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

Charles Whittaker1\*, Peter Winskill1, Marianne Sinka2, Samuel Pironon3, Claire Massey4, Ashwani Kumar5, Azra Ghani1+ & Samir Bhatt1,6+

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

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

3Insert Samuel’s affiliation here

4Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Old Road Campus, Oxford, UK.

5Insert Ashwani’s affiliation here

6Malaria Atlas Project, Oxford Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK

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

+ Denotes joint-authorship

**Significance**

Understanding the temporal profile of malaria risk, which is a product of underlying mosquito population dynamics, is essential to effective planning and control of the disease. Despite this, outstanding questions remain surrounding the drivers and determinants of these dynamics. Here, we collate a database of monthly mosquito catch data spanning y years and x locations across the Indian subcontinent. Statistical modelling and analysis of the temporal properties of these time series reveals marked variation in dynamics, both within mosquito species populations (across different locations) and between species (in the same location). Our results highlight both marked inter-species variation in both timing and extent of seasonality in population size, as well as discrete ecological structuring of these different temporal dynamics. Importantly, we demonstrate that the temporal profile of malaria risk can be clustered into a small number of policy relevant temporal modalities and provide a predictive tool to inform the planning and timing of control efforts.

**Abstract**

**Additionally, understanding the temporal dynamics (including the start, duration and end) of malaria transmission in a given location represents a vital input to optimising control strategies – our work yields new insight into the drivers of malaria’s temporal dynamics blah blah blah.**

**Background**

Responsible for an estimated 216 million cases and 454,000 deaths in 2016 alone1, malaria represents one of the most serious infectious diseases globally2. The transmission dynamics (perennial or annual, hypo, meso or hyperendemic etc) of this disease are characterised by extensive variation between settings, a phenomenon that has led to its description as “a thousand different diseases”3. This variation is a product of a diversity of ecological and human processes that together shape and structure the epidemiology of malaria infection across different settings.

Perhaps one of the factors most relevant to determining the nature and dynamics of malaria transmission is the mosquito vector; specifically, members of the *Anopheles* genus responsible for transmission of the *Plasmodium* parasite between humans. Of the 460 recognised members of the genus, approximately 70 possess the capacity to transmit human malaria parasites4, with 41 having been identified as dominant vector species responsible for the majority of transmission that occurs5. These vectors are highly 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 the spatial distribution (presence/absence) of these vectors including mapping their current distribution across different global regions6,7 as well as the trajectories and dynamics of these distributions over time in response to global climate change and other relevant factors8. This work has involved relating vector occurrence to a variety of different environmental variables (typically satellite derived) and has been highly useful in enabling strategically targeted surveillance and control programmes aimed at mitigating the impacts of vector borne diseases worldwide.

By contrast, less attention has been paid to trying to understand the temporal patterns of vector abundance, and how these are dynamics are shaped by the local environment. Mosquito populations are highly dynamic over time, exhibiting substantial fluctuations in size over the course of a year, a feature which in turn leads a similarly dynamic and temporally variable profile of vector borne disease risk. Rainfall is frequently considered a key determinant of these dynamics due to the requirement of the early life cycle stages of the mosquito for an aquatic habitat in which to develop and the preference many species display for transient, rain fed pools of water in which to breed9. Whilst a close relationship between the occurrence of rainfall and peaks in malaria infection in many locations has been observed10, these relationships are often far from clear-cut. Previous studies have observed substantial variability in the temporal patterns of mosquito population dynamics in different species residing in the same location (suggesting that these dynamics can be independent of rainfall under some circumstances, and influenced by other factors11), as well as within species across different locations (pick two refs where culicifacies shows different dynamics).

Instead, it appears that the influence of rainfall on mosquito breeding site availability (and by extension mosquito population dynamics) represents a complex combination of the intensity and duration of precipitation an area receives (which can either generate new breeding sites or destroy them based on the level of rainfall12–14), species specific breeding site preferences (transient or permanent water bodies, polluted or clean water sources etc, observed for multiple species across both African and Asian settings15–18) and the ecological structure of the local hydrological environment (which determines the comparative composition of permanent and temporary aquatic habitats available, how these habitats respond to rainfall and their physio-chemical characteristics19).

Together, these factors combine to shape and structure mosquito population dynamics (and by extension, the temporal dynamics of malaria risk) in a given location. Whilst these patterns are frequently concordant with rainfall, numerous settings exist where this is not the case. The population dynamics of *Anopheles gambiae* were shown to be highly concordant with rainfall during extensive entomological investigations in the Garki District of Nigeria20. *Anopheles funestus* by contrast frequently displays a lack of the marked seasonal fluctuations in population abundance that characterises other anopheline species, a feature possibly attributable to its preference for large permanent/semi-permanent bodies of fresh water as its breeding habitat18 and thought to be a substantial contributor to the perennial malaria transmission observed across parts of Eastern Africa21,22. Similarly, *Anopheles subpictus* can breed in a variety of habitats ranging from fresh to brackish water and including highly polluted sources15, a feature that likely contributes to its capacity in sustaining year round malaria transmission in the urban setting of Goa, India23. Complicating matters further, not all variation in population dynamics appears to be species-specific, with numerous instances of variation in the seasonal dynamics displayed by both *Anopheles culicifacies* and *Anophleles annularis* reported in the literature **(Refs here)**.

Mosquito population dynamics are therefore likely driven by a combination of both factors specific to particular mosquito species, as well as properties of the local hydrological environment and how these respond to incipient rainfall. Understanding these drivers and their comparative importance is crucial given that some of the most effective malaria control interventions are critically dependent on accurate timing of delivery in relation to seasonal peaks in disease risk. Seasonal Malaria Chemoprevention is a highly effective means of controlling malaria in areas with highly seasonal, short transmission seasons24, involving monthly administration of long-acting antimalarials to children under 5 (irrespective of malaria infection status) just before and then during the seasonal peak in transmission. Similarly, Indoor Residual Spraying (spraying houses with insecticides to kill mosquitoes) represents a highly effective means of achieving reductions in the burden of malaria in endemic communities25. Timing and frequency of spraying is dependent on both the properties of the insecticide used and the seasonality of transmission, with perennial settings typically requiring more frequent spraying. A better understanding of the factors underlying both the timing and extent of seasonal fluctuation in mosquito populations (and by extension, malaria risk) would therefore enable more accurate and effective deployment of these highly efficacious interventions.

Here we collate temporally disaggregated mosquito catch data from across the Indian subcontinent in order to better understand both the extent and drivers of variation in mosquito population dynamics. India is well positioned to in this respect due to the high diversity of malaria-competent vectors it is home to26, as well as its history of publishing high quality, temporally resolved and comprehensive entomological data. Using this data, we employ an array of statistical methods to characterise the observed temporal patterns and uncover structural similarities in the dynamics across populations and locations. Our results reveal pronounced heterogeneity in the extent, nature and dynamics of seasonal fluctuations across different mosquito populations. They underscore the critical interaction between abiotic and biotic factors in determining the population dynamics of this crucial disease vector and the importance of considering ecological structure when designing malaria intervention strategies.

**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. Our searches identified a total of 1945 records, with 1556 remaining after duplicate removal. Following Title and Abstract screening, a total of 281 records were retained for full text evaluation. Records that contained temporally disaggregated mosquito catch data spanning at least 12 months at a monthly (or better) temporal resolution and that had not been conducted in settings where an active trial of vector control interventions was ongoing were included. Those not satisfying these criteria were excluded, as well as both records where geolocation was not possible due to a lack of spatial information and records which contained information on the immature/larval mosquito life cycle stages only. Using these criteria, a total of 78 references were retained following Full Text Evaluation, yielding 106 distinct geolocatable areas across the Indian subcontinent. Together, these references contained a total of \_\_\_ 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**

In order to facilitate spatial interpolation between the monthly catches, and to smooth the inherent noise in the collated mosquito catch data (without distorting the legitimate temporal patterns present), we fitted a highly flexible class of stochastic models known as Gaussian Processes to each of the extracted time series. Gaussian Processes represent a distribution over functions such that any finite set of function values (our monthly timepoints in this case) have a joint Gaussian distribution. The Gaussian Process is entirely specified by its mean function, defined as:

and by its covariance function:

which defines the covariance between any two points, in doing so, encoding information about the extent to which we would expect two points (separated in time) to be similar. Mosquito catch data is frequently overdispersed, a phenomenon arising both from the inherent biological variation present in the system, as well as the fact that a number of time series were comprised of single points made up on aggregated weekly counts – both of these typically lead to high levels of overdispersion within systems and so the Gaussian Processes were fitted using a Negative Binomial likelihood in order to account for this:

where is a vector of values representing a realisation from Gaussian Process, our observed counts and represents a vector of hyperparameters involved in defining the functional form of the covariance function. Given mosquito population dynamics are typically characterised by repeating patterns occurring either seasonally or annually, a periodic kernel was used, with the following specification:

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 etc), specifies the magnitude of the covariance, and represents a lengthscale parameter further constraining the extent to which two values separated by a given distance (in time in our case) can co-vary with one another. Weakly informative priors were set on all of the parameters except for the lengthscale and the Negative Binomial Gaussian Process fitted to each of our extracted time series using the probabilistic programming language STAN, with a total of 10,000 iterations used for model fitting and parameter inference. For more details see Supplementary Information “Negative Binomial Gaussian Process – Fitting and Inference”.

**Time Series Characterisation and Clustering by Features**

To the fitted time series, we applied a series of mathematical operations to characterise their dynamics and 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 twice the mean (measuring how peaked the time series is), the distance of the first peak from January, and then 3 features arising from fitting Von Mises distributions to the fitted time series. The Von Mises distribution is a continuous probability distribution on the circle with range from 0 to , with the probability density function for the angle given by:

where is the modified Bessel function of order 0, the parameter is a measure of location (analogous to the mean of the normal distribution, describing where on the circle the distribution is clustered around) and describes the concentration of density around (and thus its inverse is a measure of dispersion, analogous to for the normal distribution. We fit two sets of Von Mises densities, one containing a single component:

and the other possessing two components, formulated as:

where in both instances represents the normalised monthly count formulated as a random variable on the circle, i.e. by defining . Fitting was carried out in R using the *optim* function and with the sum of squares as the loss function. Three features from this fitting were used to characterise the time series properties: the mean of the 1 component fitted Von Mises distribution, the number of peaks (determined by comparing the quality of fit, see Supplementary Methods for more details), and the weight (), specifying the comparative contributions of each component in the two-component fitting.

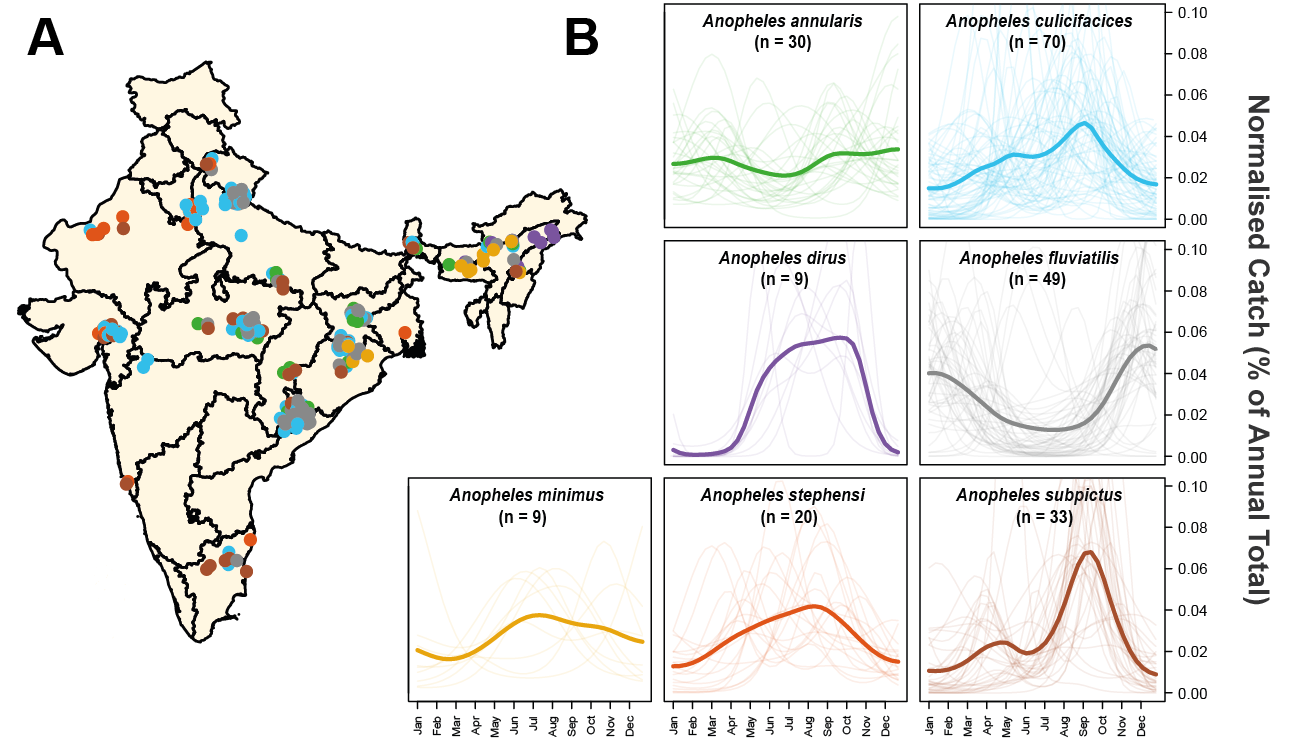
The results of applying these mathematical operations to each time series were then subject to a Principal Components analysis, allowing us to identify a low-dimensional representation of the structure present in the data amenable to visualisation but that retains the majority of the information. We then clustered the time series based on this low-dimensational representation to identify groups of time series with similar temporal features. Clustering was carried out using the k-means algorithm and using the first four principal components (which together accounted for over 95% of the total variation in the dataset).

**Statistical Modelling and Prediction of Seasonal Modality**

For each of the 106 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 ranfall and temperature satellite data (Ref)), various measures of aridity, a number of covariates describing the seasonality and extent of water bodies, the enhanced vegetation index (EVI), landcover and a number of other related variables previously used in projects mapping the distribution of *Anopheline* vectors (Ref). For further information and a complete list of the covariates used, see Supplementary Table 2. These variables were then used as covariates in a series of statistical models predicting the probability of a given location containing a particular temporal pattern. Specifically, a logistic regression model was fitted separately to one-hot-encodings for each identified temporal cluster (i.e. the response was encoded as a 1 if the time series belonged to that particular temporal cluster, and 0 otherwise), using an L1 penalty for purposes of regularisation (the so-called “LASSO”). This was implemented in R using the package glmnet. Receiver operator characteristics and AUC values were derived using k-fold cross-validation with 10 folds, using half of the available data as the training set and the other half for testing predictive accuracy.

**Results**

**Substantial Diversity in Mosquito Population Dynamics Both Within and Between Species:** A systematic review (see Methods and Supplementary Figure 1) was undertaken to identify references containing temporally disaggregated mosquito catch data spanning at least a year, that had taken place on the Indian subcontinent, and that related to one of seven dominant vector species known to transmit malaria within endemic populations. A total of \_\_\_ references, containing \_\_ time series from \_\_ locations across India were identified **(Fig 1A)**, and the data contained in these references extracted. To these time series (all at monthly resolution and spanning a period of 12 months following initial data pre-processing, see Supplementary Methods Description of Systematic Review, Data Extraction, Collation and Initial Processing for further information), we fitted flexible curves using negative binomial gaussian processes in order to interpolate between timepoints and smooth the typical overdispersion associated with mosquito catch data. We first began by plotting the fitted time series data by species. This was done in order to assess the overall extent of variation observed in mosquito population dynamics across the collated time series, and additionally, to see whether different species displayed distinct temporal patterns. Our results reveal pronounced heterogeneity between different species in their temporal population dynamics over the course of a year **(Fig 1B)**. Whereas *Anopheles dirus* populations peaked during the monsoon period (typically June to September), by contrast, in almost every instance, *Anopheles fluviatilis* populations peaked around November to February (the dry season across most of India), and reached their lowest point 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 then during the August – September period, whilst *Anopheles annularis* populations appeared to remain comparatively static across the entirety of the year. In addition to this putative species-specific variation, we also frequently observed extensive variation within species. For example, *Anopheles culicifacies* populations varied significantly in both the extent and timing of their seasonal peaks: whereas some displayed sharp peaks in the monsoon season, others more resembled *Anopheles annularis*, being present across the course of the entire year.

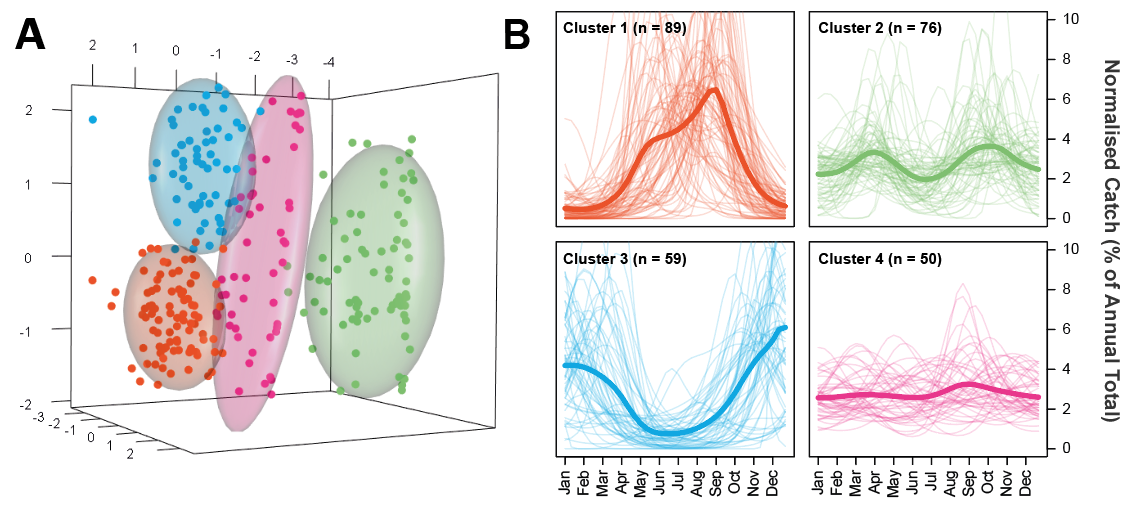


**Figure 1: Species Specific Patterns of Mosquito Population Dynamics.** Negative Binomial Gaussian Processes incorporating a periodic kernel were fitted to each of the 220 time series collected from \_\_ locations across India collated as part of the systematic review. These fitted time series (representing 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.

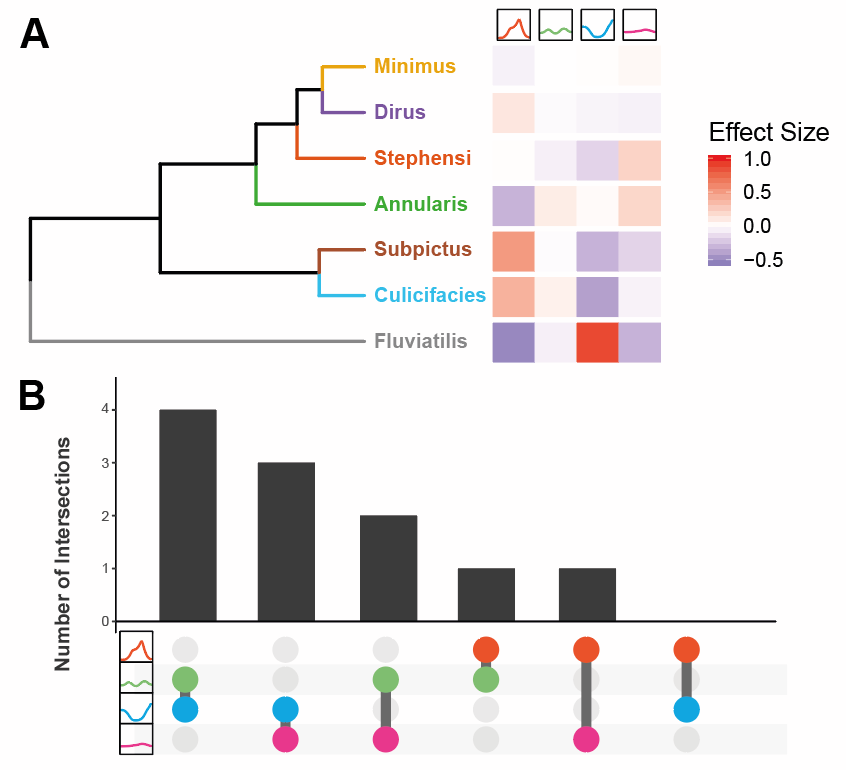
**Characterisation and Clustering of Mosquito Catch Time Series Properties Reveals Discrete Temporal Modalities:** Having identified substantial between-species variation in the temporal dynamics of mosquito populations, and the suggestion of some within-species variation also, we next asked whether the time series we had collated could be grouped into distinct clusters based on their temporal properties. Motivated by this, we applied a series of statistical operations to each time series in order to characterise its temporal features (including the number of peaks, position and magnitude of peaks, amongst other operations, see Supplementary Methods Description of Statistical Methodologies Utilised for further details). We then carried out a Principal Components Analysis followed by k-means clustering in order to cluster time series into groups with similar temporal properties **(Fig 2A)**. Our results revealed the presence of four clusters of time series with distinct temporal properties. Broadly, these could be classified into those peaking during or near to the monsoon season **(Fig 2B Cluster 1)**, time series displaying two peaks across the course of the year **(Fig 2B Cluster 2)**, those peaking during the dry season **(Fig 2B Cluster 3)**, and finally, time series displaying limited temporal variation across the course of the year **(Fig 2B Cluster 4)**. Importantly these dynamics appear to occur despite similar rainfall patterns and timings across all of the locations we examined – for each location, time-specific rainfall information was extracted from the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) dataset (Ref), and the cross-correlation between rainfall and each time series calculated **(Supplementary Figure 2)**. For Cluster 1, population dynamics were highly positively correlated with rainfall, as we would expect given the populations belonging to this cluster peak during the monsoon season. By contrast, time series belonging to Cluster 3 displayed highly negative correlations with rainfall, indicating that the observed patterns (i.e. peaks between November to February) were not a product of those areas receiving rainfall at different times of the year to other locations, and instead were genuinely representative of population dynamics that peaked during periods of dry weather. Clusters 2 and 4 showed limited correlation with rainfall, in-keeping with their flattened, less temporally dynamic profile across the course of the year.

**Mosquito Population Dynamics are Determined By A Complex Interplay of Abiotic and Biotic Factors**

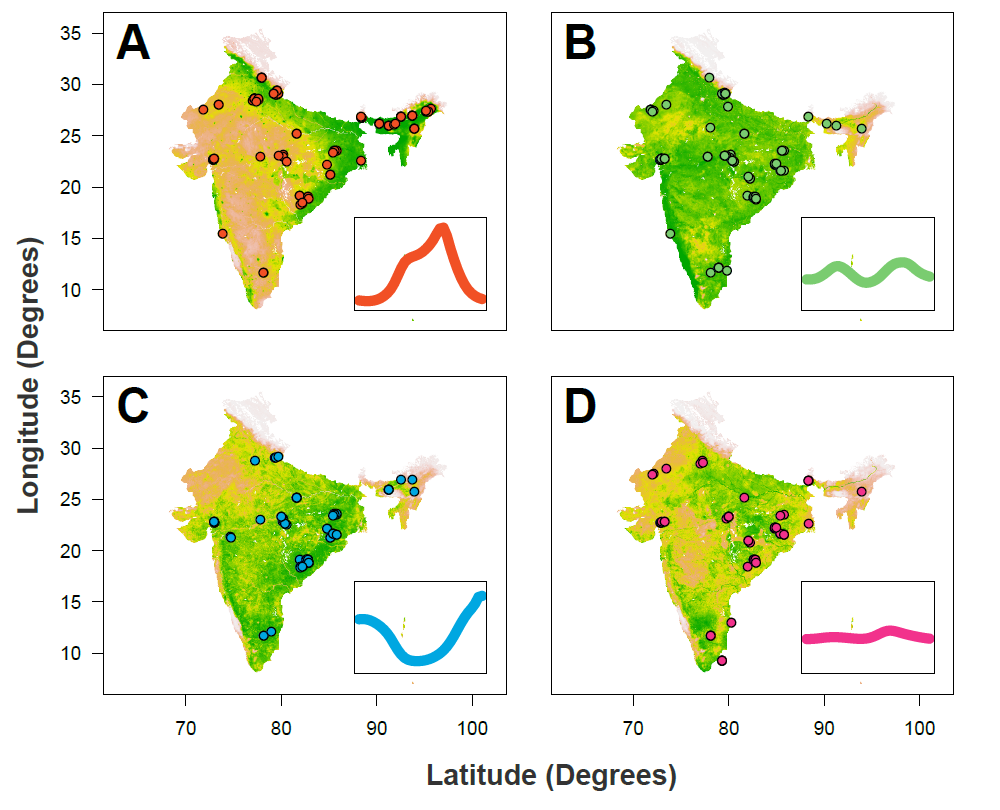
**Predictive Mapping Highlights the Extensive Variation in Seasonal Mosquito Dynamics and Malaria Risk Across India**



**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 and then clustered based on these features using the k-means algorithm. **(A)** Results of the k-means clustering algorithm for 4 clusters, as applied to the first 4 components (only 3 shown) of the 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, darker 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:**



**Figure 4:**

**Results**

**Figure 3** – Regression Model Results Predicting Seasonal Modality

* Detailing the output of model fitting including a) a visual description of some (all?) of the covariates used, b) a table containing the model estimated effects and c) some sort of bar chart detailing the comparative contribution of different components (possibly coloured by abiotic vs biotic as the main colour, and then variations on each colour to distinguish the individual covariates from one another.

**Figure 4 –** Stochastic Model Results

* Present a) simplified version of the differential rainfall driven mosquito model and b) the output of model fitting to the Das (Dev?) et al., dataset which contains lengthy time series for *Anopheles annularis*, *Anopheles culicifacies* and *Anopheles fluviatilis*. Either just the mosquito catch outputs, or also the carrying capacity for each of the three modalities.

**Figure 5 –** Malaria TransmissionModelling

* Integrate the developed mosquito model with the full Imperial model. Using this model, simulate malaria transmission under the three main modalities (seasonal peak wet season, seasonal peak dry season, static). Then simulate a small number of vector control intervention e.g. larviciding and indoor residual spraying to highlight the importance of seasonality on the efficacy of vector control interventions.
* Need to note in the paper that this doesn’t incorporate considerations of vector bionomics, but is merely intended to illustrate the capacity of seasonality to influence these interventions. Future work should and will look at incorporating bionomic considerations.

**Discussion**

**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. Study., L. H.-M. in E. A. E. & 1937, undefined. Malaria in Europe. An Ecological Study. *cabdirect.org*

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

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

6. 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

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

8. 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).

9. 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).

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

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. Fillinger, U., Sonye, G., Killeen, G. F., Knols, B. G. J. & Becker, N. The practical importance of permanent and semipermanent habitats for controlling aquatic stages of Anopheles gambiae sensu lato mosquitoes: operational observations from a rural town in western Kenya. *Trop. Med. Int. Heal.* **9**, 1274–1289 (2004).

13. 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).

14. Paaijmans, K. P., Wandago, M. O., Githeko, A. K. & Takken, W. Unexpected High Losses of Anopheles gambiae Larvae Due to Rainfall. *PLoS One* **2**, e1146 (2007).

15. Kumar Singh, R., Kumar, G. & Pradeep Kumar, Mittal Chand Dhiman, R. U Bionomics and vector potential of Anopheles subpictus as a malaria vector in India: An overview. *Int. J. Mosq. Res.* **1**, 29–37 (2014).

16. Konradsen, F., Piyaratne, M. K., Amerasinghe, F. P., Amerasinghe, P. H. & Konradsen, & F. *Physico-chemical characteristics of Anopheles culicifacies and Anopheles varuna breeding water in a dry zone stream in Sri Lanka*. *J Vect Borne Dis* **42**, (2005).

17. MINAKAWA, N. *et al.* SPATIAL DISTRIBUTION OF ANOPHELINE LARVAL HABITATS IN WESTERN KENYAN HIGHLANDS: EFFECTS OF LAND COVER TYPES AND TOPOGRAPHY. *Am. J. Trop. Med. Hyg.* **73**, 157–165 (2005).

18. 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).

19. 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).

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

21. Kaindoa, E. W. *et al.* Interventions that effectively target Anopheles funestus mosquitoes could significantly improve control of persistent malaria transmission in south-eastern Tanzania. *PLoS One* **12**, (2017).

22. Kelly-Hope, L. A., Hemingway, J. & McKenzie, F. E. Environmental factors associated with the malaria vectors Anopheles gambiae and Anopheles funestus in Kenya. *Malar. J.* **8**, 268 (2009).

23. 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).

24. 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).

25. 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

26. 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).

**Miscellaneous**

**The Ecology of Malaria Transmission – Drivers and the Importance of Understanding It**

1. Something about malaria/vector borne diseases being an huge public health burden and renewed interest in combatting them.
2. Substantial heterogeneity exists in the epidemiology of malaria and its transmission dynamics across different settings. The drivers of these divers transmission ecologies are equally diverse and include variation in the parasite composition (falciparum, vivax, increasingly knowlesi etc), as well as an array of socioeconomic factors surrounding access to healthcare, knowledge and attitudes, exposure patterns (frequently related to occupation) etc.
3. Malaria transmission dynamics are also markedly shaped by the local ecology of a region. A variety of climatic factors as well as the physical properties of the local environment significantly influence the disease’s epidemiological and ecological dynamics. This influence is typically (though not always) realised through the impact on the physiology of the mosquito vector, and by extension, on the dynamics of mosquito populations.

**Variation and Diversity in Mosquito Population Dynamics**

1. Whilst rainfall is frequently considered a key determinant of mosquito population dynamics, with the expectation that mosquito populations show tight concordance with rainfall patterns, this is not always the case (important because many malaria interventions are seasonal, or at least, most efficacious when deployed seasonally, e.g. SMC and vector control interventions such as IRS).
   1. In fact, across large parts of the world, including places such as East Africa and India, home to diverse mosquito populations, this is not the case.
      1. The exact relationship with rainfall will depend on a complex interplay between mosquito hydrological habitat breeding preferences and the biotic and abiotic structure of the local environment.
      2. Understanding these dynamics not only provides insight into the comparative influence of abiotic/biotic factors on population dynamics, but also has material consequences for disease transmission. Mosquitos are vectors of some of the world’s deadliest diseases, including dengue, chikungunya and yellow fever amongst others.

**Substantial Uncertainty Still Exists, Drivers Poorly Resolved**

1. Importance of understanding temporal dynamics not only theoretically interesting, but also vital given its relevance to disease ecology and transmission.
   1. Despite relevance to ecological theory and to malaria control, both the extent of heterogenenity in mosquito population dynamics and the drivers underpinning this heterogeneity remain poorly resolved. Previous studies have speculated driven by water availability blah blah blah, whilst others have noted across Africa that funestus/arabiensis related stuff etc etc. But understanding of comparative importance of different factors remains outstanding.
2. A substantial body of work has focussed on understanding and predicting the geographical extent of these vector populations. By contrast however, comparatively little attention has been paid to the vector’s temporal dynamics.
   1. Much work has focussed on the geographical range and distribution of these vectors (Sinka papers, Kramer’s recent Nature Microbiology one). However, presence or absence of a vector is only 1 component (though noted an absolutely necessary one) determining the extent and dynamics of disease transmission in a given area. The absolute abundance of the vector, and the temporal dynamics of the vector population being the others.
   2. This is despite the relevance to malaria transmission dynamics and hence control strategies.

**Motivations, Overview of Presented Work**

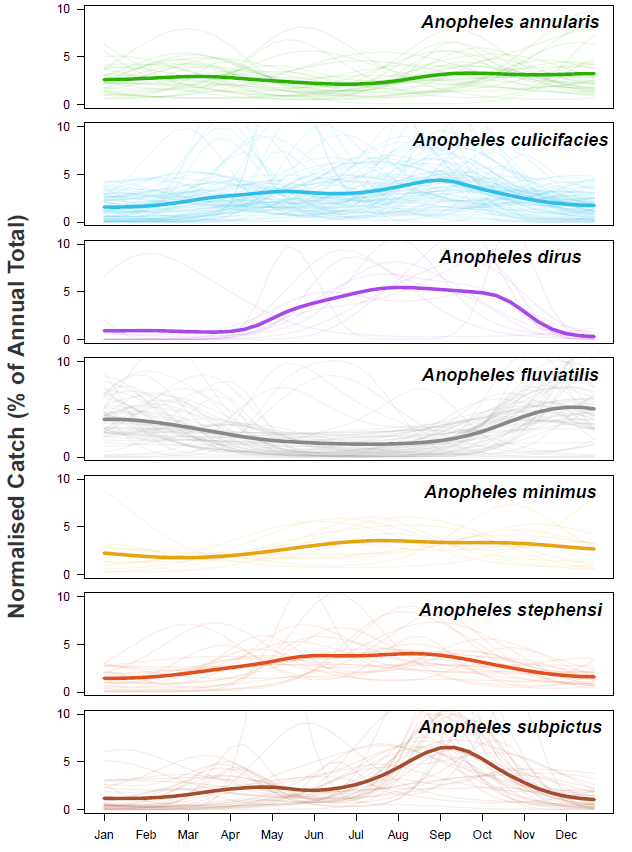
1. Here we collate and leverage a substantial collection of temporally disagreggated mosquito catch data from across the Indian subcontinent. Here we utilise Indian time series to begin to explore and answer these questions.
   1. Here, we utilise a substantial collection of mosquito time series data from across the ecologically diverse setting on India.
   2. India is well positioned to be studied and explore these dynamics because 1) there is an abundance of different mosquito species present, and similarly an abundance of different malaria vectors. Because 2) it is geographically and ecologically diverse. And 3) it has a well established history of publishing high quality, temporally resolved and comprehensive entomological data.
   3. Use this to explore variation in the extent and nature of seasonality in mosquito population dynamics, across different species and different ecological settings.
   4. Our results reveal pronounced heterogeneity in the nature and dynamics of seasonal fluctuations in mosquito populations.
   5. We utilise a range of statistical methods to assess the comparative contribution of biotic and abiotic factors to this heterogeneity, with our results indicating that a substantial proportion of the variability in dynamics is species specific.
   6. **EITHER:**
      1. We extend our results from the study sites across India to predict the dominant seasonal pattern there OR:
      2. We develop a mathematical model of mosquito population dynamics to explore the consequences of this heterogeneity. Fitting this to Das et al yada yada yada maybe some stuff about malaria control.
2. Our work highlights the diversity of population dynamics displayed by endemic mosquito species in India, even within the same environment.
   1. More broadly, our work highlights the critical interaction between abiotic and biotic factors in determining the dynamics of wildlife populations, and the importance of considering these ecological dynamics when designing malaria intervention strategies.

**Figure 1 –** Time Series Results

* Plot processed (negative binomial GP with periodic kernel, raw becomes a Supplementary Figure) time series for each of the species considered here. Either way, make sure the time series are normalised.
* Plot the time series either linearly (which I think would be best) or as radar plots (less intuitive to understand I think.

**Figure 2 –** PCA Characterisation of Time Series Features

* Some sort of composite feature visually detailing in a) the operations applied to each time series and what feature they’re extracting and b) the resulting PCA plot for the time series, with the points coloured by cluster.
* Also feature a series of plots plotting the time series belonging to each identified cluster, illustrating the capacity of our method to identify time series with distinct temporal modalities.

****

**Miscellaneous**

**Systematic Review of Indian Entomological Literature**

* Systematic review of Web of Science and PubMed. Note the number of articles available at each stage of the process. Describe the:
  + Data extraction process, namely how and what was extracted.
  + Collation and collection of satellite derived abiotic covariates, namely from Google Earth Engine. Also note that biotic components (other species presence) was inferred from the paper.
  + Geolocation process – combination of MAP’s previous work, and my own Google Earth/Indian town websites/the other websites I found.

**Time Series Fitting, Characterisationand Clustering**

* Describe the pre-processing of time series that occurred and then the negative binomial gaussian processed that was used to fit the processed time series.
* Detail each of the operations used to characterise the smoothed time series, and the PCA/clustering technique used downstream.

**Stochastic Model of Mosquito Population Dynamics – Structure and Fitting**

* Describe the basic mosquito model, as formulated in a compartmental deterministic manner, followed by an explanation of the stochastic extension.
* Describe the carrying capacity calculation, how its calculated and the different modalities the model is capable of producing.

**Malaria Transmission Model – Structure and Scenarios Considered**

* Briefly describe the full malaria transmission model but focus on the changes made, which will be to incorporate the newly developed mosquito model.
* Discuss the range of scenarios considered, focussing on 1) what parameters in the model are changed to reflect the effect of an intervention and 2) the values of these modified parameters.