**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 the Indian subcontinent to explore the factors shaping these dynamics. Our analyses reveal pronounced heterogeneity in mosquito population dynamics, both within (across different locations) and between (in the same location) species: this heterogeneity is driven by a complex interplay between mosquito species and the ecological structure of the local environment. We demonstrate that the temporal profile of malaria risk across these different locations can be categorised into a small number of unique patterns, and that these patterns can be used to create a predictive tool 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 control strategies aimed at reducing malaria transmission, enabling interventions to be deployed at times and in places where they can have the most impact. This temporal profile of malaria risk is intimately related to the dynamics of the mosquito populations underlying transmission. Many outstanding questions remain surrounding these dynamics, including the specific drivers and their dependence on the underlying ecological structure of a setting. Here we collate a data set of temporally disaggregated mosquito catch data from across the Indian subcontinent 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. We show that this despite this variation, the collated time series can be categorised into a small number of clusters characterised by distinct temporal properties – analyses of these clusters using a multinomial logistic regression-based approach 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 control efforts.

**Background**

Responsible for an estimated 228 million cases and 405,000 deaths in 20181, 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 malaria burden3, with *Plasmodium falciparum* most prevalent in African settings, and India alone accounting for almost 50% of the global *Plasmodium vivax* burden4. Transmission of these malaria causing parasites occurs via mosquito vectors belonging to the *Anopheles* genus. Of the 460 recognised Anopheline species, approximately 70 possess the capacity to transmit human malaria parasites5,6. These vectors are heterogeneously distributed across the globe, a feature that results in marked differences in the transmission dynamics and epidemiology of malaria across different contexts and ecologies.

A substantial body of work has focussed on characterising the spatial distribution (presence/absence) of these vectors including mapping their current distribution across different global regions7,8 as well as the trajectories and dynamics of these distributions over time in response to climate change and other factors9. This work has involved relating vector occurrence to a variety of different environmental variables and represents a vital input required for strategically targeted 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 dynamic over time, exhibiting substantial fluctuations in size over the course of a year10,11, a feature which in turn drives a similarly dynamic and temporally variable profile of malaria risk. Understanding the determinants of these dynamics is important given that the efficacy of some of the most effective malaria control interventions are dependent on accurate timing of delivery in relation to seasonal peaks in malaria risk. Examples include Seasonal Malaria Chemoprevention, where monthly administration of long-acting antimalarials is administered to children under 512,13 before and during the seasonal peak in transmission, and Indoor Residual Spraying14,15 where houses are sprayed with insecticides to kill mosquitoes. Effective utilisation of these interventions will be vital in helping achieve the goals of the World Health Organisation’s “High Burden, High Impact” strategy, which aims to substantially reduce malaria burden in the 11 highest burden countries comprising India and 10 countries across Africa16.

Rainfall is frequently 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 breed17. However, whilst a close relationship between the occurrence of rainfall and peaks in mosquito populations (e.g. *Anopheles gambiae*18–20 for African settings, *Anopheles culicifacies*21 and *Anopheles dirus*22 across India and south-East Asia) and malaria cases23 has been observed in many locations, the strength and extent of this relationship is highly variable. For example, *Anopheles funestus* populations frequently lack such marked seasonal fluctuations in population abundance24,25 (a feature typically attributed to its capacity to exploit large permanent/semi-permanent bodies of fresh water as breeding habitats26), whilst there are numerous reports in the literature of *Anopheles fluviatilis* populations peaking during the dry season27,28. Further compounding this complexity, not all variation in population dynamics appears to be species-specific: *Anophleles annularis* has been observed to display dynamics that range from near-perennial through to intensely seasonal11,29,30 depending on the particular location, bringing into question how generalisable relationships between rainfall and population dynamics for a given species are. These results highlight key gaps in our knowledge surrounding how rainfall interacts with the ecological structure of the local hydrological environment and species-specific breeding site preferences to shape mosquito population dynamics.

These results also suggest that ecological factors other than rainfall also likely contribute to the observed variation in mosquito population dynamics. For example, the influence of temperature on many mosquito traits (such as larval development31, biting rates and mortality rates32) have been well-characterised – however, the majority of these analyses have occurred under laboratory conditions, rather than field settings where temperature can fluctuate substantially over a variety of different timescales33. And whilst recent work in the field has suggested that considerations of both rainfall and temperature are necessary to understand observed seasonal patterns of malaria incidence34, these analyses have been restricted to only a small number of settings across sub-Saharan Africa; leaving the influence of dynamic temperature regimen on mosquito population dynamicsacross other settings largely unexplored.

A number of outstanding questions remain surrounding the influence of rainfall on mosquito populations, the role of other environmental factors, and how these factors interact with species-specific properties to 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. India is well positioned in this respect due to the high diversity of malaria-competent vectors it is home to35, as well as its history of publishing high quality, temporally resolved and comprehensive entomological data. We use these data to characterise the temporal patterns displayed by different mosquito species and identify structural differences in the dynamics both between species and across different locations. Our results reveal pronounced heterogeneity in the extent, nature and dynamics of seasonal dynamics across different mosquito populations. Exploring the drivers of these dynamics highlights the critical interaction between abiotic and species-specific factors in shaping and structuring the temporal patterns of these important disease vectors and the importance of considering ecological structure when implementing malaria intervention strategies.

**Results**

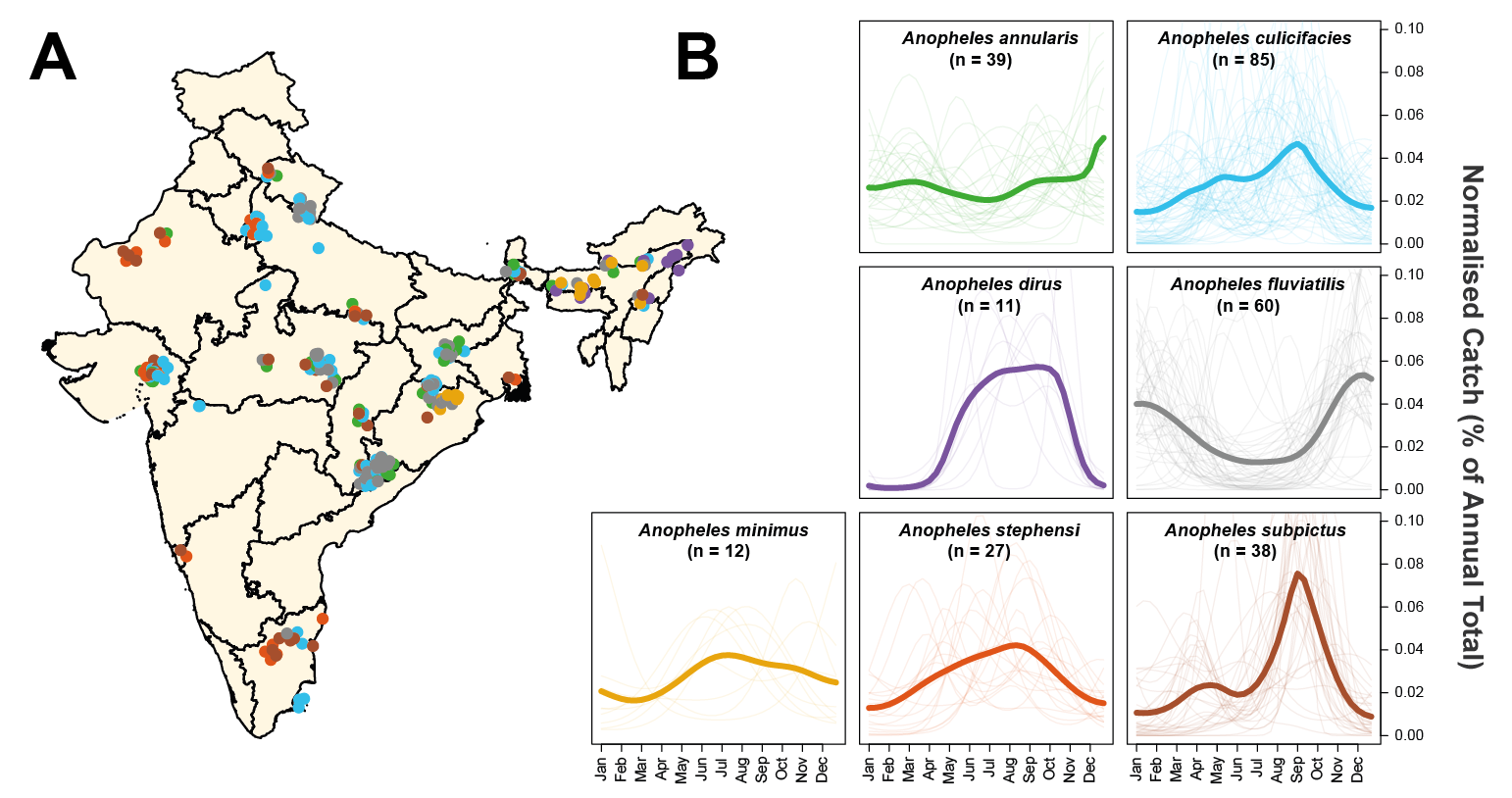
**Substantial Diversity in Mosquito Population Dynamics Both Within and Between Species:** A total of 272 time series from 117 locations across India were identified through the systematic review, which were then smoothed using a Negative Binomial Gaussian Process based framework **(Fig.1A)**. Across these smoothed time series, we observed substantial heterogeneity between different species in their temporal dynamics over the course of a year **(Fig.1B)**. Whereas *Anopheles dirus* populations tended to peak during the monsoon period (typically June to September), by contrast, in many instances, *Anopheles fluviatilis* populations peaked between November and February (the dry season across most of India), reaching their lowest density during the monsoon. A number of *Anopheles subpictus* time series showed distinct evidence of bimodality with peaks occurring both in the months of March–May and August–September, whilst a number of *Anopheles annularis* populations demonstrated perennial patterns of abundance. We also observed extensive variation in temporal dynamics within species. For example, across the 85 time-series collated for *Anopheles culicifacies*, the populations varied substantially in both the extent and timing of their seasonal peaks; whereas some displayed sharp peaks in the monsoon season, others displayed perennial characteristics similar to those observed for *Anopheles annularis*. Similar patterns were observed across the raw time series (**Supplementary Figure 1**).

**Characterisation and Clustering of Mosquito Catch Time Series Properties Reveals Discrete Temporal Modalities:** Having identified substantial variation in the temporal dynamics of mosquito populations, we next asked whether this variation could be delineated into discrete clusters of time series with distinct temporal patterns (hereafter referred to as a temporal pattern cluster) based on their statistical characteristics. Using an array of summary statistics to characterise the temporal properties of each time series followed by k-means clustering, we identified 4 temporal pattern clusters (**Figure 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 limited seasonal variation and perennial-like behaviour (Cluster 4).These varying seasonal dynamics across different locations could not be explained by rainfall pattern alone with a high positive cross-correlation product between rainfall and mosquito density in Cluster 1 (r = 0.52), a high negative correlation for Cluster 3 (r = -0.41) and low correlation for Clusters 2 and 4 (r = -0.08 and 0.03 respectively, see **Supplementary Figure 4** for further information). This indicates that the observed patterns in mosquito population dynamics are not due to differences in the timing and extent of rainfall across India – instead, they represent genuine differences in dynamics between species and across locations in how mosquito populations react and respond to incipient rainfall. Analysing the proportion of each mosquito species’ time series belonging to a particular temporal pattern cluster (**Figure 3**) revealed that for a number of species the majority of time series belonged to a single cluster. For example, *Anopheles dirus* time series were almost exclusively within Cluster 1, peaking during the monsoon season whilst *Anopheles fluviatilis* time series were almost exclusively in Cluster 3, peaking during the dry season. 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* and *Anopheles stephensi* time series appearing across all four temporal clusters - indicating the capacity for these species to adopt a wide and diverse array of temporal dynamics depending on the particular location and ecological setting.

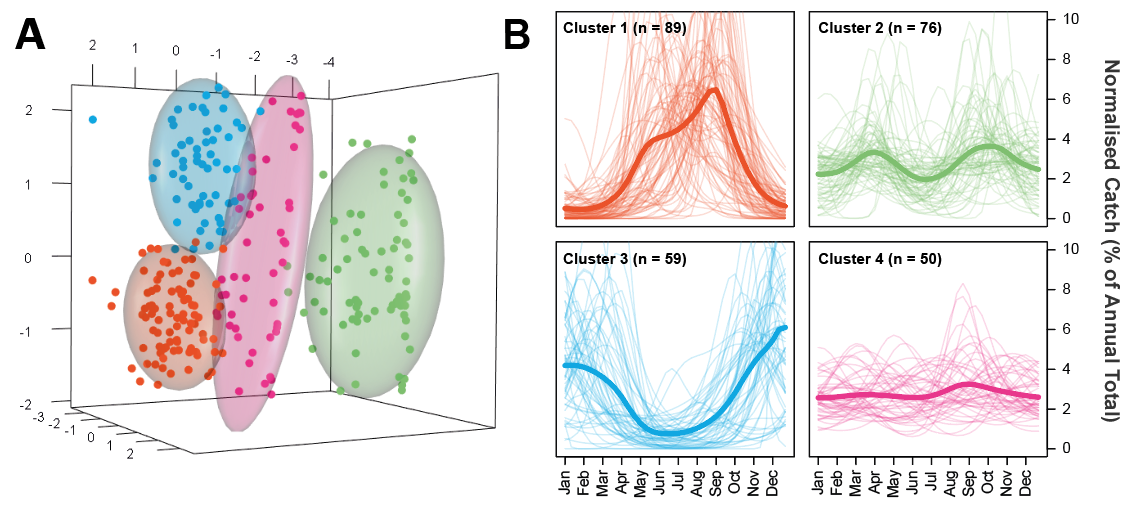
**Mosquito Population Dynamics are Determined by a Complex Interplay of Abiotic and Biotic Factors:** Given clear patterns in the collated time series, we next sought to assess the drivers of these different temporal dynamics. Using indicators for species (7 total, indicating which species a particular time series belongs to) and a suite of ecological covariates (66 total) as predictors, we fitted a multinomial logistic regression to the temporal pattern cluster labels (denoting which temporal pattern cluster each time series had been assigned to) to explore factors influencing temporal patterns and predict membership of a temporal pattern cluster. Overall, the model was able to assign each time-series into the correct cluster 60% 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 used here 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 temporal pattern cluster. Our results revealed significant and temporal pattern cluster-specific associations with predictors (**Figure 4**). Across the species coefficients, *Anopheles culicifacies* and *Anopheles subpictus* demonstrated large and positive associations with temporal cluster 1 (unimodal monsoon peaking dynamics), whereas for *Anopheles fluviatilis*, this relationship was strongly negative. Instead, *Anopheles fluviatilis* associated with temporal cluster 3 (unimodal dry season peaking dynamics). These results are in keeping with the empirical analyses presented in **Figure 3**, supporting the notion that this multinomial logistic regression framework can recover patterns of association evident in the data. Similar results were observed for the environmental predictors where a number of predictors (such as isothermality, rainfall-related patterns and the extent/dynamics of large water bodies) demonstrated highly positive or negative associations depending on the particular temporal cluster. These antagonistic patterns of association across temporal pattern clusters were borne out across an analysis of the correlation of coefficients between clusters, which revealed them to be highly negatively correlated **(Supplementary Figure 5).**

**Evidence of Discrete Ecological Structuring Across Temporal Pattern Clusters:** We next sought to assess the similarly of the coefficients for species and the environmental predictors across the temporal pattern 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*, *Anopheles dirus* and *Anopheles stephensi* all displayed similar small associations with the temporal pattern clusters suggesting the absence of a strong species-specific tendency to adopt a particular temporal pattern (and suggesting a role for ecological structure as 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 shown in **Figure 4B**) with that particular temporal cluster pattern. The 20 coefficients with the strongest association for each temporal cluster pattern were then selected. The extent of overlap in these top 20 coefficients between clusters was assessed – these results revealed minimal overlap between clusters i.e. that each temporal pattern cluster tended to associate with a set of ecological factors that were distinct and unique to that temporal pattern cluster **(Fig.5B)**. These results highlight distinct ecological structuring of the factors associated with each temporal pattern cluster, suggesting that the different temporal patterns characterising each cluster are a product of distinct ecological forces.

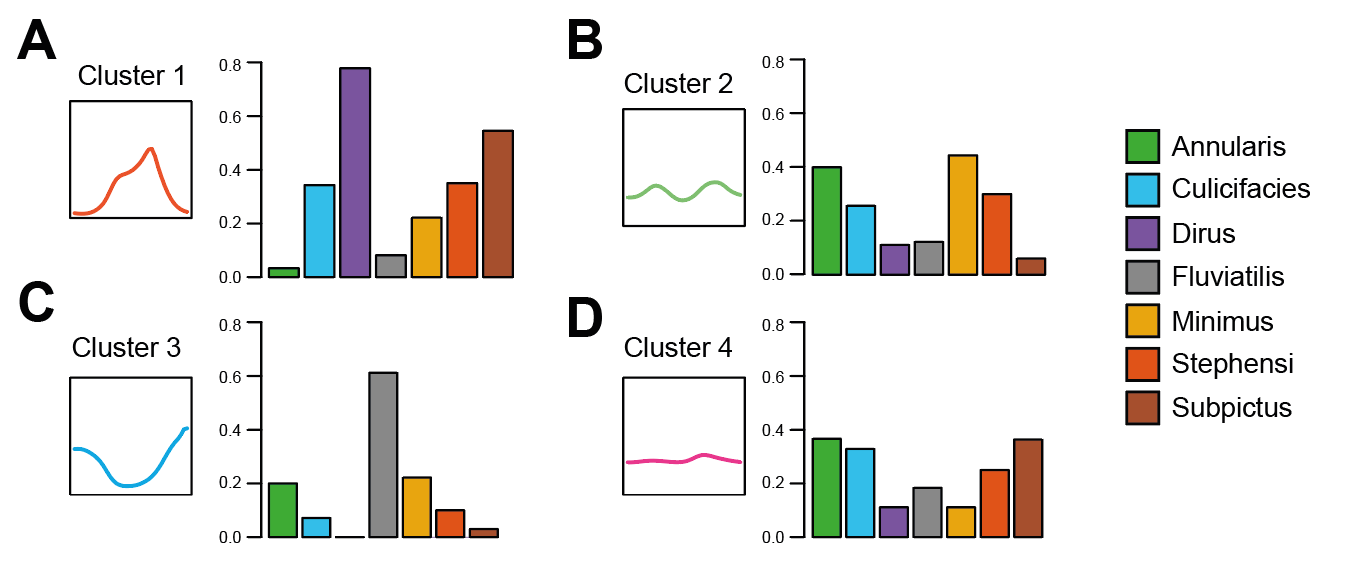
**Predictive Mapping Highlights the Extensive Variation in Seasonal Mosquito Dynamics and Malaria Risk Across India.** We next integrated the results of this multinomial modelling with recently generated spatial predictions of mosquito species presence/absence to produce predictive spatial maps of mosquito population dynamics across the Indian subcontinent. Specifically, we used this framework to generate estimates of the probability that a given location contains at least one mosquito species displaying each of the four different temporal patterns identified through the clustering approach **(Fig.5)**. Our results predict that monsoon concordant dynamics (corresponding to Cluster 1) are most likely in the North and Northeast **(Fig.5A)**. This is in contrast to the predicted spatial distribution of bimodal dynamics (Cluster 2), which are predicted to be most likely in all of 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 (Cluster 4, **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 the Indian subcontinent.



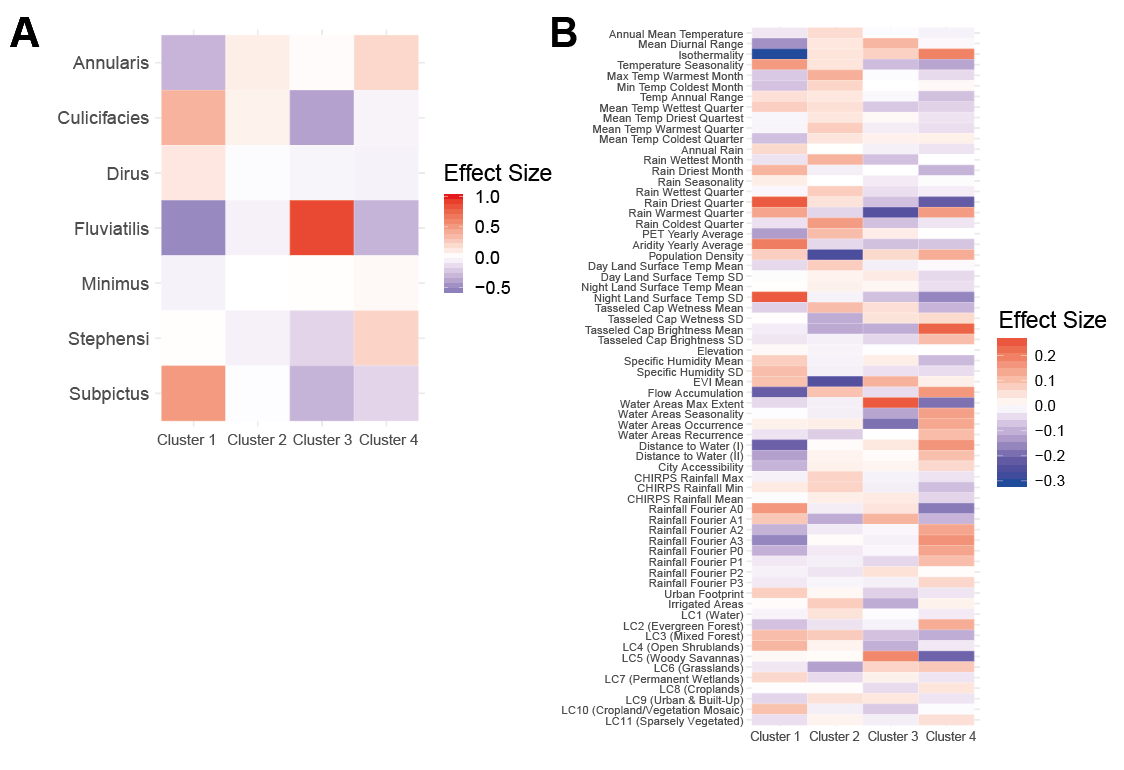
**Figure 1: Exploring and Mapping Species Specific Patterns of Mosquito Population Dynamics.** Negative Binomial Gaussian Processes incorporating a periodic kernel were fitted to each of the 272 time series collected from 118 locations across India collated as part of the systematic review. These fitted 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.

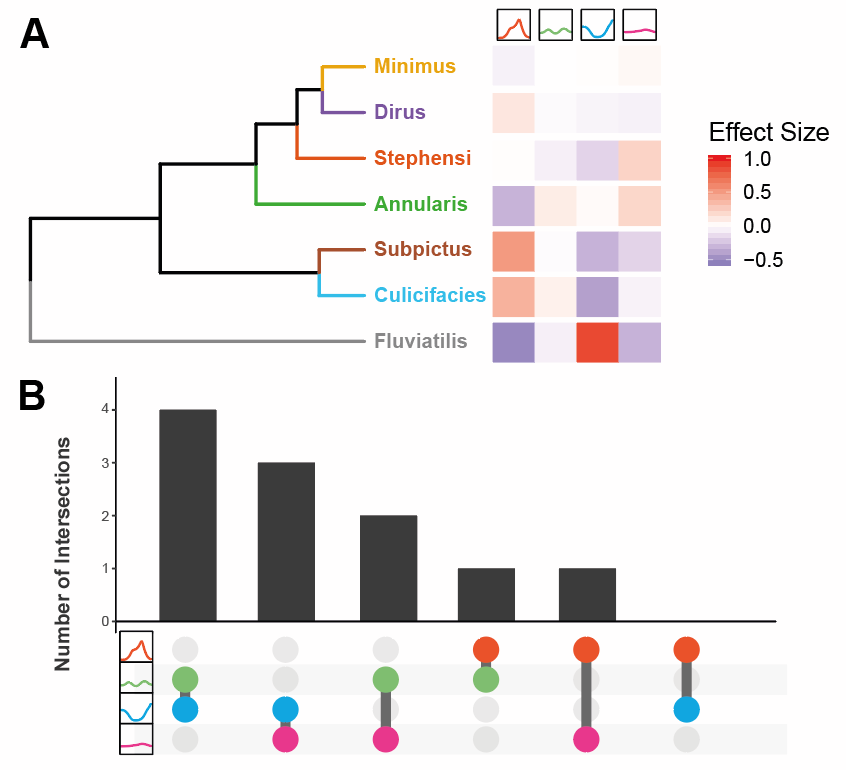


**Figure 2: Characterisation and Clustering of Time Series with Similar Temporal Properties.** A series of mathematical operations were applied to the fitted time series in order to characterise their temporal properties. A Principal Components Analysis was then carried out and the results clustered 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. **(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.

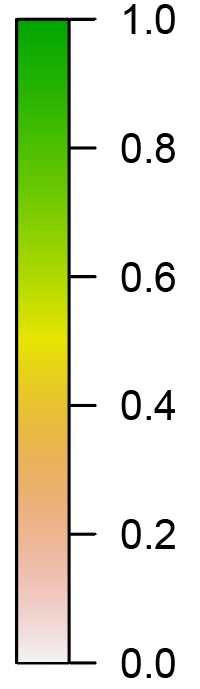
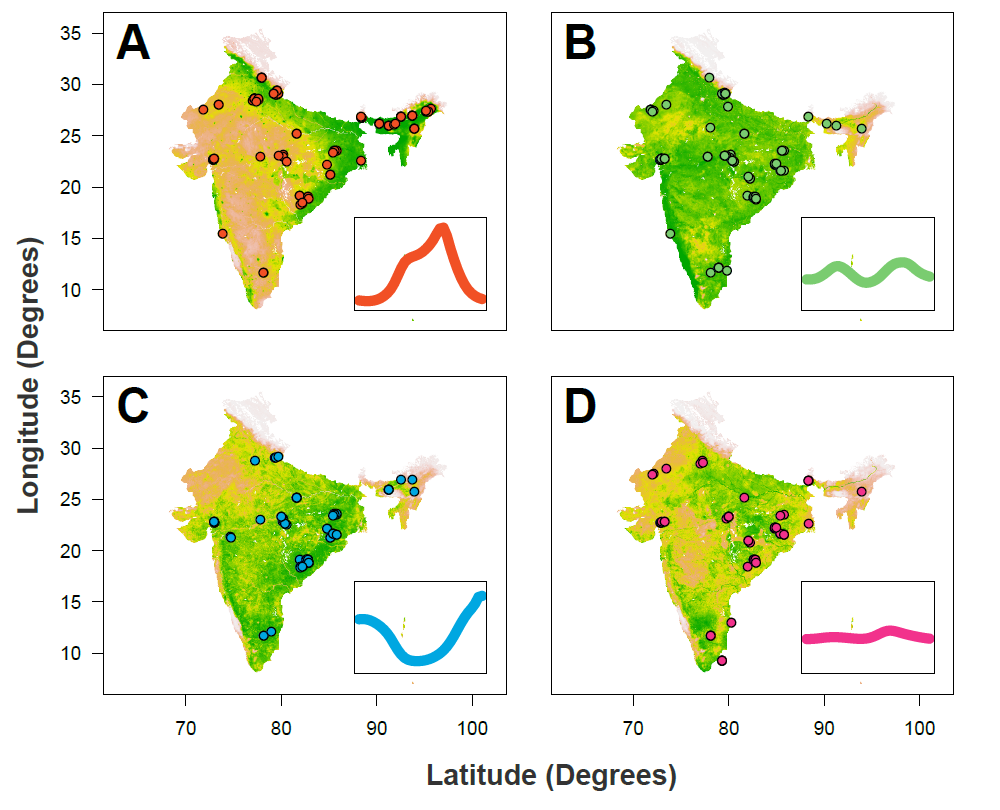


**Figure 3: The Proportion of Time Series Belonging to Each Species and Temporal Pattern Cluster.** Having clustered the time series into distinct temporal pattern clusters, we next assessed the proportion of time series for each species belonging to each temporal pattern cluster. **(A)** For each species, the proportion of time series belonging to temporal pattern 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 temporal pattern cluster 2 (bimodal dynamics). **(C)** As for A, but for temporal pattern cluster 3 (unimodal, dry season peaking dynamics). **(D)** As for A, but for temporal pattern 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 temporal pattern cluster, with this coefficient describing the strength of the association between the predictor (either species or environmental predictor) and membership of a particular temporal pattern cluster. **(A)** The set of coefficients for each species and temporal pattern cluster – red colours indicate a positive association between that species and the temporal pattern 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 temporal pattern cluster i.e. 1 coefficient for each temporal pattern cluster, displayed on the RHS and coloured according to the size 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 (66 for each of the 4 temporal pattern clusters), the 20 covariates with the strongest association with a particular temporal pattern cluster were selected. These top 20 covariates (comprising a set unique to each temporal pattern cluster) were then compared across pairs of temporal pattern clusters to assess the extent of overlap i.e. how many strongly predictive covariates were shared across temporal pattern clusters. In the above plot, the x-axis indicates the specific pairwise comparison being made e.g. between temporal cluster patterns 2 and 3, with the y axis describing the number of shared covariates in the top 20 of these two temporal cluster patterns.



**Figure 6: Predictive Maps of Seasonality Across the Indian Subcontinent.** 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 temporal pattern cluster were empirically observed.

**Discussion**

Understanding the temporal dynamics (including the start, duration and end) of malaria transmission in a given location represents a vital input to optimising malaria control strategies, enabling interventions to be deployed at times and in places where they can have the most impact. Despite this, many outstanding questions remain surrounding the drivers of these dynamics, particularly regarding their dependence on the underlying ecological structure of a setting. Here we leverage a collection of temporally disaggregated mosquito time series catch data collected from across the Indian subcontinent in order to yield new insight into these dynamics and the comparative role of abiotic and species-specific factors in shaping them. Our results reveal extensive diversity in mosquito population dynamics, with the extent of this diversity varying between species and across locations. This diversity can be broadly grouped into a small number of unique temporal patterns and demonstrate clear ecological structuring of these different clusters. Together, these results have enabled a clearer understanding of the comparative contributions of biotic (specific specific preferences) and abiotic (the broader ecological structure of the environment) factors in shaping mosquito population dynamics, and, by extension, the temporal profile of malaria risk.

One key finding is the consistent pattern of abundance peaking during the dry season observed for *Anopheles fluviatilis*. Whilst previous work has identified these dynamics36,37, our work highlights, two important features: firstly that these dynamics are observed across a high proportion of settings in which *Anopheles fluviatilis* is found (rather than being a chance finding in a single location), and secondly that these dynamics are largely restricted to *Anopheles fluviatilis*. These results are consistent with recent work that has identified streams and surrounding stagnant water as the species’ preferred breeding site38; breeding sites that become inundated and unsuitable during the monsoon season but that becoming increasingly suitable as the dry season ensues. By contrast to *Anopheles fluviatilis*, whose temporal dynamics appeared to be relatively consistent irrespective of the ecological setting, we observed extensive spatial heterogeneity in the temporal dynamics of other mosquito species. *Anopheles culicifacies*, a dominant vector thought to be responsible for the majority of malaria transmission in both central and (more recently) northeastern India39,40, was found to display a wide array of different dynamics, ranging from peaking during the monsoon season, to bimodal and even perennial behaviour depending on the sampling site. As with *Anopheles fluviatilis*, these results can likely be understood in the context of the species’ breeding habits. *Anopheles culicifacies* has been shown to be able to exploit a wide range of fresh water breeding habitats41,42, as well as brackish sources43. These contrasting results for *Anopheles culicifacies* and *Anopheles fluviatilis* highlight how the complex interplay between species specific breeding preferences (or the lack thereof) interact with ecological structure (that modulates breeding site availability and productivity) to determine the temporal dynamics of mosquito populations. Together, these results highlight important, species-specific differences in the plasticity of their dynamics, and underscore the crucial importance of considering both species composition and ecological structure into our understanding of malaria transmission.

In addition to these species-specific associations with particular temporal profiles, our results suggest a significant additional role of the environment in shaping mosquito population dynamics. For example, we found cluster-specific associations with ambient temperature and its fluctuations. Temperature is an important determinant of various mosquito traits relevant to malaria transmission – indeed, previous research has highlighted the intimate dependence on temperature of many traits ranging from rates of immature development31 through to biting and mortality rates32. Our results reveal that isothermality (i.e. constant temperature) was positively associated with perennial population dynamics (belonging to Cluster 4), and a strong negative predictor of seasonal dynamics peaking in the monsoon season (Cluster 1). Similarly, temperature seasonality (measuring the extent of temperature fluctuation over the course of a year) was positively associated with seasonal dynamics and negatively associated with perennial dynamics. Taken together, results suggest that distinct ecological processes shape mosquito population dynamics and produce different temporal profiles. This idea is further supported by examination of the cross-correlations for the regression coefficients associated with the environmental predictors used in the model across; extensive negative correlation between regression coefficients was observed between temporal pattern clusters. This negative correlation suggests that clusters are shaped by different, sometimes opposing ecological forces and supports the view of distinct ecological structuring of temporal dynamics. These findings show that mosquito population dynamics represent a complex interplay between biotic (species specific factors including but not limited to breeding site preferences) and abiotic (ecological factors including considerations of both the underlying hydrological environment as well as other factors such as the temperature profile) – and that comparative importance of these two sets of factors depends intimately on the particular species and setting. For some species (such as *Anopheles fluviatilis*), species specific factors dominate yielding dynamics that are similar across ecologically disparate locations. For others (such as *Anopheles culicifacies*), dynamics are far more plastic and subject to the underlying ecological structure of the environment, yielding the extensive diversity observed in dynamics across different locations observed for this species.

Whilst these analyses have yielded insight into the forces shaping mosquito population dynamics, it is important to note the relatively limited predictive performance of the model. There are likely multiple reasons for this, but of primary note is that the environmental covariates used here for prediction were principally developed to map and define static phenomena such as malaria prevalence44 and mosquito presence/absence35, rather than fine grained temporal processes (such as mosquito population dynamics). It is likely that different (although possibly overlapping) ecological features define the range and dynamics of mosquito populations, likely limiting the model’s predictive power. Additionally, our analyses were constrained by the incomplete and often ambiguous nature of location descriptions in papers. This precluded exact delineation of the sampling site in many instances, necessitating aggregation of covariate values up to large spatial scales– at these scales, potentially important features of the microenvironment such as smaller water sources and fine scale variation in ambient temperature (both known to be important factors regulating mosquito populations45–49) might be missed. Future work involving better spatial resolution of sampling sites and the development of novel environmental covariates that can better reflect the ecological processes governing temporal dynamics would therefore likely contribute to improved predictive performance.

Another important limitation of the work presented here is the absence of a link between the modelled mosquito temporal dynamics and the actual temporal profile of malaria risk. Whilst an association between these two quantities is well established, with their profiles typically highly correlated50,51, the nature and extent of this relationship, particularly how it varies across mosquito species, remains less clear. It can frequently be non-linear and vary with multiple interacting factors such as malaria endemicity52, mosquito abundance53 and vector competence34. Our analyses were unable to explore many of these factors: due to the extensive heterogeneity in sampling methods used to catch the mosquitoes and the reported units, we were unable to systematically explore variation in mosquito abundance reported. Similarly, the lack of accompanying epidemiological information precludes us from beginning to better resolve the competence and comparative contributions of different mosquito species to malaria transmission: in doing so, this limits our ability to translate a given temporal pattern into metrics relevant to malaria transmission, such as the Entomological Inoculation Rate (EIR). There is thought to be substantial variation in the relevance of different mosquito species to malaria transmission, with this capacity to transmit determined by a complex interplay of vector-parasite compatibility. Indeed, there is evidence to suggest that *Anopheles culicifacies* contributes disproportionately to transmission across the Indian subcontinent42 (although recent research suggests that vectors previously considered “secondary” e.g. *Anopheles subpictus* might actually have a previously unappreciated role to play in sustaining malaria transmission, particularly in setting characterised by mosquito populations with discordant temporal dynamics54). Whilst we mitigate this limitation somewhat by focussing our analyses specifically on dominant vector species previously established as relevant to malaria transmission in India39, it is not necessarily the case then that each of the mosquito species analysed here are equally relevant to malaria transmission. Future work integrating these analyses with those relating to the seasonality of case incidence (for example the work of Nguyen et al.55 applied to Madagascar) would therefore likely prove instructive.

Overall however, and despite these limitations, our work yields new insight into the drivers of the temporal processes governing malaria transmission. We show that temporal variation in mosquito populations are driven by a complex interplay of biotic and abiotic factors; factors that together yield a wide diversity of different temporal patterns and dynamics. A better understanding of these ecological factors is vital, given that understanding these dynamics in a given location represents a vital and operationally relevant input to optimising various malaria control strategies. Although important further work remains to better resolve additional factors that along with population dynamics define the temporal profile of malaria risk (such as overall mosquito abundance and vector competence), the analyses presented here provides a novel framework facilitating the synthesis of entomological data and in doing so provides new insight into the diversity of dynamics these vectors display.

**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\*” in order to identify references containing temporally disaggregated entomological data. We identified 1945 records with 1556 remaining after duplicate removal. Following Title and Abstract screening 281 records were retained for full text evaluation. We included records that contained temporally disaggregated adult mosquito catch data spanning at least 12 months with at least monthly temporal resolution, that had not been conducted in settings where an active trial of vector control interventions was ongoing, and where sufficient information to geolocate the catch site was provided. Under these criteria 78 references were retained yielding 117 distinct geolocatable areas across the Indian subcontinent. These references contained 272 time-series spanning the known malaria vectors *Anopheles annularis*, *Anopheles culicifacies*, *Anopheles dirus*, *Anopheles fluviatlis*, *Anopheles minimus*, *Anopheles stephensi* and *Anopheles subpictus*. See Supplementary Information “Data Extraction, Collation and Initial Processing” for further details.

**Time Series Fitting and Interpolation**

To smooth the noise in the collated 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. Our model was formulated as:

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 with one another. Weakly informative priors were used. Fitting was undertaken using the programming language STAN56. For more details see Supplementary Information “Negative Binomial Gaussian Process – Fitting and Inference”.

**Time Series Characterisation and Clustering by Features**

Motivated by previous work that provided a framework to statistically characterise the empirical structure of time series data57, as well as recent work characterising the seasonality of malaria case incidence55, we calculated several summary statistics for each smoothed time series in order to further 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 fitted 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 “Time Series Characterisation and Analysis” for further details. From this we obtain a series of 7 real numbers describing the various temporal properties of each time series. We then applied a Principal Components analysis (PCA) to these results to identify a lower-dimensional orthogonal 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 (temporal pattern cluster) – i.e. this clustering assigns each smoothed time series to one temporal pattern 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 ecological structure of the location. These include the BioClimatic variables (a suite of biological relevant covariates defined from monthly rainfall and temperature satellite data58), various measures of aridity59,60, a number of covariates describing the seasonality and extent of water bodies61, the enhanced vegetation index (EVI)62, landcover63 and a number of other related variables previously used in projects mapping the distribution of Anopheline vectors35. A complete list of the covariates used is in Supplementary Table 2. These variables (66 in total) and the Anopheline species (1 for each time series indicating which species the time series belonged to) were used as covariates in a penalised (L2) multinomial logistic regression model to predict the temporal pattern cluster (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 temporal pattern cluster – specifically, 1 coefficient per temporal pattern cluster and predictor, i.e. a total of k coefficients per predictor where k is the number of temporal pattern clusters. The results of these analyses were then integrated with recently produced maps of vector presence/absence to generate predictive maps of mosquito population dynamics across the Indian subcontinent (see Supplementary Information “Penalised Multinomial Logistic Regression Modelling, Evaluation of Model Accuracy and Predictive Modelling” for further detail).

**References**

1. World Health Organization. WHO | World Malaria Report 2016. *WHO* (2017).

2. Phillips, M. A. *et al.* Malaria. *Nat. Rev. Dis. Prim.* **3**, 17050 (2017).

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. Kraemer, M. U. G. *et al.* Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. *Nat. Microbiol.* **4**, 854–863 (2019).

10. 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).

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

12. 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).

13. 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).

14. 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

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

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

17. 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).

18. 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).

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

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

21. Singh, N., Mishra, A. K., Chand, S. K. & Sharma, V. P. Population dynamics of Anopheles culicifacies and malaria in the tribal area of central India. *J. Am. Mosq. Control Assoc.* **15**, 283–90 (1999).

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

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

24. 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).

25. 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).

26. 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).

27. 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).

28. 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).

29. 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).

30. 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).

31. 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).

32. 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).

33. Garske, T., Ferguson, N. M. & Ghani, A. C. Estimating Air Temperature and Its Influence on Malaria Transmission across Africa. *PLoS One* **8**, (2013).

34. 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).

35. 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).

36. 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).

37. 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).

38. 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).

39. 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

40. Akhtar, N., Nagpal, B. N., Kapoor, N., Srivastava, A. & Valecha, N. Role of an. Culicifacies as a vector of malaria in changing ecological scenario of northeastern states of India. *J. Vector Borne Dis.* **53**, 264–271 (2016).

41. 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).

42. 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).

43. 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).

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

45. Gimnig, J. E., Ombok, M., Kamau, L. & Hawley, W. A. Characteristics of Larval Anopheline (Diptera: Culicidae) Habitats in Western Kenya. *J. Med. Entomol.* **38**, 282–288 (2001).

46. Mattah, P. A. D. *et al.* Diversity in breeding sites and distribution of Anopheles mosquitoes in selected urban areas of southern Ghana. *Parasit. Vectors* **10**, 25 (2017).

47. Majambere, S., Fillinger, U., Sayer, D. R., Green, C. & Lindsay, S. W. Spatial distribution of mosquito larvae and the potential for targeted larval control in The Gambia. *Am. J. Trop. Med. Hyg.* **79**, 19–27 (2008).

48. Thomas, S. *et al.* Microclimate variables of the ambient environment deliver the actual estimates of the extrinsic incubation period of Plasmodium vivax and Plasmodium falciparum: A study from a malaria-endemic urban setting, Chennai in India. *Malar. J.* **17**, (2018).

49. Paaijmans, K. P. & Thomas, M. B. The influence of mosquito resting behaviour and associated microclimate for malaria risk. *Malaria Journal* **10**, (2011).

50. 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).

51. 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).

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

53. 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).

54. 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).

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

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

57. 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).

58. 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).

59. 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).

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

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

62. Weiss, D. J. *et al.* An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS J. Photogramm. Remote Sens.* **98**, 106–118 (2014).

63. 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).