**Seasonal Dynamics of an Emerging African Malaria Vector, *Anopheles stephensi*: Implications for Malaria Dynamics and Control**

**Population Dynamics and Seasonality of the Emerging African Malaria Vector *Anopheles stephensi*: Implications for Malaria Control**

* **Do I want to mention “Urban” somewhere???**
* **Do I want to mention “Areas where malaria is absent” somewhere???**

Charles Whittaker1\*, Ellie Sherrard-Smith1, Peter Winskill1, Gina Cuomo-Dannenburg1, Patrick G.T. Walker1, Marianne Sinka2, Ashwani Kumar3, Azra Ghani1, Samir Bhatt1,4, Thomas Churcher1 & Arran Hamlet1

1MRC Centre for Global Infectious Disease Analysis & Abdul Latif Jameel Institute for Disease and Emergency Analytics, School of Public Health, Imperial College London, London, UK

2Department of Zoology, University of Oxford, Oxford, UK

3Vector Control Research Centre, Indira Nagar, Puducherry, India

4Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

**\*Corresponding Author:** Charles Whittaker, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College, London W2 1PG, United Kingdom. Email: [charles.whittaker16@imperial.ac.uk](mailto:charles.whittaker16@imperial.ac.uk)

**Abstract**

**Reductions in malaria prevalence have been driven by extensive scale up of control interventions, but also increasing urbanisation, particularly given urban areas in SSA tend to have lower prevalence, underpinned by improved housing and reduced availability of larval habitats.**

**Urban malaria, underpinned by *Anopheles stephensi*, is immensely prevalent elsewhere. The past decade has seen reports of *Anopheles stephensi* across parts of the Horn of Africa, where it is thought to be underlying dramatic rises in malaria cases in Djibouti.**

**Significant uncertainty remains however, particularly with respect to the temporal and seasonal dynamics of the vector – the degree, extent and timing of seasonality in mosquito abundance underlies the temporal profile of malaria and disease risk and therefore has material consequences for 1) which malaria control interventions will be most effective at preventing its establishment/reducing disease burden and 2) the dynamics and speed of malaria establishment in the region.**

**Here we collate *Anopheles stephensi* longitudinal catch-data from \_\_\_ locations and \_\_\_ countries in order to better understand the population and seasonal dynamics of *Anopheles stephensi*, and further out understanding of how these dynamics might play out in the Horn of Africa.**

**Our analyses reveal pronounced variation in dynamics across locations in the extent of seasonality and timing of seasonal peaks observed for *Anopheles stephensi*, ranging from single, highly seasonal peaks, to bimodality and near-perennial patterns of annual abundance. Importantly, we discover systematic variation in seasonal dynamics between urban and rural settings, suggesting structural differences in how these environments facilitate vector abundance.**

**Integrating these seasonal profiles with a model of malaria transmission, our results reveal the degree of vector seasonality materially impacts the speed, scale up and dynamics of malaria establishment in settings where transmission is currently minimal or even absent, and thus more generally, highlights significantly uncertainty in our understanding of how *Anopheles stephensi* could influence the timing and dynamics of malaria establishment in settings across the Horn of Africa where malaria transmission is currently absent or minimal.**

**Our results underscore the significant threat this emerging, urban malaria vector poses to public health in this region of sub-Saharan Africa, as well as the continent more generally, and highlights the fundamental, urgent need for significant scale up on entomological monitoring across the region.**

**Keywords:** *Anopheles stephensi*; malaria ecology; urban malaria; population dynamics; seasonality; epidemiology.

**Introduction**

With an estimated 241 million cases and over 600,000 deaths across endemic countries in 20201, malaria represents one of the most significant infection diseases globally2. Burden of disease is concentrated in sub-Saharan Africa where an estimated 96% of malaria deaths in 2020 occurred – 80% of these in children under 53. This represents an almost 40% reduction in clinical disease since the year 2000, an achievement underpinned predominantly by significant scale-up of control interventions including insecticide-treated bednets4.

Alongside this significant expansion of control efforts, increasing urbanisation of Africa’s populace (rising from 31% to 43% between 1990 and 2018, with over 60% expected to live in urban areas by 20505) is also thought to have indirectly contributed to reductions in disease burden. Previous work has highlighted significantly lower annual Entomological Inoculation Rates (EIR) in urban compared to rural settings6,7, a finding thought to be underpinned by features factors including systematic differences in the quality of housing8,9, reduced availability and suitability of habitats for *Anopheline* breeding in urban settings10–12, better access to treatment13, and higher population densities leading to higher human to mosquito ratios (and reduced transmission)14. Whilst these trends are not always consistently identified (see e.g. surveys in Libreville, Gabon15 or Cotonou, Benin16 where prevalence of malaria is higher in urban areas than immediately surrounding locations; or previous work highlighting that *Anopheles gambiae s.s.* can adapt to breeding in polluted water characteristic of urban environment17), increasing urbanicity across sub-Saharan Africa is likely to complement planned scale-up of malaria control interventions aimed at achieving the targets outlined in the World Health Organization’s 2030 Global Technical Strategy for Malaria18.

This potential positive impact of increasing urbanization is contingent on urban settings across the continent remaining areas of comparatively low malaria transmission. However, this property of urban locations is currently under threat by the invasion and establishment of the malaria vector *Anopheles stephensi*. Found throughout South Asia, *Anopheles stephensi* is a highly efficient urban vector capable of transmitting both *Plasmodium falciparum* and *Plasmodium vivax* parasites, with this efficiency thought to be underpinned by both an increased tolerance for breeding in polluted water sources25, and a superior ability to access and utilise the manmade hydrological habitats present in urban settings20,21. The species was first identified in sub-Saharan Africa in Djibouti City in 201219 and has since been recorded in both Ethiopia20,21 and Sudan22,23, with recent work highlighting likely suitability for the species across some of the continents largest population centres comprising over 100 million people24. Whilst causality has yet to be conclusively established, its emergence is thought to have contributed to the significant resurgence of malaria transmission in Djibouti (which experienced a 10-fold increase in cases between 2013 and 2019), highlighting the potential threat establishment of this vector poses to malaria control across the Horn of Africa28, as well as wider efforts across the continent29.

The situation in Djibouti, as well as recent modelling work suggesting *Anopheles* *stephensi*’s establishment in Ethiopia could lead to a 50% increase in malaria incidence28, highlights the significant public-health threat the vector poses. Despite this, substantial uncertainty remains how its future expansion and establish might influence malaria in the region, particularly in the (predominantly urban) settings where the disease is currently largely absent. A key driver of this will be the seasonal patterns of abundance by *Anopheles stephensi*. Mosquito populations are highly temporally dynamic, often exhibiting substantial annual fluctuations in size that drive the temporal profile of disease risk. Understanding the factors underlying these dynamics is crucial given that the effectiveness of many malaria control interventions (such as seasonal malaria chemoprevention30 or indoor-residual spraying31) depends on the timing of their delivery relative to seasonal peaks in transmission. A better understanding of *Anopheles stephensi*’s seasonal dynamics is likely to have material consequences for both the nature of malaria establishment in the Horn of Africa, as well as the potential impact of currently available control interventions.

Despite this relevance, substantial uncertainty remains regarding *Anopheles stephensi*’s seasonal dynamics; numerous studies carrying out longitudinal catches are present in in the literature, but these typically only focus on a single location, precluding systematic comparison and identification of generalisable patterns. Here we collate longitudinal mosquito catch data for *Anopheles stephensi* from 6 countries and 45 unique locations across South Asia, the Middle East and the Horn of Africa in order to better understand these dynamics, and the ecological factors underpinning them. Our results highlight pronounced variation in the extent and timing of seasonality, with distinct dynamics observed across rural and urban settings. Integrating these results with a model of malaria transmission highlight the material consequences this variation has for the speed and dynamics of malaria establishment in parts of the Horn of Africa where the disease is currently (or has previously been) largely absent and underscores the crucial, urgent need for rapid scaleup of entomological monitoring for this emerging urban vector across the region.

**Methods**

**Systematic Review of *Anopheles stephensi* Literature**

We collated references from previously published systematic reviews of literature relating to *Anopheles stephensi*27,32, and updated these previous searches (both conducted in 2017) by searching *Web of Science* and *PubMed* databases from January 2017 to September 2020 for further relevant references (full details of keywords and search methodology available in **Supplementary Information**). We included all records containing temporally disaggregated adult mosquito catch data with monthly (or finer) temporal resolution spanning at least 10 months, that had not been conducted as part of vector control intervention trials, where sufficient information to geolocate the catch site to the administrative unit 2 level, and where a total of at least 25 *Anopheles stephensi* had been caught over the study period. Overall, a total of 36 references were collated containing 65 time-series from catch surveys carried out in distinct locations from across Afghanistan (n=2), Djibouti (n=1), India (n=32), Iran (n=17), Myanmar (n=5) and Pakistan (n=8). See **Supplementary Information** for further details.

**Gaussian Processing Fitting and Smoothing of Time Series Data**

Per previous work32, we fitted the following Gaussian Process model to smooth these noisy mosquito count data, using a Negative Binomial likelihood to account for overdispersion in the data:

where is a distribution over functions from a zero-mean Gaussian Process with covariance function , with the covariance between two timepoints and defined according to the kernel function . Given that mosquito population dynamics are typically characterised by repeating patterns occurring either seasonally or annually, a periodic kernel function was used to define the covariance between pairs of points, which depends on both the Euclidean distance in time between points as well as the hyperparameters , and , where represents the period over which we would expect points to show similar dynamics (i.e. a period of twelve would imply we expect points separated by 12 months to be most similar), specifies the magnitude of the covariance, and  represents a lengthscale parameter further constraining the extent to which two values separated by a given time can co-vary. (x) are function evaluations at times , are the observed mosquito counts indexed by timepoint . Fitting was undertaken with the probabilistic programming language STAN33.

**Characterisation of Temporal Properties and Clustering of Similar Time-Series**

Motivated by previous work providing a framework to statistically characterise the empirical structure of time-series data34, we calculated a number of summary statistics for each smoothed time-series to characterise their temporal properties. These include the median of the period () from the Negative Binomial Gaussian Process fitting (that describes the dominant modality of annual or sub-annual temporal variation present in the data), the proportion of points in each smoothed time-series where the total catch was greater than 1.65x the mean (providing a measure of how peaked the time-series is), the distance of the first peak from January (providing an indication of timing), the proportion of the total catch that was found in any 4 month period (in-keeping with previous definitions of malaria seasonality35), and then 3 features arising from fitting 1 and 2 component Von-Mises distributions (a continuous probability distribution representing the maximum entropy distribution for circular/repeating data) to the smoothed time-series. These 3 features were the mean of the 1 component Von-Mises distribution (providing an indication of where in the year the region of highest catch numbers were), the number of peaks (determined by comparing the quality of fit for 1 and 2 component Von-Mises distributions, roughly corresponding to asking whether assuming the time-series was unimodal or bimodal provided the best fit), and the weight (), specifying the comparative contributions of each component in the two-component fitting. See **Supplementary Information** for further details. We therefore obtain for each time-series a set of 7 real numbers providing a reduced representation of their temporal properties – to this, we then applied a Principal Components Analysis to identify a lower-dimensional representation of the structure present in the data amenable to visualisation and implemented k-means clustering to identify clusters of time-series with similar temporal properties.

**Random Forest Prediction of Seasonality and Peaks**

For each of the 65 time-series, we extracted a suite of environmental variables derived from satellite data that together describe the ecological structure of the location the entomological survey was carried out in, including the *BioClimatic* variables (a suite of biological relevant covariates defined from monthly rainfall and temperature satellite data36) as well as measures of landcover and urbanicity amongst others (a complete list can be found in **Supplementary Table X**). These environmental variables were then used as predictors within a Random-Forest based classification framework aimed at predicting whether a given time-series was either “more seasonal” (possessing a single distinct seasonal peak, as defined by \_\_\_\_) or “less seasonal” (lacking a distinct single seasonal peak, which typically translated to either possessing two distinct seasonal peaks or more perennial patterns of transmission – see **Supplementary Figure X** for the individual fitted profiles for each time-series), with these definitions defined per previous work34 based on the results of fitting and comparing 1 and 2 component Von-Mises distributions to each smoothed time-series**.** These models were fit using the software package *Ranger*37, implemented in the *tidymodels* framework for R38, with 6-fold cross-validation utilised to optimise hyperparameter combinations; presented results are based on averaging the results of 25 separate iterations of cross-validation and model fitting (to account for stochasticity in model fitting), and any predictions made using out-of-bag estimates in all instances. Further details on defining the outcome variable and the Random Forest modelling framework can be found in the **Supplementary Information**.

**Modelling Malaria Establishment and Transmission**

We integrated the temporal profiles of *Anopheles stephensi* abundance into a deterministic version of a well-established compartmental model of *Plasmodium falciparum* malaria transmission and disease39–41 to explore the implications of *Anopheles stephensi* establishment and seasonality on the dynamics of malaria transmission, with a particular focus on areas where malaria transmission is currently absent or only minimally present. In brief, the human population is split into either susceptible or infected individuals, with those infected either being asymptomatic, currently suffering from clinical disease, possessing a sub-microscopic infection, having been treated, or currently in a period of prophylaxis following treatment. Mosquito populations are divided according to different life-stages, with larvae, pupae and adults all simulated, and subdivided into either susceptible, exposed or infectious (after the extrinsic incubation period, EIP) classes. The model explicitly takes into account heterogeneity in transmission, age-dependent mosquito biting rates, and the acquisition of natural immunity following exposure. Full details of the model can be found in **Supplementary Information**. For the purposes of exploring the impact of *Anopheles stephensi*, we integrate the collated temporal profiles of mosquito abundance, and integrate these into this modelling framework, making a number of assumptions regarding 1) the rapidity with which *Anopheles stephensi* populations scale and establish in the Horn of Africa; and 2) the final average annual density level the *Anopheles stephensi* population reaches – sensitivity of our conclusions to these assumptions are presented in both the main text and **Supplementary Figures X** and **X**.

**Results**

**Extensive Diversity In Temporal Dynamics Across the Collated *Anopheles stephensi* Time-Series:** A total of 65 time-series from X unique locations across Afghanistan, Djibouti, India, Iran, Myanmar and Pakistan were identified (**Fig.1A)**. These noisy time-series were then smoothed using a negative binomial gaussian process-based framework (see **Fig.1B** for a single example time-series from each country, see **Supp Fig.1** for all individual time-series). Substantial variation in temporal dynamics was observed across the collated time-series in terms of the degree and timing of seasonality – this ranged from highly seasonal dynamics with a single, clear seasonal peak (e.g. as with Afghanistan example in **Fig.1B**) to more perennial patterns of abundance (see Pakistan example in **Fig.1B**) and even bimodal annual population dynamics with two peaks observed across the course of a single year (as with the Iran example in **Fig.1B**).

**Statistical Characterisation and Clustering of Temporal Properties Highlights Distinct Temporal Archetypes:** An array of summary statistics were calculated for each time series to characterise their temporal properties (see **Supplementary Methods *Time Series Characterisation and Analysis*** and **Supp Fig.2**). A reduced representation of this variation was then generated using principal components analysis (PCA) and the result then clustered using the k-means algorithm to cluster the time-series into discrete groups of time-series that all share similar temporal patterns. Our results highlight two distinct clusters of time-series, with each group characterised by distinct and different temporal patterns **(Fig. 2A)** and that differ in the extent and degree of their seasonality (here defined as the percentage of total vector density that occurs across any continuous 3-month period, **Fig. 2B**). Cluster 1 time-series typically had clear, single seasonal peaks and were more seasonal on average (57% of total vector density in a 3-month period) than Cluster 2 time-series, which had less seasonal (more perennial) patterns of annual abundance (average 36% vector density in any consecutive 3-month period), including time-series with two peaks across the course of a single year. These results were sensitive to our choice of the number of clusters the k-means algorithm was specified to identify – specifying 4 clusters instead of 2 resulted in further disaggregation of the 49 time-series in Cluster 1 into 3 separate clusters, each characterised by a single seasonal peak, but which differed in the timing of their seasonal peak relative to the timing of peaks in monthly rainfall (**Supp Fig 3**). Despite differing significantly in vector abundance seasonality **(Fig. 2C, top panel, p<0.001)**, there was no significant difference across the Cluster 1 and Cluster 2 time-series and their respective locations in terms of rainfall seasonality **(Fig. 2C, bottom panel, p=0.59)**. This suggests that the differences in vector seasonality across the two clusters was not being driven by differences in the timing and seasonality of rainfall across the surveyed locations.

**Random-Forest Modelling and Prediction of Seasonal Dynamics Highlights Urbanicity As a Key Driver:** In order to explore the ecological factors underpinning this observed variation in *Anopheles stephensi* seasonality, we fitted a random forest-based classification modelling framework to a suite of satellite-derived environmental covariates in order to predict cluster membership (either Cluster 1 or Cluster 2, as defined in **Fig 2)**. Due to the significant class size imbalance between Cluster 1 (n=49) and Cluster 2 (n=16), we up-sampled the Cluster 2 data using the smote algorithm in order to generate balanced class sizes. Model discriminative and predictive performance was good – across the 25 iterations of random forest model fitting carried out, the mean AUC was 0.89 (indicating good predictive performance, **Fig. 3A**) and on average, the model was able to correctly classify Cluster 1 time-series X% of the time, and Cluster 2 time-series Y% of the time. Annual temperature seasonality and rainfall seasonality (both bioclimatic variables describing the extent of annual fluctuations in the two environmental quantities), population per square kilometre and a number of measures of landcover (specifically LC30 which corresponds to \_\_\_\_\_\_, and LC20 which corresponds to \_\_\_\_\_\_) were all highly predictive of Cluster membership **(Fig. 3B)**. Areas with lower population density were more likely to belong to Cluster 2, as were areas with low values of LC30 (i.e. minimal land occupied by \_\_\_\_\_). By contrast, areas with high values of LC20 (i.e. substantial fraction of land covered by \_\_\_\_) were more likely to belong to Cluster 1, as were areas in which rainfall was strongly seasonal (see **Supp Fig. 4** for the full array of partial dependence plots for all covariates used in the analysis).

Motivated by this strong association with population per square kilometre, we manually collated whether each time-series had been carried out in a rural or urban setting (as defined by the authors of each study) and examined the association between rurality/urbanicity and temporal dynamics **(Fig. 3C)**. There was a near-significant association with rurality/urbanicity and cluster membership (chi squared test, p=0.07) – of the 25 time-series from studies carried out in urban settings, a significant majority (n =22) were assigned to Cluster 1 (88%) and only 3 to Cluster 2 (12%). By contrast, time-series carried from rural locations were more evenly spread between the two clusters – of the 37 time-series from studies in rural settings, 24 were assigned to Cluster 1 (65%) and 13 assigned to Cluster 2 (35%).

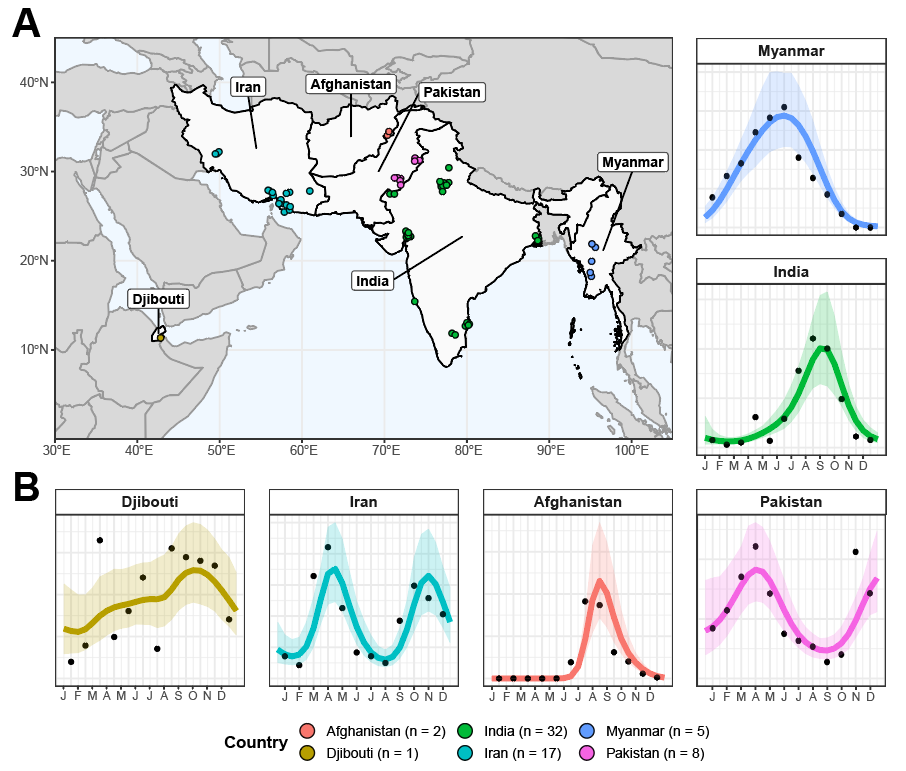
These results are qualitatively robust to the up-sampling procedure that we applied. Model predictive performance and variable importance rankings were similar when no up-sampling was applied to the dataset (mean AUC of 0.81, still indicating good predictive performance, see **Supp Fig. 5** for the ROC and variable importance plots), though average predictive accuracy on Cluster 2 (X%) was significantly lower than predictive accuracy on Cluster 1 time-series (Y%). Model performance and relative ranking variable importance ordering was similarly retained when fitting the model and explicitly holding out a small subset of the data to evaluate model performance (n=7 time-series, see **Supp Fig. 6** for both results with and without upsampling). Overall study catch size was highly variable between the studies **(Supp Fig 7)**, and whilst the median catch size across the two clusters did not differ (Moody’s Median Test, p=\_\_), the mean catch size did differ (t-test, p=\_\_). We therefore carried out an additional sensitivity analysis including study average monthly catch as a covariate in the random forest model (**Supp Fig 8)** – predictive performance and relative rankings of variable importance remained similar despite this addition.

**Exploring Possible Seasonal Dynamics and Surveillance of *Anopheles stephensi* Across the Horn of Africa:** We next collated the same satellite-derived environmental covariates described above for countries across the Horn of Africa where *Anopheles stephensi* has been found. These covariates were then inputted into the trained random forest classification model to predict potential temporal dynamics (as denoted by Cluster membership) of *Anopheles stephensi* across the Horn of Africa region **(Fig. 4A)**. Our results highlight distinct geographical areas across the region considered more likely by the model to belong to Cluster 1 (more seasonal, **Fig. 4B**) and Cluster 2 (less seasonal), as well as substantial areas of significant uncertainty. We next asked what consequences this uncertainty about the degree and timing of *Anopheles stephensi* seasonality might have on entomological surveillance of the vector across the region. Specifically, what the probability of missing *Anopheles stephensi* in entomological catches might be as a function of the number of consecutive months sampled (with start month selected randomly i.e., assuming no knowledge of *Anopheles stephensi’s* temporal dynamics). Our results highlight that in instances where sites are sampled for a limited number of months, there is a significant risk of missing *Anopheles stephensi* – even at 5 months of consecutive entomological sampling ranging from X-Y% (mean Z%) for Cluster 1 time-series and A-C% (mean Y%) for Cluster 2 time-series, with the exact value dependent on the specific time-series (and *Anopheles stephensi* temporal profile) being considered **(Fig. 4C)**. Our results highlight that in the absence of a detailed understanding of *Anopheles stephensi*’s anticipated temporal dynamics, sampling for a limited number of months poses a significant risk of missing an already present *Anopheles stephensi*, and erroneously concluding it is absent.

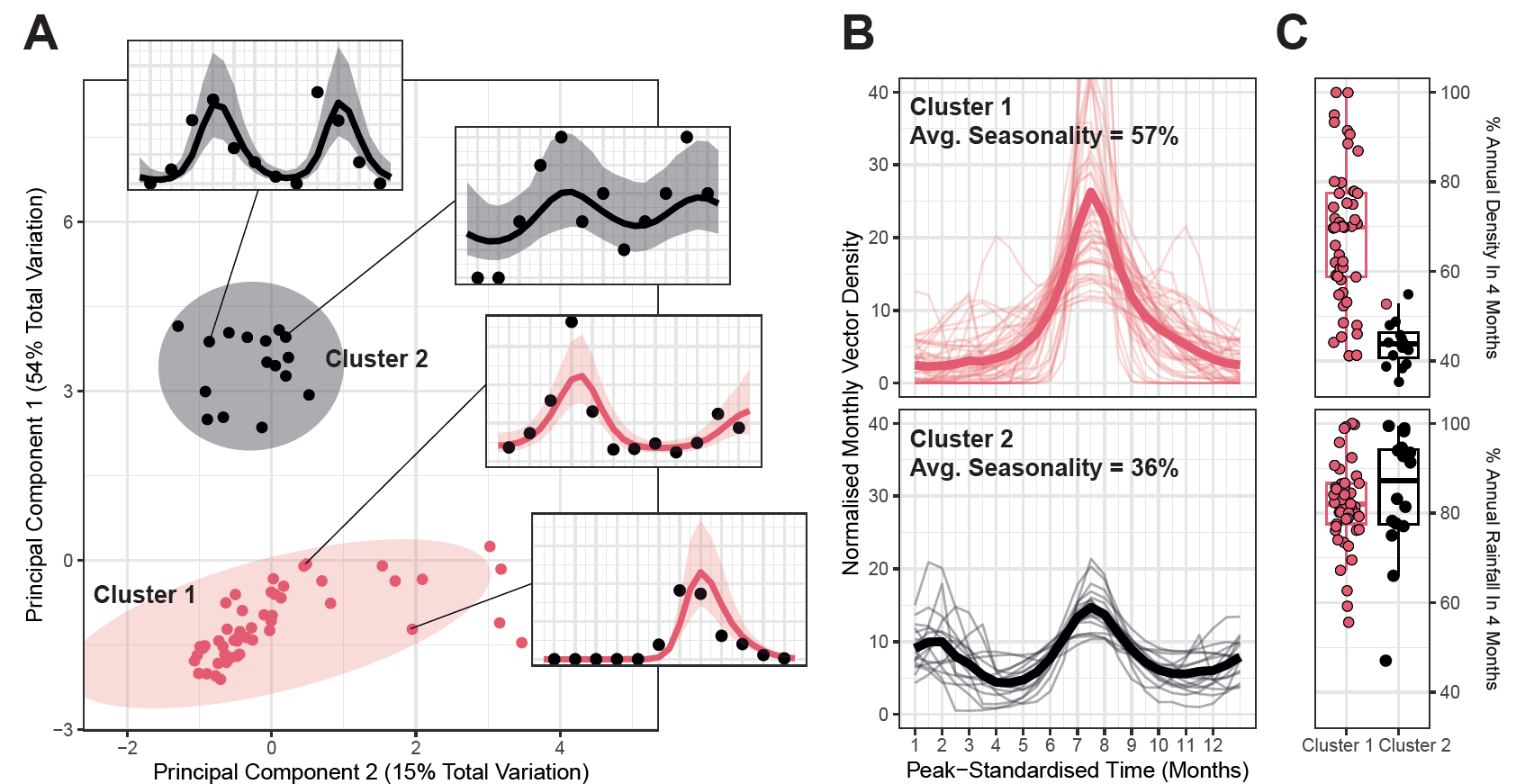
**Modelling the Impact of *Anopheles stephensi* Seasonality On the Impact of Vector Control Measures:** We next adapted an existing model of malaria transmission (see ***Supplementary Information*** for further details) to directly incorporate the different temporal profiles of *Anopheles stephensi* abundance that had been collated, and used the model to explore how variation in *Anopheles stephensi* temporal dynamics influences the impact of the indoor residual spraying (IRS), a key malaria control intervention. We explore three different IRS compounds (each of which is commonly used throughout the Horn of Africa), with a range of different functional half-lives (with respect to mosquito killing potential) following spraying **(Fig. 5A)** and model the impact of a single annual round of IRS. Across each of the 65 time-series assessed, pirimiphos methyl consistently had the highest impact, reducing annual malaria incidence by X% in the 12 months following spraying, compared to Y% and Z% on average for clothiandin and bendiocarb respectively **(Fig. 5B)**. The overall magnitude of reduction in malaria burden was dependent on the degree of seasonality, with IRS having the most impact in highly seasonal settings, and the least impact in settings where the degree of seasonality was minimal **(Fig. 5C** and **Fig. 5D)** – the influence of seasonality on impact was largest for bendiocarb (the insecticide with the shortest half-life), with reduction in malaria burden ranging from only A% in the lowest seasonality setting, through to as high as B% in the most seasonal setting, a C fold difference. Important to note however is that achieving this impact is dependent on optimal timing of intervention delivery relative to any seasonal peaks, which in turn requires sufficient understanding of the vector’s population dynamics and when peaks in vector density are likely to occur.

**Discussion (800-1000 words)**

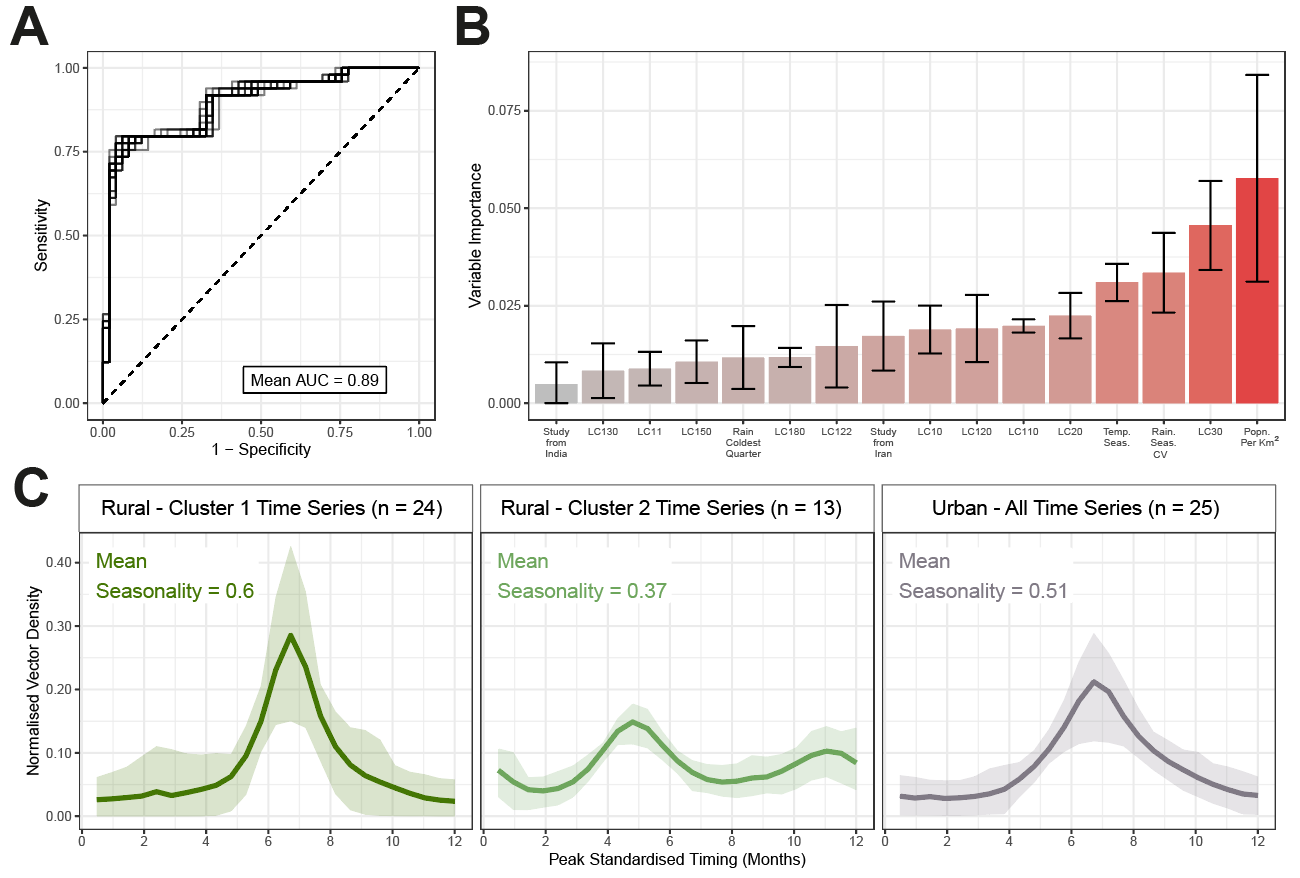
**Total (~3500-4000 words)**



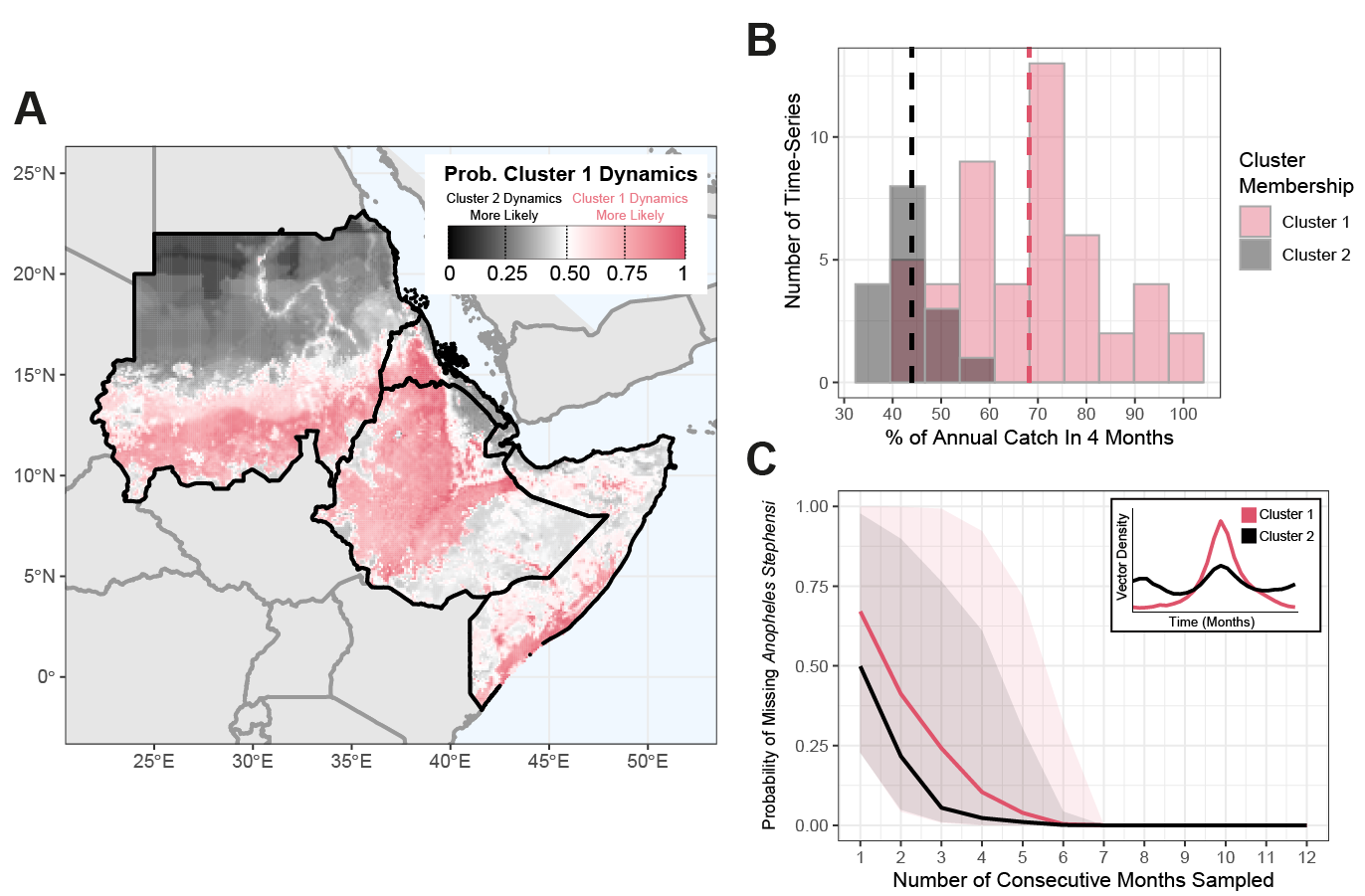
**Figure 1: Sources and Locations of *Anopheles stephensi* Time-Series Data and Examples for Each Country. (A)** Map of the geographical range over which collated time-series had been carried out in, with countries where studies had been carried out in highlighted in light grey, and the locations of individual studies indicated by points, coloured according to the country they were carried out in (Afghanistan = red, Djibouti = yellow, India = green, Iran = turquoise, Myanmar = blue and Pakistan = pink). **(B)** Example *Anopheles stephensi* time-series from each country, with the empirical monthly mosquito catch (black points) and fitted gaussian process curves (mean = coloured line, ribbon = 95% Bayesian Credible Interval) for each, coloured according to country. The x-axis indicates the month of sampling, the y-axis normalised annual vector density (i.e. arbitrary units).



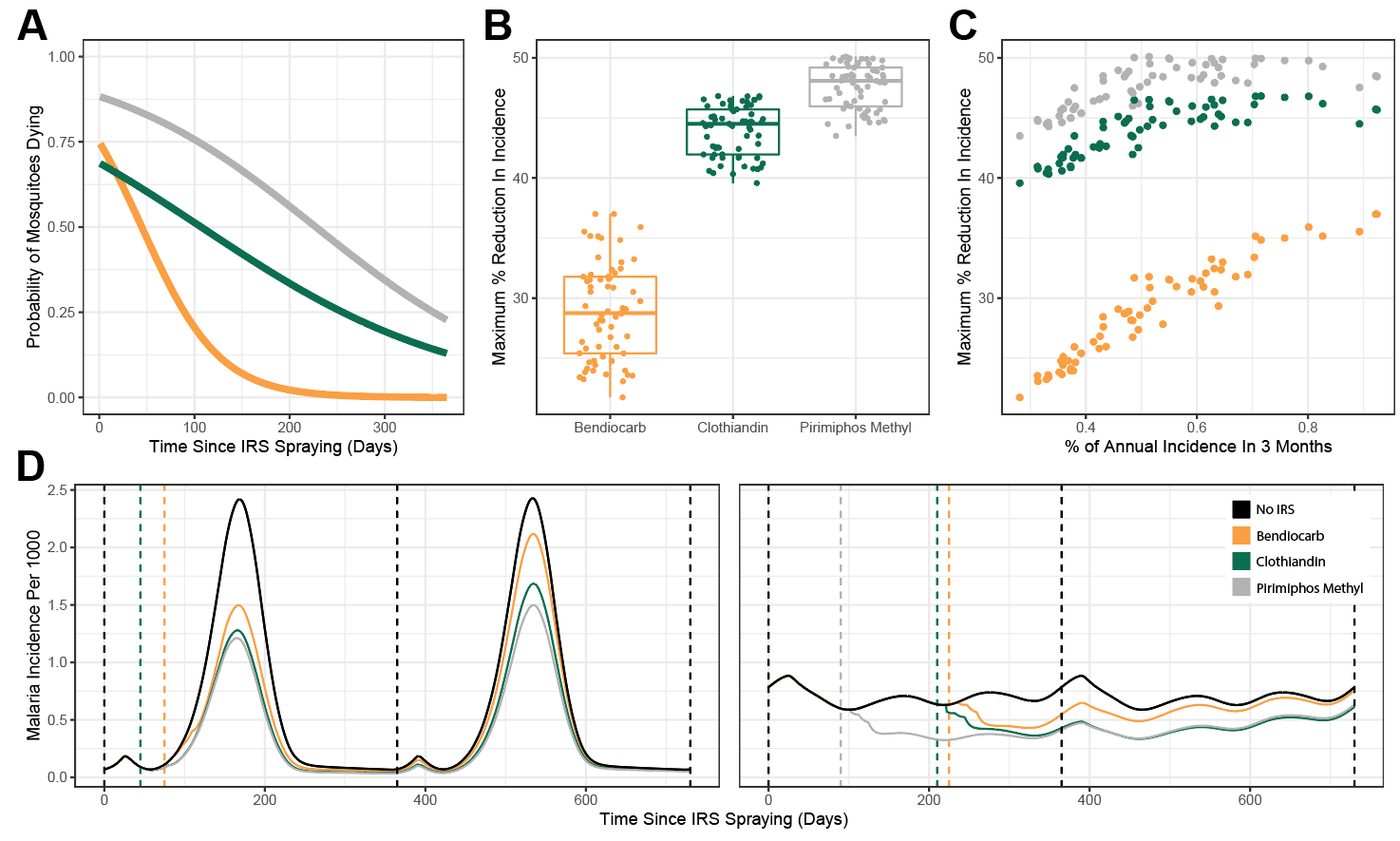
**Figure 2: Characterisation and Clustering to Identify Time-Series with Similar Temporal Properties. (A)** Results of principal components analysis (PCA) and k-means clustering for 2 clusters. Points indicate individual time-series, with point colour indicating cluster membership. Ellipsoids demarcate the 75th quantile of the density associated with each cluster. Principal components 1 and 2 are plotted, together explaining 69% of the total variation in temporal properties across the time-series. **(B)** Time-series belonging to each cluster. Pale lines represent individual time-series, brighter line the mean of all the time-series belonging to that cluster – in all cases vector density is normalised to sum to 1 over the course of the year, and time-standardised so that the highest vector density for each time-series is arbitrarily set to occur at month 7. **(C)** Boxplots of the percentage of annual total mosquito catch (top) and annual total rainfall (bottom) for each time-series. Rainfall data comes from the *CHIRPS* dataset37 and is specific to study location and time-period. Each point indicates an individual time-series.



**Figure 3: Random Forest Prediction of Temporal Cluster Membership.** A random forest-based classification modelling framework was used to predict membership of the temporal cluster (either Cluster 1 or Cluster 2, as defined in **Fig 2)** and explore the ecological factors underpinning variation in *Anopheles stephensi* seasonality. **(A)** Receiver-operator curve (ROC) for each of the 25 individual iterations of random forest model fitting carried out, with results for each displayed as grey lines. The mean AUC across these 25 iterations was 0.89. **(B)** Variable importance plot for the covariates included in the random forest model – bar height indicates the mean variable importance across the 25 individual iterations of random forest fitting, with error bars representing the 95% confidence interval. **(C)** Collated *Anopheles stephensi* time-series, disaggregated according to urbanicity and cluster membership. Cluster 1 and Cluster 2 time-series from rural locations are plotted separately; all time series for time-series carried out in urban locations (22 belonging to Cluster 1 and only 3 belonging to Cluster 2) are plotted together. Coloured line indicates the mean and ribbon indicates the 90% range spanned by the group of time series belonging to each displayed grouping.



**Figure 4: Predicting the Possible Seasonal Dynamics of *Anopheles stephensi* Across the Horn of Africa. (A)** Environmental covariates were collated across countries in the Horn of Africa where *Anopheles stephensi* has been found, and the fitted random forest classification model from **Fig 3** used to predict potential temporal dynamics. Map shows the probability of temporal dynamics belonging to Cluster 1, with pink corresponding to Cluster 1 dynamics being more likely than Cluster 2, black indicating Cluster 2 dynamics are more likely than those for Cluster 1, and white indicating both are equally likely. **(B)** Histogram of the percentage of total annual catch in any continuous 4-month period for each time-series, coloured according to the cluster the time-series belong to. **(C)** The probability of missing *Anopheles stephensi* in an entomological catch survey (y-axis) as a function of the number of consecutive months randomly sampled, assuming the start-month is picked at random. Coloured lines are the mean results across all time-series belonging to each cluster, with the shaded area indicating the range spanned by all time-series belonging to each cluster.



**Figure 5: Modelling the Public-Health Impact of Indoor Residual Spraying (IRS) and How This Is Impacted by *Anopheles stephensi* Seasonality. (A)** Probability of mosquitoes dying upon exposure to each IRS compound – yellow indicates bendiocarb, green indicates clothiandin and grey indicates pirimiphos methyl. **(B)** Percentage reduction in annual incidence (with optimal timing of IRS delivery), for each of the IRS compounds considered. Individual points correspond to specific time-series. **(C)** The relationship between percentage reduction in annual malaria incidences and the overall seasonality of malaria incidence in the setting (as modelled and implied by each *Anopheles stephensi* temporal profile). **(D)** Malaria incidence modelling results for a highly seasonal (left) and less seasonal (right) setting. Black lines indicate endemic dynamics in the absence of any IRS, coloured solid lines indicate incidence following a single IRS round (with timing of the round indicated by the coloured dashed lines).

**References**

1. Organization, W. H. & Others. World malaria report 2021. (2021).

2. Vos, T. *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1204–1222 (2020).

3. Malaria. https://www.who.int/news-room/fact-sheets/detail/malaria.

4. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).

5. Nations, U. Revision of world urbanization prospects. *United Nations: New York, NY, USA* **799**, (2018).

6. Doumbe-Belisse, P. *et al.* Urban malaria in sub-Saharan Africa: dynamic of the vectorial system and the entomological inoculation rate. *Malar. J.* **20**, 364 (2021).

7. Robert, V. *et al.* Malaria transmission in urban sub-Saharan Africa. *Am. J. Trop. Med. Hyg.* **68**, 169–176 (2003).

8. Trape, J. F. & Zoulani, A. Malaria and urbanization in central Africa: the example of Brazzaville. Part III: Relationships between urbanization and the intensity of malaria transmission. *Trans. R. Soc. Trop. Med. Hyg.* **81 Suppl 2**, 19–25 (1987).

9. Killeen, G. F., Govella, N. J., Mlacha, Y. P. & Chaki, P. P. Suppression of malaria vector densities and human infection prevalence associated with scale-up of mosquito-proofed housing in Dar es Salaam, Tanzania: re-analysis of an observational series of parasitological and entomological surveys. *Lancet Planet Health* **3**, e132–e143 (2019).

10. De Silva, P. M. & Marshall, J. M. Factors contributing to urban malaria transmission in sub-saharan Africa: a systematic review. *J. Trop. Med.* **2012**, 819563 (2012).

11. Awolola, T. S., Oduola, A. O., Obansa, J. B., Chukwurar, N. J. & Unyimadu, J. P. Anopheles gambiae s.s. breeding in polluted water bodies in urban Lagos, southwestern Nigeria. *J. Vector Borne Dis.* **44**, 241–244 (2007).

12. Kasili, S. *et al.* Entomological assessment of the potential for malaria transmission in Kibera slum of Nairobi, Kenya. *J. Vector Borne Dis.* **46**, 273–279 (2009).

13. Weiss, D. J. *et al.* Global maps of travel time to healthcare facilities. *Nat. Med.* **26**, 1835–1838 (2020).

14. Romeo-Aznar, V., Paul, R., Telle, O. & Pascual, M. Mosquito-borne transmission in urban landscapes: the missing link between vector abundance and human density. *Proc. Biol. Sci.* **285**, (2018).

15. Mourou, J.-R. *et al.* Malaria transmission in Libreville: results of a one year survey. *Malar. J.* **11**, 40 (2012).

16. Wang, S.-J. *et al.* Rapid Urban Malaria Appraisal (RUMA) IV: epidemiology of urban malaria in Cotonou (Benin). *Malar. J.* **5**, 45 (2006).

17. Klinkenberg, E., McCall, P., Wilson, M. D., Amerasinghe, F. P. & Donnelly, M. J. Impact of urban agriculture on malaria vectors in Accra, Ghana. *Malar. J.* **7**, 151 (2008).

18. Organization, W. H. & Others. Global technical strategy for malaria 2016-2030. 2021 update. *World Health Organization: Geneva* (2021).

19. Batra, C. P., Adak, T., Sharma, V. P. & Mittal, P. K. Impact of urbanization on bionomics of An. culicifacies and An. stephensi in Delhi. *Indian J. Malariol.* **38**, 61–75 (2001).

20. Thomas, S. *et al.* Overhead tank is the potential breeding habitat of Anopheles stephensi in an urban transmission setting of Chennai, India. *Malar. J.* **15**, 274 (2016).

21. Kumar, A. & Thavaselvam, D. Breeding habitats and their contribution to Anopheles stephensi in Panaji. *Indian J. Malariol.* **29**, 35–40 (1992).

22. Faulde, M. K., Rueda, L. M. & Khaireh, B. A. First record of the Asian malaria vector Anopheles stephensi and its possible role in the resurgence of malaria in Djibouti, Horn of Africa. *Acta Trop.* **139**, 39–43 (2014).

23. Balkew, M. *et al.* Geographical distribution of Anopheles stephensi in eastern Ethiopia. *Parasit. Vectors* **13**, 35 (2020).

24. Tadesse, F. G. *et al.* Anopheles stephensi Mosquitoes as Vectors of Plasmodium vivax and falciparum, Horn of Africa, 2019. *Emerg. Infect. Dis.* **27**, 603–607 (2021).

25. Ahmed, A. *et al.* Invasive malaria vector anopheles stephensi mosquitoes in Sudan, 2016-2018. *Emerg. Infect. Dis.* **27**, 2952–2954 (2021).

26. Ahmed, A., Khogali, R., Elnour, M.-A. B., Nakao, R. & Salim, B. Emergence of the invasive malaria vector Anopheles stephensi in Khartoum State, Central Sudan. *Parasit. Vectors* **14**, 511 (2021).

27. Sinka, M. E. *et al.* A new malaria vector in Africa: Predicting the expansion range of Anopheles stephensi and identifying the urban populations at risk. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 24900–24908 (2020).

28. Hamlet, A. *et al.* The potential impact of Anopheles stephensi establishment on the transmission of Plasmodium falciparum in Ethiopia and prospective control measures. *bioRxiv* (2021) doi:10.1101/2021.08.19.21262272.

29. Feachem, R. G. A. *et al.* Malaria eradication within a generation: ambitious, achievable, and necessary. *Lancet* **394**, 1056–1112 (2019).

30. ACCESS-SMC Partnership. Effectiveness of seasonal malaria chemoprevention at scale in west and central Africa: an observational study. *Lancet* **396**, 1829–1840 (2020).

31. Tukei, B. B., Beke, A. & Lamadrid-Figueroa, H. Assessing the effect of indoor residual spraying (IRS) on malaria morbidity in Northern Uganda: a before and after study. *Malar. J.* **16**, 4 (2017).

32. Whittaker, C. *et al.* The ecological structure of mosquito population seasonal dynamics. *bioRxiv* (2021) doi:10.1101/2021.01.09.21249456.

33. Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).

34. Fulcher, B. D., Little, M. A. & Jones, N. S. Highly comparative time-series analysis: the empirical structure of time series and their methods. *J. R. Soc. Interface* **10**, 20130048 (2013).

35. Cairns, M. *et al.* Estimating the potential public health impact of seasonal malaria chemoprevention in African children. *Nat. Commun.* **3**, 881 (2012).

36. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1‐km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

37. Wright, M. N. & Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *arXiv [stat.ML]* (2015).

38. Kuhn, M. & Wickham, H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. *Boston, MA, USA. [(accessed on 10 December 2020)]* (2020).

39. Griffin, J. T., Ferguson, N. M. & Ghani, A. C. Estimates of the changing age-burden of Plasmodium falciparum malaria disease in sub-Saharan Africa. *Nat. Commun.* **5**, 3136 (2014).

40. Challenger, J. D. *et al.* Predicting the public health impact of a malaria transmission-blocking vaccine. *Nat. Commun.* **12**, 1494 (2021).

41. Griffin, J. T. *et al.* Reducing Plasmodium falciparum malaria transmission in Africa: a model-based evaluation of intervention strategies. *PLoS Med.* **7**, e1000324 (2010).