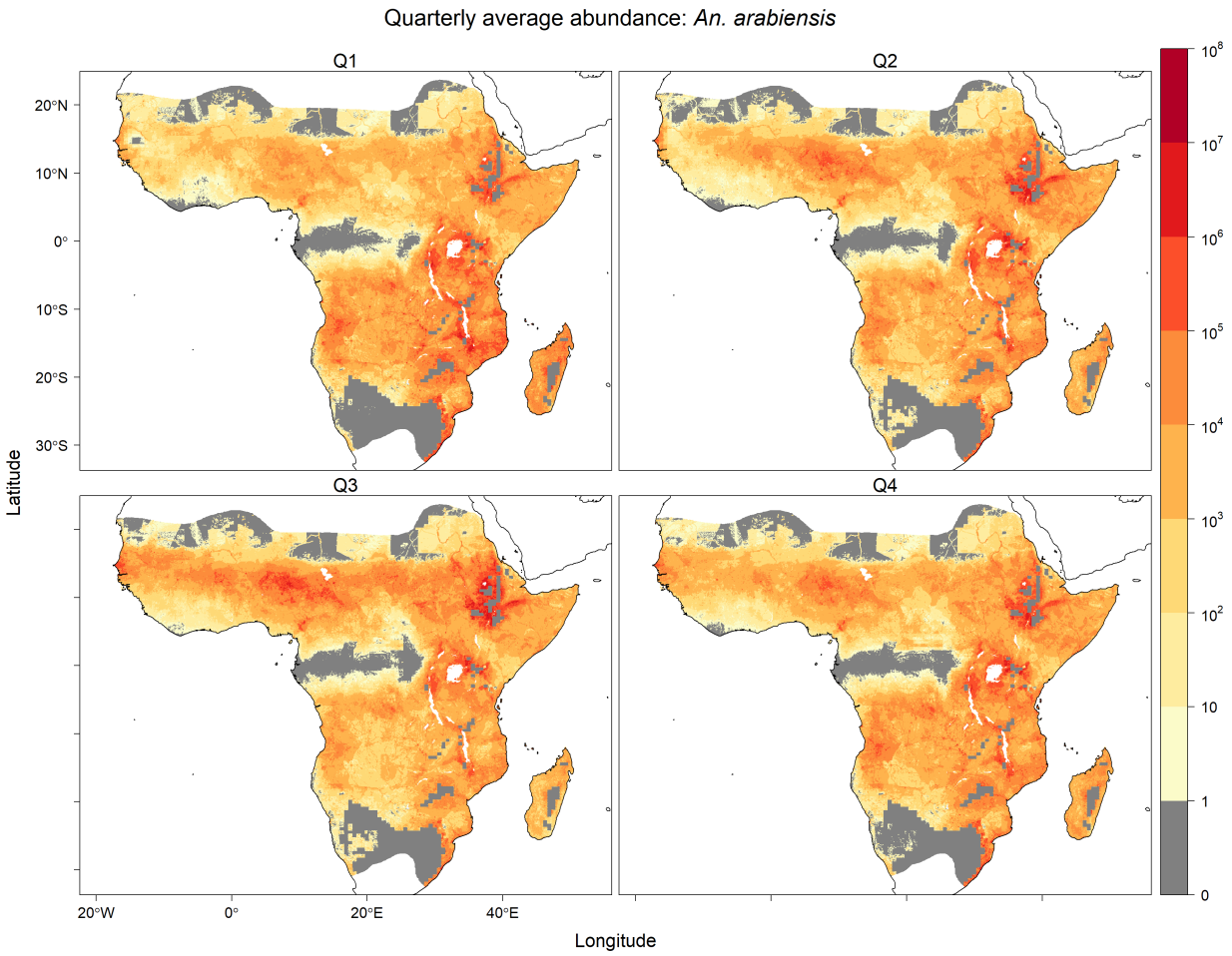


Figure 2: Equilibrium vector abundance of *An. arabiensis* per 25 square kilometres by quarter (Q1: January–March, Q2: April–June, Q3: July–September, Q4: October–December) averaged over years 2002–2020. Predictions are masked to the potential range of *An. gambiae s.l.* [24] provided by the Malaria Atlas Project (`https://malariaatlas.org/`) under the terms of a Creative Commons Attribution 4.0 International License available at `https://creativecommons.org/licenses/by/4.0/`. Note logarithmic scale.

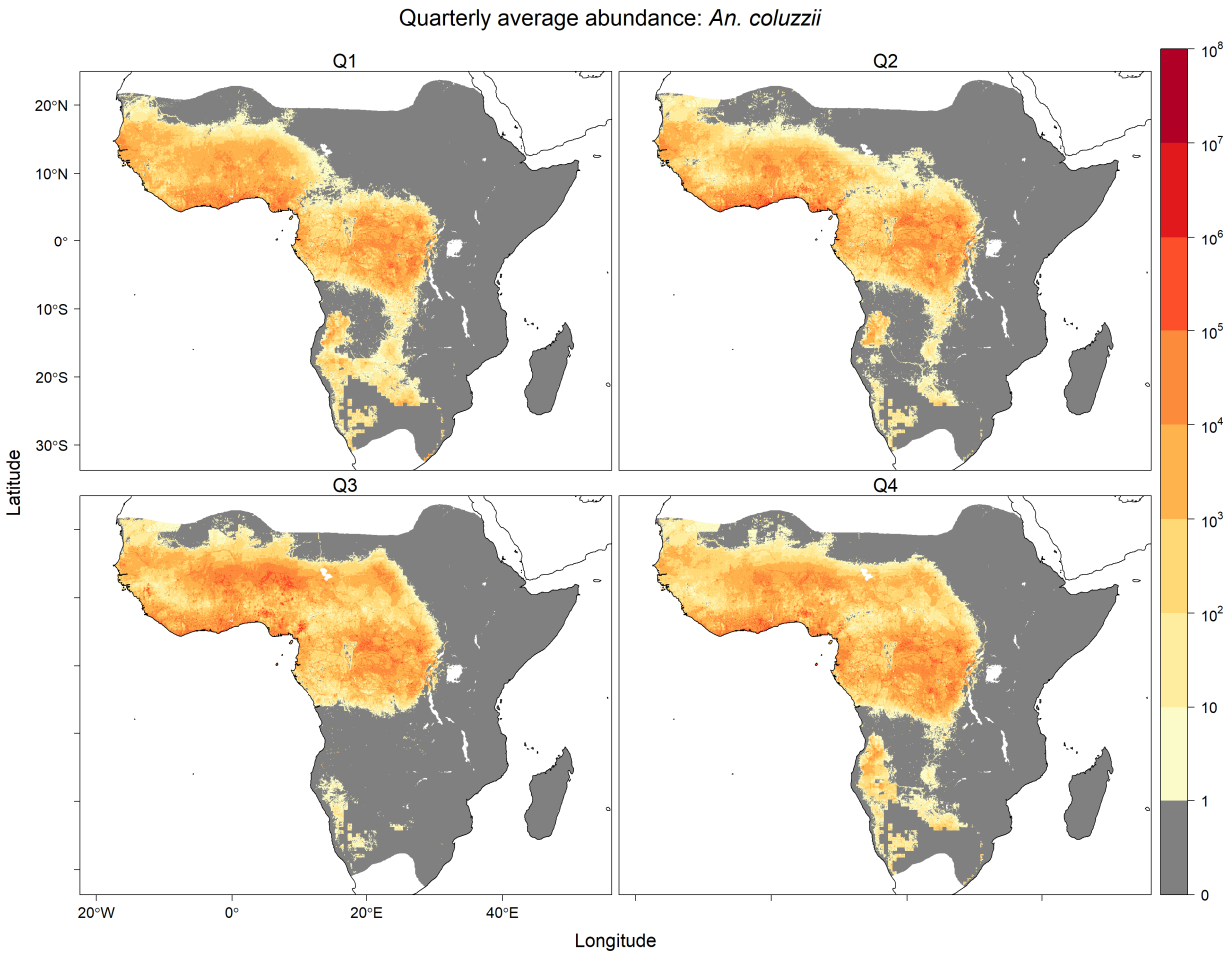


Figure 3: Equilibrium vector abundance of *An. coluzzii* per 25 square kilometres by quarter (Q1: January–March, Q2: April–June, Q3: July–September, Q4: October–December) averaged over years 2002–2020. Predictions are masked to the potential range of *An. gambiae s.l.* [24] provided by the Malaria Atlas Project (`https://malariaatlas.org/`) under the terms of a Creative Commons Attribution 4.0 International License available at `https://creativecommons.org/licenses/by/4.0/`. Note logarithmic scale.

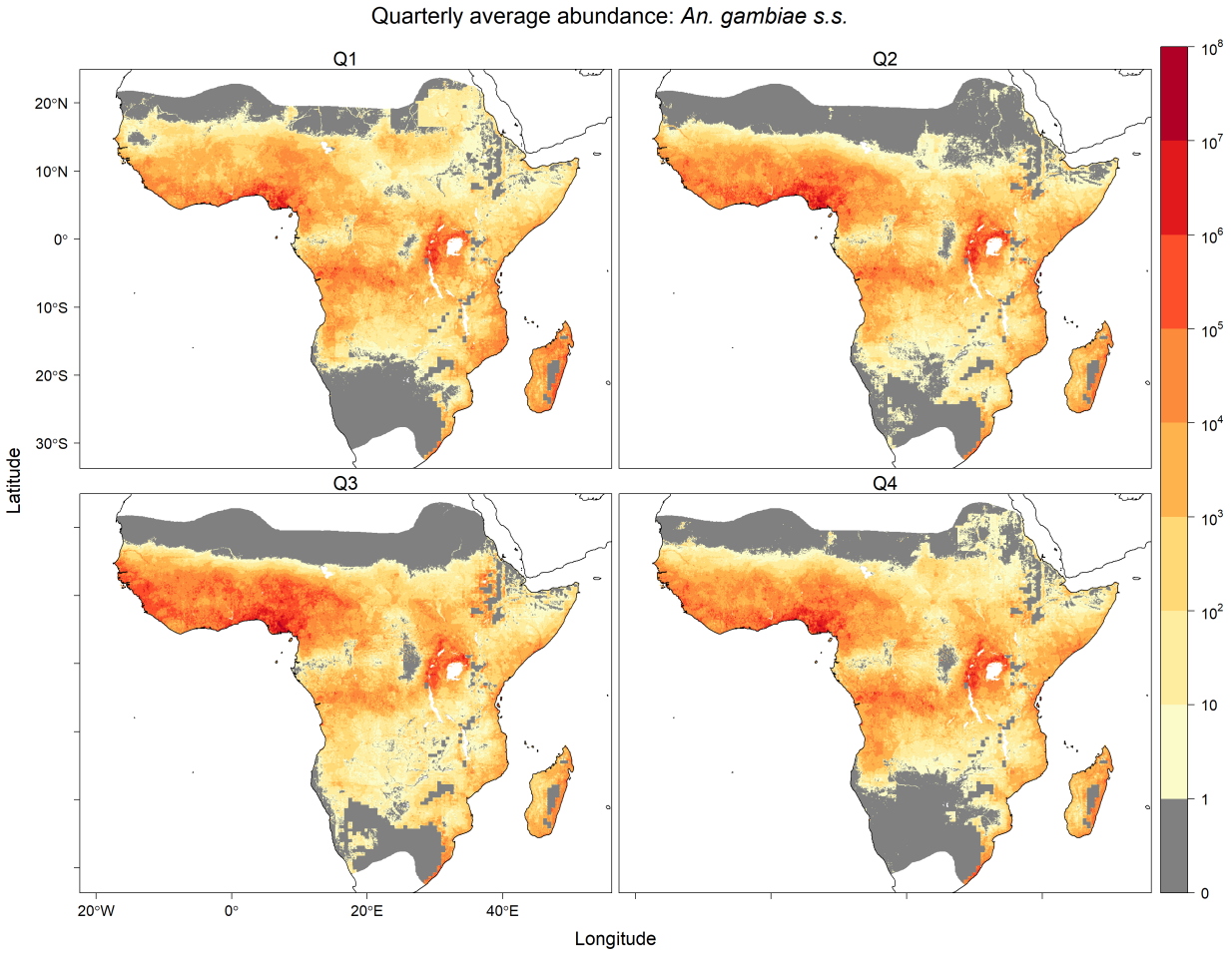


Figure 4: Equilibrium vector abundance of *An. gambiae* s.s. per 25 square kilometres by quarter (Q1: January–March, Q2: April–June, Q3: July–September, Q4: October–December) averaged over years 2002–2020. Predictions are masked to the potential range of *An. gambiae s.l.* [24] provided by the Malaria Atlas Project (`https://malariaatlas.org/`) under the terms of a Creative Commons Attribution 4.0 International License available at `https://creativecommons.org/licenses/by/4.0/`. Note logarithmic scale.

Comparative Evaluation of Equilibrium Abundance Predictions

The predicted daily species equilibrium vector abundances were evaluated by comparison to previous empirical analyses of large spatial scale prediction of species in the *An. gambiae* species complex from four distinct perspectives: i) species distribution models of species occurrence, ii) relative abundance models, iii) observed entomological inoculation rate (EIR) and iv) effective population size estimates from genotype data. Studies in i) and ii) that incorporate species occurrence data may not be entirely independent of the data used for the multinomial GLM model of species relative abundance in this study. However, the comparisons iii) and iv) provide independent empirical comparisons. None of the comparative studies allow for intra-annually varying covariates, and so comparisons considered only the mean equilibrium vector abundance (Figure 1) averaged over daily, and seasonal (e.g., Figures 2, 3 and 4), variation for the 2002–2020 prediction period.

i) Species Distribution Models. Species distribution models based on species occurrence do not estimate abundance but instead "potential distribution" [26] or the relative probability of presence (e.g., [24]) and so are not directly comparable with this study. Qualitative comparison is still possible for the several species distribution models applied within the *An. gambiae* complex [22, 24, 26, 87, 88]. Of these, only Wiebe et al. [24] predict relative probabilities of presence for all three species based on annually varying covariates. The species distribution model predictions in that study [24] are clipped to independently derived species range limits, but comparison within the limits suggests broad qualitative agreement with the average equilibrium vector abundance predictions of Figure 1: *An. arabiensis* estimates are highest in the northern, eastern and southern regions (see also [22, 26, 88]); *An. coluzzii* estimates are highest in western Africa; and *An. gambiae s.s.* estimates are highest in western Africa and near Lake Victoria with high estimates also occurring in central Africa and the eastern coast of Madagascar. The range limits imposed by Wiebe et al. [24] exclude *An. gambiae s.s.* from the northeastern portion of the range for *An. gambiae s.l.* In this region, Figure 1 shows lower densities extending into the Horn of Africa (the Somali Peninsula and eastern Ethiopia) compared to *An. arabiensis*. Species distribution models that permit predictions into the Horn of Africa suggest either presence [22, 87] or absence [88] of *An. gambiae s.s.*

ii) Relative Abundance Models. Extrapolations from relative abundance estimates of *An. arabiensis* versus the aggregate of *An. coluzzii* and *An. gambiae s.s.* collated for Nigeria, Niger and the Senegambia in an early study by Lindsay et al. [28] suggest strong similarities with Figure 1. A similar pattern is predicted by Sinka et al. [27]. A discrepancy is noted in southern Tanzania and Mozambique, where the relative abundance of *An. arabiensis* is higher in the two comparative studies than in Figure 1. The extrapolated predictions of the Lindsay et al. study [28] suggest higher relative abundance of *An. gambiae s.s.* in some areas of Ethiopia, unlike the predictions of Figure 1 or the study by Sinka et al. [27] that accord with empirical observations of *An. arabiensis* rather than *An. gambiae s.s.* in this area [89]. Although restricted to comparisons of *An. coluzzii* and *An. gambiae s.s.* relative abundance based on species occurrence data from western and central Africa, Tene Fossog et al. [29] similarly predict higher estimates of the former species across the Sahel and higher estimates

of *An. gambiae s.s.* in the West Sudanian savannah. A discrepancy occurs in central Africa, where there is a region of higher abundance predicted for *An. coluzzii* compared to *An. gambiae s.s.* in Figure 1, whereas the reverse is suggested by the predictions of Tene Fossog et al. [29]. As noted previously, the predicted relative abundance in this area is a region of extrapolation in Figure 1 with associated higher uncertainty (Figure S24b) due to the lack of available genotyped samples. Consideration of field collected larvae reared to emergence in the laboratory suggests the presence of both species in central Africa [54].

iii) Entomological Inoculation Rate. Outside of areas with highly successful vector control and malaria case intervention, Rogers et al. [22] suggest a strong relationship between vector abundance and the entomological inoculation rate (EIR) that is equal to the vector abundance multiplied by the human feeding and sporozoite rates divided by the human population. Proceeding therefore with the predicted species equilibrium vector abundances, where ITN use is incorporated, allows comparison with observations of EIR across East Africa before (years 2000–2010) and after (years 2011–2022) widespread increases in ITN use [90]. For this comparison, EIR is estimated from the predicted species equilibrium vector abundances (Figure 1) averaged over 2002–2020 in each grid cell east of 28°E longitude and south of 4°N latitude. Human feeding rates were estimated from the daily feeding probabilities reported in the Methods and the median sporozoite rates *An. gambiae s.l.* reported for both time periods by Msugupakulya et al. [90]. There is substantial variability in these estimates and also the EIR observations [90] that include uncertainty in the sporozoite rate and proportion of EIR attributable to various vector species. Comparisons are therefore based on central 0.80 probability intervals constructed from estimates of the 0.10 and 0.90 quantiles. Using observations of EIR [90], the central 0.80 interval for annualised EIR attributable to *An. gambiae s.l.* in East Africa is (0.15, 175) for the 2000–2010 period and (0, 20) for the 2011–2022 period (see Supplementary Information in [90]). The corresponding 0.80 central intervals for annualised EIR based on the predicted species equilibrium vector abundances were (0, 353) in 2002 and (0, 114) in 2020. It should be noted that occurrences of zero sporozoite rates, which are observed for *An. gambiae s.l.* [90], will underestimate vector abundance based on observed EIR. The equilibrium based estimates of EIR, which assumed median sporozoite rates, were indeed somewhat higher than observed EIR. Despite the requisite assumptions and caveats, the studies suggest overlapping 0.80 central intervals for EIR in each time period with decreasing upper bounds since widespread increases in ITN use.

iv) Effective Population Size. A pan-African estimate of mean total effective population size for *An. coluzzii* and *An. gambiae s.s.* obtained from allele frequency data is between 1.6×10^9 and 3.1×10^9 after accounting for seasonal fluctuations that are bounded by a minimum of 3.1×10^7 and a maximum of 6.2×10^9 [91]. The genetic study uses genotype data collected between 2000 and 2012 with the majority of samples collected since 2009 [92]. In 2011, ITN use was widely increasing in East Africa, as noted in iii) above. Regarding the predicted overall equilibrium vector abundance aggregated over the entirety of the *An. gambiae s.l.* species range across sub-Saharan Africa, multiplying the predicted average daily aggregate vector equilibrium abundances for adult female *An. coluzzii* and *An. gambiae s.s.*

by a factor of two to allow for a roughly equal sex ratio suggested a total population average daily estimate that decreased from 8.6×10^{10} individuals in 2002 to 5.3×10^{10} individuals in 2014 (Figures S19 and S20 depict the 2002 and 2014 anomalies from Figure 1 for *An. coluzzii* and *An. gambiae s.s.*). Assuming a ratio of total to effective population size of 0.10 [93] suggests an effective population size of 5.3×10^9 based on the combined predicted equilibrium abundances of *An. coluzzii* and *An. gambiae s.s.* in 2014, which is outside the estimated range for the mean effective population size although of the same order of magnitude and below the seasonal maximum effective population size estimate.

Estimated Niche Overlap and Interspecific Competition

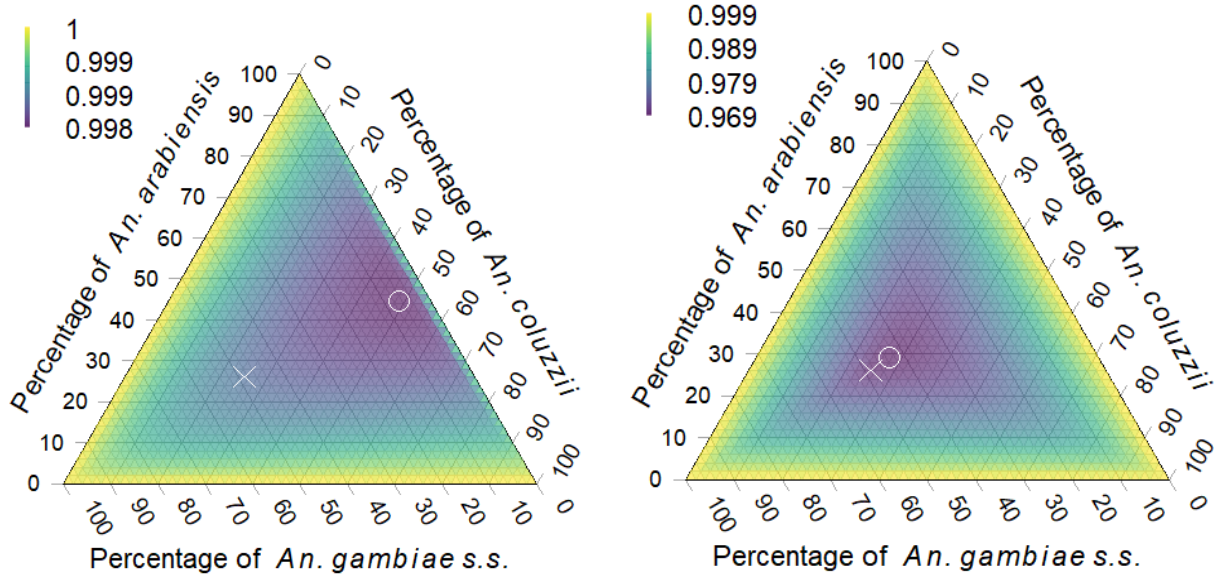
In Burkina Faso for the years 2002–2020, *An. gambiae s.s.* was predicted to be the dominant species (Figure 5) compared to the other two species each with approximately 25% species composition. The species with the greatest niche overlap was *An. gambiae s.s.* (Table 1). The estimated competitive interactions in $\hat{\boldsymbol{A}}$ that involved *An. gambiae s.s.* were the strongest and of the same magnitude (Table 1). The competitive effect of *An. gambiae s.s.* on *An. arabiensis* and *An. coluzzii* was 39% greater than the reciprocal effects of the latter species on the former. The equilibrium that maximised resilience was further away from the observed mean abundance $\bar{\boldsymbol{x}}$ for the Pianka niche overlap estimates (Figure 5a) compared to the optimised competition estimates (Figure 5b). As noted in the Methods, the Pianka index could be considered as a descriptive measure of niche overlap rather than an estimate of competitive interactions [31]. In contrast, for the optimised estimate of interspecific competition ($\hat{\boldsymbol{A}}$), the observed mean abundance approximated the most resilient species composition.

Table 1: Estimated interspecific competitive interactions in Burkina Faso for the years 2002–2020 using the Pianka niche overlap ($\boldsymbol{\alpha}$) and optimised ($\hat{\boldsymbol{A}}$) methods: *An. arabiensis* (Aa) versus *An. coluzzii* (Ac), *An. arabiensis* versus *An. gambiae s.s.* (Ag), and *An. coluzzii* versus *An. gambiae s.s.*

	Aa vs Ac	Aa vs Ag	Ac vs Ag
Pianka	0.94	0.43	0.51
Optimised	0.23	0.32	0.32

Discussion

Management of malaria requires knowledge of the abundance of important vector species [1, 94] such as members of the *An. gambiae s.l.* species complex [6] to allow efficient malaria control by conventional methods [8] or potential new technologies including genetic biocontrol [7]. The spatiotemporal distribution of vector species abundances, and hence threat of malaria, will change with environmental factors [79] and hence potentially also with climate change [26]. Here, statistically estimated environmental relationships are used to give insights into the predicted abundances of the three important malaria vectors of *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* across their combined ranges of sub-Saharan



(a) $\boldsymbol{A} = \boldsymbol{\alpha}$.

(b) $\boldsymbol{A} = \hat{\boldsymbol{A}}$.

Figure 5: Spectral radius for all possible equilibria where species coexist given (a) Pianka niche overlap estimate $\boldsymbol{A} = \boldsymbol{\alpha}$ and (b) optimised estimate $\boldsymbol{A} = \hat{\boldsymbol{A}}$ for the interspecific competitive interactions. Points show the location of species equilibrium with lowest spectral radius (○) and the location of the predicted mean equilibrium relative abundances (×). Note change in colour scale for (a) versus (b).

Africa, and also evaluate the potential for interspecific interactions through niche overlap or competition.

Meteorological factors, such as precipitation and relative humidity, and human factors, such as population density and use of insecticide treated nets (ITNs), were both linked to the relative abundances (species composition) of *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* and the per capita abundance of *An. gambiae s.l.* These analyses of species relative abundance from genotyped species occurrence data and abundance of *An. gambiae s.l.* from count data were combined to produce joint predictions of daily vector equilibrium abundances for the years 2002–2020 (Figures 1). Predictions showed marked seasonal changes in West Africa for predicted equilibrium abundances of all three species (Figures 2, 3, 4), and a latitudinal progression for *An. arabiensis* and *An. gambiae s.s.* abundances that tracked the wet seasons of higher latitudes. The seasonal variation is explained by the intra-annually varying meteorological covariates of temperature, relative humidity and precipitation.

Predicted equilibrium abundances also varied among years (Figures S18, S19, S20), where for example abundances of *An. arabiensis* and *An. gambiae s.s.* were higher across much of the species ranges in 2002 compared to 2020. Many factors contributed to interannual variation. During this time period, interannual variability in meteorological conditions occurred with more cold days and precipitation for some regions in East Africa, for example, at the end of the 2002–2020 prediction period compared to the beginning [95, 96]. Human population density also increased overall with increased urbanisation and built-up areas [97],

while ITN coverage by household in sub-Saharan Africa increased from 5% in 2000 to 70% in 2022 [3]. The equilibrium abundance of *An. gambiae s.l.* was predicted to peak at intermediate levels of human population (Figure S23), and ITN coverage was linked with decreased *An. gambiae s.l.* per capita abundance (Table S5). The relative abundances of *An. coluzzii* and *An. gambiae s.s.* versus *An. arabiensis* were linked to a negative interaction between ITN coverage and year of sample (Table S3). It is possible that the negative interaction between year and ITN coverage for *An. coluzzii* and *An. gambiae s.s.* may be linked to insecticide resistance that is developing with time for some insecticides in certain regions [72], but interpretations of ITN efficacy must be balanced against the observational nature of this study. For *An. coluzzii*, the interaction was indeed small relative to positive estimates for temporal trend and ITN coverage, such that *An. coluzzii* would be predicted to increase across years at the expense of *An. arabiensis* if all other covariates were held constant. The trend of increasing *An. coluzzii* relative abundance could be caused in part to climatic or land use factors, but it could also be explained by the relatively recent species characterisation of *An. coluzzii* and *An. gambiae s.s.* combined with improved genotyping capabilities of the *An. gambiae* species complex [29, 30, 42]. For example, the occurrence of *An. coluzzii* in northwestern Kenya has recently been shown using whole genome sequencing of historical samples dated from 2006 and 2019, where taxonomic characterisation based on previous molecular methods limited to identification of alternative species would have missed detection of *An. coluzzii* [69]. The multinomial GLM estimates (Table S3) suggested that the relative abundance of *An. gambiae s.s.* would generally decrease compared to *An. arabiensis* in later years with high ITN coverage. Temporal shifts in species composition away from *An. gambiae s.s.* towards *An. arabiensis* in areas with high ITN use has been attributed to the outdoor feeding and resting behaviour of the latter species [6, 67, 98].

The predicted equilibrium vector abundance at a given site is a potential abundance that may not be realised for many important ecological factors not considered in this model due to local conditions not accounted for in this continental scale analysis. Predictions based on the model assumptions were therefore comparatively evaluated against four alternative empirical modelling perspectives with general agreement: i) species distribution models based on species occurrence records [22, 24, 26, 87, 88], ii) relative abundance models [27–29], iii) entomological inoculation rate (EIR) estimates in eastern and southern Africa [90] and iv) effective population size estimates from genotype data [91]. Overall, results showed qualitatively similar patterns, although it should be noted that none of the comparisons in i)-ii) provide species-specific estimates of vector abundances. The comparisons iii) and iv), however, provide independent empirical estimates of EIR and vector abundances that are not dissimilar from estimates based on predicted species vector equilibrium abundances.

Analyses suggested that the three species can coexist (Figure 5) despite evidence of substantial niche overlap and interspecific competition (Table 1) across the shared species range. The estimates for niche overlap among the three species (Table 1) for Burkina Faso during 2002–2020 followed a similar pattern to those reported by Pombi et al. [30] in a localised study that derived the estimates from field observations at three sites in central Burkina Faso where all three species coexist. The niche overlap estimates from this study also included the southern region of the country where *An. gambiae s.s.* predominates [99]. Both studies estimated the highest Pianka niche overlap indices between *An. arabiensis* and *An. coluzzii* and similar niche overlaps for comparisons that included *An. gambiae s.s.* (Table

1). The estimates of interspecific competition that maximised interspecific competition and resilience suggested substantial competition was possible in Burkina Faso (Table 1), where all three species occur.

Statistical studies that develop dynamic population models for anophelines from time series data in Africa have targeted species complexes such as *An. gambiae s.l.* or *An. funestus* [20, 21, 100]. Such studies would additionally benefit from geographically and temporally extensive time series data at the species level. Ideally, daily times series of species-level abundances collected over long durations would be available from many sites geographically dispersed across sub-Saharan Africa. Then delayed density dependent models, such as described in the Methods, could be implemented within a state space modelling framework to estimate the parameters of interest while allowing for non-equilibrium dynamics. Without such time series data, this study used smoothed moving averages of the meteorological covariates to approximate daily variation among environmental carrying capacities and equilibrium abundances of the species based on maximum likelihood estimates. The predicted equilibrium is spatially and temporally varying due to changing environmental conditions that modify the Grinellian niche of the constituent species [101], which in turn is modified by interspecific competition to produce the realised niche [102]. However, the predicted equilibrium is only a measure of potential abundance: the species may be limited by dispersal capability or source-sink metapopulation dynamics such that it is either above or below the predicted equilibrium abundance over the long run [11, 103–107]. Moreover, species abundances do not exactly attain potential equilibrium due to changing environment, dispersal and immigration. The proposed density dependent model may be elaborated to account for immigration and emigration, see [12] for a non-delayed example, as well as delayed maturation with introgression among the constituent species of a species complex such as *An. gambiae s.l.* [80].

Given the sparse data available for this study across the broad spatial and temporal scope of interest, and predicted changes of species abundance both within and among years, sophisticated analyses for causal attribution among cumulative effects were commensurately limited. Higher quality longitudinal abundance data at the species level of sufficient subseasonal temporal frequency would greatly help statistical evaluation of competing causal hypotheses, such as vector control, malaria interventions, land use and climate. The predictions and results of this study accounted for intra-annual shifts in species composition, but could not explicitly control for many important covariates. For example, factors such as land use impact not only larval habitat but also the feeding habit of *An. arabiensis* that feeds on livestock to a greater degree than *An. coluzzii* and *An. gambiae s.s.* [23]. Inclusion of spatiotemporal covariates of livestock density in particular would assist predictions of *An. arabiensis* that exhibits greater zoophily biting habit. Improved estimation of the relationship between species abundances and human feeding rates would also require abundance data for each of the three species, whereas most of the spatially extensive and longitudinal data available for this study were limited to counts for the entire species complex. Although the ecologically important and spatially and temporally varying covariate of livestock density was not explicitly represented, the lower probability of human host feeding by *An. arabiensis* compared to the other species was incorporated into the predictions. The abundance estimates also accounted for collection methods, but the analysis of the relative abundance data did not incorporate sampling effects because species identification by genotyping were

often based on samples collected from the field by multiple methods that did not document, for example, indoor versus outdoor collection. Hence, predictions of *An. arabiensis* relative abundance must be interpreted cautiously and particularly so in regional areas with high livestock density relative to human inhabitant density.

The modelling approach used here elaborated previous relative abundance models [27–29] to consider the three species of *An. arabiensis*, *An. coluzzii* and *An. gambiae*. The daily predictions of equilibrium abundance at the continental scale for the years 2002–2020 are only approximations of potential mosquito abundance at any given place and time. Daily varying meteorological covariates were used to predict the relative abundances of these species and also the per capita vector abundance of *An. gambiae s.l.* across sub-Saharan Africa for the first time. These statistical models used covariate and response data collated to a grid of 5 km $\times$ 5 km cells for analysis and prediction at the continental scale. Other regional and continental scale studies use this same resolution [23, 24, 27, 29, 54] as a suitable choice for the ecology and dispersal of species in the *An. gambiae* species complex [23, 29, 54]. At finer resolutions, meteorological factors such as precipitation and temperature can for some locales be more important predictors of *An. gambiae s.s.* abundance counts than microscale covariates such as the number of occupants per home or distance to larval habitat [108]. Climatological factors remain important predictors for resolutions that are coarser than the choice of 5 km $\times$ 5 km cells [26, 28]. The choice of 5 km $\times$ 5 km resolution, however, does not deny the known importance of microscale habitat variability to the risk of malaria transmission [73], and is more suitable for interpretation at coarser national or sub-national levels [24].

Conclusion

Meteorological factors, such as precipitation and relative humidity, and human factors, such as population density and insecticide treated net coverage, were found at the continental scale to be important spatially and temporally varying predictors of vector equilibrium abundances of the primary malaria vector species within the *Anopheles gambiae s.l.* species complex: *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* Daily varying meteorological covariates such as precipitation and relative humidity led to substantial variation in the predicted seasonal equilibrium abundances of the three species. ITN use was negatively associated with per capita vector abundance of the *An. gambiae* species complex. An interaction with ITN use and year predicted a net positive estimate on the ratio of *An. gambiae s.s.* versus *An. arabiensis* around 2002, but this prediction shifted to a net negative against *An. gambiae s.s.* around 2020. A longer term trend of increasing relative abundance of *An. coluzzii* was also suggested, and was possibly attributable to advances in genetic methods of species identification within the *An. gambiae* species complex that occurred over the observational period. Human population density was also an important predictor that suggested a negative associated with per capita vector abundance approaching the threshold for urban areas. Across sub-Saharan Africa, the observational analyses suggested evidence for niche overlap and competitive interactions among the three species. Additional time series data of mosquito vector species abundance, collected at fine temporal resolutions of monthly sampling intervals or higher frequencies and across broad geographical regions, are necessary

to improve discernment of ecological interactions among species and the effectiveness of vector control interventions designed to measurably reduce the burden of mosquito-borne disease.

Supplementary information

The reproducible code, data sources and licenses are summarised in the R package **AgslPredict** [39] provided as a supplement for this article.

Acknowledgements

The authors thank Scott Foster, Skipton Woolley and the anonymous reviewers for helpful comments. This article incorporates elicited information provided by the following scientific experts: Adedapo Adeogun, Nigerian Institute of Medical Research, Yaba, Lagos, Nigeria; Guillaume Carissimo, Agency for Science, Technology and Research (A*STAR), Singapore; George Christophides, Imperial College London, London, UK; Carlo Costantini, Institut de Recherche pour le Développement (IRD), Montpellier, France; Ibrahima Dia, Institut Pasteur de Dakar, Dakar, Senegal; Mawlouth Diallo, Institut Pasteur de Dakar, Dakar, Senegal; Wamdaogo Moussa Guelbeogo, Centre National de Recherche et de Formation sur le Paludisme, Ouagadougou, Burkina Faso; Stephen Higgs, Kansas State University, Manhattan, Kansas, USA; Charles M. Mbogo, KEMRI Centre for Geographic Medicine Research–Coast, Kenya; Juma Mcha, Malaria Elimination Program, Zanzibar, United Republic of Tanzania; Givemore Munhenga, National Institute for Communicable Diseases and University of Witwatersrand, Johannesburg, South Africa; Martin Rono, KEMRI Centre for Geographic Medicine Research–Coast, Kilifi, Kenya; N'fale Sagnon, Centre National de Recherche et de Formation sur le Paludisme, Ouagadougou, Burkina Faso; Dana L. Vanlandingham, Kansas State University, Manhattan, Kansas, USA. The authors thank the experts for the generous contribution of their domain expertise.

Declarations

- Funding. This work was funded by a grant to the Foundation for the National Institutes of Health from the Bill & Melinda Gates Foundation (INV-008525).

- Conflict of interest/Competing interests. The authors declare no conflicts of interest.

- Ethics approval and consent to participate. This article incorporates elicited information with participants providing their consent to participate under ethics approval by the CSIRO Social Science Human Research Ethic Committee (approval number 136/19).

- Consent for publication. The authors provide their consent for publication.

- Data availability. The data sources and licenses are summarised in the R package **AgslPredict** [39] provided as a supplement for this article.

- Materials availability. Not applicable.
- Code availability. The reproducible code is provided in the R package **AgslPredict** [39] provided as a supplement for this article.
- Author contribution. **G. R. Hosack:** Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, software, visualization, writing–original draft, writing–review and editing. **M. El-Hachem:** Conceptualization, data curation, investigation, methodology, software, validation, visualization, writing–review and editing. **A. Ickowicz:** Conceptualization, investigation, methodology, writing–review and editing. **N. J. Beeton:** Conceptualization, investigation, visualization, writing–review and editing. **A. Wilkins:** Conceptualization, writing–review and editing. **K. R. Hayes:** Conceptualization, funding acquisition, visualization, writing–review and editing. **S. S. C. Rund:** Data curation, writing–review and editing. **S. A. Kelly:** Data curation, writing–review and editing. **M. A. McDowell:** Data curation, writing–review and editing.

References

[1] Ross, R. The Prevention of Malaria. 2nd ed. London: John Murray; 1911.

[2] Macdonald, G. The Epidemiology and Control of Malaria. London: Oxford University Press; 1957.

[3] WHO. World malaria report 2024: addressing inequity in the global malaria response. Geneva: World Health Organization; 2024.

[4] Kahamba, NF, Finda, M, Ngowo, HS, Msugupakulya, BJ, Baldini, F, Koekemoer, LL, et al. Using ecological observations to improve malaria control in areas where *Anopheles funestus* is the dominant vector. Malaria Journal. 2022;21:158.

[5] James, S, Santos, M. The promise and challenge of genetic biocontrol approaches for malaria elimination. Tropical Medicine and Infectious Disease. 2023;8(4):201.

[6] Takken, W, Charlwood, D, Lindsay, SW. The behaviour of adult *Anopheles gambiae*, sub-Saharan Africa's principal malaria vector, and its relevance to malaria control: a review. Malaria Journal. 2024;23:161.

[7] Hancock, PA, North, A, Leach, AW, Winskill, P, Ghani, AC, Godfray, HCJ, et al. The potential of gene drives in malaria vector species to control malaria in African environments. Nature Communications. 2024;15:8976.

[8] Ochomo, E, Rund, SSC, Mthawanji, RS, Antonio-Nkondjio, C, Machani, M, Samake, S, et al. Mosquito control by abatement programmes in the United States: perspectives and lessons for countries in sub-Saharan Africa. Malaria Journal. 2024;23:8.

[9] Smith, DL, Battle, KE, Hay, SI, Barker, CM, Scott, TW, McKenzie, FL. Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. PLoS Pathogens. 2012;8:e1002588.

[10] May, RM. Stability and Complexity in Model Ecosystems. 2nd ed. Princeton, New Jersey: Princeton University Press; 1974.

[11] Soberón, J. Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters. 2007;10:1115-23.

[12] Thompson, PL, Guzman, LM, De Meester, L, Horváth, Z, Ptacnik, R, Vanschoenwinkel, B, et al. A process-based metacommunity framework linking local and regional scale community ecology. Ecology Letters. 2020;23:1314-29.

[13] Smith, DL, McKenzie, FE. Statics and dynamics of malaria infection in *Anopheles* mosquitoes. Malaria journal. 2004;3:13.

[14] Reiner, RC, Jr, Perkins, TA, Barker, CM, Niu, T, Chaves, LF, Ellis, AM, et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. Journal of the Royal Society Interface. 2013;10:20120921.

[15] Diekmann, O, Heesterbeek, JAP, Metz, JAJ. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology. 1990;28:365-82.

[16] Parham, PE, Michael, E. Modeling the effects of weather and climate change on malaria transmission. Environmental Health Perspectives. 2010;118:620-6.

[17] Dietz, K, Molineaux, L, Thomas, A. A malaria model tested in the African savannah. Bulletin of the World Health Organization. 1974;50:347-57.

[18] Molineaux, L, Gramiccia, G. The Garki Project. Geneva: World Health Organization; 1980.

[19] Garrett-Jones, C. The human blood index of malaria vectors in relation to epidemiological assessment. Bull World Health Org. 1964;30:241-61.

[20] White, MT, Griffin, JT, Churcher, TS, Ferguson, NM, Basáñez, MG, Ghani, AC. Modelling the impact of vector control interventions on *Anopheles gambiae* population dynamics. Parasites and Vectors. 2011;4:153.

[21] Yé, Y, Hoshen, M, Kyobutungi, C, Louis, VR, Sauerborn, R. Local scale prediction of *Plasmodium falciparum* malaria transmission in an endemic region using temperature and rainfall. Global Health Action. 2009;2:1923.

[22] Rogers, DJ, Randolph, SE, Snow, RW, Hay, SI. Satellite imagery in the study and forecast of malaria. Nature. 2002;415:710-5.

[23] Sinka, ME, Bangs, MJ, Manguin, S, Coetzee, M, Mbogo, CM, Hemingway, J, et al. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic prècis. Parasites & Vectors. 2010;3:117.

[24] Wiebe, A, Longbottom, J, Gleave, K, Shearer, FM, Sinka, ME, Massey, NC, et al. Geographical distributions of African malaria vector sibling species and evidence for insecticide resistance. Malaria Journal. 2017;16:85.

[25] Sinka, ME, Bangs, MJ, Manguin, S, Rubio-Palis, Y, Chareonviriyaphap, T, Coetzee, M, et al. A global map of dominant malaria vectors. Parasites & Vectors. 2012;5:69.

[26] Drake, JM, Beier, JC. Ecological niche and potential distribution of *Anopheles arabiensis* in Africa in 2050. Malaria journal. 2014;13:213.

[27] Sinka, ME, Golding, N, Massey, NC, Wiebe, A, Huang, Z, Hay, SI, et al. Modelling the relative abundance of the primary African vectors of malaria before and after the implementation of indoor, insecticide-based vector control. Malaria Journal. 2016;15:142.

[28] Lindsay, SW, Parson, L, Thomas, CJ. Mapping the ranges and relative abundance of the two principal African malaria vectors, *Anopheles gambiae sensu stricto* and *An. arabiensis*, using climate data. Proceedings of the Royal Society of London, Series B. 1998;265:847-54.

[29] Tene Fossog, B, Ayala, D, Acevedo, P, Kengne, P, Ngomo Abeso Mebuy, I, Makanga, B, et al. Habitat segregation and ecological character displacement in cryptic African malaria mosquitoes. Evol Appl. 2015;8:326-45.

[30] Pombi, M, Kengne, P, Gimonneau, G, Tene-Fossog, B, Ayala, D, Kamdem, C, et al. Dissecting functional components of reproductive isolation among closely related sympatric species of the *Anopheles gambiae* complex. Evolutionary Applications. 2017;10:1102-20.

[31] Pianka, ER. Niche overlap and diffuse competition. Proceedings of the National Academy of Sciences USA. 1974;71:2141-5.

[32] May, RM. Some notes on estimating the competition matrix, α. Ecology. 1975;56:737-41.

[33] Lounibos, LP. Competitive displacement and reduction. Journal of the American Mosquito Control Association. 2007;23:276-82.

[34] Walsh, KA, Kavishe, DR, Duggan, LM, Tarimo, LJ, Msoffe, RV, Manase, E, et al. Blood host preferences and competitive inter-species dynamics within an African malaria vector species complex inferred from signs of animal activity around aquatic larval habitats. bioRxiv. 2024. Available from: `https://www.biorxiv.org/content/early/2024/08/07/2024.08.04.606513`.

[35] Mwema, T, Lukubwe, O, Joseph, R, Maliti, D, Iitula, I, Katokele, S, et al. Human and vector behaviors determine exposure to *Anopheles* in Namibia. Parasites and Vectors. 2022;15:436.

[36] Aryaprema, VS, Steck, MR, Peper, ST, Xue, Rd, Qualls, WA. A systematic review of published literature on mosquito control action thresholds across the world. PLoS Neglected Tropical Diseases. 2023;17:e0011173.

[37] North, AR, Godfray, HCJ. Modelling the persistence of mosquito vectors of malaria in Burkina Faso. Malaria Journal. 2018;17:140.

[38] Malaria Atlas Project; 2021. Accessed 17 December 2021. URL `https://malariaatlas.org/`.

[39] Hosack, GR, El-Hachem, M. AgslPredict: Code and data for predictions of *Anopheles gambiae* sensu lato; 2025. R package version 0.5.0. doi:10.5281/zenodo.16990774. Available from: `https://github.com/csiro-risk-assessment/AgslPredict`.

[40] Giraldo-Calderón, GI, Harb, OS, Kelly, SA, Rund, SS, Roos, DS, McDowell, MA. VectorBase.org updates: bioinformatic resources for invertebrate vectors of human pathogens and related organisms. Current Opinion in Insect Science. 2022;50:100860.

[41] VectorBase. Popbio Map Legacy. Release 64.; 2023. Accessed: 22 April 2024. `https://vectorbase.org/common/downloads/popbio-map-legacy/VectorBase-popbio-map-rel64-July2023-Samples.csv.gz`.

[42] Coetzee, M, Hunt, R, Wilkerson, R, Torre, A, Coulibaly, M, Besansky, N. *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. Zootaxa. 2013;3619:246-74.

[43] Service, MW. Mosquito (Diptera: Culicidae) dispersal – the long and short of it. Journal of Medical Entomology. 1993;34:579-88.

[44] VectorBase. Popbio Map Legacy Abundance with Zeroes. Release 64.; 2023. Accessed 7 June 2024. `https://vectorbase.org/common/downloads/popbio-map-legacy/VectorBase-popbio-map-rel64-July2023-Abundance-with-zeroes.csv.gz`.

[45] Brown, O, Flegg, JA, Weiss, DJ, Golding, N. A global mathematical model of climatic suitability for *Plasmodium falciparum* malaria. Malaria Journal. 2024;23:306.

[46] NASA POWER. National Aeronautics and Space Administration (NASA) Langley Research Center. NASA Earth Science/Applied Science Program (LaRC) Prediction of Worldwide Energy Resource (POWER) Project; 2024. Available from: `https://power.larc.nasa.gov/`.

[47] Sparks, AH. nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. The Journal of Open Source Software. 2018 oct;3(30):1035.

[48] Sparks, AH. nasapower: NASA-POWER Data from R; 2024. R package version 4.2.1. Available from: https://CRAN.R-project.org/package=nasapower.

[49] Massicotte, P, South, A. rnaturalearth: World Map Data from Natural Earth; 2023. R package version 1.0.1. Available from: https://CRAN.R-project.org/package=rnaturalearth.

[50] South, A, Michael, S, Massicotte, P. rnaturalearthdata: World Vector Map Data from Natural Earth Used in 'rnaturalearth'; 2024. R package version 1.0.0. Available from: https://CRAN.R-project.org/package=rnaturalearthdata.

[51] NOAA NCEI. ETOPO 2022 15 Arc-Second Global Relief Model. Accessed 13 June 2024; 2022. Available from: https://doi.org/10.25921/fd45-gt74.

[52] Messager, ML, Lehner, B, Cockburn, C, Lamouroux, N, Pella, H, Snelder, T, et al. Global prevalence of non-perennial rivers and streams. Nature. 2021;594:391-7.

[53] Lehner, B, Messager, ML, Korver, MC, Linke, S. Global hydro-environmental lake characteristics at high spatial resolution. Scientific Data. 2022;9:351.

[54] De Marco, CM, Virgillito, C, Frosi, L, Santarelli, G, Filipponi, F, Manica, M, et al. Habitat drivers and predicted distribution shifts of *Anopheles coluzzii* underclimate change: Results from the systematic review. Science of the Total Environment. 2025;992:179939.

[55] Robert, V, Macintyre, K, Keating, J, Trape, JF, Duchemin, JB, Warren, M, et al. Malaria transmission in urban sub-Saharan Africa. The American Journal of Tropical Medicine and Hygiene. 2003;68:169-76.

[56] Hay, SI, Guerra, CA, Tatem, AJ, Atkinson, PM, Snow, RW. Urbanization, malaria transmission and disease burden in Africa. Nature Reviews Microbiology. 2005;3(1):81-90.

[57] Doumbe-Belisse, P, Kopya, E, Ngadjeu, CS, Sonhafouo-Chiana, N, Talipouo, A, Djamouko-Djonkam, L, et al. Urban malaria in sub-Saharan Africa: dynamic of the vectorial system and the entomological inoculation rate. Malaria Journal. 2021;20:364.

[58] European Commission. A recommendation on the method to delineate cities, urban and rural areas for international statistical comparisons. Prepared by the European Commission – Eurostat and DG for Regional and Urban Policy – ILO, FAO, OECD, UN-Habitat, World Bank; 2020.

[59] Bertozzi-Villa, A, Bever, CA, Koenker, H, Weiss, DJ, Vargas-Ruiz, C, Nandi, AK, et al. Maps and metrics of insecticide-treated net access, use, and nets-per-capita in Africa from 2000-2020. Nature Communications. 2021;12:3589.

[60] Malaria Atlas Project. Africa Insecticide Treated Net Use; 2025. Accessed 31 March 2025. URL https://malariaatlas.org/.

[61] Malaria Atlas Project. Africa IRS Coverage; 2025. Accessed 31 March 2025. URL https://malariaatlas.org/.

[62] Wilson, AL, Boelaert, M, Kleinschmidt, I, Pinder, M, Scott, TW, Tusting, LS, et al. Evidence-based vector control? Improving the quality of vector control trials. Trends in Parasitology. 2015.

[63] WHO. Achieving and maintaining universal coverage with long-lasting insecticidal nets for malaria control. World Health Organization, Global Malaria Programme; 2017. Available from: https://www.who.int/publications/i/item/WHO-HTM-GMP-2017.20.

[64] Galick, DS, Vaz, LM, Ondo, L, Iyanga, MM, Bikie, FEE, Avue, RMN, et al. Reconsidering indoor residual spraying coverage targets: A retrospective analysis of high-resolution programmatic malaria control data. Proceedings of the National Academy of Sciences. 2025;122(16):e2421531122.

[65] WHO. Indoor residual spraying : use of indoor residual spraying for scaling up global malaria control and elimination: WHO position statement. World Health Organization. Global Malaria Programme and World Health Organization. Malaria Unit.; 2007. Available from: https://iris.who.int/handle/10665/69386.

[66] Antonio-Nkondjio, C, Demanou, M, Etan, J, Bouchite, B. Impact of cyfluthrin (Solfac EW050) impregnated bed nets on malaria transmission in the city of Mbandjock: lessons for the nationwide distribution of long-lasting insecticidal nets(LLINs) in Cameroon. Parasites and Vectors. 2013;6:10.

[67] Russell, TL, Lwetoijera, DW, Maliti, D, Chipwaza, B, Kihonda, J, Charlwood, JD, et al. Impact of promoting longer-lasting insecticide treatment of bed nets upon malaria transmission in a rural Tanzanian setting with pre-existing high coverage of untreated nets. Malaria Journal. 2010;9:187.

[68] Mutuku, FM, King, CH, Mungai, P, Mbogo, C, Mwangangi, J, Muchiri, EM, et al. Impact of insecticide-treated bed nets on malaria transmission indices on the south coast of Kenya. Malaria Journal. 2011;10:356.

[69] Kamau, L, Bennett, KL, Ochomo, E, Herren, J, Agumba, S, Otieno, S, et al. The *Anopheles coluzzii* range extends into Kenya: detection, insecticide resistance profiles and population genetic structure in relation to conspecific populations in West and Central Africa. Malaria Journal. 2024;23:122.

[70] Fahrmeir, L, Tutz, G. Multivariate Statistical Modelling Based on Generalized Linear Models. Springer-Verlag New York Inc.; 1994.

[71] Venables, WN, Ripley, BD. Modern Applied Statistics with S. 4th ed. New York: Springer; 2002.

[72] Hancock, P, Hendriks, C, Tangena, JA, Gibson, H, Hemingway, J, Coleman, M, et al. Mapping trends in insecticide resistancephenotypes in African malaria vectors. PLOS Biology. 2020;18:e3000633.

[73] Keating, J, Macintyre, K, Mbogo, CM, Githure, JI, Beier, JC. Characterization of potential larval habitats for *Anopheles* mosquitoes in relation to urban land-use in Malindi, Kenya. International Journal of Health Geographics. 2004;3:1-13.

[74] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2024. Available from: `https://www.R-project.org/`.

[75] Marschner, IC. glm2: Fitting generalized linear models with convergence problems. The R Journal. 2011;3:12-5.

[76] Shanks, GD, Hay, SI, Omumbo, JA, Snow, RW. Malaria in Kenya's western highlands. Emerging Infectious Diseases. 2005;11:1425-32.

[77] Bayoh, MN, Lindsay, SW. Effect of temperature on the development of the aquatic stages of *Anopheles gambiae* sensu stricto (Diptera: Culicidae). Bulletin of Entomological Research. 2003;93:375-81.

[78] Paaijmans, KP, Imbahale, SS, Thomas, MB, Takken, W. Relevant microclimate for determining the development rate of malaria mosquitoes and possible implications of climate change. Malaria Journal. 2010;9:196.

[79] Ukawuba, I, Shaman, J. Inference and dynamic simulation of malaria using a simple climate-driven entomological model of malaria transmission. PLOS Computational Biology. 2022;18:e1010161.

[80] Hosack, GR, Beeton, NJ, Ickowicz, A, Peel, D, Wilkins, A, Dambacher, JM, et al. Risk Assessment for Controlling Mosquito Vectors with Engineered Nucleases: Paternal Male Bias Construct. Report No. EP2022-4945. Hobart, Australia: CSIRO; 2023. Available from: `https://doi.org/10.25919/2t8h-5k81`.

[81] Hosack, GR. Prior elicitation for generalised linear models and extensions. Bayesian Analysis. 2024. Advance Publication 1 - 30. Available from: `https://doi.org/10.1214/24-BA1472`.

[82] Huestis, DL, Dao, A, Diallo, M, Sanogo, ZL, Samake, D, Yaro, AS, et al. Windborne long-distance migration of malaria mosquitoes in the Sahel. Nature. 2019;574:404-8.

[83] Hosack, GR, El-Hachem, M, Beeton, NJ. Stability of difference equations with interspecific density dependence, competition, and maturation delay. Bulletin of Mathematical Biology. 2025. Forthcoming at: `https://doi.org/10.1007/s11538-025-01515-0`.

[84] Beverton, RJH, Holt, SJ. On the Dynamics of Exploited Fish Populations. No. 11 in Fish and Fisheries Series; 1957. Republished by Springer Science+Business Media Dordrecht, 1993.

[85] Leslie, PH, Gower, JC. The properties of a stochastic model for two competing species. Biometrika. 1958;45:316-30.

[86] Beddington, JR, Free, CA, Lawton, JH. Concepts of Stability and Resilience in Predator-Prey Models. Journal of Animal Ecology. 1976;45:791-816.

[87] Levine, RS, Peterson, AT, Benedict, MQ. Geographic and ecologic distributions of the *Anopheles gambiae* complex predicted using a genetic algorithm. American Journal of Tropical Medicine and Hygiene. 2004;70:105-9.

[88] Moffett, A, Shackelford, N, Sarkar, S. Malaria in Africa: vector species' niche models and relative risk maps. PloS one. 2007;2:e824.

[89] Adugna, F, Wale, M, Nibret, E. Review of *Anopheles* mosquito species, abundance, and distribution in Ethiopia. Journal of Tropical Medicine. 2021;2021:6726622.

[90] Msugupakulya, BJ, Urio, NH, Jumanne, M, Ngowo, HS, Selvaraj, P, Okumu, FO, et al. Changes in contributions of different *Anopheles* vector species to malaria transmission in east and southern Africa from 2000 to 2022. Parasites & Vectors. 2023;16:408.

[91] Khatri, BS, Burt, A. Robust estimation of recent effective population size from number of independent origins in soft sweeps. Molecular Biology and Evolution. 2019;36:2040-52.

[92] *Anopheles gambiae* 1000 Genomes Consortium. Genetic diversity of the African malaria vector *Anopheles gambiae*. Nature. 2017;552:96-100.

[93] Frankham, R. Effective population size/adult population size ratios in wildlife: a review. Genetical Research. 1995;66:95-107.

[94] Macdonald, G. The analysis of equilibrium in malaria. Tropical Diseases Bulletin. 1952;49:813-29.

[95] Mbawala, JR, Li, H, Ndabagenga, DM, Zeng, J, Mlonganile, PN. Assessment of the spatio-temporal trends of annual extreme temperature indices over Tanzania during the period of 1982–2022. Journal of Geoscience and Environment Protection. 2024;12:33-50.

[96] Mohammed, EA, Zhi, X, Adbela, KA. Extreme weather patterns in Ethiopia: analyzing extreme temperature and precipitation variability. Atmosphere. 2025;16:133.

[97] Sarfo, I, Qiao, J, Yeboah, E, Puplampu, DA, Kwang, C, Fynn, IEM, et al. Meta-analysis of land use systems development in Africa: Trajectories, implications, adaptive capacity, and future dynamics. Land Use Policy. 2024;144:107261.

[98] Mwangangi, JM, Mbogo, CM, Orindi, BO, Muturi, EJ, Midega, JT, Nzovu, J, et al. Shifts in malaria vector species composition and transmission dynamics along the Kenyan coast over the past 20 years. Malaria Journal. 2013;12:13.

[99] Traoré, A, Badolo, A, Guelbeogo, MW, Sanou, A, Viana, M, Nelli, L, et al. Anopheline species composition and the 1014F-genotype in different ecological settings of Burkina Faso in relation to malaria transmission. Malaria Journal. 2019;18:165.

[100] Ngowo, HS, Okumu, FO, Hape, EE, Mshani, IH, Ferguson, HM, Matthiopoulos, J. Using Bayesian state-space models to understand the population dynamicsof the dominant malaria vector, *Anopheles funestus* in rural Tanzania. Malaria Journal. 2022;21:161.

[101] Grinnell, J. The niche relationships of the California thrasher. The Auk. 1917;34:427-33.

[102] Huchinson, GE. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology. 1957;22:415-27.

[103] Pulliam, HR. On the relationship between niche and distribution. Ecology Letters. 2000;3:349-61.

[104] Rogers, DJ, Randolph, SE. Studying the global distribution of infectious diseases using GIS and RS. Nature Reviews Microbiology. 2003;1:231-7.

[105] Peterson, AT. Ecologic niche modeling and spatial patterns of disease transmission. Emerging Infectious Diseases. 2006;12:1822-6.

[106] Elith, J, Leathwick, JR. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics. 2009;40:677-97.

[107] Yates, KL, Bouchet, PJ, Caley, MJ, Mengersen, K, Randin, CF, Parnell, S, et al. Outstanding challenges in the transferability of ecological models. Trends in Ecology and Evolution. 2018;33:790-802.

[108] Walker, M, Winskill, P, Basáñez, MG, Mwangangi, JM, Mbogo, C, Beier, JC, et al. Temporal and micro-spatial heterogeneity in the distribution of *Anopheles* vectors of malaria along the Kenyan coast. Parasites & Vectors. 2013;6:311.