**The Ecological Structure of Mosquito Population Dynamics: Insights from India, Consequences for Malaria Control**

Charles Whittaker1\*, Peter Winskill1, Marianne Sinka2, Samuel Pironon3, Claire Massey4, Peter W Gething5, Daniel J Weiss5, Michele Nguyen6, Ashwani Kumar7, Azra Ghani1 & Samir Bhatt1

1Department of Infectious Disease Epidemiology, Imperial College, London, UK

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

3Royal Botanic Gardens Kew, Richmond, West Sussex, UK

4Big Data Institute, University of Oxford, Old Road Campus, Oxford, UK.

5Malaria Atlas Project, Telethon Kids Institute, Perth, Australia

6Asian School of the Environment, Nanyang Technological University, Singapore

7National Institute of Malaria Research, Field Unit, Campal, Goa, India

\*Corresponding Author: [charles.whittaker16@imperial.ac.uk](mailto:charles.whittaker16@imperial.ac.uk)

**Significance**

Effective planning and control of malaria requires an understanding of the underlying mosquito population dynamics that determine the temporal profile of malaria risk. Here, we collate a database of monthly mosquito catch data spanning 40 years and 117 unique locations across India to explore the factors shaping these dynamics. Our analyses reveal pronounced heterogeneity in mosquito population dynamics, both within (across different locations) and across (in the same location) species: this heterogeneity is driven by a complex interplay between species-specific factors and the ecological structure of the local environment. Despite this variation, the temporal patterns of mosquito abundance across these different locations can be categorised into a small number of clusters, each characterised by distinct temporal properties. Based on these results, we create a tool to predict mosquito population seasonality in a given location, to inform the planning and timing of control efforts.

**Abstract**

Understanding the temporal dynamics (including the start, duration and end) of malaria transmission is key to optimising various control strategies, enabling interventions to be deployed at times when they can have the most impact. This temporal profile of malaria risk is intimately related to the dynamics of the mosquito populations underlying transmission. However, many outstanding questions remain surrounding these dynamics, including the specific drivers and their dependence on the ecological structure of a setting. Here we collate mosquito time-series catch data from across India in order to better understand these dynamics and the factors shaping them. Our analyses reveal pronounced heterogeneity in mosquito population dynamics, both within (across different locations) and between (in the same location) species. Despite this variation, we show that these time-series can be clustered into a small number of categories characterised by distinct temporal properties. Exploration of these categories highlights that a complex interplay of species-specific factors and the ecological structure of the local environment shape the temporal dynamics (including timing and extent of seasonality) of mosquito populations. The results of these analyses are then integrated with spatial predictions of species presence/absence in order to generate predictive maps of mosquito population seasonality across India, to inform the planning and timing of malaria control efforts.

**Background**

With an estimated 200 million cases and over 600,000 deaths in 20171, malaria represents one of the most serious infectious diseases globally2. Nineteen countries in sub-Saharan Africa along with India account for almost 85% of the global burden3, with *Plasmodium falciparum* most prevalent in African settings, and India alone accounting for almost 50% of the global *Plasmodium vivax* burden4. Transmission occurs via mosquito vectors belonging to the *Anopheles* genus – these vectors are heterogeneously distributed across the globe5,6, a feature that results in marked differences in the transmission dynamics of malaria across different ecological contexts.

Much work has focussed on characterising the global spatial distribution (presence/absence) of these malaria vectors7,8. This work represents a vital input to surveillance and control programmes aimed at mitigating the impacts of vector borne diseases worldwide. By contrast, less attention has been paid to understanding the temporal patterns of vector abundance, and how these dynamics are shaped by the local environment. Mosquito populations are highly temporally dynamic, exhibiting substantial annual fluctuations in size9,10 that drive the temporal profile of malaria risk. Understanding the determinants of these dynamics is important given that the efficacy of many malaria control interventions (such as seasonal malaria chemoprevention11,12 and indoor-residual spraying13,14) depends on the timing of their delivery in relation to seasonal peaks in malaria risk. Effective utilisation of these interventions will be vital for achieving the goals of the World Health Organisation’s “High Burden, High Impact” strategy, which aims to substantially reduce malaria in India and the ten African nations with the highest global burden15.

Rainfall is considered a key determinant of mosquito temporal dynamics due to the requirement of an aquatic habitat for the early life cycle stages, with many species displaying a preference for transient, rain-fed pools of water in which to breed16. However, whilst a close relationship has been observed between rainfall occurrence and peaks in mosquito populations (e.g. *Anopheles gambiae*17–19 for African settings and *Anopheles dirus*20 across India and south-East Asia) and malaria cases21, the extent of this relationship is variable. *Anopheles funestus* populations frequently lack marked seasonal fluctuations in population abundance22,23, whilst *Anopheles fluviatilis* populations have been reported to peak during the dry season in certain contexts25,26. Further compounding this complexity, not all variation in population dynamics appears to be species-specific – *Anophleles annularis* population dynamics range from near-perennial through to intensely seasonal10,27,28, bringing into question how generalisable relationships between rainfall and population dynamics are.

Ecological factors other than rainfall are also likely to contribute to variation in mosquito population dynamics. Temperature has a marked influence on many mosquito traits (including larval development29, biting rates and mortality rates30), with recent field-based work suggesting that considerations of both rainfall and temperature are necessary to understand seasonal patterns of malaria incidence31. However, these analyses have been restricted to a small number of settings across sub-Saharan Africa; leaving the influence of dynamic temperature regimen on mosquito population dynamicsin other ecological settings largely unexplored.

These results highlight a number of outstanding questions surrounding the environmental factors and species-specific properties that shape and structure mosquito population dynamics. Using India as a case study, we collate a dataset of temporally disaggregated mosquito catch data from across the country to better understand the extent and drivers of variation in mosquito population dynamics. We use these data to characterise the temporal patterns displayed by different mosquito species and identify pronounced heterogeneity in the extent and nature of seasonal dynamics, both between species and across different locations. Exploring the drivers of these dynamics highlights the critical interaction between abiotic and species-specific factors. It also underscores the importance of considering both species composition and ecological structure when implementing malaria control interventions.

**Results**

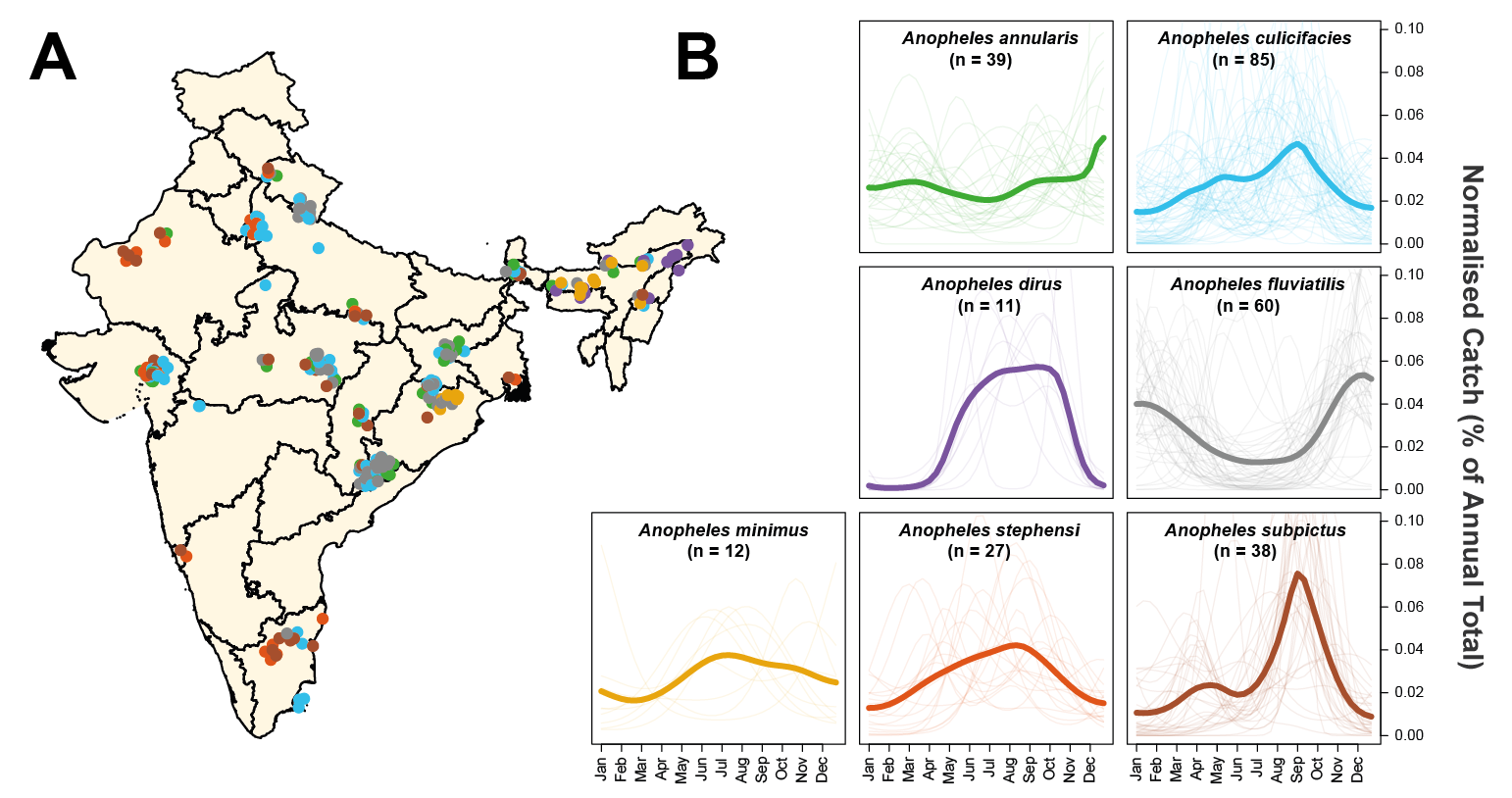
**Substantial Diversity in Mosquito Population Dynamics Within and Between Species:** A total of 272 time-series from 117 locations across India were identified through the systematic review **(Fig.1A)**. These time-series were then smoothed using a Negative Binomial Gaussian Process based framework. Substantial heterogeneity was observed between different species in their temporal dynamics over the course of a year **(Fig.1B)**. *Anopheles dirus* populations tended to peak during the monsoon period (typically June to September), whilst many *Anopheles fluviatilis* populations peaked between November and February (the dry season across most of India), reaching their lowest density during the monsoon. By contrast, a number of *Anopheles annularis* time-series demonstrated perennial patterns of abundance. We also observed extensive variation in temporal dynamics within species. Across the 85 time-series collated for *Anopheles culicifacies*, populations varied substantially in both the extent and timing of their seasonal peaks; this ranged from sharp peaks in the monsoon season to perennial characteristics more similar to those observed for *Anopheles annularis*.

**Characterisation of Mosquito Catch Time-Series Properties Reveals Distinct Temporal Patterns:** We next asked whether this variation could be delineated into discrete clusters of time-series with distinct temporal patterns based on their statistical characteristics. Using an array of summary statistics to characterise temporal properties followed by k-means clustering, we identified 4 clusters of time-series with similar temporal patterns (**Fig.2)**. These clusters included time-series peaking during the monsoon season (Cluster 1), displaying bimodal characteristics (Cluster 2), peaking in the dry season (Cluster 3) or displaying perennial patterns of abundance (Cluster 4). These variable seasonal dynamics across different locations were not due to differences in the timing and extent of rainfall across India. We observed a high positive cross-correlation product between rainfall and mosquito density for Cluster 1 (r=0.52), but a negative correlation for Cluster 3 (r=-0.41) and low correlation for Clusters 2 and 4 (r=-0.08 and 0.03 respectively, **Supplementary Fig.4**). This suggests that the observed patterns in represent genuine differences between species and across locations in how mosquito populations respond to rainfall. Analysing the proportion of each mosquito species’ time-series belonging to a particular cluster (**Fig.3**) revealed that for a number of species the majority of time-series belonged to a single cluster. *Anopheles dirus* time-series were almost exclusively within Cluster 1 (monsoon season peaking) whilst *Anopheles fluviatilis* time-series were almost exclusively in Cluster 3 (dry season peaking). This is suggestive of strong species-specific tendencies linking these two species to particular temporal dynamics. In contrast, other species showed extensive plasticity in the temporal patterns observed, with *Anopheles culicifacies* time-series appearing across all four clusters - indicating the capacity for these species to adopt a wide and diverse array of temporal dynamics depending on the particular ecological setting.

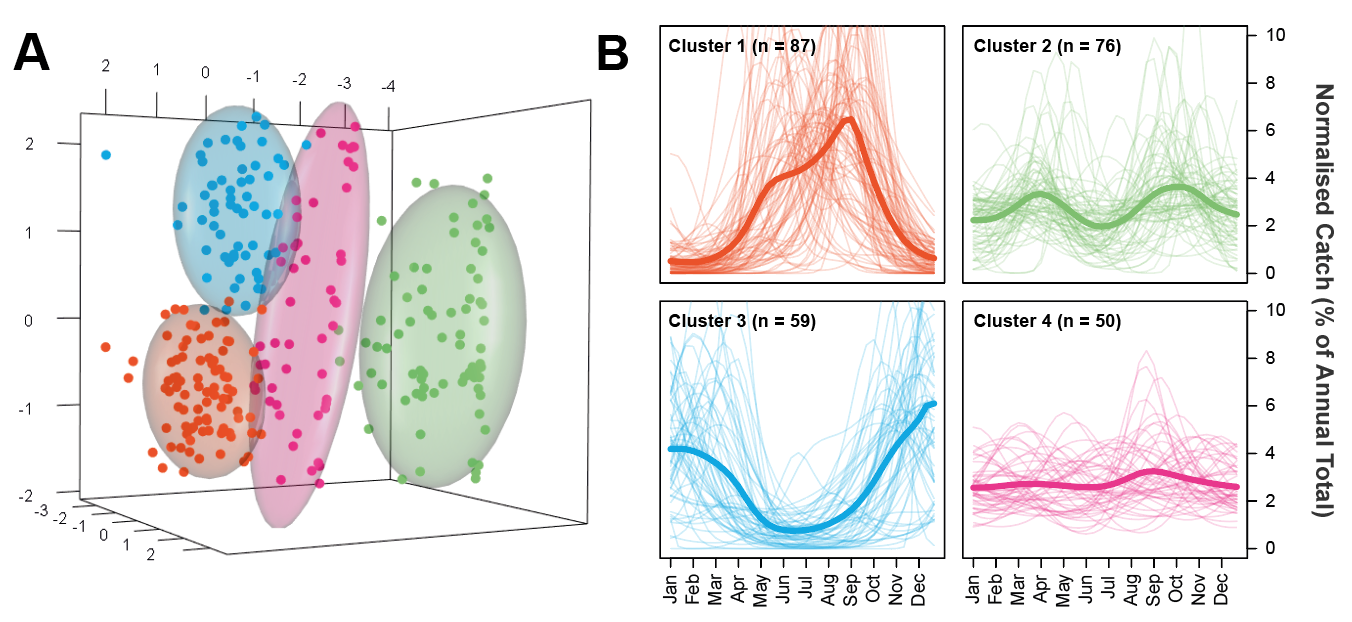
**Mosquito Population Dynamics are Determined by a Complex Interplay of Abiotic and Biotic Factors:** We next sought to assess the underlying drivers of these different temporal dynamics. Using binary indicators for species (7 total, indicating which species a particular time-series belongs to) and a suite of ecological covariates (25 total) as predictors, we fitted a multinomial logistic regression to the cluster labels (denoting which cluster each time-series had been assigned to) to explore factors influencing temporal patterns and predict membership of a cluster. Overall, the model was able to assign each time-series into the correct cluster 58% of the time (compared to 25% expected for a completely random classifier). We next assessed the magnitude and direction of the regression coefficients – the multinomial logistic framework produces one coefficient estimate for each cluster and predictor (a total of 4 coefficients per predictor), with that coefficient defining the strength of the association between a predictor and a particular cluster. Our results revealed cluster-specific associations with predictors (**Figure 4**). Across the species coefficients, *Anopheles culicifacies* and *Anopheles subpictus* demonstrated large and positive associations with cluster 1 (monsoon peaking dynamics), whereas for *Anopheles fluviatilis*, this relationship was strongly negative. Instead, *Anopheles fluviatilis* was associated with temporal cluster 3 (dry season peaking dynamics). These results are in-keeping with the empirical analyses presented in **Fig.3** and support the notion that this multinomial logistic regression framework can recover patterns of association evident in the data. Similar results were observed for the environmental covariates where a number (such as isothermality, rainfall-related patterns and the extent/dynamics of large water bodies) demonstrated highly positive or negative associations depending on the particular cluster considered.

**Evidence of Discrete Ecological Structuring Across Temporal Pattern Clusters:** We next sought to assess the similarity of the coefficients for species and the environmental covariates across the clusters more systematically. We employed a hierarchical clustering approach to examine the species coefficient values and identified significant structuring **(Fig.5A).** *Anopheles culicifacies* and *Anopheles subpictus* clustered together, indicating that these species displayed similar associations with different temporal patterns (i.e. a strong positive association with Cluster 1 peaking during the monsoon, a strong negative association with Cluster 3 and only weak associations with Clusters 2 and 4). *Anopheles fluviatilis* by contrast displayed the opposite pattern (a strong positive association with Cluster 3 and a strong negative association with Cluster 1). *Anopheles minimus* and *Anopheles dirus* displayed weak associations with all clusters, suggesting (a) the absence of a strong species-specific tendency to adopt a particular temporal pattern and (b) that ecological structure may be the primary determinant of their dynamics. The ecological coefficients for each cluster pattern were ranked according to the strength of the association (either positive or negative, with the values displayed in **Figure 4B**) with that particular cluster. The 15 coefficients with the strongest association for each cluster were then selected. The extent of overlap in these top 15 coefficients between clusters was then assessed – these results revealed that each cluster (comprising time-series with similar temporal properties) tended to associate with a unique set of ecological factors **(Fig.5B)**. These mutually exclusive patterns of association with environmental covariates across clusters were borne out across an analysis of the correlation of coefficients between clusters, which revealed them to be highly negatively correlated **(Supplementary Fig.5)**.

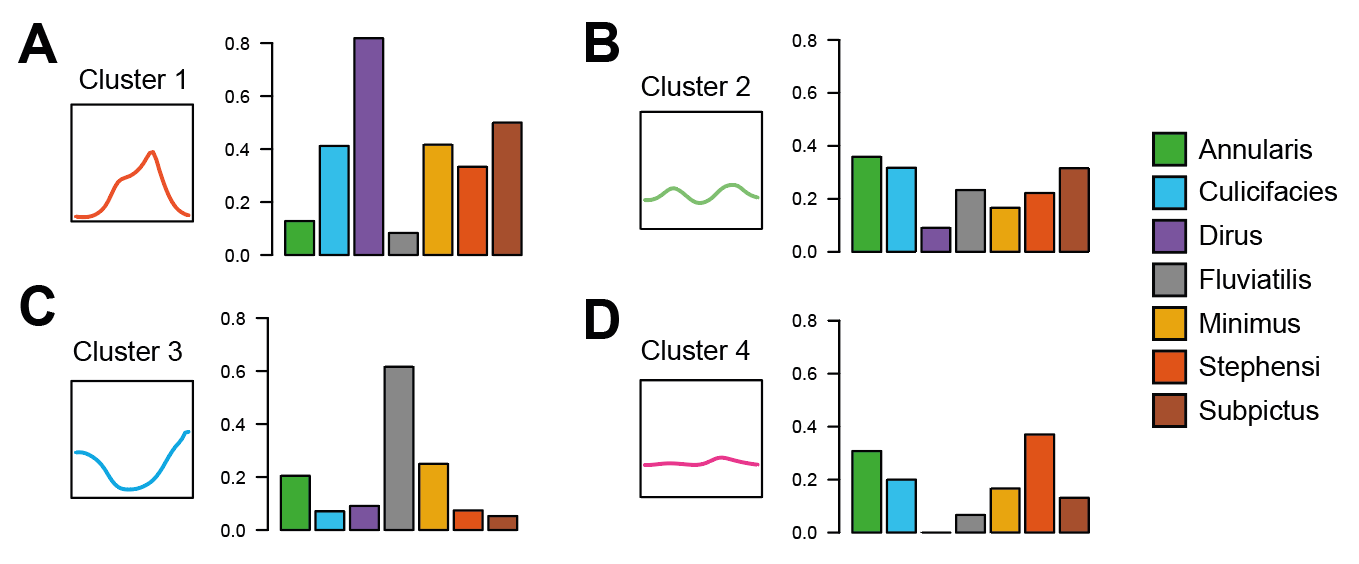
**Predictive Mapping Highlights the Extensive Variation in Seasonal Mosquito Dynamics and Malaria Risk Across India.** We next integrated these results with spatial predictions of mosquito species presence/absence to produce predictive maps of mosquito population dynamics across India; specifically, to generate estimates of the probability that a given location contains ≥1 mosquito species displaying a particular temporal pattern **(Fig.5)**. Our results predict that monsoon peaking dynamics (Cluster 1) are most likely in the North and Northeast **(Fig.5A)**. This contrasts with the predicted spatial distribution of bimodal dynamics (Cluster 2), which are predicted to be more likely across central India and less likely in the Northeast. Dynamics involving peaks during the dry season tracks the predicted spatial distribution of *Anopheles fluviatilis* closely and is predicted to be most probable across central India **(Fig**.**5C)** – a similar pattern was observed for spatial predictions of perennial dynamics **(Fig.5D)**. Together these results suggest that spatial variability in both species occurrence and environmental factors together generate complex patterns of mosquito temporal dynamics across India.



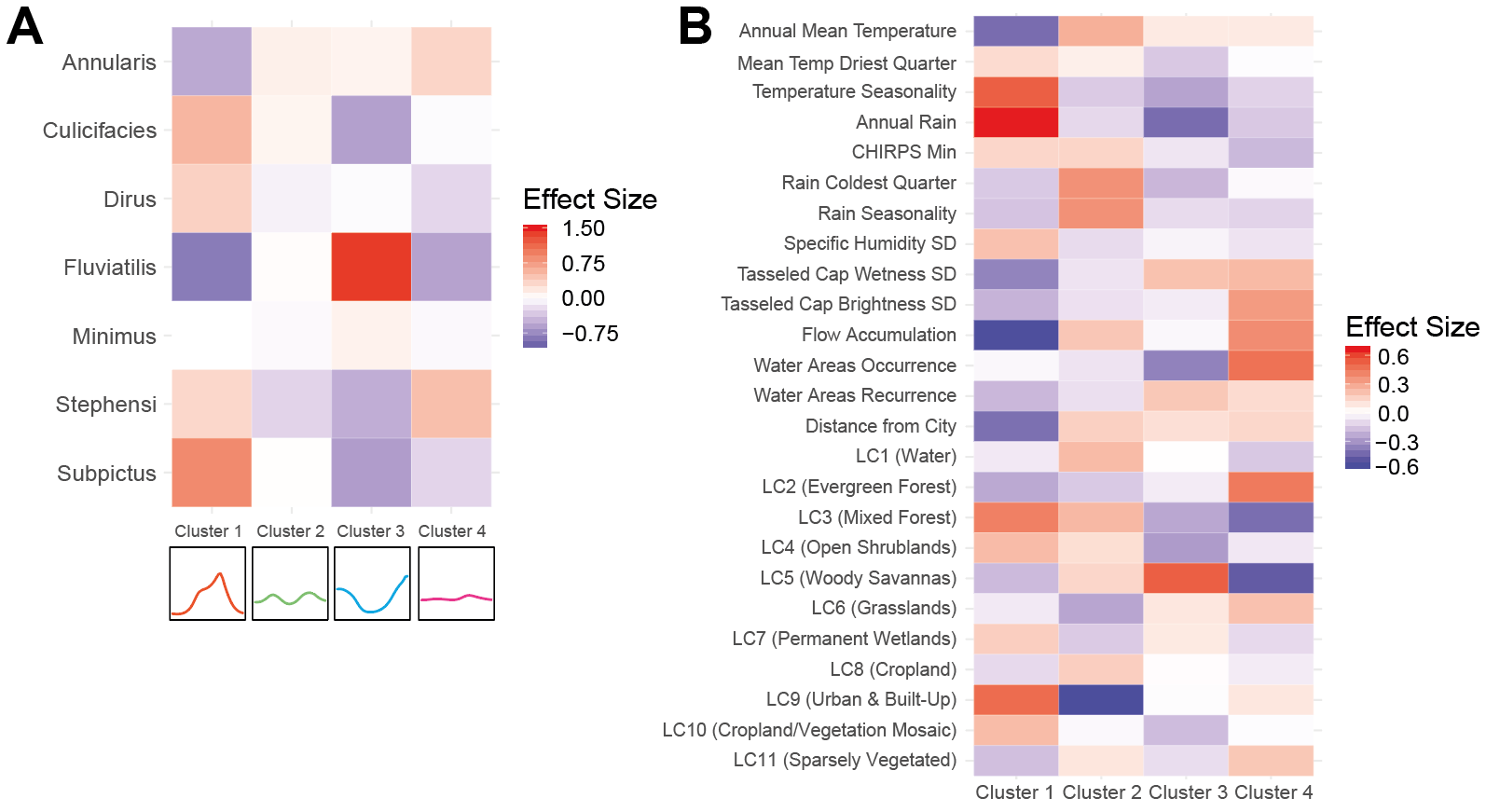
**Figure 1: Exploring and Mapping Species Specific Patterns of Mosquito Population Dynamics.** Negative Binomial Gaussian Processes incorporating a periodic kernel were fitted to each of the 272 time-series collected from 118 locations across India collated as part of the systematic review. These fitted time-series (representing monthly catches over the course of a year) were then normalised and the results plotted here, disaggregated by species. **(A)** Map of India showing the different locations for which time-series data was available. Points represent a single collected time-series, coloured according to species. **(B)** Normalised, Gaussian process fitted time-series disaggregated by species. In all instances, pale lines represent a single time-series for that particular species, and the brighter line is the mean of all of the time-series belonging to that species, evaluated at that particular timepoint. Although a number of mosquito species show patterns of population dynamics with peaks around India’s monsoon season (typically June – September), there is substantial variation in the extent and nature of the observed dynamics, with many time-series (and species) displaying different patterns.



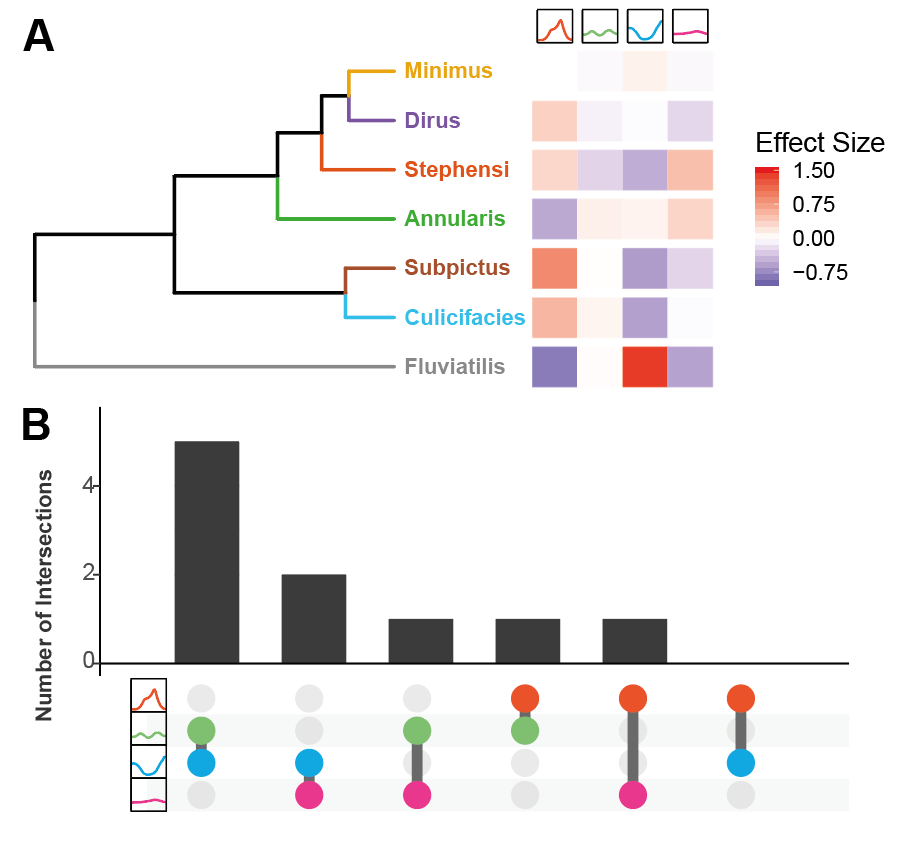
**Figure 2: Characterisation and Clustering of Time-Series with Similar Temporal Properties.** A series of mathematical operations were applied to the fitted time-series in order to characterise their temporal properties. A Principal Components Analysis was then carried out and the results clustered using the k-means algorithm. **(A)** Results of the k-means clustering algorithm for 4 clusters applied to a Principal Components Analysis. Colour of the points refers to cluster membership, coloured ellipsoids demarcate the 75th quantile of the density associated with each cluster. First 3 principal components are plotted, explaining 53%, 15% and 14% of the overall variation, respectively. **(B)** Plots displaying the time-series belonging to each cluster. Pale lines represent individual time-series, brighter line represents the mean of all the time-series belonging to that cluster, evaluated at each timepoint. Characterisation and clustering in this way revealed distinct groups of time-series that share similar temporal properties.

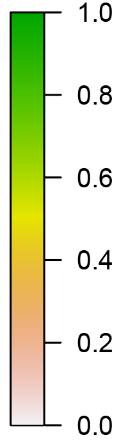
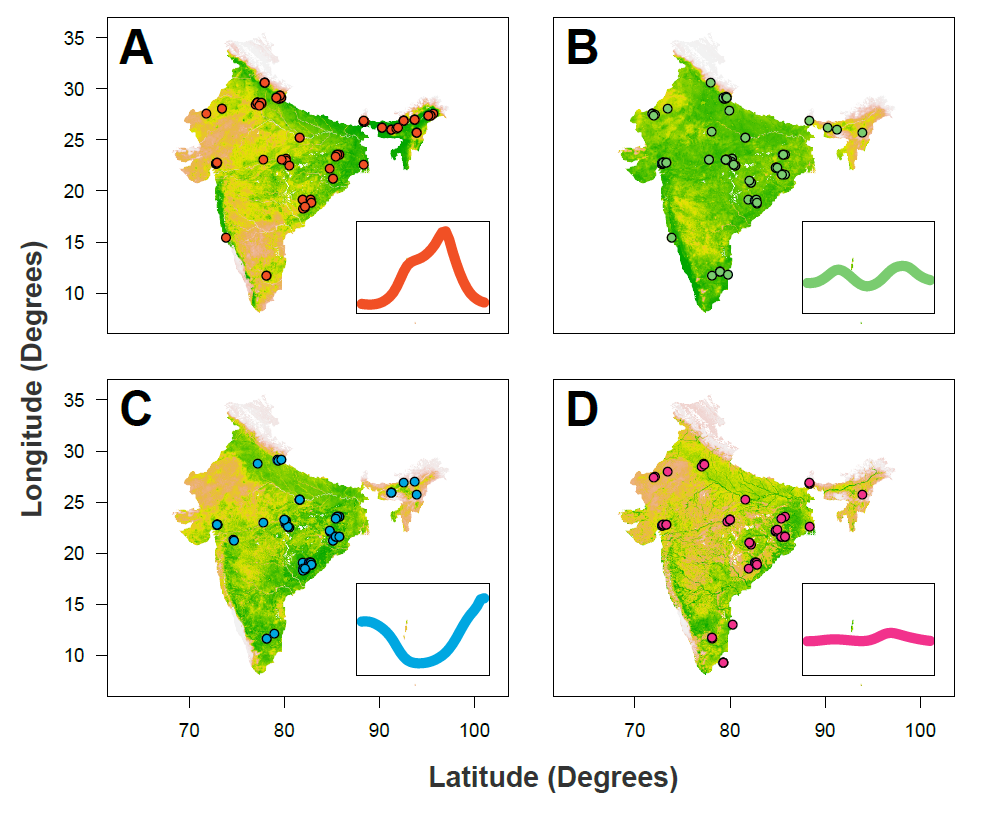


**Figure 3: The Proportion of Time-Series Belonging to Each Species and Cluster.** Having clustered the time-series into distinct clusters with distinct temporal properties, we next assessed the proportion of time-series for each species belonging to each cluster. **(A)** For each species, the proportion of time-series belonging to cluster 1 (unimodal, monsoon season peaking dynamics) – different coloured bars indicate different species (see legend) and y axis corresponds to the proportion of time-series (for a given species) belonging to that cluster. **(B)** As for A, but for cluster 2 (bimodal dynamics). **(C)** As for A, but for cluster 3 (unimodal, dry season peaking dynamics). **(D)** As for A, but for cluster 4 (perennial-like dynamics).



**Figure 4: The Drivers of Distinct Temporal Patterns – Coefficient Values from the Multinomial Logistic Regression.** Using both species and a suite of environmental variables as predictors, a multinomial logistic regression framework was used to drivers of mosquito population dynamics. This was fitted within a Bayesian framework using the probabilistic programming language STAN. The output of this multinomial logistic regression is a single coefficient per predictor and cluster, with this coefficient describing the strength of the association between the predictor (either species or environmental predictor) and membership of a particular cluster. **(A)** The set of coefficients for each species and cluster – red colours indicate a positive association between that species and the cluster, whilst blue colours indicate a negative association. Intensity of the colour indicates the strength of the association. **(B)** As for **(A)** but for the 66 environmental predictors also included in the multinomial logistic regression model.

**Figure 5: Exploring the Role of Species-Specific Tendencies and Ecological Structure of Mosquito Temporal Dynamics. (A)** Hierarchical clustering of the multinomial logistic regression results for species: the dendogram arranges mosquito species according to their relatedness as defined by the set of coefficient values for each species (describing the strength of the association between that species and the particular cluster i.e. 1 coefficient for each cluster, displayed on the RHS and coloured according to the size and sign of the coefficient). **(B)** Upset plot summarising the multinomial logistic regression results for the environmental covariates used as predictors. For each set of ecological covariate coefficients arising from the multinomial logistic regression (25 for each of the 4 clusters), the 15 covariates with the strongest association with a particular cluster were selected. These top 15 covariates (comprising a set unique to each cluster) were then compared across pairs of clusters to assess the extent of overlap i.e. how many strongly predictive covariates were shared across clusters. In the above plot, the x-axis indicates the specific pairwise comparison being made e.g. between clusters 2 and 3, with the y axis describing the number of shared covariates in the top 15 of these two clusters.



**Figure 6: Predictive Maps of Seasonality Across India.** The results of the multinomial logistic regression were integrated with recently generated maps describing the probability of presence/absence for different *Anopheline* species. Together, these were used to generate estimates of a given area possessing at least one mosquito species with a particular temporal profile (as defined by the previously described clusters). **(A)** Results of this analysis for Cluster 1 (the “monsoon peak” cluster) – red dots describe the locations in which a mosquito species with a temporal profile assigned to Cluster 1 were found. **(B)** As for A, but for the “bimodal” cluster. **(C)** As for A, but for the “peak in dry season” cluster. **(D)** As for A, but for the “perennial” cluster. In all cases, the map colour describes the probability of a given area containing one or more mosquito species displaying that pattern of temporal dynamics. The coloured points indicate locations where a mosquito species displaying temporal dynamics belonging to that cluster were empirically observed.

**Discussion**

Understanding the temporal dynamics of malaria transmission represents a vital input to effective deployment of malaria control interventions, yet many outstanding questions remain surrounding the drivers of these dynamics. Here we leverage a collection of temporally disaggregated mosquito time-series catch data from across India to explore these dynamics and the comparative role of abiotic and species-specific factors in shaping them. Our results reveal extensive variation in mosquito population dynamics between species and across locations. Analysis of this variation has revealed a complex interplay between biotic (species-specific preferences) and abiotic (the broader ecological structure of the environment) factors in shaping mosquito population dynamics. Importantly, the comparative importance of these factors depends intimately on the setting and mosquito species being considered. Where environmental factors dominate, we show that distinct ecological forces structure and determine population dynamics to produce a diverse array of different temporal profiles.

In a manner largely independent of the ecological setting, *Anopheles fluviatilis* populations typically peaked peaking during the dry season. Whilst previous work has identified these dynamics33,34, our work highlights the consistency of this observation across locations, showing that these dynamics are largely restricted to *Anopheles fluviatilis*. These results align with previous work that has indicated streams and surrounding stagnant water to be the species’ preferred breeding site35. Such breeding sites are unsuitable during the monsoon season when flooding occurs but suitable as the dry season ensues. By contrast, *Anopheles culicifacies* displayed a wide array of temporal dynamics depending on the sampling site. These ranged from peaking during the monsoon to bimodal and even perennial behaviour. As with *Anopheles fluviatilis*, these results can likely be understood in the context of breeding habits, with the species able to exploit a wide range of both fresh-water36,37 and brackish38 sources. These contrasting results for *Anopheles fluviatilis* (where dynamics are similar across ecologically disparate locations) and *Anopheles culicifacies* (where dynamics are far more plastic and shaped by the environment) highlight how a complex interplay between species-specific breeding preferences (or the lack thereof) can interact with ecological structure to determine the temporal dynamics of mosquito populations.

In addition to these species-specific associations, our results suggest a significant role of the environment in shaping mosquito population dynamics. For example, temperature seasonality was positively associated with monsoon peaking seasonal dynamics (Cluster 1) but negatively associated with all other temporal profiles. Interestingly, and by contrast, rainfall seasonality was negatively associated with monsoon peaking dynamics (albeit weakly). Temperature is an important determinant of various mosquito traits relevant to malaria transmission, with previous research supporting a significant influence of temperature on many mosquito life-history traits29,39. Despite this, rainfall is frequently considered to be the primary driver of mosquito population dynamics – these results suggest that, contrary to expectations derived from primarily African settings, seasonal fluctuations in temperature, more so than rainfall, contribute to the marked seasonal dynamics observed across mosquito populations belonging to Cluster 1.

In contrast to the seasonal dynamics of Cluster 1, the perennial patterns of abundance observed for Cluster 4 were most strongly associated with flow accumulation and water area occurrence (acting as proxies for proximity to rivers and static bodies of water respectively). These factors were negatively associated with all other temporal profiles. This is consistent with reports indicating that static water sources may provide sites available for oviposition and mosquito breeding year round40,41. As well as these hydrological and climactic associations, we also observed a significant influence of landcover patterns on temporal dynamics. Urbanicity (measured by the two covariates Landcover and Distance to City) was consistently and positively associated with rainfall concordant, monsoon peaking dynamics and negatively associated with other temporal profiles. This is possibly due to the diverse array of physical features present in cities (ranging from tyres to wells and overhead tanks) that are able to hold water following rainfall, and which have previously been characterised as breeding sites for a range of mosquito species42,43. However, reports of perennial patterns of abundance in other urban centres suggests that city-specific approaches to management of water containers may also shape mosquito population dynamics43. Together, these results demonstrate clear structuring of the factors shaping mosquito population dynamics and highlight that unique sets of ecological factors drive different temporal profiles.

**PARAGRAPH ABOUT BREEDING HABITS, PREFERENCES AND SIBLING SPECIES**

It is important to note that factors other than mosquito dynamics are also involved in defining the temporal profile of malaria risk. Whilst an association between the size of mosquito populations and case numbers is well established44,45, the nature of this relationship remains less clear. Interactions between malaria endemicity46, mosquito abundance47 and vector competence31 can lead to non-linear dynamics that can be further modified by human behavioural factors such as migration or occupational practices48. Due to heterogeneity in mosquito sampling methods and limitations on the extent of entomological data describing relevant malaria metrics such as sporozoite positivity, we were unable to explore many of these factors. Similarly, the lack of accompanying epidemiological information precludes us from better resolving the comparative contributions of different mosquito species to malaria transmission. This limits our ability to translate temporal patterns of mosquito populations into metrics relevant to malaria transmission, such as the Entomological Inoculation Rate (EIR). Whilst we mitigate this limitation somewhat by focussing our analyses specifically on dominant vector species previously established as relevant to malaria transmission in India50, it is not necessarily the case that each mosquito species analysed here is equally relevant to malaria transmission. Future work integrating these analyses with those exploring seasonality of case incidence (c.f. Nguyen et al.51) would therefore likely prove instructive.

Overall, our work yields new insight into the drivers of the temporal processes governing malaria transmission. We show that temporal variation in mosquito populations is driven by a complex interplay of biotic and abiotic factors, with the comparative importance of these two sets of factors depending intimately on the particular species and setting. In doing so, this work underscores the crucial importance of integrating both species composition and ecological structure into our understanding of the temporal profile of malaria risk – a crucial and operationally relevant input for optimising the delivery of malaria control interventions.

**Methods**

**Systematic Review of Indian Entomological Literature**

Web of Science and PubMed databases were searched on 17th October 2017 using the keywords “India” AND “Anophel\*” to identify references with temporally disaggregated entomological data. We identified 1945 records with 1556 remaining after removing duplicates. Following Title and Abstract screening 281 records were retained for full text evaluation. We included records containing temporally disaggregated adult mosquito catch data with monthly (or finer) temporal resolution spanning at least 12 months that had not been conducted as part of vector control intervention trials, and where sufficient information to geolocate the catch site was provided. 78 references were retained that yielded 117 geolocatable areas across India. These references contained 272 time-series spanning the malaria vectors *Anopheles annularis*, *culicifacies*, *dirus*, *fluviatlis*, *minimus*, *stephensi* and *subpictus*. See Supplementary Information for further details.

**Time-Series Fitting and Interpolation**

To smooth the noise in the mosquito catch data we fitted a Gaussian Process model to each of the extracted time-series, using a Negative Binomial likelihood to account for overdispersion in the data:

where is a distribution of functions from a zero-mean Gaussian Process with covariance function ,(x) are function evaluations at times , are the observed mosquito counts indexed by timepoint , and and represent a vector of hyperparameters involved in defining the overdispersion of the Negative Binomial distribution and the functional form of the covariance function respectively. 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, defined as:

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. Weakly informative priors were used. Fitting was undertaken using the programming language STAN52.

**Time-Series Characterisation and Clustering by Features**

Motivated by previous work providing a framework to statistically characterise the empirical structure of time-series data53 and work characterising the seasonality of malaria case incidence51, we calculated several summary statistics for each smoothed time-series to characterise their temporal properties. These include the Kullback-Liebler divergence (measuring the divergence of the time-series from a uniform distribution), the median of the period () from the Negative Binomial Gaussian Process fitting (informing the dominant temporal modality present in the data), the proportion of points greater than 1.65x the mean (measuring how peaked the time-series is), the distance of the first peak from January, and then 3 features arising from fitting 1 and 2 component Von-Mises distributions to the smoothed time-series: specifically, the mean of the 1 component Von-Mises distribution, the number of peaks (determined by comparing the quality of fit for 1 and 2 component Von-Mises distributions), and the weight (), specifying the comparative contributions of each component in the two-component fitting. See Supplementary Information for further details. From this we obtain a series of 7 real numbers describing the temporal properties of each time-series. We then applied a Principal Components Analysis to these results 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 features – i.e. this clustering assigns each smoothed time-series to one cluster.

**Statistical Modelling and Prediction of Seasonal Modality**

For each of the 117 study locations we extracted a suite of environmental variables derived from satellite data that together describe the location’s ecological structure. These include the BioClimatic variables (a suite of biological relevant covariates defined from monthly rainfall and temperature satellite data54), various measures of aridity55,56, a number of covariates describing the seasonality and extent of water bodies57, landcover58 and a number of other variables previously used in defining the global distribution of Anopheline vectors32. A complete list of the covariates used is in Supplementary Table 2. These covariates (25 in total) and a covariate for Anopheline species (1 for each time-series indicating which species it belonged to) were used as covariates in a penalised (L2) multinomial logistic regression model predicting the cluster (of time-series with similar temporal properties, assigned based on the results of the k-means clustering) a particular time-series belonged to. Fitting this model yielded regression coefficients describing the strength of association between a species/environmental variable and membership of a particular cluster – specifically, 1 coefficient per cluster and predictor, i.e. a total of k coefficients per predictor where k is the number of clusters. The results of these analyses were then integrated with recently produced maps of vector presence/absence (as part of work conducted with the Humbug Project (<http://humbug.ac.uk/>), funded through a Google Impact Challenge grant). Together, this facilitated generation of predictive maps of mosquito population dynamics across the India (see Supplementary Information for further detail).

**References**

1. Weiss, D. J. *et al.* Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study. *Lancet* **394**, 322–331 (2019).

2. Roth, G. A. *et al.* Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1736–1788 (2018).

3. World malaria report 2019. Available at: https://www.who.int/news-room/feature-stories/detail/world-malaria-report-2019. (Accessed: 24th December 2019)

4. Battle, K. E. *et al.* Mapping the global endemicity and clinical burden of Plasmodium vivax, 2000–17: a spatial and temporal modelling study. *Lancet* **394**, 332–343 (2019).

5. Warrell, D. A. & Gilles, H. M. *Essential Malariology*. (CRC Press, 2002).

6. Hay, S. I. *et al.* Developing Global Maps of the Dominant Anopheles Vectors of Human Malaria. *PLoS Med.* **7**, e1000209 (2010).

7. Sinka, M. E. Global Distribution of the Dominant Vector Species of Malaria. in *Anopheles mosquitoes - New insights into malaria vectors* (InTech, 2013). doi:10.5772/54163

8. Sinka, M. E. *et al.* A global map of dominant malaria vectors. *Parasit. Vectors* **5**, 69 (2012).

9. Koenraadt, C. J. ., Githeko, A. . & Takken, W. The effects of rainfall and evapotranspiration on the temporal dynamics of Anopheles gambiae s.s. and Anopheles arabiensis in a Kenyan village. *Acta Trop.* **90**, 141–153 (2004).

10. Das, M. K. *et al.* Malaria epidemiology in an area of stable transmission in tribal population of Jharkhand, India. *Malar. J.* **16**, 181 (2017).

11. Wilson, A. L. & IPTc Taskforce. A Systematic Review and Meta-Analysis of the Efficacy and Safety of Intermittent Preventive Treatment of Malaria in Children (IPTc). *PLoS One* **6**, e16976 (2011).

12. Ross, A., Maire, N., Sicuri, E., Smith, T. & Conteh, L. Determinants of the cost-effectiveness of intermittent preventive treatment for malaria in infants and children. *PLoS One* **6**, (2011).

13. Pluess, B., Tanser, F. C., Lengeler, C. & Sharp, B. L. Indoor residual spraying for preventing malaria. *Cochrane Database Syst. Rev.* (2010). doi:10.1002/14651858.CD006657.pub2

14. Health Organization, W. *AN OPERATIONAL MANUAL FOR INDOOR RESIDUAL SPRAYING (IRS) FOR MALARIA TRANSMISSION CONTROL AND ELIMINATION SECOND EDITION INDOOR RESIDUAL SPRAYING*.

15. WHO | High burden to high impact: a targeted malaria response. *WHO* (2019).

16. Gimnig, J. E., Ombok, M., Kamau, L. & Hawley, W. A. Characteristics of larval anopheline (Diptera: Culicidae) habitats in Western Kenya. *J. Med. Entomol.* **38**, 282–8 (2001).

17. Appawu, M. *et al.* Malaria transmission dynamics at a site in northern Ghana proposed for testing malaria vaccines. *Trop. Med. Int. Heal.* **9**, 164–170 (2004).

18. Okello, P. E. *et al.* Variation in malaria transmission intensity in seven sites throughout Uganda. *Am. J. Trop. Med. Hyg.* **75**, 219–225 (2006).

19. White, M. T. *et al.* Modelling the impact of vector control interventions on Anopheles gambiae population dynamics. *Parasit. Vectors* **4**, 153 (2011).

20. Obsomer, V., Defourny, P. & Coosemans, M. The Anopheles dirus complex: spatial distribution and environmental drivers. *Malar. J.* **6**, 26 (2007).

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

22. Cohuet, A. *et al.* High Malaria Transmission Intensity Due to Anopheles funestus (Diptera: Culicidae) in a Village of Savannah–Forest Transition Area in Cameroon . *J. Med. Entomol.* **41**, 901–905 (2004).

23. Mendis, C. *et al.* Anopheles arabiensis and An. funestus are equally important vectors of malaria in Matola coastal suburb of Maputo, southern Mozambique. *Med. Vet. Entomol.* **14**, 171–180 (2000).

24. Sinka, M. E. *et al.* The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. *Parasit. Vectors* **3**, 117 (2010).

25. Sahu, S. S. *et al.* Bionomics of anopheles fluviatilis and anopheles culicifacies (Diptera: Culicidae) in relation to malaria transmission in east-central India. *J. Med. Entomol.* **54**, 821–830 (2017).

26. Misra, S. P., Nandi, J., Narasimham, M. V & Rajagopal, R. D Malaria transmission in Nagaland, India. Part-I--Anophelines and their seasonality. *J. Commun. Dis.* **25**, 62–66 (1993).

27. Singh, R. K., Haq, S., Kumar, G. & Dhiman, R. C. U Bionomics and vectorial capacity of Anopheles annularis with special reference to India: a review. *J. Commun. Dis.* **45**, 1–16 (2013).

28. Das, N. G., Gopalakrishnan, R., Talukdar, P. K. & Baruah, I. Diversity and seasonal densities of vector anophelines in relation to forest fringe malaria in district Sonitpur, Assam (India). *J. Parasit. Dis.* **35**, 123–128 (2011).

29. Bayoh, M. N. & Lindsay, S. W. Effect of temperature on the development of the aquatic stages of Anopheles gambiae sensu stricto (Diptera: Culicidae) . *Bull. Entomol. Res.* **93**, 375–381 (2003).

30. Shapiro, L. L. M., Whitehead, S. A. & Thomas, M. B. Quantifying the effects of temperature on mosquito and parasite traits that determine the transmission potential of human malaria. *PLOS Biol.* **15**, e2003489 (2017).

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

32. Sinka, M. E. *et al.* The dominant Anopheles vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. *Parasit. Vectors* **4**, 89 (2011).

33. Gunasekaran, K., Jambulingam, P., Sadanandane, C., Sahu, S. S. & Das, P. K. D Reliability of light trap sampling for Anopheles fluviatilis, a vector of malaria. *Acta Trop.* **58**, 1–11 (1994).

34. Sahu, S. S. *et al.* Bionomics of Anopheles fluviatilis and Anopheles culicifacies (Diptera: Culicidae) in Relation to Malaria Transmission in East-Central India. *J. Med. Entomol.* **54**, 821–830 (2017).

35. Dash, S. *et al.* Preferential breeding habitats of sibling species complexes of Anopheles fluviatilis in east-central India. *~ 91 ~ Int. J. Mosq. Res.* **5**, 91–95 (2018).

36. Surendran, S. N. & Ramasamy, R. Some characteristics of the larval breeding sites of Anopheles culicifacies species B and E in Sri Lanka. *J. Vector Borne Dis.* **42**, 39–44 (2005).

37. Barik, T. K., Sahu, B. & Swain, V. A review on Anopheles culicifacies: From bionomics to control with special reference to Indian subcontinent. *Acta Tropica* **109**, 87–97 (2009).

38. Jude, P. J., Dharshini, S., Vinobaba, M., Surendran, S. N. & Ramasamy, R. Anopheles culicifacies breeding in brackish waters in Sri Lanka and implications for malaria control. *Malar. J.* **9**, (2010).

39. Shapiro, L. L. M., Whitehead, S. A. & Thomas, M. B. Quantifying the effects of temperature on mosquito and parasite traits that determine the transmission potential of human malaria. *PLoS Biol.* **15**, (2017).

40. Minakawa, N., Dida, G. O., Sonye, G. O., Futami, K. & Njenga, S. M. Malaria vectors in Lake Victoria and adjacent habitats in Western Kenya. *PLoS One* **7**, (2012).

41. Kumar, A. *et al.* Anopheles subpictus carry human malaria parasites in an urban area of Western India and may facilitate perennial malaria transmission. *Malar. J.* **15**, (2016).

42. Lin, C.-H., Schiøler, K. L., Ekstrøm, C. T. & Konradsen, F. Location, seasonal, and functional characteristics of water holding containers with juvenile and pupal Aedes aegypti in Southern Taiwan: A cross-sectional study using hurdle model analyses. *PLoS Negl. Trop. Dis.* **12**, e0006882 (2018).

43. 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**, (2016).

44. Bashar, K. & Tuno, N. Seasonal abundance of Anopheles mosquitoes and their association with meteorological factors and malaria incidence in Bangladesh. *Parasit. Vectors* **7**, 442 (2014).

45. GALARDO, A. K. R. *et al.* Seasonal abundance of anopheline mosquitoes and their association with rainfall and malaria along the Matapí River, Amapí, Brazil. *Med. Vet. Entomol.* **23**, 335–349 (2009).

46. Churcher, T. S., Trape, J. F. & Cohuet, A. Human-to-mosquito transmission efficiency increases as malaria is controlled. *Nat. Commun.* **6**, (2015).

47. 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. R. Soc. B Biol. Sci.* **285**, (2018).

48. Cohen, J. M. *et al.* Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. *Malar. J.* **12**, (2013).

49. Dev, V. & Sharma, V. P. The Dominant Mosquito Vectors of Human Malaria in India. in *Anopheles mosquitoes - New insights into malaria vectors* (InTech, 2013). doi:10.5772/55215

50. Dev, V. & Sharma, V. P. The Dominant Mosquito Vectors of Human Malaria in India. in *Anopheles mosquitoes - New insights into malaria vectors* (InTech, 2013). doi:10.5772/55215

51. Nguyen, M. *et al.* Mapping malaria seasonality: a case study from Madagascar. (2019).

52. Carpenter, B. *et al.* Stan: A Probabilistic Programming Language. *J. Stat. Softw.* **76**, (2017).

53. 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**, (2013).

54. 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).

55. Zomer, R. J., Trabucco, A., Bossio, D. A. & Verchot, L. V. Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agric. Ecosyst. Environ.* **126**, 67–80 (2008).

56. Zomer, R., Bossio, D., Trabucco, A. & Yuanjie, L. *Trees and water: smallholder agroforestry on irrigated lands in Northern India*. (2007).

57. Lehner, B. & Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* **296**, 1–22 (2004).

58. Friedl, M. A. *et al.* MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **114**, 168–182 (2010).