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

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 predictions of relative abundance for each species or aggregates of the species. This study incorporated relative abundance data at the species level and abundance data at the species complex level to estimate and predict the carrying capacities of each species. Statistical models with interpretable parameters were used to gain insight into how factors such as meteorological conditions and human density affect the spatial and temporal predictions.

Results: At the continental scale, precipitation and temperature were found to be important spatially and temporally varying predictors of equilibrium abundances of these three dominant malaria vector species in Africa. The equilibrium abundance of *An. gambiae* sensu stricto was predicted to have increased across much of its range during the period 2002–2020, and particularly so in the Lake Victoria basin. In contrast, *An. arabiensis* was predicted to have decreased around Lake Victoria and southeastern Africa but increased elsewhere in eastern Africa. A predicted increasing inter-annual trend for *An. coluzzii* was possibly caused by changing species identification methods over the observational period.

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

Background

The reduction of mosquito abundance by vector control measures is an effective means to lower or eliminate human malaria [1–3]. The evaluation of conventional and potential novel vector control strategies on disease transmission dynamics thus benefits from prediction and estimation of vector abundances [4, 5]. The species complex *Anopheles gambiae* sensu lato (s.l.) is an important contributing cause of the malaria burden in Africa [6]. Here, the predicted abundances of the three dominant malaria vectors with the broadest geographical distribution are investigated: *An. arabiensis*, *An. coluzzii* and *An. gambiae* sensu stricto (s.s.). The distributions of these three species are strongly linked to environmental covariates such as local meteorological conditions and climate [7–10] with overlapping niches in some regions [11].

Geographical predictions of *An. gambiae* s.l. based on species distribution models and niche modelling [12] often use presence-absence or presence only data to predict the species niche or relative probability of occurrence [8, 9, 13, 14]. Relative abundance data at the species level is also often available for *An. gambiae* species complex field collections because the member species are not morphologically distinguishable, and so molecular identification from subsamples of field collections is required. Sinka et al. [15] 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 relative abundance data to predict the relative abundance of *An. arabiensis* versus the aggregate of the two species *An. coluzzii* and *An. gambiae* s.s. Tene Fossog et al. [16] analyse relative species frequencies of *An. coluzzii* and *An. gambiae* s.s. to predict relative probabilities of occurrence in western Africa. Niche overlap, which is a function of relative abundance data [17], is investigated by Pombi et al. [11] for *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. in certain regions. Niche overlap is sometimes suggested as an index of the strength of interspecific competition [18], although this correlative index that is based on species relative abundances may diverge from underlying competitive relationships among species [17]. It must also be observed that malaria transmission dynamics are driven by mosquito abundance [19], not just relative abundance or species composition. Mathematical models of long term or equilibrium abundance, for example, have long served as useful guides for vector control and malaria interventions [1, 2].

Equilibrium abundance is a useful concept that allows for the possibility that the true or actual abundance of a species population may diverge from its carrying capacity because of local environmental or anthropogenic perturbations, dispersal capability or source-sink

metapopulation dynamics, or through interspecific competition [20–22]. Here, the equilibrium abundances of *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. are predicted across sub-Saharan Africa from 2002 to 2020. Statistical relationships are established with environmental factors such as precipitation and temperature [23–25] to enable predictions. The spatial approach of Sinka et al. [15] is expanded to account for spatio-temporal predictors of the relative abundances of *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. Moreover, the analysis of relative abundance data available at the species level is combined with analysis of abundance data available at the taxonomic level of *An. gambiae* s.l. to predict potential species abundance, or specifically species equilibrium abundances that are related to the species carrying capacities. In this approach, the species equilibrium abundances are modelled as responding to spatial, temporal and spatio-temporal covariates such as precipitation and temperature, but the equilibrium abundances not directly observed because of observation error. The collection method, for example, is accounted for in the model of abundance to address this source of observation bias. The predicted equilibrium abundances of *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. are related to meteorological conditions of precipitation and temperature, as well as human population density and other geographical factors. These predictions are 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 details of the relative abundance data, abundance data, spatial covariates and spatio-temporal covariates are found in the supplementary `R` package **AgslPredict** for this paper [26]. The code for analyses and graphical results of relative abundance predictions, abundance predictions and competition estimates are also available in **AgslPredict**.

Data

The genotype data used for relative abundance of the three species *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. were downloaded from `VectorBase.org` [27, 28] and spatially subsetted to the spatial scope (Table S1, Figure S1). 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" [29]. 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 (1%) or available only to years (1%) were filtered out of the dataset. Information on time of collection varied across the samples. For example, about 14% of samples were reported to the month, 52% were reported to the day, and 23% were reported as a duration of days. The start and end dates of these records with durations between days 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. About 74% of the sample

durations were one day or less (i.e., collection during a day or overnight). Samples with durations greater than 31 days (10%) were filtered out of the dataset. *An. coluzzii* and *An. gambiae* s.s. were not distinguished prior to 2001 [11] 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 unit of analysis. The frequency counts of the three species identified for each unit sample were then recorded. Sample units with shared collection dates and collection methods were then aggregated to 5 km $\times$ 5 km grid cells across the spatial extent, which resulted in 1,738 samples of relative abundance. The 5 km $\times$ 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 [30].

The abundance data for *An. gambiae* s.l. were downloaded from `VectorBase.org` [27, 31] and spatially subsetted to the spatial scope (Table S2, Figure S2). A search for abundance of *An. gambiae* s.l. datasets showed that most data are only taxonomically resolved to the level of the species complex. Statistical analysis for abundance therefore used the total observed abundances of *An. gambiae* s.l. as the response variable for the abundance analysis. Observation methods that could be categorised to specific collection methods (see Table 1) were retained for analysis. Samples that were identified as occurring outdoors were assigned a value of one for the "outdoors" covariate, those labelled indoors were assigned a zero value, and those that were indeterminant were assigned an intermediate value of 0.5 for the outdoors covariate. Records were then subsetted to samples identified as adult female mosquitoes. About 2% of the remaining samples had indeterminate sampling dates, so they were excluded from the analysis. The available records extended from 1997 to 2021. The `VectorBase.org` data for *An. gambiae* s.l. abundance were sparsely distributed over Africa across the observation period both geographically and through time. Sites with long term longitudinal observations often also included long durations of days without observations. As a measure of sampling effort, samples were defined using the same procedure as applied to the relative abundance data. Observations were aggregated to 5 km $\times$ 5 km cells that captured the typical home range of a female mosquito 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).

Meteorological data for daily precipitation totals, average daily temperature and average daily 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 [32] using the R package **nasapower** [33, 34]. These data are provided at a resolution of 5/8° longitude by 0.5° latitude. The meteorological covariates of a meteorological grid cell were assigned to all those 5 km $\times$ 5km grid cells with centroids that occurred within the corresponding meteorological grid cell. The meteorological data are summarised 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 from 5 km $\times$ 5 km grid cell centroids to coasts (Figure S12) and rivers (Figure S13) were calculated from data provided by Natural Earth that were obtained using the R packages **rnaturalearth** [35] and **rnaturalearthdata** [36]. Elevation data at 15 arcseconds resolution were obtained from the National Centers for Environmental

Information, National Oceanic and Atmospheric Administration [37], and averaged to obtain mean elevation for each 5 km $\times$ 5km grid cell (Figure S14). Human population data in 2015 at 30 arcseconds resolution were obtained from Landscan [38], and averaged to obtain mean human population density for each 5 km by 5 km grid cell (Figure S15).

Analysis of Relative Species Abundance

Sinka et al. [15] use multinomial generalised additive models of relative abundance data to predict relative proportions of *An. funestus*, *An. arabiensis* and the combination of the two species *An. coluzzii* + *An. gambiae* s.s. Each analysis used five spatial covariates selected from geographic summary statistics of vegetation, elevation, temperature and precipitation. An additional model investigated the inclusion of intervention data interpreted from various sources. Here, in contrast to [15], 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. [15], and ii) increase the temporal and geographic dispersion of collections for relative abundance. A greater number of covariates were therefore allowed into this analysis, notably the inclusion of temporal and spatio-temporal covariates such as temperature, daily precipitation and daily relative humidity that were absent from the analyses of Sinka et al. [15]. However, intervention data were not included as explicit covariates because of the difficulty with summarising cumulative intervention efficacy (e.g., [39]) and effort on vector abundances (see [15] for discussion) across the entirety of the spatial and temporal scope of this analysis. Method of collection was also not included for the relative abundance analysis because the most commonly recorded description was simply "catch of live specimens" (27% of samples; collection method was incorporated into the analysis for the abundance data, see next section). However, a linear trend was incorporated into the model that would allow capture of a trend in species compositional change related to increasing overall vector control intervention effort across Africa during the observation period.

The dates listed in the filtered `VectorBase.org` relative abundance 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. Thus, each meteorological predictor was constructed based on the one-month averages preceding the start and end dates of the sample. Geospatial covariates included longitude, the absolute value of latitude as measure of insolation, elevation, distance to nearest river, distance to coastline and human population density.

Multinomial logit generalised linear models (GLMs), see [40], were analysed using R package **nnet** [41] 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). Let $\boldsymbol{\nu}(s,t)$ denote the set of covariates at site or grid cell s and time t with *An. arabiensis* as the baseline or reference category of the multinomial

logit GLM. Then the statistical model was defined by

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

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 $\boldsymbol{\gamma}$ is composed of the unknown coefficients. Backwards stepwise model selection was initiated from a quadratic surface model using the `R` package **MASS** based on the Bayesian information criterion (BIC) [41]. The quadratic surface model was a second order approximation, that is, each spatial and spatio-temporal covariate appeared as a squared term, and a linear term, along with all pairwise interactions of the latter. The temporal covariate of year of sample collection was also included to account for any annual trend in species composition, such as might arise from changing taxonomic identification practices over the dataset duration period or trend in vector control interventions. The model with the lowest BIC value was chosen as the final model.

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

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 1), and whether or not a collection method was documented to occur outdoors. One project that had "CDC light trap" as the listed collection method was scored indoors using the methods description from the original study [42]. Projects with collection methods that were uncertain with respect to indoor or outdoor sampling were assigned a value of 0.5 for outdoors.

The total observed abundance $y(s,t)$ of adult female *An. gambiae* s.l. at "site" (grid cell) s and time t was assumed to depend on observation method as described above, and also precipitation and human population. The absolute value of latitude as a measure of insolation and elevation were included as interactions. The statistical model for total abundance per indoor human given collection method k was described by a negative binomial GLM,

$$\begin{aligned} y(s,t,k) &\sim \text{NegativeBinomial}\left(\mu(s,t,k)u(s,t,k), \theta\right) \\ \log \rho(s,t,k) &= \boldsymbol{w}(s,t,k)^\top \boldsymbol{\zeta} \\ \log f(s,t) &= \boldsymbol{v}(s,t)^\top \boldsymbol{\beta}, \\ \log \mu(s,t,k) &= \log \rho(s,t,k) + \log f(s,t) \\ \mathbb{E}[y(s,t,k) \mid \mu(s,t,k)u(s,t,k), \theta] &= \mu(s,t,k)u(s,t,k), \\ \mathbb{V}[y(s,t,k) \mid \mu(s,t,k)u(s,t,k), \theta] &= \mu(s,t,k)u(s,t,k) + [\mu(s,t,k)u(s,t,k)]^2/\theta, \end{aligned}$$

where θ is the overdispersion parameter and $u(s,t,k)$ is the estimated effort for observation method k at site s on day t, and is used as an offset. The covariate vector $\boldsymbol{w}(s_i, t_i, k)$ with unknown coefficients $\boldsymbol{\zeta}$ encodes the observation method according to Table 1 that

Table 1: **Coding for collection methods reported in abundance data.** Collection methods described by VectorBase are grouped by row: *ABC*, animal baited catch ("animal baited net trap catch" or "animal biting catch - outdoors"); *APS*, "artificial pit shelter"; *LTC*, light trap catch ("CDC light trap catch" or "CDC miniature light trap model 512"); *ETC*, "exit trap catch"; *ILTC*, "indoor light trap catch"; *MBC*, "man biting catch"; *MBCI*, "man biting catch - indoors"; *MBCO*, "man biting catch - outdoors"; *OLTC*, "outdoor light trap catch"; and *PSC*, "pyrethrum spray catch". The collection method are scored by column with respect to light traps (*light*), exit traps (*exit*), pyrethrum spray catch (*PSC*), animal baited catch (*animal*), pit shelter (*pit*) and outdoor deployment (*outdoor*). The latter category is scored at an intermediate value of 0.5 where the collection method was indeterminate with respect to indoor versus outdoor sampling.

Method	Light	Exit	PSC	Animal	Pit	Outdoor
ABC	0	0	0	1	0	1.0
APS	0	0	0	0	1	1.0
ETC	0	1	0	0	0	0.0
ILTC	1	0	0	0	0	0.0
LTC	1	0	0	0	0	0.5
MBC	0	0	0	0	0	0.5
MBCI	0	0	0	0	0	0.0
MBCO	0	0	0	0	0	1.0
OLTC	1	0	0	0	0	1.0
PSC	0	0	1	0	0	0.0

alters 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 $\mu(s,t,k)$. If $\boldsymbol{w}(s,t,k)$ is a zero vector then the corresponding collection method is MCBI with $\rho(s,t,k = MCBI) = 1$ and $\mu(s,t,k = MCBI) = f(s,t)$.

The actual 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 coefficients $\boldsymbol{\beta}$. The covariate vector $\boldsymbol{v}(s,t)$ incorporated the lagged effect of precipitation on larval habitat. For example, the carrying capacity for *An. gambiae* is related to weekly total precipitation by [24]. Here, the immature phase of the life cycle is assumed to be about 10 days [25]. 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. This precipitation covariate was therefore spatially and temporally varying.

A second order approximation was considered for the running mean of precipitation so that both a linear and squared term for the precipitation running mean was incorporated. The linear term shared linear interactions with absolute value of latitude as a measure of insolation, elevation that relates to annual average temperature, and the spatially and temporally varying running mean for temperature that was constructed in the same way as the precipitation covariate. These interactions allowed for modification of the impact of

precipitation on adult female *An. gambiae* s.l. abundance due to site specific and temporally changing temperature. The spatial covariate of human population density was included as a linear term to allow per capita mean vector abundance, as defined relative to indoor human landing catches, to change proportionally as human population density increases. For estimation, the covariates were scaled to range between zero and one based on the ranges. The negative binomial GLM was fit in R v4.4.0 [43] using the R package **glm2** [44] and function `glm.nb` in package **MASS** [41].

Predicted Species Equilibrium Abundance

The equilibrium abundance of the three species *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. was predicted for the time period from 1 January 2002 to 31 December 2020 across the potential spatial range of *An. gambiae* s.l. [9, 45]. 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 ($MCBI$), such that $f(s,t) = \mu(s,t,k = MBCI)$. Abundance predictions of *An. gambiae* s.l. are best related to precipitation in regions where temperature is not a limiting factor [46]. A temperature threshold rule was therefore applied. Laboratory experiments suggest that larval emergence of *An. gambiae* s.l. does not occur if daily water temperatures are less than 18°C or more than 34°C [47]. Air temperatures tend to be lower than water temperatures in *Anopheles gambiae* s.l. larval habitats [48], although the exact relationship between water and air temperatures can vary depending on local environmental factors [10]. About 1% of data observations included average 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 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 predicted mean abundance $f(s,t) = \mu(s,t,k = MBCI)$ is the average 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 [49] accounting for species differences among *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. using a prior elicitation method for GLMs [50]. 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 [49]. These estimates reflect the greater degree of zoophilic feeding habit of *An. arabiensis* [6]. Let $\xi(s,t)$ denote the average abundance of adult female *An. gambiae* s.l. per human after adjustment for the daily probability of human feeding. Note that there are more adult female mosquitoes than those that may be observed feeding on humans. Using the predicted total 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 the per capita *An. gambiae* s.l. abundance adjusted by human feeding probability 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 abundances $\boldsymbol{x}(s,t)$ are 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 $h(s)$ is the human population at site s. Predictions were thresholded to non-urban areas because the three species of interest are not important vectors in central urban areas [6, 51]. The threshold of 1500 human inhabitants on average per km^2 corresponds to urban areas [52]. If $h(s) > 1500$ humans then $x_m(s,t) = 0$ for all species at site s.

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 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 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 [49] for an elaborated model to address these considerations including wind-assisted advection, sensu [53]). The predicted species abundances $\boldsymbol{x}(s,t)$ are therefore considered predictions of equilibrium species 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. [11] estimate the niche overlap among the three species *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. in Burkina Faso. The pairwise estimates of niche overlap α_{ij} between species i and species j use the method of Pianka [17]. Applied to the predictions over all grid cells and days, the niche overlap 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 daily abundance $\bar{x}_m(s)$ at a site s 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 [18]. The niche overlap index has been suggested as a measure of interspecific competition [18], although Pianka [17] lists limitations with this interpretation.

A contrasting approach estimates interspecific competition by identifying an underlying dynamic model [54] that is a multispecies version of the Beverton-Holt population model [55] or two-species Leslie-Gower model [56] 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 [54] 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 abundance $\bar{x}_m$ of species m over all sites and times is an approximation of its predicted equilibrium abundance, such that $\bar{\boldsymbol{x}} \approx \hat{\boldsymbol{x}}^*$, then following [54] 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. Stability is measured by the spectral radius, or largest eigenvalue by magnitude, of the linearised system around the equilibrium [54]: 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. The symmetric positive definite structure of the niche overlap method is retained for $\boldsymbol{A}$ in this approach.

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 [25] for *An. gambiae* and adjusted for equal sex ratio. See also [54] for an example using the empirical results of Pombi et al. [11] for Burkina Faso only. Adult daily survival probabilities across the species range of *An. gambiae* s.l. were elicited from domain experts [49] accounting for species differences among *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. using a prior elicitation method for GLMs [50]. 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 [49]. The estimate of competition does not depend on $\boldsymbol{q}$ in the method of [54], so that the above assumptions allowed estimation of the interspecific density dependent coefficients A_{ij} with $i \neq j$.

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 BIC value of the full model was 50,626 with 112 degrees of freedom. The backwards stepwise selection resulted in a best fitting final model having a BIC value of 50,532 with 100 degrees of freedom. The meteorological covariates of precipitation, temperature and relative humidity were all retained as quadratic relationships, with both linear and quadratic terms retained as useful predictors of relative abundance, in the best fitting model. The estimated coefficients of covariates for relative abundance also indicated a strong increasing annual trend in the relative abundance of *An. coluzzii*, which also decreased from east to west with increasing longitude. The relative abundances of both *An. coluzzii* and *An. gambiae* s.s. showed a decreasing trend with higher latitudes. Due to interactions, this latitudinal trend was diminished by increasing elevation but enhanced by human population density. Human population density was evidently an important predictor as it also showed a negative interaction with longitude for the log odds of *An. coluzzii* and *An. gambiae* s.s.

Predictions from the multinomial GLM are averaged through time and presented across the spatial scope in Figure S16. The associated predicted variance on the linear predictor scale is presented in Figure S17. Higher relative abundance was predicted for *An. coluzzii* relative to *An. arabiensis* in coastal West Africa. Greater relative abundance was predicted for *An. gambiae* s.s. versus *An. arabiensis* in coastal West Africa and some regions near Lake Victoria. With less certainty, higher relative abundance for *An. gambiae* s.s. was predicted in some areas of Madagascar. Higher relative abundance for *An. gambiae* s.s. versus *An. arabiensis* was also predicted in the northern Sahel, the Horn of Africa, and the Arabian Peninsula, but these predictions were associated with high uncertainty. The regions of prediction with high uncertainty had comparatively few relative abundance data sources (Figure S1).

The parameter estimates for the negative binomial GLM are provided in Table S4. Among the observation methods of Table 1, pit traps were estimated as the least efficient collection method. Precipitation was a strong predictor with abundance peaking at an intermediate value within the range of observed precipitation: the linear coefficient for precipitation was 15.6 and the corresponding squared term was -12.12. The quadratic relationship with precipitation in the absence of interactions was therefore concave with peak per capita abundance predicted at 0.64, which was above the midpoint of the precipitation covariate that was scaled between 0 and 1 by its observed range. Negative interactions were estimated between precipitation and temperature, which also varied both spatially and temporally along with precipitation, and also the spatial covariates of elevation and distance from the equator. Taking these interactions into account by setting the covariates to their respective midpoints in turn suggested a peak per capita abundance that occurred at about half the precipitation level in the absence of interactions (0.31). A negative association was suggested between per capita mosquito female abundance and human population density, which was expected since *An. gambiae* s.l. is not an important vector in highly urbanised areas [6, 51]. Dense human

population centres have fewer *An. gambiae* s.l. mosquitoes due to reduced larval habitat compared to less urbanised environments, and so the per capita ratio of *An. gambiae* s.l. vectors to humans was estimated to decrease with greater human population density and increasing urbanisation. The highest overall per capita abundance was predicted near the equator (Figure S18a) with some areas of high uncertainty (Figure S18b) located in regions with comparatively few abundance data sources (Figure S2).

Predicted Equilibrium Species Abundances

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

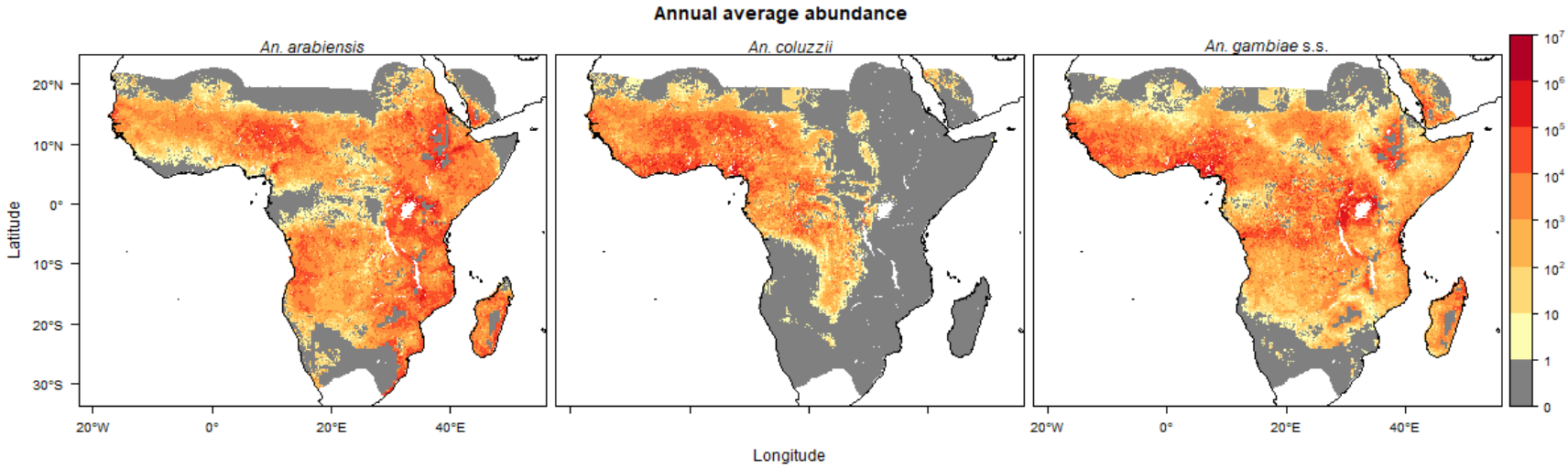


Figure 1: Annual average of predicted equilibrium abundances of dominant malaria vector species of *An. gambiae* s.l. per 25 square kilometres over years 2002–2020. Predictions are masked to the potential range of *An. gambiae* s.l. [9] provided by the Malaria Atlas Project (`https://malariaatlas.org/`) under the terms of a Creative Commons Attribution 4.0 International License available at `https://creativecommons.org/licenses/by/4.0/`. Note logarithmic scale.

The predicted equilibrium abundance varied both by season and across the years for all three species. The predicted quarterly average equilibrium abundance for *An. arabiensis* (Figure 2) was higher in parts of eastern Africa in Q1 (January – March) compared to Q3 (July – September). The predicted range for *An. coluzzii* contracted towards western Africa and also increased in areal density in Q3 compared to other quarters (Figure 3). High areal densities in western Africa were also predicted for *An. gambiae* s.s. in Q3 compared to Q1 (Figure 4).

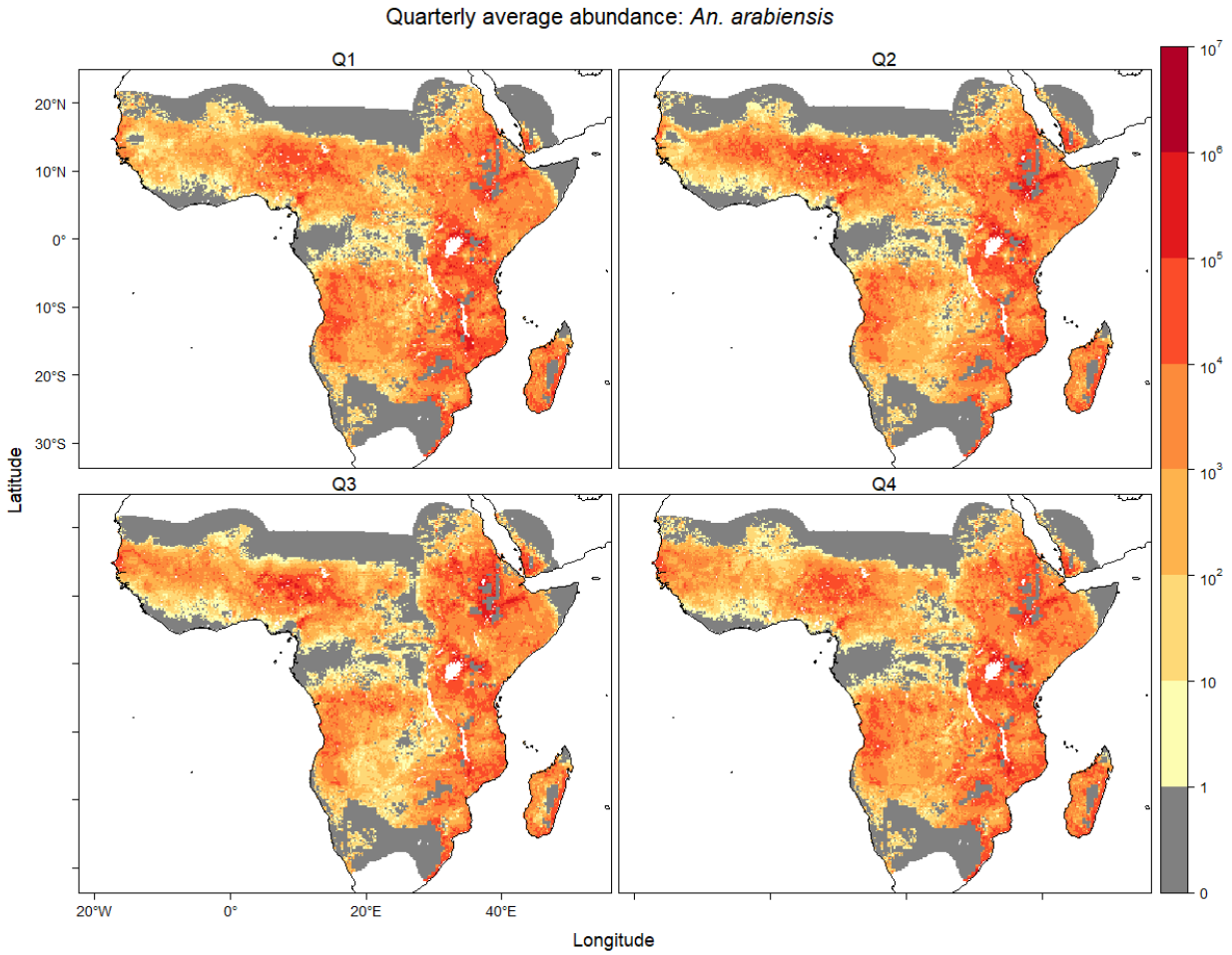


Figure 2: Annual average of total potential equilibrium abundance of *An. arabiensis* per 25 square kilometres by quarter (Q1: January–March, Q2: April–June, Q3: July–September, Q4: October–December) over years 2002–2020. Predictions are masked to the potential range of *An. gambiae* s.l. [9] provided 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.

Across four years separated by six year intervals between 2002–2020, the areal density of *An. arabiensis* was predicted to increase in eastern Africa, except for the southeastern coastal region where decreased areal densities were predicted (Figure 5). The equilibrium

abundance of *An. coluzzii* was generally predicted to increase compared to the initial year of 2002 (Figure 6). Predictions for *An. gambiae* s.s. suggested widespread increases in abundance with particularly high increases predicted for the Lake Victoria basin (Figure 7).

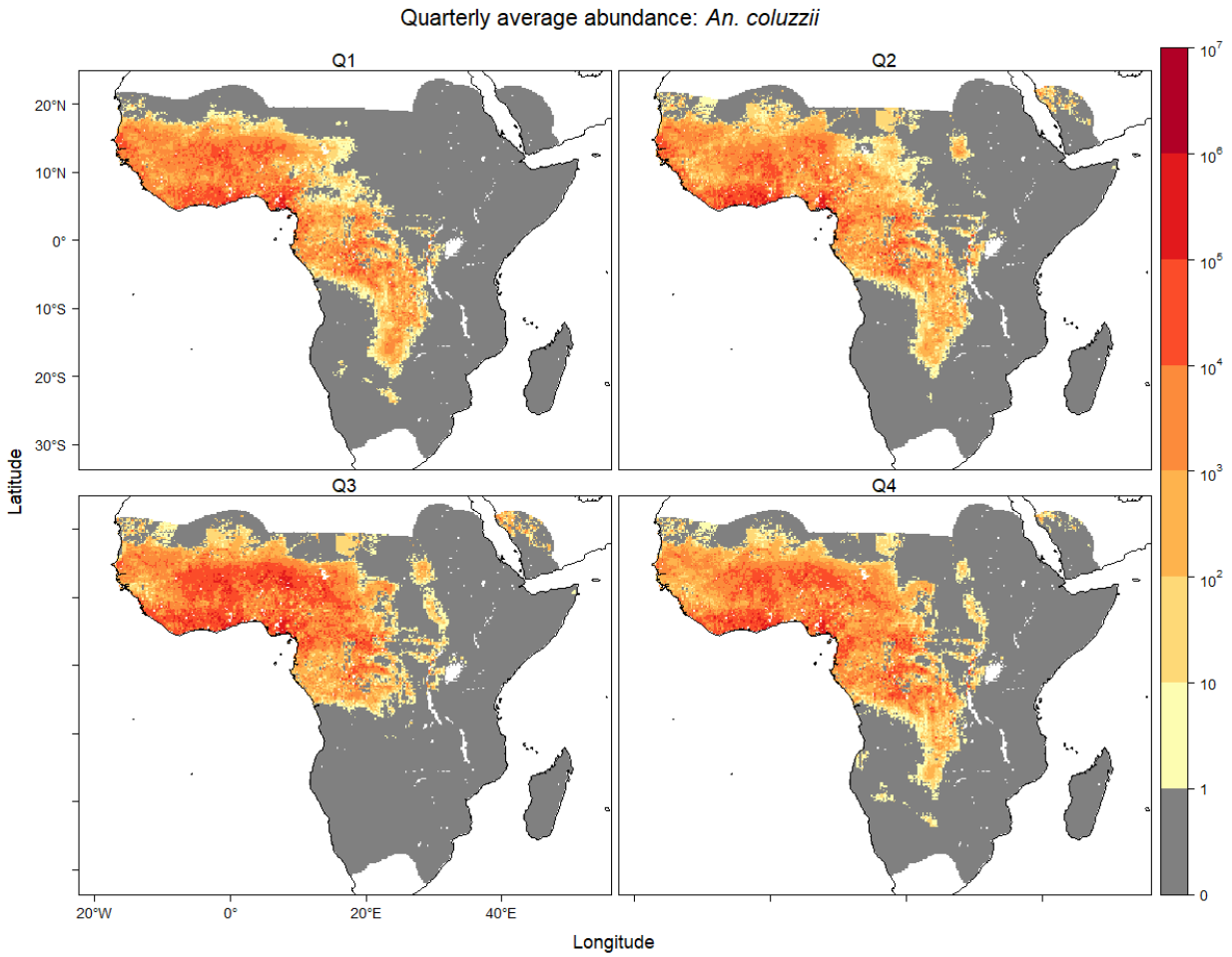


Figure 3: Annual average of total potential equilibrium abundance of *An. coluzzii* per 25 square kilometres by quarter (Q1: January–March, Q2: April–June, Q3: July–September, Q4: October–December) over years 2002–2020. Predictions are masked to the potential range of *An. gambiae* s.l. [9] provided 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.

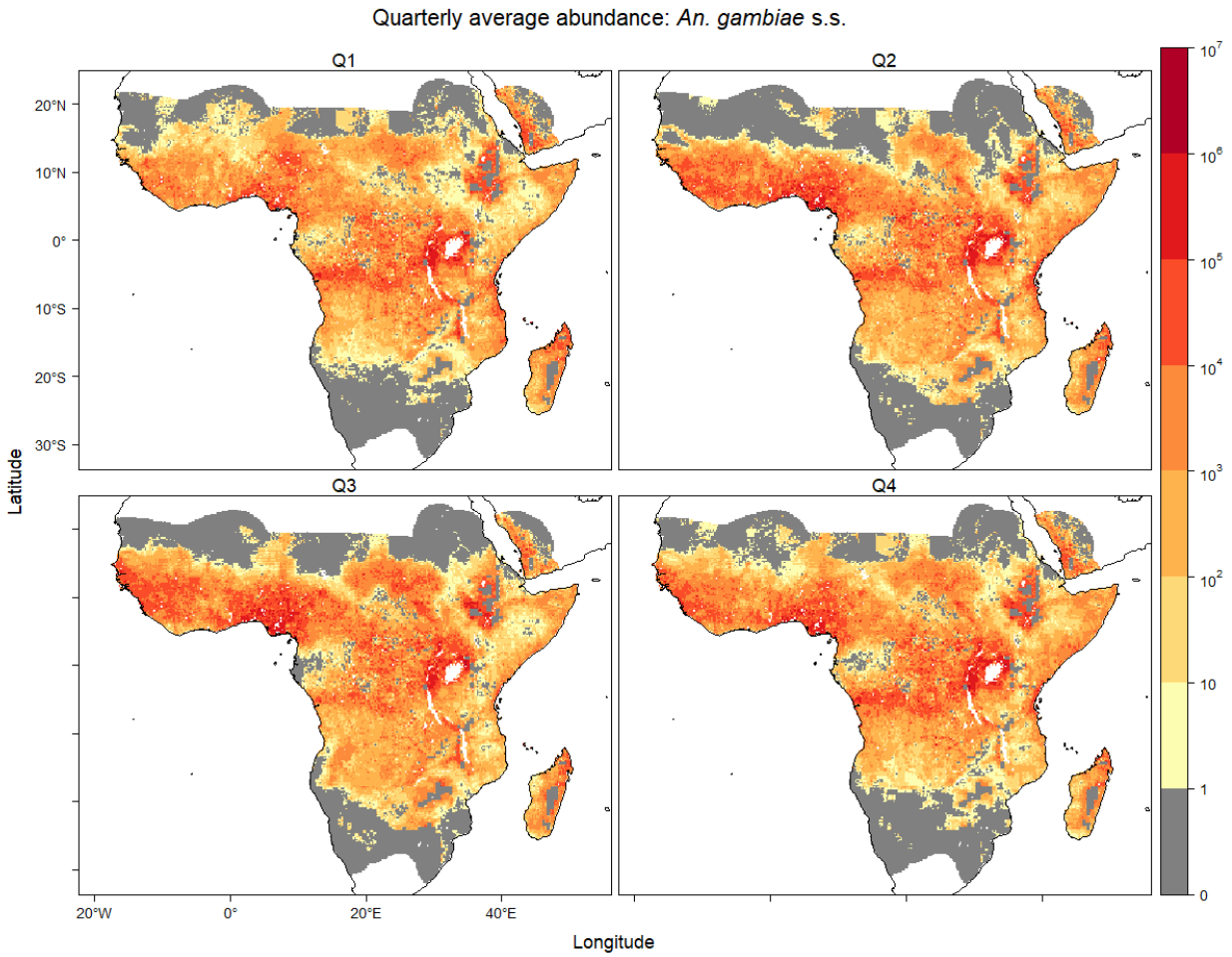


Figure 4: Annual average of total potential equilibrium abundance of *An. gambiae* s.s. per 25 square kilometres by quarter (Q1: January–March, Q2: April–June, Q3: July–September, Q4: October–December) over years 2002–2020. Predictions are masked to the potential range of *An. gambiae* s.l. [9] provided 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.

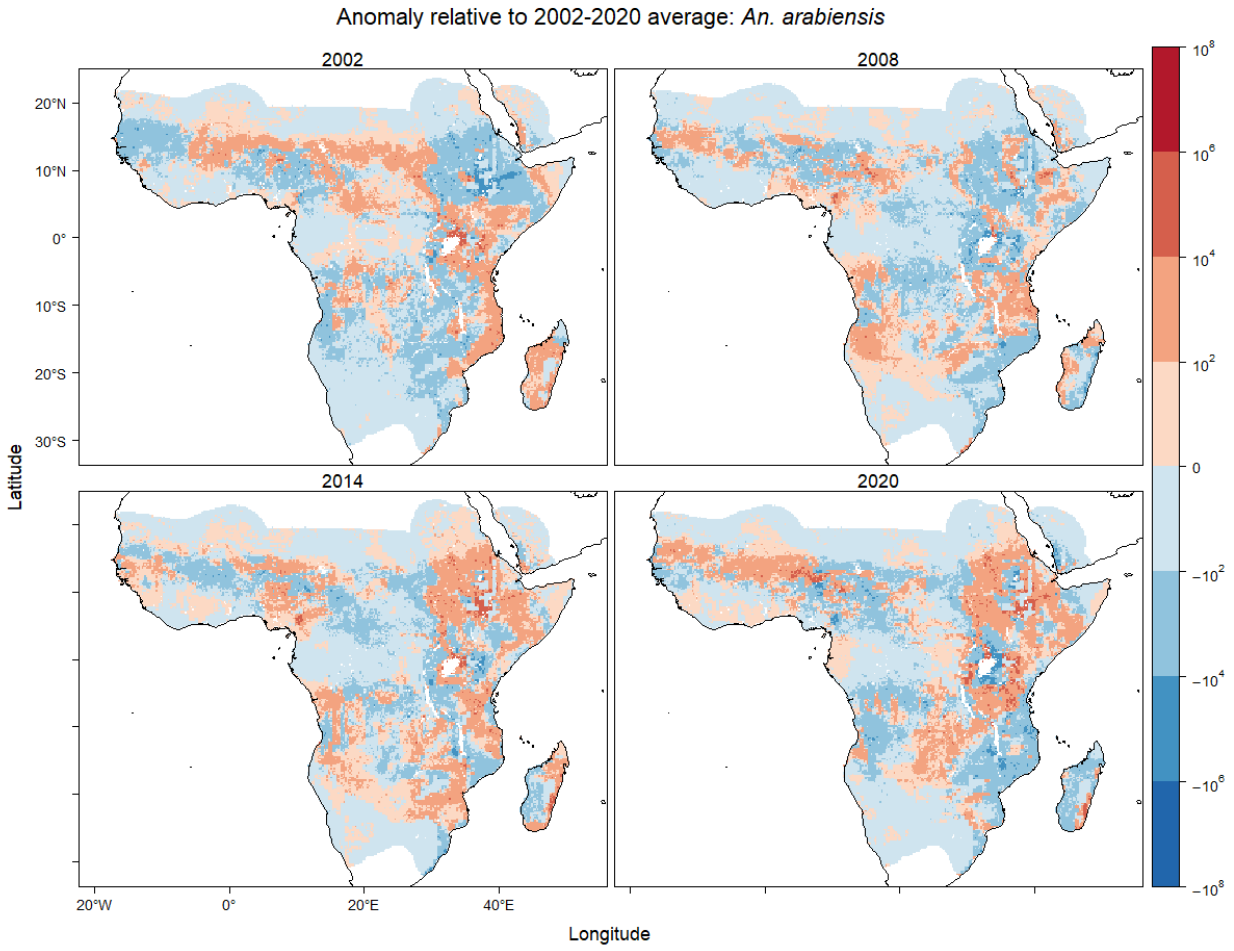


Figure 5: Annual average anomalies of total potential equilibrium abundance of *An. arabiensis* relative to its predicted average per 25 square kilometres over years 2002–2020. Predictions are masked to the potential range of *An. gambiae* s.l. [9] provided by the Malaria Atlas Project (`https://malariaatlas.org/`) under the terms of a Creative Commons Attribution 4.0 International License available at `https://creativecommons.org/licenses/by/4.0/`.

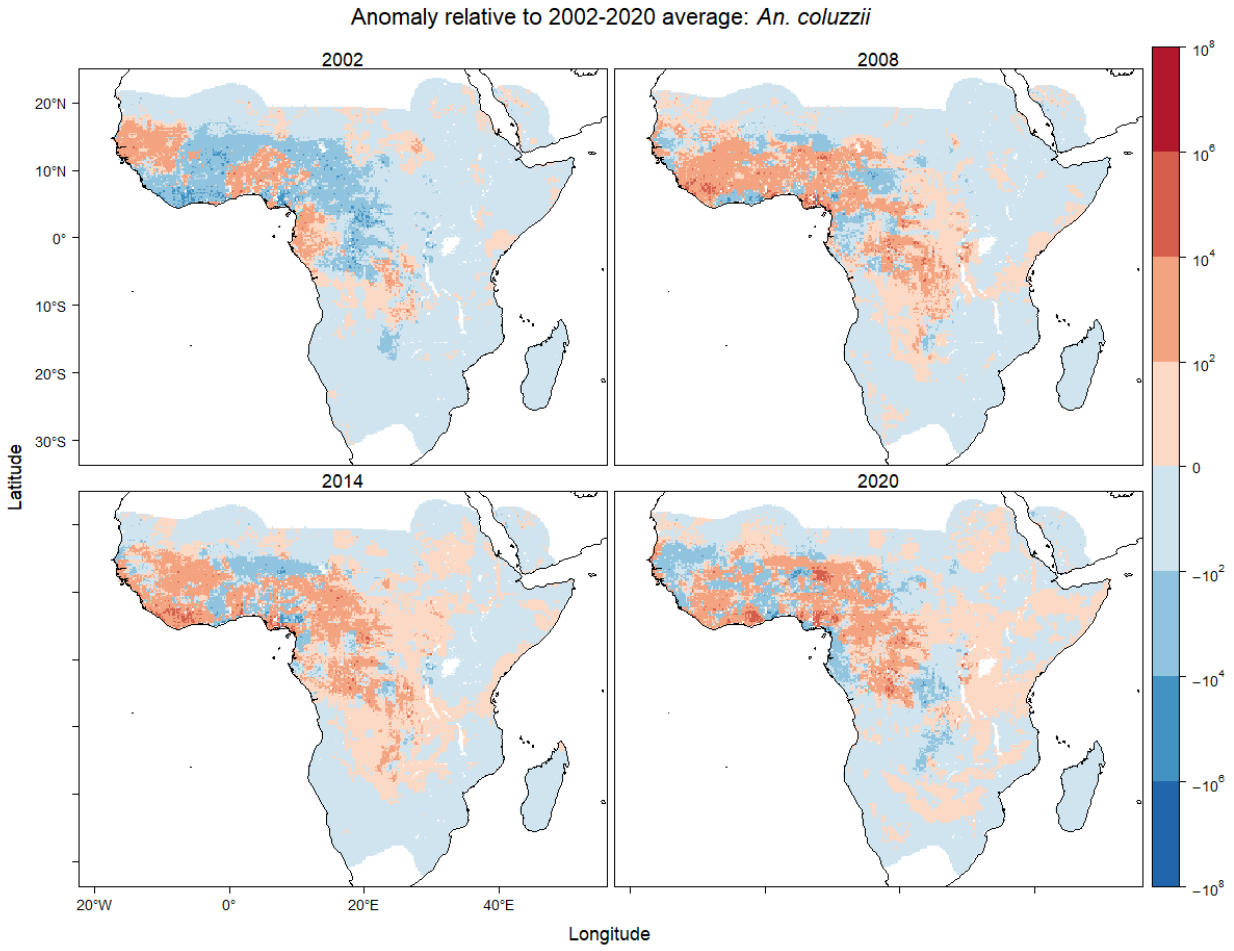


Figure 6: Annual average anomalies of total potential equilibrium abundance of *An. coluzzii* relative to its predicted average per 25 square kilometres over years 2002–2020. Predictions are masked to the potential range of *An. gambiae* s.l. [9] provided by the Malaria Atlas Project (`https://malariaatlas.org/`) under the terms of a Creative Commons Attribution 4.0 International License available at `https://creativecommons.org/licenses/by/4.0/`.

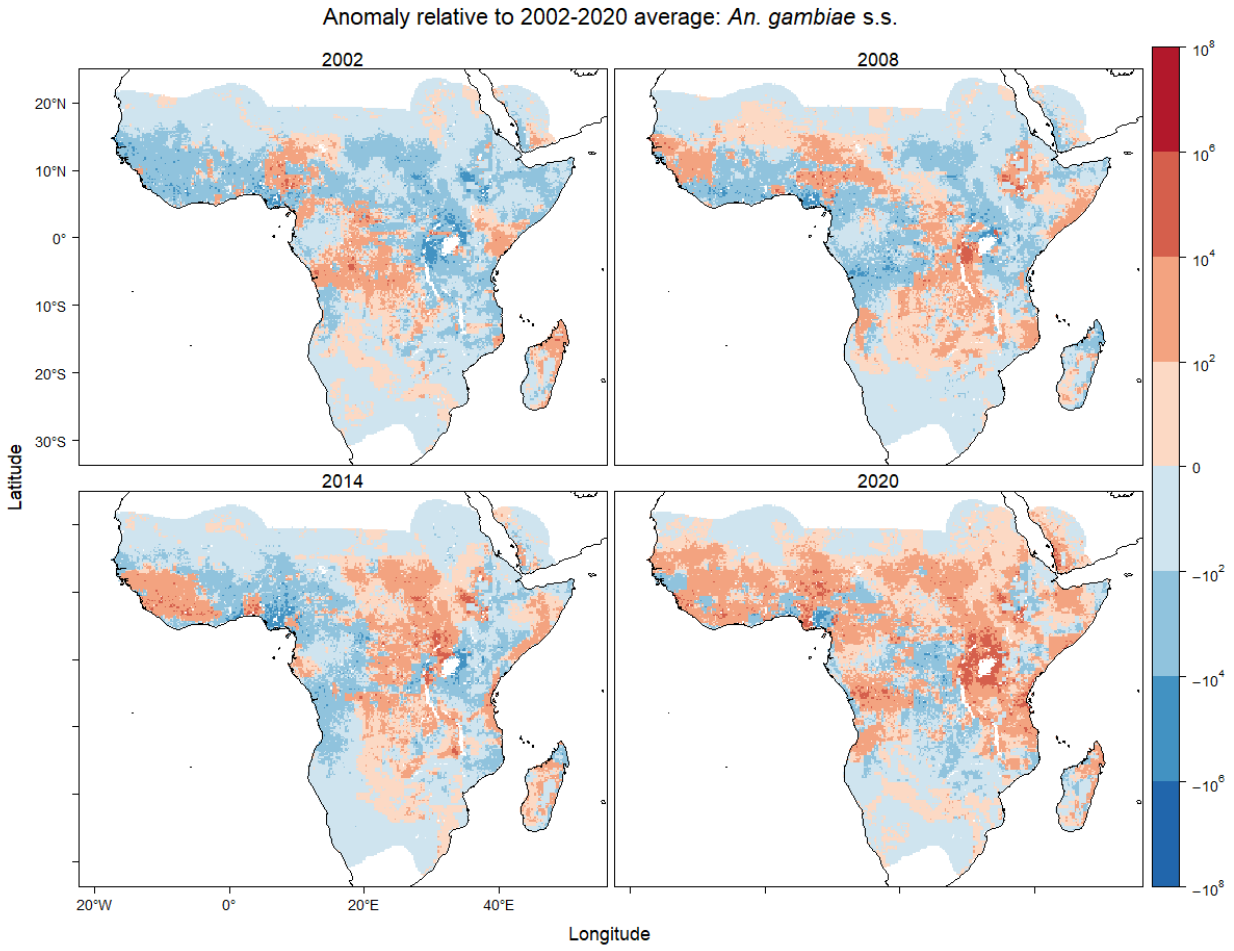


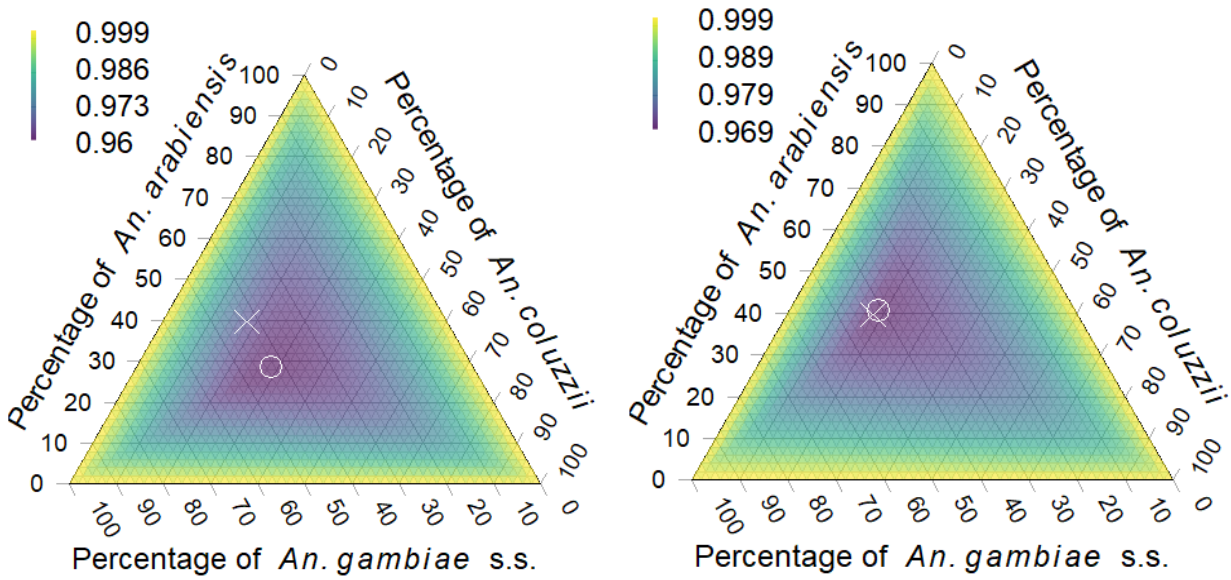
Figure 7: Annual average anomalies of total potential equilibrium abundance of *An. gambiae* s.s. relative to its predicted average per 25 square kilometres over years 2002–2020. Predictions are masked to the potential range of *An. gambiae* s.l. [9] provided by the Malaria Atlas Project (`https://malariaatlas.org/`) under the terms of a Creative Commons Attribution 4.0 International License available at `https://creativecommons.org/licenses/by/4.0/`.

Estimated Niche Overlap and Interspecific Competition

The Pianka estimates of niche overlap and the optimised interspecific competition estimates are compared in Table 2. The species with the greatest niche overlap was *An. gambiae* s.s. The strongest interspecific competition was estimated between *An. arabiensis* and *An. gambiae* s.s., which exhibited the greatest degree of spatial overlap (Figure 1). The two remaining pairwise relationships of optimised interspecific competition estimates were of the same magnitude. The equilibrium that maximised stability was further away from the observed equilibrium $\bar{\boldsymbol{x}}$ for the Pianka niche overlap estimates (Figure 8a) compared to the optimised competition estimates (Figure 8b). For the optimised estimate ($\hat{\boldsymbol{A}}$), the observed equilibrium approximately corresponded to the most stable species composition.

Table 2: Estimated competitive interactions over the species range of *An. gambiae* s.l. 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
Niche overlap ($\boldsymbol{\alpha}$)	0.14	0.24	0.24
Competition ($\hat{\boldsymbol{A}}$)	0.22	0.37	0.23



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

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

Figure 8: Spectral radius for all possible interior equilibrium species abundances 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 overall equilibrium (×). Note change in colour scale for (a) versus (b).

Discussion

Management of malaria requires knowledge of the distribution and abundance of important vector species, such as members of the *An. gambiae* s.l. species complex, to allow efficient vector abundance control by conventional methods [5] or potential new technologies including genetic biocontrol [4]. The spatio-temporal distribution of vector species abundances, and hence threat of malaria, will change with environmental factors [10] and hence potentially also with climate change [8]. In this work, we investigated these environmental relationships 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 Africa, and also evaluated the potential for interspecific interactions through niche overlap or competition.

Here, analyses of separate relative abundance data and abundance data were combined to produce joint predictions of the species potential equilibrium abundances. Interpretable predictive models were used to help learn about ecological relationships. The spatially and temporally varying meteorological factors of precipitation, relative humidity and temperature were found to be important predictors of relative abundance (Figure S16) over the observational period. Precipitation and temperature were also predictive of the per capita total abundance of *An. gambiae* s.l. The resulting predictions of equilibrium species abundances exhibited greater geographical overlap between *An. arabiensis* and *An. gambiae* s.s., whereas *An. coluzzii* was spatially restricted to western Africa (Figure 1). Across the species range of *An. gambiae* s.l. of sub-Saharan Africa, the three species varied intra-annually by seasons (Figures 2, 3 and 4) and inter-annually (Figures 5, 6 and 7) over the prediction period from 2002 to 2020. These temporally changing predictions are attributable in the analysis to changing meteorological conditions because the remaining covariates included in the statistical models varied either spatially or temporally but not both. The predictions suggested that changing climate conditions are associated with increased equilibrium abundance of *An. gambiae* s.s. in many areas, such as the region near Lake Victoria that experienced anomalously low temperatures and high relative humidity at the end of the prediction period (Figures S10, S11), at the expense of *An. arabiensis*. The species *An. arabiensis* was also predicted to decrease in southeastern Africa, but increase elsewhere in East Africa during the prediction period. Analyses suggested that the three species can coexist (Figure 8) despite evidence of substantial niche overlap and interspecific competition (Table 2) across the shared species range.

The species equilibrium abundances predicted by the model provided relative results similar to analyses restricted to relative abundance data [15], with relatively high abundance for *An. arabiensis* predicted in the northern, southern and eastern regions of the *An. gambiae* s.l. species range. In the northeastern region of the *An. gambiae* s.l. species range, the predicted relative abundance for *An. arabiensis* in this study (Figure S16) was low in some areas where its dominance was predicted by [15]. Occasionally low relative abundance for *An. arabiensis* was however also predicted by [23] in this area. The spatial pattern of low and high abundance areas for *An. arabiensis* (Figure 1) was similar to predicted patterns of potential niche indices [8]. The abundance predictions for all three species (Figure 1) were geographically similar to the predicted relative probabilities of occurrence provided by [9], where high abundances and high relative probabilities of occurrence overlapped in

many regions for each species. One area of discrepancy was the Horn of Africa and the Arabian Peninsula, where *An. gambiae* s.s. was predicted in abundance (Figure 1) but had low predicted relative probabilities of occurrence in [9]. However, the relative abundance predictions for *An. gambiae* s.s. in these regions were accompanied by high uncertainty (Figure S17) because of the sparse data (Figure S1) available there.

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 the delayed density dependent model proposed by [54] and 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 daily time series data, this study used smoothed moving average covariates of the meteorological covariates that determine the environmental carrying capacities and equilibrium abundances of the species. The equilibrium abundance of a species is the long term species abundance predicted in the absence of perturbations. The predicted equilibrium is spatially and temporally varying due to changing environmental conditions that modify the Grinellian niche of the constituent species [57], which in turn is modified by interspecific competition to produce the realised niche [58]. 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 [21, 59]. 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 [22] 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. [49].

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 sub-seasonal 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 inter-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. [6]. Although the ecologically important and spatially and temporally varying covariate of livestock density was not explicitly represented, the lower human host feeding rate for the former 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 [28]. 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 increased relative abundance of *An. arabiensis* versus the species pair of *An. gambiae* s.s. plus *An. coluzzii* are linked to the distribution of long lasting insecticidal nets in a study [15] focused on spatial analysis. The study noted that the effect of vector control

efforts should consider both the initial impact as well as long term effects. However, vector control activities varied through both time and space in Africa over the observational period, and furthermore the primary endpoint for malaria control in most of sub-Saharan Africa is human health rather than vector control [5]. The cumulative effects of vector control efforts also depend on insecticide resistance, land use, and climate [39, 60–62]. In contrast to [15], the change in species composition was in this study elaborated to the species level for *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. and allowed for temporal covariates including interannual trends in species composition. During the observational period, the use of insecticide treated nets (ITNs) in Africa doubled from 24 in use per person in 2010 to 44 in use per person in 2020 [63], and ITN coverage by household in sub-Saharan Africa increased from 5% in 2000 to 70% in 2022 [3]. However, the relative abundance of *An. coluzzii* was predicted to increase instead of *An. arabiensis*. Although this trend of increasing *An. coluzzii* relative abundance could be caused in part to climatic or land use factors, 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 [11, 16, 29]. That is, the estimated trend of increasing *An. coluzzii* relative abundance in this study could be explained by observational factors rather than underlying changes in actual species composition over the observational period. 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* [64]. The predictions for relative abundance therefore fixed the year covariate to the last year of the prediction period, 2020, so that all temporal changes in predicted species equilibrium abundances were attributable to the meteorological covariates of precipitation, temperature and relative humidity.

Regarding the predicted overall equilibrium abundance across the entirety of the *An. gambiae* s.l. species range, multiplying the predicted average daily aggregate total for adult *An. coluzzii* and *An. gambiae* s.s. by a factor of two to allow for a roughly equal sex ratio suggested an estimate of 1.87×10^{10} average daily abundance of these two species across sub-Saharan Africa. For comparison, an estimate of 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 [65]. Assuming a ratio of total to effective population size of 0.10 [66] suggested an effective population size of 1.87×10^9 based on the combined predicted equilibrium abundances of *An. coluzzii* and *An. gambiae* s.s. The estimates for niche overlap among the three species (Table 2) were, in comparison, less than those reported by Pombi et al. [11] in a study that derived the estimates from field observations in a region of Burkina Faso where all three species coexist. In contrast, the niche overlap estimates from this study included the whole of the range of *An. gambiae* s.l. in regions where not all species were predicted to occur with non-negligible abundances. As a consequence, the niche overlap estimates were lower in this study because of its broader geographic scale. The interspecific competition estimates (Table 2) suggested that the three species could exhibit substantial interspecific competition while maintaining stable coexistence at the overall predicted equilibrium abundances.

Conclusion

Meteorological factors such as precipitation and temperature were found at the continental scale to be important spatially and temporally varying predictors of equilibrium abundances of the dominant malaria vector species within the *Anopheles gambiae* s.l. species complex: *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. Human population density was also an important predictor that was negatively associated with per capita vector abundance. A longer term trend of increasing relative abundance of *An. coluzzii* was also found, and was possibly attributable to advances in genetic methods of species identification within the *An. gambiae* species complex that occurred over the observational period. The equilibrium abundance of *An. gambiae* s.s. was predicted to increase across much of its range during the prediction period, such as around the Lake Victoria basin, due to changing meteorological conditions. In contrast, *An. arabiensis* was predicted to decrease around Lake Victoria and southeastern Africa but increase elsewhere in eastern Africa. 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 **Agsl-Predict** [26] provided as a supplement for this article.

Acknowledgements

The authors thank Scott Foster and Skipton Woolley 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 D'eveloppement (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** [26] provided as a supplement for this article.
- Materials availability. Not applicable.
- Code availability. The reproducible code is provided in the R package **AgslPredict** [26] 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 2023. Geneva: World Health Organization; 2023.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

[12] Sillero, N, Arenas-Castro, S, Enriquez-Urzelai, U, Vale, CG, Sousa-Guedes, D, Martínez-Freiría, F, et al. Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. Ecological Modelling. 2021;456:109671.

[13] Barker, JR, MacIsaac, HJ. Species distribution models applied to mosquitoes: Use, quality assessment, and recommendations for best practice. Ecological Modelling. 2022;472:110073.

[14] Adeogun, A, Babalola, AS, Okoko, OO, Oyeniyi, T, Omotayo, A, Izekor, RT, et al. Spatial distribution and ecological niche modeling of geographical spread of *Anopheles gambiae* complex in Nigeria using real time data. Scientific Reports. 2023;13:13679.

[15] 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.

[16] 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.

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

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

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

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

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

[22] 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.

[23] 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.

[24] 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.

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

[26] Hosack, GR, El-Hachem, M. AgslPredict: Code and data for predictions of *Anopheles gambiae* sensu lato; 2024. R package version 0.2.0. Available from: `https://github.com/csiro-risk-assessment/AgslPredict`.

[27] 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.

[28] 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`.

[29] 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.

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

[31] 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.

[32] 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/.

[33] 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.

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

[35] 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.

[36] 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.

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

[38] Bright, E, Rose, A, Urban, M. LandScan Global 2015 [Data set]. Oak Ridge National Laboratory; 2016. Accessed 20 September 2022. Available from: https://doi.org/10.48690/1524210.

[39] Zhou, G, Afrane, YA, Dixit, A, Atieli, HE, Lee, MC, Wanjala, CL, et al. Modest additive effects of integrated vector control measures on malaria prevalence and transmission in western Kenya. Malaria Journal. 2013;12:256.

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

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

[42] 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.

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

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

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

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

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] Sinka, ME, Pironon, S, Massey, NC, Longbottom, J, Hemingway, J, Moyes, CL, et al. A new malaria vector in Africa: Predicting theexpansion range of *Anopheles stephensi* and identifying the urban populations at risk. Proceedings of the National Academy of Sciences, USA. 2020;117:24900-8.

[52] 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.

[53] 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.

[54] Hosack, GR, El-Hachem, M, Beeton, NJ. Stability of difference equations with interspecific density dependence, competition, and maturation delay. arXiv. 2025;arXiv:2502.13426 [q-bio.PE]. Available from: https://doi.org/10.48550/arXiv.2502.13426.

[55] 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.

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

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

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

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

[60] Munga, S, Yakob, L, Mushinzimana, E, Zhou, G, Ouna, T, Minakawa, N, et al. Land use and land cover changes and spatiotemporal dynamics of anopheline larval habitats during a four-year period in a highland community of Africa. American Journal of Tropical Medicine and Hygiene. 2009;81:1079.

[61] Sougoufara, S, Doucouré, S, Backé Sembéne, PM, Harry, M, Sokhna, C. Challenges for malaria vector control in sub-Saharan Africa: Resistance and behavioral adaptations in *Anopheles* populations. Journal of Vector Borne Diseases. 2017;54:4-15.

[62] Anoopkumar, A, Aneesh, EM. A critical assessment of mosquito control and the influence of climate change on mosquito-borne disease epidemics. Environment, Development and Sustainability. 2022;24:8900-29.

[63] Malaria Atlas Project. Interventions – ITN Use (in 100 People); 2024. Accessed 24 October 2024. URL `https://data.malariaatlas.org/trends?year=2020&metricGroup=Interventions&geographicLevel=admin0&metricSubcategory=ITN&metricType=rate&metricName=use`.

[64] 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.

[65] 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.

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