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

Charles Whittaker1\*, Ellie Sherrard-Smith1, Peter Winskill1, Gina Cuomo-Dannenburg1, Patrick G.T. Walker1, Marianne Sinka2, Samuel Pironon3, 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

3Biodiversity Informatics and Spatial Analysis Department, Royal Botanic Gardens Kew, Richmond, Surrey, UK

4Vector Control Research Centre, Indira Nagar, Puducherry, India

5Section 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)

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

**Abstract**

Increasing urbanisation has contributed to reductions in malaria burden across sub-Saharan Africa with urban areas typically experiencing lower levels of transmission. Recent invasion and proliferation of the urban malaria vector *Anopheles stephensi* across the Horn of Africa threatens this progress. Understanding the vector’s seasonal dynamics in these settings represents a crucial input to its control, given that mosquito seasonality shapes the temporal profile of malaria risk, and in turn which interventions will be most effective. Here we collate longitudinal catch-data for the species from across South Asia and the Middle East to better understand these dynamics and explore their implications for malaria transmission in urban settings across the Horn of Africa where the disease is currently largely absent. Our analyses reveal pronounced variation across *Anopheles stephensi* populations in the extent and timing of seasonality, ranging from single, highly seasonal peaks to bimodality and near-perennial patterns of annual abundance. We show that this diversity is underpinned by different patterns of land use, as well as variation in seasonal patterns of rainfall and temperature across locations. Dynamics frequently differ between rural and urban settings, suggesting structural differences in how these environments shape patterns of vector abundance. Integrating these seasonal profiles with a model of malaria transmission, we show that the impact of IRS is highest in settings with the highest seasonality, but that crucially, achieving this impact requires an understanding of when *Anopheles stephensi* populations are likely to peak. Our results highlight a critical need to better understand the vector’s temporal dynamics across the Horn of Africa, but also a lack of studies from the region able to directly inform these questions, underscoring the need for longitudinal entomological monitoring of the vector across the region.

**Introduction**

There has been an estimated 40% reduction in the burden of malaria since 2000, an achievement underpinned predominantly by significant scale-up of control interventions1. Alongside this expansion of control efforts, increasing urbanisation of Africa’s populace (rising from 31% to 43% between 1990 and 2018, with >60% expected to live in urban areas by 20502) is also thought to have indirectly contributed to reductions in disease burden. Previous work has found significantly lower annual Entomological Inoculation Rates (EIR) in urban compared to rural settings3,4. This is thought to be underpinned by factors including differences in the quality of housing5,6, reduced availability and suitability of habitats for *Anopheline* breeding in urban settings7–9, better access to treatment10, and higher population densities leading to higher human to mosquito ratios (and reduced transmission)11. Whilst these trends are not always consistently identified (e.g. surveys where prevalence of malaria is higher in urban areas than surrounding locations12,13; or previous work highlighting that *Anopheles gambiae s.s.* can adapt to breeding in polluted water characteristic of urban environment14), 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 Malaria15.

This impact of increasing urbanization on disease burden is contingent on their remaining areas of comparatively low malaria transmission. This phenomenon is currently under threat by the invasion and establishment of the highly efficient urban malaria vector *Anopheles stephensi*. Found throughout South Asia, of the three known forms (“type”, “intermediate” and “mysorensis”), the mysorensis form is predominantly found in rural settings, and typically possesses a low vectorial capacity (due to its zoophilic behaviour16). By contrast, the type and intermediate forms represent highly efficient vectors capable of transmitting both *Plasmodium falciparum* and *Plasmodium vivax*, with their ability to proliferate in urban locations distinguishing this species from other malaria vectors in the region. This efficiency as an urban vector is thought to be underpinned by an increased tolerance for breeding in polluted water sources17, and a superior ability to utilise the manmade hydrological habitats present in urban settings18,19. The species was first identified in sub-Saharan Africa in Djibouti City in 201220 and has since been recorded in both Ethiopia21,22 and Sudan23,24, with recent work highlighting likely suitability for the species across some of the continent’s largest population centres comprising over 100 million people25. 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 Africa26 and the continent more generally27.

Experiences in Djibouti to date highlights the significant public-health threat this vector potentially poses. Despite this, substantial uncertainty remains regarding how its establishment might influence malaria control 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 chemoprevention28 or indoor-residual spraying29) depends on the timing of their delivery relative to seasonal peaks in transmission. A better understanding of *Anopheles stephensi*’s seasonal dynamics is therefore likely to have material consequences for the effective control of malaria across the Horn of Africa.

Substantial uncertainty remains regarding *Anopheles stephensi*’s seasonal dynamics; studies carrying out longitudinal catches are present in in the literature, but 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 across South Asia and the Middle East in order to better understand these dynamics. 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 highlights how this variation will influence the efficacy of malaria control efforts in parts of the Horn of Africa where the disease is currently (or has previously been) largely absent and underscores the need for rapid scaleup of entomological monitoring across the region.

**Methods**

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

We collated references from published systematic reviews of literature relating to *Anopheles stephensi*25,30, and updated these previous searches by searching *Web of Science* and *PubMed* from January 2017 to September 2020. 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, with sufficient information to geolocate the catch site, and where at least 25 *Anopheles stephensi* had been caught over the study period. A total of 36 references were collated containing 65 time-series from surveys carried 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 to Time Series Data**

Per previous work30, we fitted the following Gaussian Process-based model to smooth these noisy mosquito count time-series, using a Negative Binomial likelihood to account for overdispersion:

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 . A periodic kernel function was used to capture the seasonally/annually repeating patterns often observed in mosquito populations, specified by the hyperparameters , and . 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 covariance magnitude, and  represents a lengthscale parameter further constraining the extent to which two values separated by a given time co-vary. (x) are function evaluations at times , are the observed mosquito counts indexed by timepoint . Model fitting was carried out within a Bayesian framework, using the probabilistic programming language STAN31.

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

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 gaussian process fitting, the relative entropy between the time-series and a uniform distribution, the proportion of months in each smoothed time-series where the catch was greater than 1.65x the mean, the difference in timing between peaks in vector density and rainfall, the proportion of the total catch that was found in any 4 month period (in-keeping with previous definitions of malaria seasonality32), 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. See **Supplementary Information** for further details. From this, we 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 Cluster Membership**

For each time-series, we extracted a suite of environmental variables derived from satellite data describing the ecological structure of the location the survey was carried out in, including the *BioClimatic* variables33 and measures of landcover and urbanicity (see **Supplementary Table X** for complete list). These environmental variables were then used as predictors within a Random-Forest based classification framework aimed at predicting the cluster membership (i.e. results of the k-means algorithm described above)**.** These models were fitted using the software package *Ranger*34, implemented in the R *tidymodels* framework35, with 6-fold cross-validation utilised to optimise hyperparameter combinations. Results are based on averaging the results of 25 separate iterations of cross-validation and model fitting and predictions made using out-of-bag model estimates. There were significant imbalances in class size across the clusters and so we carried out upsampling using the SMOTE36 algorithm (though results are also presented without upsampling). Further details can be found in the **Supplementary Information**.

**Modelling *Anopheles stephensi*-Driven Malaria Dynamics and Control**

We integrated the *Anopheles stephensi* time-series into a deterministic model of *Plasmodium falciparum* malaria transmission and disease37–39 to explore the implications of *Anopheles stephensi* seasonality on control of the disease; with a focus on settings in the Horn of Africa where malaria is currently largely absent. Specifically, we use the modelling framework to understand how variation in seasonality of the mosquito might influence the impact of indoor residual spraying (IRS), a key vector control intervention. Full details of the model and the scenarios explored can be found in **Supplementary Information** but briefly, the model considers both humans and mosquitoes explicitly, and incorporates a range of features specific to malaria, including heterogeneity in transmission and exposure, age-dependent mosquito biting rates, and the acquisition of natural immunity following exposure.

**Results**

**Diversity In Temporal Dynamics Across the Collated *Anopheles stephensi* Time-Series:** A total of 65 time-series from studies 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 example time-series from each country, **Supp Fig.1** for all time-series). Substantial variation in temporal dynamics was observed across in terms of the degree and timing of seasonality – ranging from highly seasonal dynamics with a single seasonal peak (e.g. Afghanistan example in **Fig.1B**) to more perennial patterns of abundance (Pakistan example in **Fig.1B**) and bimodal population dynamics with two peaks observed during a single year (Iran example in **Fig.1B**).

**Statistical Characterisation and Clustering of Temporal Properties Highlights Distinct Archetypes:** An array of summary statistics were calculated for each time series to characterise their temporal properties (see **Supp Fig.2**), followed by k-means clustering of the results to cluster the time-series into discrete groups sharing similar temporal patterns. Our results highlight two distinct clusters, with each cluster characterised by distinct temporal patterns and degree of their seasonality (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 Cluster 1 and Cluster 2 time-series in terms of rainfall seasonality **(Fig.2C**, bottom panel, p=0.59).

**Random-Forest Modelling of Seasonal Dynamics Highlights Urbanicity As a Key Factor:** We fitted a random forest-based classification framework to 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 to generate balanced classes. Across the 25 iterations of random forest model fitting, the mean AUC was 0.89 (indicating good predictive performance, **Fig.3A**) and on average, the model was able to correctly classify Cluster 1 and Cluster 2 time-series equally well (83% and 85% of the time respectively). Population per km2, temperature and rainfall seasonality, and a number of measures of landcover (specifically LC30 which corresponds to mosaic cropland/natural vegetation, and LC20 which corresponds to irrigated or post-flooding cropland) were all highly predictive of Cluster membership (**Fig.3B)**. Time-series from surveys in locations with lower population density were more likely to belong to Cluster 2 (i.e. less seasonal), as were areas with high values of LC20 (i.e. land predominantly occupied by irrigated or post-flooding cropland). By contrast, areas with high values of LC10 and LC30 (i.e. substantial fraction of land covered either rainfed cropland or a mosaic of cropland and natural vegetation) were more likely to belong to Cluster 1 (i.e. more seasonal), as were areas in which rainfall was strongly seasonal (see **Supp Fig.4** for partial dependence covariate plots). Examining the association between rurality/urbanicity (as defined by the authors of each study) and temporal dynamics (**Fig.3C**), we observed a near-significant association with rurality/urbanicity and cluster membership (chi-squared test, p=0.07). 88% (n=22/25) time-series from urban settings were assigned to Cluster 1, compared to only 65% (n=24/37) from rural settings. Only 12% (3/25) time-series from urban settings were assigned to Cluster 1, compared to only 35% (n=13/37) from rural settings.

Model predictive performance and variable importance rankings were similar when no up-sampling was applied to the dataset (AUC=0.81, see **Supp Fig.5**), though average predictive accuracy on Cluster 2 (50%) was lower than for Cluster 1 time-series (94%). Model performance and variable importance ordering also remained similar 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**). 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=0.47), the mean catch size did differ (t-test, p=0.025). We therefore carried out a sensitivity analysis including study average monthly catch as a covariate in the random forest model (**Supp Fig.8)** – predictive performance and variable importance remained similar. The ability of the model to predict seasonality (percentage of vector catch in any 3-month period) was more modest, although estimates were positively correlated (r = 0.43, see **Supp Fig.9**) and measures of landcover including LC10, LC20 and LC30, as well as temperature and rainfall seasonality were all strongly associated predictors of seasonality, as in the classification modelling.

**Exploring Possible Seasonal Dynamics and Surveillance of *Anopheles stephensi* Across the Horn of Africa:** We collated the same satellite-derived environmental covariates for countries across the Horn of Africa and integrated them with the trained random forest classification model to predict potential temporal dynamics (as denoted by Cluster membership) of *Anopheles stephensi* across the 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. Specifically, what the probability of missing *Anopheles stephensi* in entomological surveys 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). 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**). 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 *Anopheles stephensi* and erroneously concluding it is absent.

**Modelling the Impact of *Anopheles stephensi* Seasonality On Vector Control Measures:** Integrating the collated temporal profiles of *Anopheles stephensi* abundance that had been collated with a model of malaria transmission, we explored 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 commonly used throughout the Horn of Africa), with a range of different functional half-lives (**Fig.5A**) and model the impact of a single annual round of IRS. Across the *Anopheles stephensi* temporal profiles considered, pirimiphos methyl consistently had the highest impact, reducing annual malaria incidence by 47.6% on average in the 12 months following spraying, compared to 43.9% and 28.9% on average for clothiandin and bendiocarb (**Fig.5B**). Overall impact was highly 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 21.7% in the lowest seasonality setting, through to as high as 37.0% in the most seasonal setting, a 1.7 fold difference. We assume in all instances optimal timing of IRS delivery relative to seasonal peaks, which in turn requires sufficient understanding of the vector’s dynamics and when peaks in density are going to occur.

**Discussion**

Invasion and establishment of *Anopheles stephensi* across the Horn of Africa represents an urgent threat to malaria control in the region. Understanding the temporal profile of malaria risk this vector might produce, particularly in urban settings across the region where malaria has historically been largely absent or only minimally present, will represent a crucial input to effective deployment of control interventions aimed at mitigating the potential impact of *Anopheles stephensi*’s arrival. Collating data from across South Asia and the Middle East, we identify extensive diversity across *Anopheles stephensi* populations in the extent and nature of their seasonal dynamics. This variation is associated with a wide array of ecological factors, including patterns of land-use and temporal fluctuations in rainfall and temperature. Perhaps most crucially, we find evidence of distinct temporal dynamics across rural and urban settings – this variation has material consequences for the efficacy of interventions aimed at controlling the threat this vector poses to urban areas across the Horn of Africa.

Our analyses have identified several ecological factors associated with *Anopheles stephensi* seasonal dynamics. Patterns of land-use (e.g. whether the land is rain-fed or irrigated), as well as the seasonality of temperature and rainfall were identified as key drivers of the extent and nature of *Anopheles stephensi* seasonality. This is consistent with previous work that has identified features of the local hydrological environment (particularly whether it is predominantly rain-fed or static, perennially available bodies of water) as a key factor shaping the seasonality of diverse anopheline species across India30. Similarly, previous work has identified temperature as a key driver of mosquito population dynamics, due to its impact on an array of mosquito life-history traits including biting rate, lifespan and fecundity (amongst several others)40,41. Our results therefore highlight the importance of considering both the hydrological environment (including patterns of land use and their interaction with rainfall and other hydrological features), as well seasonal fluctuations in temperature when trying to understand seasonal patterns of mosquito abundance.

Perhaps most notably, our analyses identified population per km2 as a key predictor of cluster membership, with high population density being strongly associated with Cluster 1 dynamics (i.e. more seasonal patterns of abundance), a finding that was also observed when stratifying surveys according to whether they had been carried out in urban or rural settings. This potential disparity in temporal dynamics across rural and urban settings will likely have implications for the public health impact of different control interventions. Our results suggest that urban *Anopheles stephensi* populations are likely to display seasonal dynamics, supporting the utility of interventions like IRS in these settings – the same is not necessarily true in rural settings, where a range of seasonal profiles including more perennial patterns of abundance were observed. Implementing these measures and achieving sufficient IRS (and ITN) coverage in urban settings is likely to prove challenging, given the historical absence of large-scale vector control campaigns. If these barriers can be surmounted, such measures are likely to be impactful however, given the results presented here as well as previous modelling work that has identified low altitude urban areas with minimal levels of pre-existing transmission as the areas likely to experience the largest increase in disease burden (due to high population densities, absence of existing vector control and minimal human population immunity26).

Our work has also highlighted the paucity of currently available entomological data from the Horn of Africa region, particularly the absence of longitudinal studies surveying the same site over multiple months and the risk this poses for erroneously concluding *Anopheles stephensi*’s absence. Indeed, our analyses demonstrate that a lack of a detailed understanding of the vector’s dynamics (precluding targeting of entomological surveys to periods of highest vector density) risks missing *Anopheles stephensi*’s presence. Longitudinal surveys elucidating these dynamics would therefore be useful in enabling subsequent refinement and timing of shorter surveys aimed at detecting presence only, even before considering the additional information these surveys would provide on temporal dynamics that can facilitate the effective targeting and timing of control interventions such as IRS. Such surveillance is vital – unnoticed proliferation of an invasive anopheline vector has previously been observed with *Anopheles arabiensis* in North-Eastern Brazil, where extensive spread prior to detection subsequently led to a significant malaria epidemic across the region42.

There are a number of important limitations to the work presented here. Firstly, we do not formally include considerations of insecticide resistance in our model of malaria transmission. Insecticide resistance is well-documented for *Anopheles stephensi*, including populations across Afghanistan44, Iran45, Pakistan46 and India47. Recent populations assayed in Ethiopia showed resistance to insecticides of all four major insecticide classes48,49, suggesting that pyrethroid-only ITNs and IRS (both already in use across the country) might have limited impact at controlling malaria transmitted by *Anopheles stephensi*. Relatedly, we do not consider uncertainty in *Anopheles stephensi* bionomic properties (such as timing of biting, or whether resting occurs predominantly indoors or outdoors), which will further modulate the impact of interventions such as IRS whose killing is mediated primarily through indoors resting following feeding. Significant variation in *Anopheles stephensi*’s bionomic properties across settings has previously been identified50, including a propensity for crepuscular biting and resting outside of houses compared to African anopheline species25. Previous work has identified these factors as key drivers of malaria intervention impact51, and more specifically, the impact of control interventions aimed at mitigating the public health threat this vector poses26. Whilst the aim of the work here is not to provide specific estimates of intervention impact, and instead highlight how seasonality modulates impact and underscore an understanding of seasonal dynamics as a crucial input to optimising control interventions, these limitations highlight the urgent need for a more detailed characterisation of *Anopheles stephensi* in the settings across the Horn of Africa where it is now present, in order to more precisely quantify how its bionomic properties and insecticide resistance profile might erode intervention impact.

We also assume that the inferred relationships linking environmental features to temporal dynamics will translate from settings in South Asia and the Middle East to the Horn of Africa. Indeed, our results highlight significant plasticity and variation in *Anopheles stephensi*’s seasonal dynamics depending on the setting, and therefore the extent to which our results will extrapolate to new settings beyond *Anopheles stephensi*’s historical rangeremains unclear. Relatedly, due to the limited amount of data available and the wide geographical range over which the collated studies were conducted, we cannot rule out possible spatial confounding in shaping the associations inferred – though exploratory analysis of the distribution of locations stratified by rural/urban status and cluster assignment did not reveal obvious patterns of spatial confounding (**Supp Fig.10**). One factor that we were unable to consider is the possibility of variation between *Anopheles stephensi* forms (“type”, “intermediate” and “mysorensis”) in their geographical range and temporal dynamics. Identification of the particular *Anopheles stephensi* form is challenging, often requiring close visual examination52 or molecular methods53. Availability of this data was limited, and we therefore lack the ability to disaggregate time-series by the specific *Anopheles stephensi* form caught. It therefore remains unclear whether the variation in temporal dynamics observed across the time-series is due to inherently flexible dynamics that are shaped by distinct environmental drivers, or instead maybe arising from different *Anopheles stephensi* forms.

Despite these limitations, our work highlights significant variation in temporal dynamics across surveyed *Anopheles stephensi* populations; variation that is shaped by distinct ecological factors, can differ pronouncedly between urban and rural settings, and which has material consequences for the potential effectiveness of vector control interventions. Our work also underscores the need to better understand the vector’s dynamics in settings where it has newly established. Indeed, the trajectory of *Anopheles stephensi*’s establishment and subsequent dynamics in Horn of Africa remains deeply unclear and the paucity of published studies from the region underscores the need for studies longitudinally surveying locations where *Anopheles stephensi* has recently arrived. This will be vital to understanding the patterns of seasonal variation the vector displays, a crucial and operationally relevant input for optimising the delivery of malaria control interventions aiming to mitigate the impact of this invasive vector.

**Data and Code Availability**

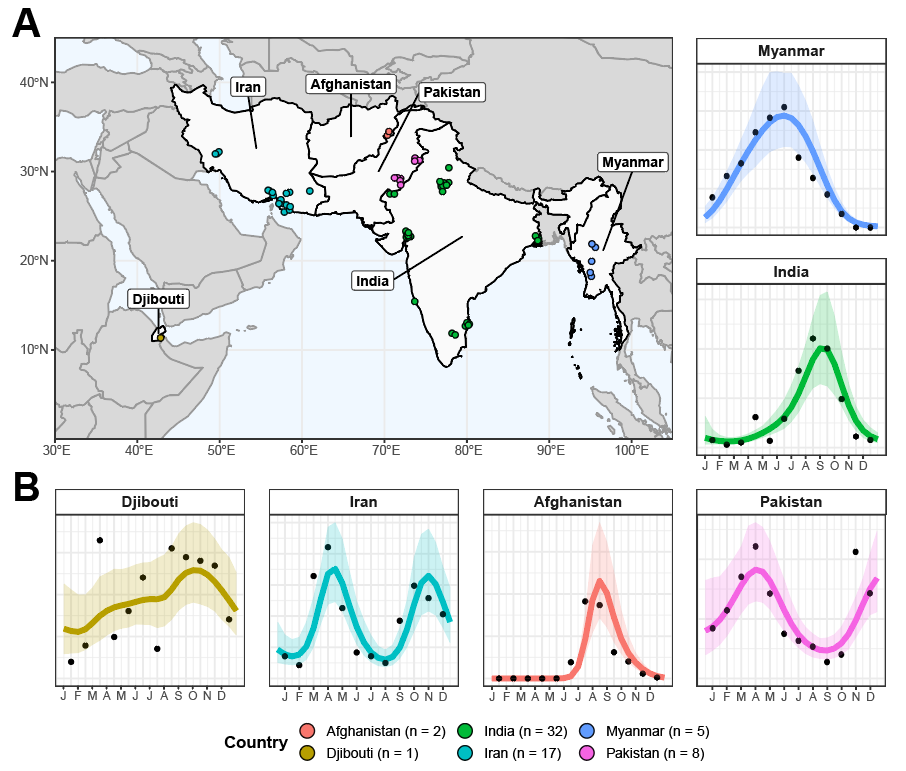
All data collated as part of this study and the code required to reproduce these analyses can be found at the following link: <https://github.com/cwhittaker1000/stephenseasonality>. A copy of the data is also available with the manuscript supplementary information. A static, archived version of the repository can be found via the following DOI: \_\_\_\_\_\_\_ and associated GitHub Release: \_\_\_\_\_\_\_\_.

**Acknowledgements and Funding Sources**

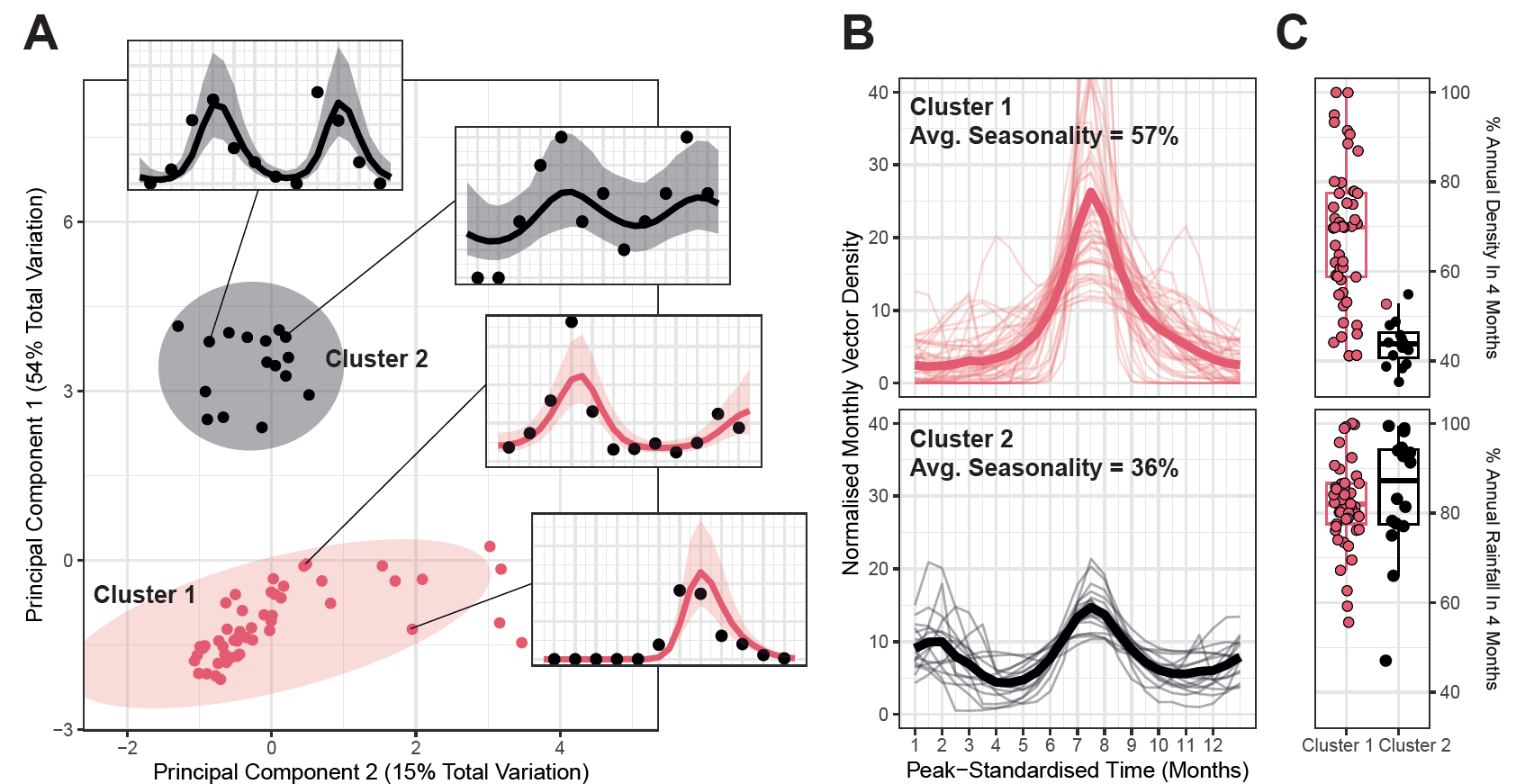
C.W. is supported by a Medical Research Council Doctoral Training Partnership PhD Studentship. SB & AG both acknowledge grant support from the Bill and Melinda Gates Foundation. *-more here-*

**Author Contributions**

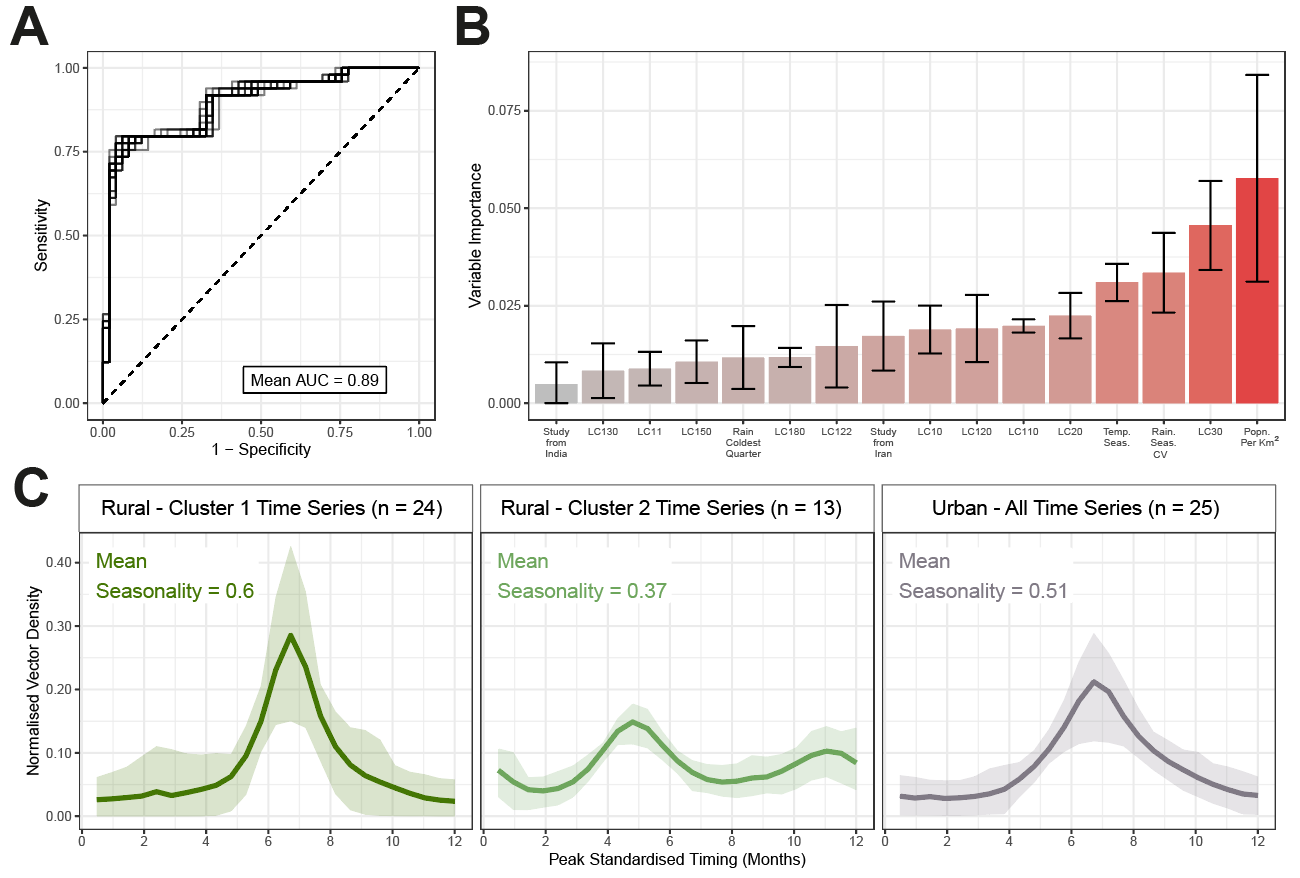
CW and AH conceived the study. TC, ESS and SB contributed to the design of the study. CW carried out the systematic review. CW and SB developed the underlying statistical framework, with input on the analyses from AH, AG, TC, GCD, PGTW, PW and ESS. CW wrote the first draft manuscript, with all authors providing feedback and suggestions during manuscript drafting. All authors approved the final version of the manuscript.



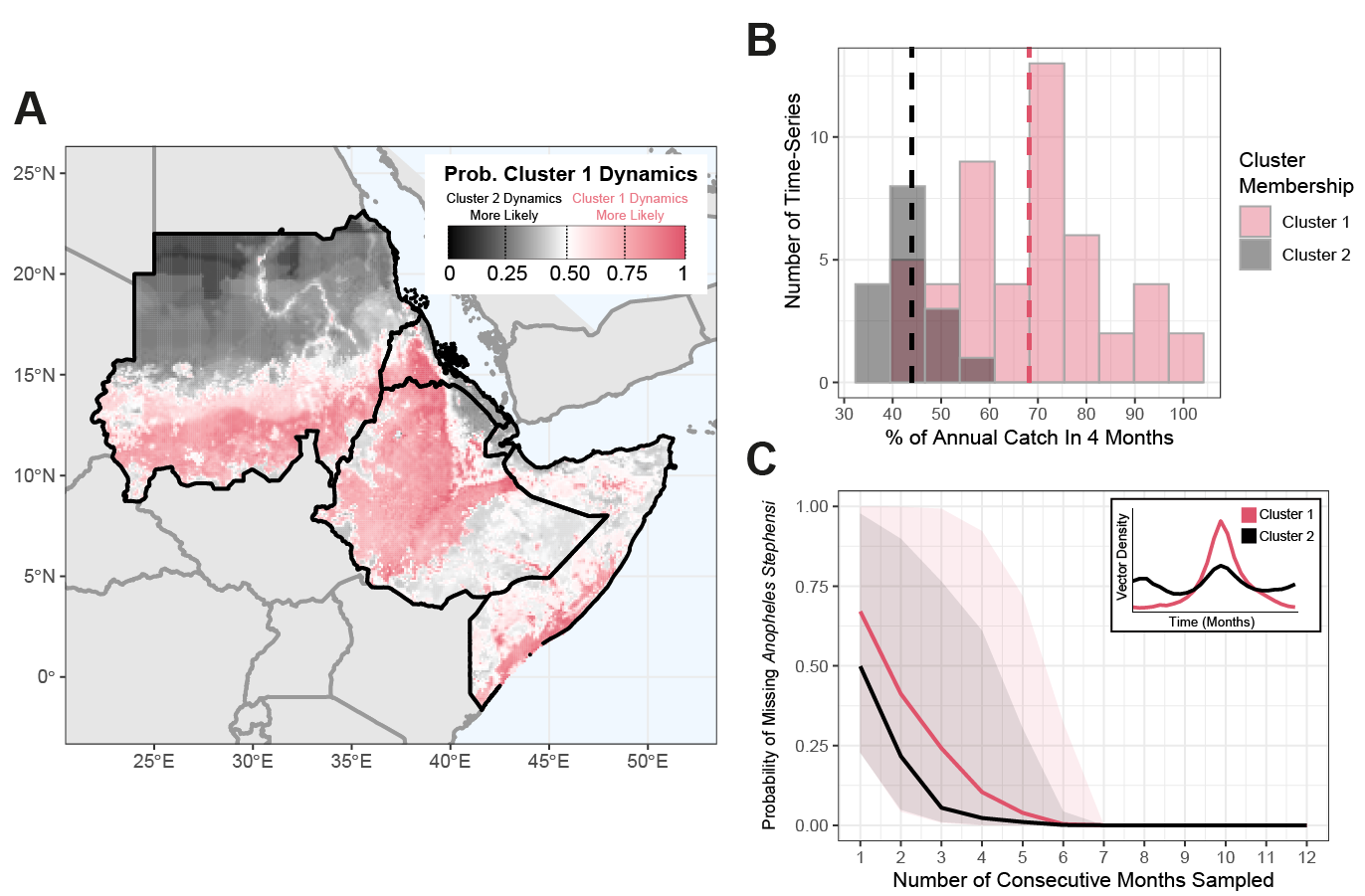
**Figure 1: Sources and Locations of *Anopheles stephensi* Time-Series Data and Examples for Each Country. (A)** Map of the geographical range over which collated time-series had been carried out in, with countries where studies had been carried out in highlighted in light grey, and the locations of 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).



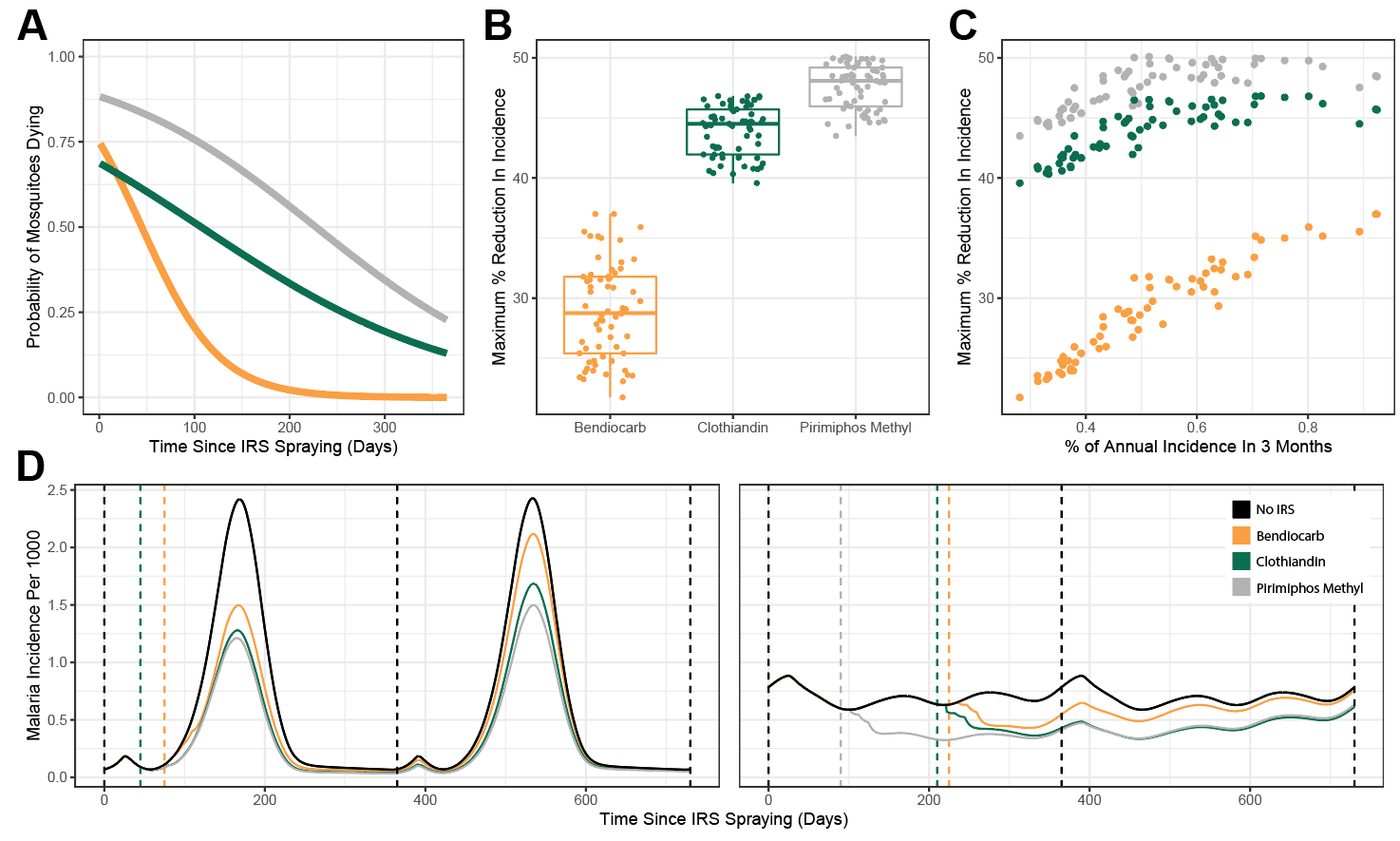
**Figure 2: Characterisation and Clustering to Identify Time-Series with Similar Temporal Properties. (A)** Results of principal components analysis (PCA) and k-means clustering for 2 clusters. Points indicate individual time-series, with point colour indicating cluster membership. Ellipsoids demarcate 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. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).

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

3. 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).

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

5. 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).

6. 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).

7. 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).

8. 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).

9. 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).

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

11. 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).

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

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

14. 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).

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

16. Subbarao, S. K., Vasantha, K., Adak, T., Sharma, V. P. & Curtis, C. F. Egg-float ridge number in Anopheles stephensi: ecological variation and genetic analysis. *Med. Vet. Entomol.* **1**, 265–271 (1987).

17. 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).

18. 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).

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

20. 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).

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

22. 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).

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

24. 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).

25. 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).

26. 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.

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

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

29. 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).

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

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

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

33. 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).

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

35. 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).

36. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).

37. 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).

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

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

40. Beck-Johnson, L. M. *et al.* The importance of temperature fluctuations in understanding mosquito population dynamics and malaria risk. *R Soc Open Sci* **4**, 160969 (2017).

41. Mordecai, E. A. *et al.* Thermal biology of mosquito-borne disease. *Ecol. Lett.* **22**, 1690–1708 (2019).

42. Killeen, G. F., Fillinger, U., Kiche, I., Gouagna, L. C. & Knols, B. G. J. Eradication of Anopheles gambiae from Brazil: lessons for malaria control in Africa? *Lancet Infect. Dis.* **2**, 618–627 (2002).

43. Enayati, A., Hanafi-Bojd, A. A., Sedaghat, M. M., Zaim, M. & Hemingway, J. Evolution of insecticide resistance and its mechanisms in Anopheles stephensi in the WHO Eastern Mediterranean Region. *Malar. J.* **19**, 258 (2020).

44. Safi, N. H. Z. *et al.* Status of insecticide resistance and its biochemical and molecular mechanisms in Anopheles stephensi (Diptera: Culicidae) from Afghanistan. *Malar. J.* **18**, 249 (2019).

45. Vatandoost, H. & Hanafi-Bojd, A. A. Indication of pyrethroid resistance in the main malaria vector, Anopheles stephensi from Iran. *Asian Pac. J. Trop. Med.* **5**, 722–726 (2012).

46. Ali Khan, H. A., Akram, W. & Lee, S. Resistance to Selected Pyrethroid Insecticides in the Malaria Mosquito, Anopheles stephensi (Diptera: Culicidae), From Punjab, Pakistan. *J. Med. Entomol.* **55**, 735–738 (2018).

47. Tiwari, S., Ghosh, S. K., Ojha, V. P., Dash, A. P. & Raghavendra, K. Reduced susceptibility to selected synthetic pyrethroids in urban malaria vector Anopheles stephensi: a case study in Mangalore city, South India. *Malar. J.* **9**, 179 (2010).

48. Yared, S. *et al.* Insecticide resistance in Anopheles stephensi in Somali Region, eastern Ethiopia. *Malar. J.* **19**, 180 (2020).

49. Balkew, M. *et al.* An update on the distribution, bionomics, and insecticide susceptibility of Anopheles stephensi in Ethiopia, 2018-2020. *Malar. J.* **20**, 263 (2021).

50. Massey, N. C. *et al.* A global bionomic database for the dominant vectors of human malaria. *Sci. Data* **3**, 160014 (2016).

51. Sherrard-Smith, E. *et al.* Mosquito feeding behavior and how it influences residual malaria transmission across Africa. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 15086–15095 (2019).

52. Nagpal, B. N., Srivastava, A., Kalra, N. L. & Subbarao, S. K. Spiracular indices in Anopheles stephensi: a taxonomic tool to identify ecological variants. *J. Med. Entomol.* **40**, 747–749 (2003).

53. Chavshin, A. R. *et al.* Molecular characterization, biological forms and sporozoite rate of Anopheles stephensi in southern Iran. *Asian Pac. J. Trop. Biomed.* **4**, 47–51 (2014).