Local human movement patterns and land use impact exposure to zoonotic malaria in Malaysian Borneo

Kimberly M Fornace*[1], Neal Alexander[2], Tommy R Abidin[3], Paddy M Brock[4], Tock H Chua[3], Indra Vythilingam[5], Heather M. Ferguson[4], Benny O. Manin[3], Meng L. Wong[5], Sui Hann Ng[3], Jon Cox[1], Chris J Drakeley[1]

1. Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK
2. Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
3. Department of Pathobiology and Medical Diagnostics, Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia
4. Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, UK
5. Parasitology Department, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

*** Corresponding author:**

Kimberly M Fornace

Kimberly.Fornace@lshtm.ac.uk

London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK

+44 (0) 20 7927 2010

Abstract: (150 max)

Human movement into insect vector and wildlife reservoir habitats determines zoonotic disease risks; however, few data are available to quantify the impact of land use on pathogen transmission. Here, we utilise GPS tracking devices and novel applications of ecological methods to develop fine-scale models of human space use relative to land cover to assess exposure to the zoonotic malaria *Plasmodium knowlesi* in Malaysian Borneo. Combining data with spatially explicit models of mosquito biting rates, we demonstrate the role of individual heterogeneities in local space use in disease exposure. At a community level, our data indicate that areas close to both secondary forest and houses have the highest probability of human *P. knowlesi* exposure, providing quantitative evidence for the importance of ecotones. Despite higher biting rates in forests, incorporating human movement space use into exposure estimates illustrates the importance of intensified interactions between pathogens, insect vectors and people around habitat edges.

Keywords (min 3):

Disease ecology, spatial epidemiology, *Plasmodium knowlesi*, malaria, human movement, land use

Introduction:

Environmental change and human encroachment into wildlife habitats are key drivers in the emergence and transmission of zoonotic diseases (1, 2). Individual movements into different habitats influence exposure to disease vectors and animal reservoirs, determining risk and propagation of vector-borne diseases (3-5). Increased contact between these populations is theorised to drive increases of the zoonotic malaria *Plasmodium knowlesi* in Malaysian Borneo, now the main cause of human malaria within this region. *P. knowlesi* is carried by long- and pig-tailed macaques (*Macaca fascicularis* and *M. nemestrina*) and transmitted by the *Anopheles leucospryphus* mosquito group, both populations highly sensitive to land cover and land use change (6). Although higher spatial overlap between people, macaques and mosquito vectors likely drives transmission, the impact of human movement and land use in determining individual infection risks is poorly understood (7).

The emergence of the zoonotic malaria *Plasmodium knowlesi* has been positively associated with both forest cover and historical deforestation (8, 9). However, out of necessity, statistical approaches to assess environmental risk factors for *P. knowlesi* and other infectious diseases typically evaluate relationships between disease metrics and local land cover surrounding houses or villages. While an individual may spend most of their time within the vicinity of their residence, this area does not necessarily represent where they are most likely to be exposed to a disease. This is supported by varying associations between *P. knowlesi* occurrence and landscape variables at different distances from households, ranging from 100m to 5km, likely partially due to human movement into different surrounding habitats (8, 10). Although land cover variables describing physical terrestrial surfaces are frequently incorporated into disease models, land use is rarely quantified. Land use is commonly defined as “the arrangements, activities, and inputs that people undertake in certain land cover types” (11). Places with similar types of land cover may be used very

71 differently, with the activities and frequencies with which people visit these places determining the
72 spatial distribution of disease (1).
73
74 Mathematical modelling studies have revealed the importance of spatial variation in contact rates
75 due to the movement of individuals through heterogeneous environments with varying transmission
76 intensity (12). A multi-species transmission model of *P. knowlesi* highlighted the role of mixing
77 patterns between populations in different ecological settings in determining the basic reproductive
78 rate and subsequent modelling studies illustrate the sensitivity of this disease system to population
79 densities of both people and wildlife hosts (7, 13). However, although mechanistic models have been
80 extended to explore the potential importance of these heterogeneities in disease dynamics, there
81 are inherent constraints on model complexity and most models make simplistic assumptions about
82 the habitat uses of different populations.
83
84 Empirical data on human population movement is increasingly available, allowing assessment of the
85 impact of mobility on infectious disease dispersion and risks (5). On larger spatial scales, mobile
86 phone data has revealed the role of human migration in the transmission of infectious diseases such
87 as malaria, dengue and rubella (14-16).Although this data can provide insights into long range
88 movements, spatial resolution of this data is limited, particularly in areas with poor or no mobile
89 coverage, such as forested areas (17). Alternatively, the advent of low-cost GPS tracking devices
90 allows quantification of fine-scale movements, demonstrating marked heterogeneity in individual
91 movement and risk behaviours (3, 18). Combining these data with detailed data on land cover and
92 vector dynamics can provide new insights into how landscapes affect *P. knowlesi* transmission.
93
94 Previous studies of *P. knowlesi* have relied on questionnaire surveys, identifying self-reported travel
95 to nearby plantations and forest areas as a risk factor for *P. knowlesi* and other malaria infections
96 (e.g. (19-21)). However, the resultant spatial range and frequency of these movements remain

unknown and the definition of different habitat types is entirely subjective. Further, little is known about differences in local movement patterns in different demographic groups. While infections in male adults have been linked to forest and plantation work, it is unknown whether infections reported in women and young children are likely to arise from exposure to similar environments (22). The main mosquito vector in this area, *An. balabacensis*, is primarily exophagic and has been identified in farm, forest and village areas near houses (23, 24). Macaque populations are reported in close proximity to human settlements and molecular and modelling studies suggest transmission remains primarily zoonotic in this area (7, 25, 26). A case control study detected higher abundances of *An. balabacensis* near *P. knowlesi* case housesholds, suggesting the possibility of peri-domestic transmission (24). Understanding the importance of these habitats is essential to effectively target intervention strategies and predict impacts of future environmental changes.

Key questions remain about where individuals are likely to be exposed to *P. knowlesi* and how landscape determines risk. Functional ecology approaches allow the distribution of different populations to be modelled based on biological resources and relate transmission to landscape and environmental factors (27). Within wildlife ecology, numerous methods have been developed to estimate utilisation distributions (UDs), the probability of an individual or species being within a specific location during the sampling period (28). Although these methods traditionally rely on kernel density smoothing, kernel density estimates may not actually reflect time individuals spend in a specific location if there is substantial missing data or irregular time intervals. Alternatively, biased random bridges (BRBs) improve on these methods by estimating the utilisation distribution as a time-ordered series of points, taking advantage of the autocorrelated nature of GPS tracks to bias movement predictions towards subsequent locations in a time series (29). This allows for interpolation of missing values and adjustment for spatial error to estimate utilisation distributions representing both the intensity (mean residence time per visit) and frequency of individual visits to specific locations. By integrating these estimates of individual space use with detailed spatial and

environmental data in a Bayesian framework, fine-scale patterns of human land use can be predicted and overlaid with spatiotemporal models of mosquito distribution. This allows exploration of how landscape composition, as well as configuration and connectivity between habitats, impacts human exposure to *P. knowlesi* and other vector-borne and zoonotic diseases.

Focusing on one aspect of land use, human movement and time spent within different land cover types, we explored the role of heterogeneity in local space use on disease exposure. Rolling cross-sectional GPS tracking surveys were conducted in two study areas with on-going *P. knowlesi* transmission in Northern Sabah, Malaysia (Matunggong and Limbuak (30)). We aimed to characterise local movement patterns and identify individuals and locations associated with increased *P. knowlesi* exposure risks by: 1. analysing individual movement patterns and developing predictive maps of human space use relative to spatial and environmental factors, 2. modelling biting rates of the main vector *An. balabacensis*, and 3. assessing exposure risks for *P. knowlesi* based on predicted mosquito and human densities (Figure 1) Integrating these three approaches allowed a uniquely spatially explicit examination of disease risk.

Figure 1. Analysis methods used to estimate individual and community-level exposure to *P. knowlesi* sporozoite positive *An. balabacencis* bites

Methods

Study site

This study was conducted in two rural communities in Northern Sabah, Malaysia: Matunggong, Kudat (6°47N, 116°48E, population: 1260) and Limbuak, Pulau Banggi (7°09N, 117°05E, population: 1009) (Figure 2). These areas were the focus for integrated entomology, primatology and social science studies for risk factors for *P. knowlesi* (https://www.lshtm.ac.uk/research/centres-projects-

groups/monkeybar), with clinical cases and submicroscopic infections reported from both sites and *P. knowlesi* sero-prevalence estimated as 6.8% and 11.7% in Matunggong and Limbuak respectively (30).

Demographic data and GPS locations of primary residences were collected for all individuals residing in these areas (30). Potential spatial and environmental covariates for these sites were assembled from ground-based and remote-sensing data sources (Supplementary File 1). The enhanced vegetation index (EVI) was used to capture temporal changes in vegetation levels; this index captures photosynthetic activity and has higher sensitivity in high biomass areas compared to the normalised difference vegetation index (NDVI) frequently used. Due to the high cloud cover within this area, EVI at a high spatial resolution could not be obtained for all time periods. Instead, EVI data at a lower spatial but higher temporal resolution was used and monthly averages were calculated from all available cloud-free data and resampled to 30m per pixel (31).

Figure 2. A. Location of study sites and tracked houses (households with one or more individual GPS tracked) and survey houses (households with only questionnaire data collected and used for prediction) in B. Matunggong, Kudat and C. Limbuak, Banggi; description of land cover classification and survey methodology in (30)

GPS tracking survey

A minimum of 50 participants per site were targeted in a rolling cross-sectional survey (32). During pre-defined two-week intervals, randomly selected participants from comprehensive lists of eligible community members were asked to carry a QStarz BT-QT13000XT GPS tracking device (QStarz, Taipei, Taiwan) programmed to record coordinates continuously at one-minute intervals for at least 14 days regardless of individual movement. Individuals were excluded if they were not primarily residing in the study area, under 8 years old or did not consent. Trained fieldworkers visited the

participant every two days to confirm the device was functioning, replace batteries and administer questionnaires on locations visited and GPS use. Fieldworkers recorded whether the device was working and if the individual was observed carrying the GPS device to assess compliance. Individuals were excluded from analysis if insufficient GPS data were collected (less than 33% of sampling period) or individuals were observed not using the device for two or more visits.

Human space use

Biased random bridges were used to calculate individual utilisation distributions, the probability of an individual being in a location in space within the sampled time period (29). Within this study, large proportions of GPS fixes were missed due to technical issues with batteries and GPS tracking; biased random bridges were used to interpolate between known locations and adjust for missing data, using the time series GPS data to provide a more accurate estimate of space use. Utilisation distributions were calculated separately for each individual for all movement and night-time only movements (6pm – 6am).

To fit biased random bridges, we estimated the maximum threshold between points before they were considered uncorrelated (T_{max}) as 3 hours based on typical reported activity times. The minimum distance between relocations (L_{min}), the distance below which an individual is considered stationary, was set at 10m to account for GPS recording error based on static tests. Finally, the minimum smoothing parameter (h_{min}), the minimum standard deviation in relocation uncertainty, was set as 30m to account for the resolution of habitat data and capture the range of locations an individual could occupy while being recorded at the same place (28, 29). Estimates of the core utilisation area (home range) were based on the 99th percentile, representing the area with a 99% cumulative probability distribution of use by the sampled individual.

To assess relationships between space use and environmental factors and develop predictive maps of community space use, we fit resource utilisation functions, regression models in which the utilisation distributions are used as the response variable, improving on models using raw GPS count points as the response when there is location uncertainty and missing data (33). The probability density function (utilisation distribution) per individual was rasterised to 30m^2 grid cells and environmental and spatial covariates extracted for each grid cell. Potential environmental covariates included distance to the individual's own house, distance to closest house, distance to roads, land use class (forest, agriculture, cleared or water), distance to forest edge, elevation and slope (Supplementary File 1). Resource utilisation was modelled as a Bayesian semi-continuous (hurdle) model with two functionally independent components, a Bernoulli distribution for the probability of individual *i* visiting a specific grid cell *j* (ω_{ij}) and a gamma distribution for the utilisation distribution in grid cells visited (y_{ij}) (34, 35). For each individual, we defined absences to be all grid cells with a utilisation distribution less than 0.00001, indicating a very low probability the individual visited this grid cell during the study period. We included all presences (grid cells with a utilisation distribution > 0.00001) and randomly subsampled equal numbers of absences (grid cells not visited) for each individual as including equal numbers of presences and absences can improve predictive abilities of species distribution models (36). The utilisation distribution for grid cells visited is defined as:

$$y_{ij} = \begin{Bmatrix} \text{Gamma}\left(\frac{\mu_{ij}^2}{\sigma^2}, \frac{\sigma^2}{\mu_{ij}}\right) \text{with probability } 1 - \phi_{ij} \\ 0 \text{ with probability } \phi_{ij} \end{Bmatrix}$$

Where the mean of y_{ij} is given by:

$$\mu_{ij} = \text{E}\left(y_{ij} | X_{ij}^T\right) = \left(1 - \omega_{ij}\right)\upsilon_{ij}$$

The full model was specified as:

$$\omega_{ij} \sim \text{Bernoulli}\left(\phi_{ij}\right)$$

With the linear predictor for the Bernoulli model specified as:

$$\text{logit}(\phi_{ij}) = \beta_0 + X_{ij}^T\beta_i + \gamma_j$$

Where β_0 represents the intercept, $X_{ij}^T\beta_i$ represents a vector of covariate effects and γ_j represents the additive terms of random effects for individual. For the Gamma component, σ^2 is the variance and the linear predictor υ_{ij} is specified as:

$$\log(\upsilon_{ij}) = \alpha_0 + X_{ij}^T\alpha_i + \varphi_j$$

With α_0 representing the intercept, $X_{ij}^T\alpha_i$representing a vector of coordinates and φ_j representing the random effects. Weakly informative normal priors specified as Normal (0,1/0.01) were used for all intercepts and coefficients. Bayesian inference was implemented using integrated nested Laplace approximation (INLA) (37). This approach uses a deterministic algorithm for Bayesian inference, increasing computational efficiency relative to Markov chain Monte Carlo and other simulation-based approaches (34). We did not explicitly include spatial autocorrelation as several distance-based covariates were included (e.g. distance from own house) (33). Predictive models used data for all individuals aged 8 or over residing in these communities (Table 1) and models were limited to land areas within 5km of households included in the study site. Separate models were fit for each site.

Exposure to infected vectors

To estimate vector biting rates, we assembled data from 328 nights of human landing catches (HLCs) conducted with 5km of the Matunggong study site while GPS tracking was on-going, including: monthly longitudinal surveillance (23), investigations surrounding households of cases and controls

(24), and environmentally stratified outdoor catches (38) (Supplementary File 2). We limited this data to counts of *An. balabacensis*, the primary *knowlesi* vector which comprises over 95% of *Anopheles* caught in this region. As one experiment only collected mosquitoes for 6 hours, we fit a linear model of all available data vs totals after 6 hour catches to estimate the total numbers of *An. balabacensis* which would have been caught over 12 hours for these data (R^2= 0.85). Plausible environmental covariates were assembled, including land use type, slope, aspect, elevation, topographic wetness index, EVI, population density and average monthly temperature and rainfall. To select variables for inclusion, Pearson correlation analysis was used to assess multicollinearity between selected environmental variables. As topographic slope and TWI had a strong negative correlation, only TWI was included in the analysis. The autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to explore correlation between time lags.

A Bayesian hierarchical spatiotemporal model was implemented using counts of *An. balabacensis* bites as the outcome, denoted as m_{it}; j = 1…n; t = 1…n; where j indexes location and t indexes month. The log number of person-nights per catch was included as an offset to adjust for numbers of catchers conducting HLCs during different experiments. As the data were overdispersed, a negative binomial distribution was used to model m_{it}. The linear predictor was specified as:

$$\log(\mu_{jt}) = \log(N_{jt}) + Z_0 + D_{jt}^T Z + w_j + e_t$$

Where N_{ijt} represents the number of person-nights for each HLC catch, Z_0 represents the intercept, $D_{jt}^T Z$ represents a vector of covariates, w_j is the spatial effect and e_t is the temporal effect. The temporal effect e_t was included as a fixed effect, random effect or temporally structured random walk model of order 1 in candidate models (39). The spatial effect w_j was modelled as a Matern covariance function between locations s_j and s_k:

$$W \sim Multivariate\ Normal\ (0, \Sigma)$$

$$\Sigma_{hk} = Cov(\xi(s_h), \xi(s_k)) = Cov(\xi_h, \xi_k) = \frac{\sigma^2}{\Gamma(\lambda)2^{\lambda-1}}(\kappa||s_h - s_k||)^{\lambda}K_{\lambda}(\kappa||s_h - s_k||)$$

271

272 Where $||s_h - s_k||$ denotes the Euclidean distance between locations s_h and s_k, ξ(s) is the latent
273 Gaussian field accounting for spatial correlation, σ^2 is the spatial process variance and K_λ is a
274 modified Bessel function of the second kind and order λ > 0. κ is a scaling parameter related to *r*, the
275 distance at which spatial correlation becomes negligible, by $r = \sqrt{8\lambda}/\kappa$. A stochastic partial
276 differential equations (SPDE) approach was used, representing the spatial process by Gaussian
277 Markov random fields (GMRF) by partitioning the study area into non-intersecting triangles (40). This
278 approach represents the covariance matrix Σ by the inverse of the precision matrix *Q* of the GMRF
279 (34, 40). Prior distributions were specified on fixed effects and hyperparameters. A vague normal
280 prior distribution was used for the intercept. Weakly informative priors were used for fixed effects
281 specified as *N(1,1/0.01)*. Priors for spatial hyperparameters were specified as range *r ~ N(10, 1/0.01)*
282 and standard deviation *σ ~ N(0.1, 1/0.01)* as described by Lindgren and Rue (39).

283

284 As these vectors are rarely reported indoors (24) and HLCs were primarily conducted outside, we
285 excluded areas within houses for calculations of exposure risks. The proportion of infectious
286 mosquitoes, *c*, was parameterised using a beta distribution for *P. knowlesi* sporozoite rates within
287 this site; with only 4 out of 1524 collected mosquitoes positive, it was not possible to look at
288 variations of infection rates by time and space. Spatially explicit exposure risks were calculated as
289 derived quantity from human resource utilisation, mosquito biting rate models and probability of *P.*
290 *knowlesi* sporozoite positivity. Individual exposure risk was explored using a simple exposure
291 assessment model where the number of infected bites received by an individual is the sum of bites
292 by infected vector across all locations visited, with the number of infectious bites received by
293 individual *i* in month *t* as:

$$r_{it} = c \sum_{j=1}^{J} y_{ij} m_{jt}$$

Where *j* indexes the grid cells visited, y_{ij} is the utilisation distribution, m_{jt} is the number of bites per individual in that cell and month, and *c* is the proportion of infectious mosquitoes (4). To evaluate places associated with exposure for the entire community, we calculated the number of infectious bites per grid cell each month as:

$$r_{jt} = c \sum_{i=1}^{I} Y_{ij} m_{jt}$$

Where Y_{ij} is the predicted utilisation distribution for all individuals within the community per grid cell *j*. All analyses were conducted in R version 3.5, with Bayesian models implemented using Integrated Nested Laplace Approximation (INLA) (37). Model fit was assessed using deviance information criteria (DIC) and area under the receiver operating curve (AUC), root mean square error (RMSE) or conditional predictive ordinate (CPO) (41).

Ethics approval

This study was approved by the Medical Research Sub-Committee of the Malaysian Ministry of Health (NMRR-12-537-12568) and the Research Ethics Committee of the London School of Hygiene and Tropical Medicine (6531). Written informed consent was obtained from all participants or parents or guardians and assent obtained from children under 18.

Results:

Between February 2014 and May 2016, 285 consenting people participated in the GPS tracking study with 243 included in the final analysis including 109 in Limbuak and 134 in Matunggong (Table 1). The most commonly reported occupation was farm or plantation work (n=73), primarily conducted within the immediate vicinity of the house. A total of 3,424,913 GPS points were collected,

representing 6,319,885 person-minutes of sampling time. Median sampling duration was 16.27 days (IQR 13.72 – 19.97), with points recorded for a median of 59.1% (IQR: 46.9% - 71.1%) of the sampling duration. Maximum distances travelled ranged from no travel outside the house to 116km, with a median distance travelled of 1.8km. Utilisation distributions (UDs), the probability of an individual being in a location in space within a given time (Figure 3), varied by gender and occupation . Individuals at the more rural Limbuak site covered larger distances (Table 2), with the largest distances covered by individuals reporting primary occupations of fishing (n=5) and office work (n=9). Although substantial differences were reported in all movements (24 hour sampling) between seasons, no seasonal differences were observed in human movements during peak *Anopheles* biting times (6pm-6am).

Table 1. Baseline characteristics of study site communities and sampled populations

	Matunggong		Limbuak	
	Sampled	Community*	Sampled	Community*
N	134	958	109	633
Gender				
Male, % (n)	51.5% (69)	46.1% (442)	47.7% (52)	46.1% (292)
Women, % (n)	48.5% (65)	53.9% (516)	52.3% (57)	53.9% (341)
Age in years, median (IQR)	31 (17 – 53)	32.5 (8 – 51)	29 (15 – 46)	30 (15 – 47)
Main occupation, % (n)				
Farming	29.9% (40)	28.6% (274)	7.3% (8)	10.2% (65)
Plantation work	10.4% (14)	8.6% (82)	10.1% (11)	7.6% (48)
Student	26.1% (35)	27.7% (265)	26.6% (29)	21.0% (133)
Other	6.7% (9)	9.1% (87)	15.6% (17)	14.4% (91)
No employment/ housewife	26.9% (36)	26.1% (250)	40.4% (44)	46.8% (296)

* Community includes all individuals eligible for these surveys (residents ages 8 and over)

Figure 3. Human movement relative to habitat. A. Example of GPS tracks from a 22-year-old male plantation worker in Matunggong over aerial imagery, B. Probability density of an individual utilisation distribution calculated from GPS tracks

334 **Table 2.** Home range estimates by demographic group and site

	Area of 99% UD for all movement (hectares) Median (IQR)	Area of 99% UD from 6pm – 6am (hectares) Median (IQR)
Demographic group		
Men	32.09 (7.07, 148.93)	4.50 (2.79, 19.53)
Women	74.25 (12.24, 320.74)	6.08 (2.79, 24.17)
Children (under 15)	26.01 (6.39, 151.94)	3.83 (2.79, 8.73)
Occupation		
Farming	29.34 (8.15, 324.38)	6.75 (2.79, 19.80)
Plantation work	49.14 (9.72, 201.33)	4.59 (2.79, 27.72)
Fishing	442.49 (40.07, 1189.00)	227.16 (4.05, 465.14)
Office work	96.80 (63.61, 256.75)	13.63 (2.88, 20.14)
Other	19.98 (6.30, 26.82)	2.97 (2.61, 18.27)
No employment/ housewife	43.38 (11.97, 157.59)	3.60 (2.79, 19.12)
Site		
Limbuak	99.99 (24.57, 387.54)	7.74 (2.88, 58.05)
Matunggong	12.02 (3.94, 85.55)	2.97 (2.70, 11.77)
Season		
Dry (February – July)	28.62 (5.45, 252.45)	4.19 (2.79, 19.60)
Wet (August – January)	54.90 (17.23, 160.99)	4.64 (2.79, 19.35)

335

336

337

338 For both study areas, we developed models of community space use during peak mosquito biting
339 hours (6pm – 6am), in the form of resource utilisation functions, predictions of time- and space-
340 specific UDs on the basis of spatial and environmental variables (28). Between 6pm – 6am, human
341 space use (UDs) was mostly predictable and negatively correlated with distance from the individual's
342 house, other houses, roads and slope. The AUC for presence/ absence models was 0.936 for
343 Matunggong and 0.938 for Limbuak and RMSE for the overall model was 0.0073 and 0.0043 for
344 Matunggong and Limbuak respectively. While individuals were more likely to use areas further away
345 from forests in the Matunggong site, human space use was positively correlated with proximity to
346 forests in the Limbuak site (Table 3). Despite marked differences between different demographic
347 groups and seasons observed during 24 hour movements, these factors did not improve the
348 predictive power of the model for movements between 6pm and 6am.

349

Table 3. Estimated coefficients for fixed effects of resource utilisation functions (6pm – 6am)

	Matunggong			Limbuak		
	Mean	SD	95% CI	Mean	SD	95%CI
Probability of presence/ absence						
Intercept	3.383	0.839	3.218, 3.547	3.571	0.104	3.368, 3.775
Distance from own house (km)	-0.954	0.006	-0.966, -0.942	-0.543	0.003	-0.548, -0.539
Distance from forest (km)	5.997	0.177	-5.650, 6.344	-1.845	0.050	-1.944, -1.746
Distance from road (km)	-5.552	0.057	-5.663, -5.441	-3.656	0.019	-3.694, -3.618
Distance from houses (km)	-0.504	0.030	-0.563, -0.444	0.176	0.007	0.162, 0.189
Elevation (100 MSL)	-0.710	0.025	-0.759, -0.662	-1.268	0.037	-1.340, -1.197
Slope (degrees)	-0.0244	0.002	-0.028, -0.021	-0.009	0.001	-0.012, -0.006
Utilisation distributions for locations present						
Intercept	-6.846	0.866	-8.549, -5.147	-5.676	1.017	-7.673, -3.681
Distance from own house (km)	-0.583	0.004	-0.590, -0.576	-0.308	0.002	-0.311, -0.305
Distance from forest (km)	12.012	0.199	11.621, 12.403	-1.771	0.049	-1.868, -1.675
Distance from road (km)	-0.833	0.054	-0.939, -0.728	-1.532	0.011	-1.554, -1.511
Distance from houses (km)	-0.819	0.023	-0.864, -0.773	-0.239	0.006	-0.249, -0.228
Elevation (100 MSL)	0.664	0.027	0.610, 0.718	-0.297	0.003	-0.303, -0.297
Slope (degrees)	-0.021	0.002	-0.024, -0.018	-0.034	0.001	-0.036, -0.031

Between August 2013 and December 2015, 4814 *An. balabacensis* were caught from 328 sampling nights in 155 unique locations. The median biting rate was 2.1 bites per night per person, ranging from 0 – 28 bites per person per night (Figure 4). Despite monthly variation, including temporal autocorrelation did not improve model fit (Table 4). Although no associations were identified between land classification and vector density in this site, models identified positive relationships with enhanced vegetation indices (EVI) and negative associations with distance to forest and human population density (Table 5). Of 1524 mosquitoes tested for *Plasmodium* sporozoites, the median sporozoite rate was 0.24% (95% CI: 0.09 – 0.58%).

Table 4. Model selection statistics for mosquito biting rates

Model		DIC*	Marginal Likelihood	Model complexity*	RMSE*	Mean log-score (CPO)
M1	No spatial or temporal effect	2367.03	-1196.61	4.12	4.99	3.61
M2	Spatial effect only	2292.97	-1175.47	40.03	4.42	4.16
M3	Spatial effect + month as	2282.88	-1173.68	43.99	4.24	3.90

	fixed effect					
M4	Spatial effect + month as random effect	2222.89	-1155.91	50.28	4.05	3.61
M5	Spatial effect + month as random walk	2225.43	-1167.79	47.55	4.09	3.63

Table 5. Posterior rate ratio estimates and 95% Bayesian credible interval (BCI) for model 4 of mosquito biting rates

Covariate	95% BCI Rate Ratio		
	Mean	2.5%	97.5%
Population density	0.963	0.916	1.004
EVI	3.185	1.185	8.532
Distance to forest (100m)	0.926	0.871	0.976
Spatial range (km)	3.120	0.514	6.926

Figure 4. Mosquito biting rates. A. *An. balabacensis* biting rate per person-night from data collected in Matunggong, B. Predicted mean *An. balabacensis* biting rates per month from spatiotemporal models, C. Predicted number of bites for all individuals residing in Matunggong by distance from secondary forest, and by D. Distance from households

For individuals included in the GPS tracking study in Matunggong, where both human movement and entomology data was available, we calculated exposure risks as a derived quantity from utilisation distributions and mosquito biting rate models. Exposure varied markedly between individuals, with an overall 150-fold difference in predicted mean probabilities of infected bites per night (range: 0.00005-0.0078) (Table 6). No clear differences were observed between genders, age groups or occupations of individuals sampled and there was no association between risk and distance travelled.

382 **Table 6.** Probabilities of infected bites per person per night for sampled individuals in Matunggong

383 by demographic characteristics

	Predicted infectious bites per night (median (IQR))
Demographic group	
Men	0.00157 (0.000804, 0.00289)
Women	0.00219 (0.000864, 0.00307)
Children (under 15)	0.00131 (0.000812, 0.00330)
Occupation	
Farming	0.00180 (0.00101, 0.00362)
Plantation work	0.00216 (0.000680, 0.00278)
Student	0.00143 (0.000915, 0.00304)
Other	0.00225 (0.000852, 0.00302)
No employment/ housewife	0.00142 (0.000297, 0.00263)

384

385

386 Using the resource utilisation function with demographic and spatial data for all individuals in
387 Matunggong, we predicted community-wide space use and estimated exposure to infected
388 mosquitoes (Figure 5). The predicted number of person nights per grid cell for the entire community
389 ranged from 0 to 12.79 (median: 0.01, IQR: 0.0004 – 0.99), with the mean probability of a
390 community member exposed to an infected bite per grid cell of 0.00082 (IQR: 0.00001, 0.00050).
391 Although over 43% of the study site is forest and relatively high biting rates were predicted in forests
392 during the study period (mean: 1.94, range: 0.04 – 12.59), this habitat was rarely used by people in
393 the evenings, with less than 8% of predicted person-nights in forests. Models only based on
394 mosquito biting rates and not including human space use predicted 42% of infectious bites occurred
395 in forested areas and only 8.6% of bites occurring within 100m of houses (Figure 5C). In contrast,
396 when space use patterns are included, over 91% of predicted infected bites were predicted within
397 500m of houses (Figure 5D). Highest exposure risks were consistently found near forest edges and in
398 close proximity to households, despite spatial and temporal heterogeneity and model uncertainty
399 (Figure 4).

400

Figure 5. A. Land use in Matunggong site, B. Predicted number of person- nights for entire community per grid cell, C. Predicted mosquito biting rates, D. Predicted infected bites per grid cell

Discussion

This study highlights the importance of human space use in different land cover types in determining exposure to zoonotic and vector-borne diseases such as *P. knowlesi*. Although *P. knowlesi* has previously been associated with forest exposure (e.g. (19)) and higher biting rates have been reported in forest interiors (23), this novel approach incorporating both mosquito and human space use data provides a new perspective on peri-domestic transmission, with more than 90% of infectious bites predicted in areas surrounding households at forest edges. This study additionally demonstrates the utility of ecological methods to understand human movement and identify geographical areas associated with higher contact with disease vectors.

Within these communities, local movement patterns during peak vector times were largely predictable and could be explained by spatial and environmental factors. However, despite this finding, there was substantial variation in predicted exposure between individuals as a result of heterogeneity in habitats used. No significant differences in exposure were predicted between men and women, with individuals with high exposure risks identified across occupational and age groups. Although this finding differs from clinical reports, a comprehensive survey within this community identified equal proportions of men and women exposed to *P. knowlesi* as evidenced by specific antibody responses and data on asymptomatic infections suggests higher numbers of non-clinical infections in women (30, 42). While infrequent events or long-range movements (such as hunting trips) may contribute to these differences in clinical cases and may not have been captured within

this two-week sampling period within the study site, this analysis highlights the importance of routine movements into local environments in shaping exposure risks.

This improved understanding of how local human land use is related to exposure risk has important implications for surveillance and control programmes. Malaria control programmes often rely on interventions within the house, such as insecticide treated bednets and indoor residual spraying; however, movements outside during peak biting times illustrate the importance of also targeting outdoor transmission. The identification of areas where exposure is likely to occur can further be used to refine interventions; for example, although insecticide treated hammocks have been proposed for deep forest environments, larval source management may be more appropriate to target environments in close proximity to houses. Although initial *P. knowlesi* cases were primarily identified in adult men living and working in forests (20), this study illustrates the potential importance of peri-domestic habitats in transmission and provides quantitative insight on mixing between people and infected mosquitoes in forest fringe areas. As Malaysia moves towards malaria elimination, surveillance systems are incorporating novel focal investigation methods, including monitoring changes in local land use and populations at risk (43). In additional to routine vector surveillance, this study highlights the need to incorporate measures of human space when defining risk zones.

Even with the large and highly detailed movement dataset analysed, this study was limited by the availability of mosquito data; as human landing catch data were assembled from other studies, there was not uniform spatial and temporal coverage of the study site increasing uncertainty. The limited mosquito data availability precluded development of mosquito biting rate models for Limbuak and other outlying islands. An additional limitation to estimating mosquito biting rates was the difficulty obtaining spatially and temporally resolute remote sensing data for predictors due to high cloud cover (44). As few positive mosquitoes were identified, uniform estimates of sporozoite rates based

on available data were used across the Matunggong site; if further data was available, these models could be refined to incorporate estimates of human and macaque density, mosquito biting preferences in different habitats and infection levels in all hosts (45). Additionally, as this study was designed to quantitatively estimate time spent in different landscapes, further studies could explore other aspects of land use, such as the purposes of travel, activities undertaken or practices used to modify or management land cover.

Despite these limitations, this is the first large-scale study to utilise GPS tracking data and ecological methods to create fine-scale maps of exposure risk. This study highlights the importance of incorporating heterogenous patterns of human space use into disease models, as the majority of human exposure may occur in areas with lower vector biting rates but greater probabilities of human use. Further, results quantitatively illustrate the importance of forest edges and local habitat in *P. knowlesi* transmission and can inform understanding of other zoonotic and vector-borne diseases.

Acknowledgements

We would like to thank the MONKEYBAR project team and the participants in Sabah, Malaysia for their help with these studies. We acknowledge the Medical Research Council, Natural Environmental Research Council, Economic and Social Research Council and Biotechnology and Biosciences Research Council for funding received for this project through the Environmental and Social Ecology of Human Infectious Diseases Initiative, grant no. G1100796. This work was additionally supported by the National Socio-Environmental Synthesis Center (SESYNC).

Source code. R scripts for fitting biased random bridges (with simulated GPS data), spatiotemporal models of mosquito biting rates and semi-continuous resource utilisation models

Supplementary file 1. Data sources for assessed spatial and environmental covariates

Supplementary file 2. Data sources of mosquito biting data

References:

1. Lambin EF, Tran A, Vanwambeke SO, Linard C, Soti V. Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts. Int J Health Geogr. 2010;9:54.
2. Patz JA, Daszak P, Tabor GM, Aguirre AA, Pearl M, Epstein J, et al. Unhealthy landscapes: Policy recommendations on land use change and infectious disease emergence. Environ Health Perspect. 2004;112(10):1092-8.
3. Stoddard ST, Forshey BM, Morrison AC, Paz-Soldan VA, Vazquez-Prokopec GM, Astete H, et al. House-to-house human movement drives dengue virus transmission. Proc Natl Acad Sci U S A. 2013;110(3):994-9.
4. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Paz Soldan V, Kochel TJ, Kitron U, et al. The role of human movement in the transmission of vector-borne pathogens. PLoS Negl Trop Dis. 2009;3(7):e481.
5. Pindolia DK, Garcia AJ, Wesolowski A, Smith DL, Buckee CO, Noor AM, et al. Human movement data for malaria control and elimination strategic planning. Malar J. 2012;11(205).
6. Moyes CL, Shearer FM, Huang Z, Wiebe A, Gibson HS, Nijman V, et al. Predicting the geographical distributions of the macaque hosts and mosquito vectors of Plasmodium knowlesi malaria in forested and non-forested areas. Parasit Vectors. 2016;9(1):242.
7. Imai N, White MT, Ghani AC, Drakeley CJ. Transmission and control of Plasmodium knowlesi: a mathematical modelling study. PLoS Negl Trop Dis. 2014;8(7):e2978.
8. Fornace KM, Abidin TR, Alexander N, Brock P, Grigg MJ, Murphy A, et al. Association between Landscape Factors and Spatial Patterns of Plasmodium knowlesi Infections in Sabah, Malaysia. Emerg Infect Dis. 2016;22(2):201-8.
9. Shearer FM, Huang Z, Weiss DJ, Wiebe A, Gibson HS, Battle KE, et al. Estimating Geographical Variation in the Risk of Zoonotic Plasmodium knowlesi Infection in Countries Eliminating Malaria. PLoS Negl Trop Dis. 2016;10(8):e0004915.
10. Brock PM, Fornace KM, Grigg MJ, Anstey NM, William T, Cox J, et al. Predictive analysis across spatial scales links zoonotic malaria to deforestation. Proceedings of the Royal Society B. 2019;286(1894).
11. IPCC. Land use, land-use change, and forestry. Watson RT, Noble IR, Bolin B, Ravindranath NH, Verado DJ, Dokken DJ, editors. UK: Cambridge University Press; 2000.
12. Acevedo MA, Prosper O, Lopiano K, Ruktanonchai N, Caughlin TT, Martcheva M, et al. Spatial heterogeneity, host movement and mosquito-borne disease transmission. PLoS One. 2015;10(6):e0127552.
13. Yakob L, Lloyd AL, Kao RR, Ferguson HM, Brock PM, Drakeley C, et al. Plasmodium knowlesi invasion following spread by infected mosquitoes, macaques and humans. Parasitology. 2018;145(1):101-10.
14. Chang HH, Wesolowski A, Sinha I, Jacob CG, Mahmud A, Uddin D, et al. Mapping imported malaria in Bangladesh using parasite genetic and human mobility data. Elife. 2019;8.
15. Wesolowski A, Qureshi T, Boni MF, Sundsoy PR, Johansson MA, Rasheed SB, et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. Proc Natl Acad Sci U S A. 2015;112(38):11887-92.
16. Wesolowski A, Metcalf CJ, Eagle N, Kombich J, Grenfell BT, Bjornstad ON, et al. Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. Proc Natl Acad Sci U S A. 2015;112(35):11114-9.
17. Wesolowski A, Buckee CO, Engo-Monsen K, Metcalf CJE. Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data. J Infect Dis. 2016;214(suppl_4):S414-S20.
18. Vazquez-Prokopec GM, Bisanzio D, Stoddard ST, Paz-Soldan V, Morrison AC, Elder JP, et al. Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. PLoS One. 2013;8(4):e58802.

19. Grigg MJ, Cox J, William T, Jelip J, Fornace KM, Brock PM, et al. Individual-level factors associated with the risk of acquiring human Plasmodium knowlesi malaria in Malaysia: a case control study. Lancet Planetary Health. 2017;1:e97-104.
20. Singh B, Kim Sung L, Matusop A, Radhakrishnan A, Shamsul SS, Cox-Singh J, et al. A large focus of naturally acquired Plasmodium knowlesi infections in human beings. Lancet. 2004;363(9414):1017-24.
21. Yasuoka J, Levins R. Impact of deforestation and agricultural development on anopheline ecology and malaria epidemiology. Am J Trop Med Hyg. 2007;76(3):450-60.
22. Barber BE, William T, Dhararaj P, Anderios F, Grigg MJ, Yeo TW, et al. Epidemiology of Plasmodium knowlesi malaria in north-east Sabah, Malaysia: family clusters and wide age distribution. Malar J. 2012;11:401.
23. Wong ML, Chua TH, Leong CS, Khaw LT, Fornace K, Wan-Sulaiman WY, et al. Seasonal and Spatial Dynamics of the Primary Vector of Plasmodium knowlesi within a Major Transmission Focus in Sabah, Malaysia. PLoS Negl Trop Dis. 2015;9(10):e0004135.
24. Manin BO, Ferguson HM, Vythilingam I, Fornace K, William T, Torr SJ, et al. Investigating the Contribution of Peri-domestic Transmission to Risk of Zoonotic Malaria Infection in Humans. PLoS Negl Trop Dis. 2016;10(10):e0005064.
25. Lee KS, Divis PC, Zakaria SK, Matusop A, Julin RA, Conway DJ, et al. Plasmodium knowlesi: reservoir hosts and tracking the emergence in humans and macaques. PLoS Pathog. 2011;7(4):e1002015.
26. Chua TH, Manin BO, Daim S, Vythilingam I, Drakeley C. Phylogenetic analysis of simian Plasmodium spp. infecting Anopheles balabacensis Baisas in Sabah, Malaysia. PLoS Negl Trop Dis. 2017;11(10):e0005991.
27. Hartemink N, Vanwambeke SO, Purse BV, Gilbert M, Van Dyck H. Towards a resource-based habitat approach for spatial modelling of vector-borne disease risks. Biol Rev Camb Philos Soc. 2015;90(4):1151-62.
28. Papworth SK, Bunnefeld N, Slocombe K, Milner-Gulland EJ. Movement ecology of human resource users: using net squared displacement, biased random bridges and resource utilization functions to quantify hunter and gatherer behaviour. Methods in Ecology and Evolution. 2012;3(3):584-94.
29. Benhamou S. Dynamic approach to space and habitat use based on biased random bridges. PLoS One. 2011;6(1):e14592.
30. Fornace KM, Herman LS, Abidin TR, Chua TH, Daim S, Lorenzo PJ, et al. Exposure and infection to Plasmodium knowlesi in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. PLoS Negl Trop Dis. 2018;12(6):e0006432.
31. DAAC NL. MODIS/ Terra Vegetation Indices 16-Day L3 Global 250m Grid SIN V006. Sioux Falls, South Dakota: USGS Earth Resources Observation and Science (EROS) Center.
32. Johnston R, Brady HE. The rolling cross-section design. Electoral Studies. 2002;21(2):283-95.
33. Hooten MB, Hanks EM, Johnson DS, Alldredge MW. Reconciling resource utilization and resource selection functions. J Anim Ecol. 2013;82(6):1146-54.
34. Blangiardo M, Cameletti M. Spatial and Spatio-temporal Bayesian Models with R-INLA. Chichester, United Kingdom: John Wiley & Sons, Ltd; 2015.
35. Sadykova D, Scott BE, De Dominicis M, Wakelin SL, Sadykov A, Wolf J. Bayesian joint models with INLA exploring marine mobile predator-prey and competitor species habitat overlap. Ecol Evol. 2017;7:5212-26.
36. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution. 2012;3:327-38.
37. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society. 2009;71(Part 2):319-92.

580 38. Ng SH, Homathevi R, Chua TH. Mosquitoes of Kudat: species composition and their medical
581 importance (Diptera: Culicidae). Serangga. 2016;21(2):149-62.
582 39. Lindgren F, Rue H. Bayesian Spatial Modelling with R-INLA. Journal of Statistical Software.
583 2015;63(19).
584 40. Lindgren F, Rue H, Lindstrom J. An explicit link between Gaussian fields and Gaussian Markov
585 random fields: the stochastic partial differential equation approach. Statistical Methodology.
586 2011;73(4):423-98.
587 41. Held L, Schrodle B, Rue H. Posterior and Cross-validatory Predictive Checks: A Comparison of
588 MCMC and INLA. In: Kneib T, Tutz G, editors. Statistical Modelling and Regression Structures:
589 Physica-Verlag HD; 2010.
590 42. Fornace KM, Nuin NA, Betson M, Grigg MJ, William T, Anstey NM, et al. Asymptomatic and
591 Submicroscopic Carriage of Plasmodium knowlesi Malaria in Household and Community Members of
592 Clinical Cases in Sabah, Malaysia. J Infect Dis. 2015.
593 43. Sektor Penyakit Bawaan Vektor Bahagian Kawalan Penyakit Kementarian Kesihatan
594 Malaysia. Garis Panduan Pencegahan Malaria Re-Introduction di Malaysia. Putra Jaya: Kementerian
595 Kesihatan Malaysia; 2016. Contract No.: KKM/TM/620/2016.
596 44. Weiss DJ, Mappin B, Dalrymple U, Bhatt S, Cameron E, Hay SI, et al. Re-examining
597 environmental correlates of Plasmodium falciparum malaria endemicity: a data-intensive variable
598 selection approach. Malar J. 2015;14:68.
599 45. Yakob L, Bonsall MB, Yan G. Modelling knowlesi malaria transmission in humans: vector
600 preference and host competence. Malar J. 2010;9:329.

601

602

Human Movement

Human GPS tracks collected

- Two week sampling of randomly selected individuals (over age 8)

Biased random bridges (individual space use)

- Estimate size of area used per individual for all movements and 6pm – 6am
- Compare between individuals

Resource utilization functions (community space use)

- Identify spatial and environmental characteristics associated with increased human space use
- Develop predictive models of human space use during peak mosquito biting times

Mosquito Ecology

Human landing catches

- Numbers of *An. balabacensis*
- Identification of *P. knowlesi* sporozoite positive mosquitoes

Spatiotemporal model of *An. balabacensis* biting rates

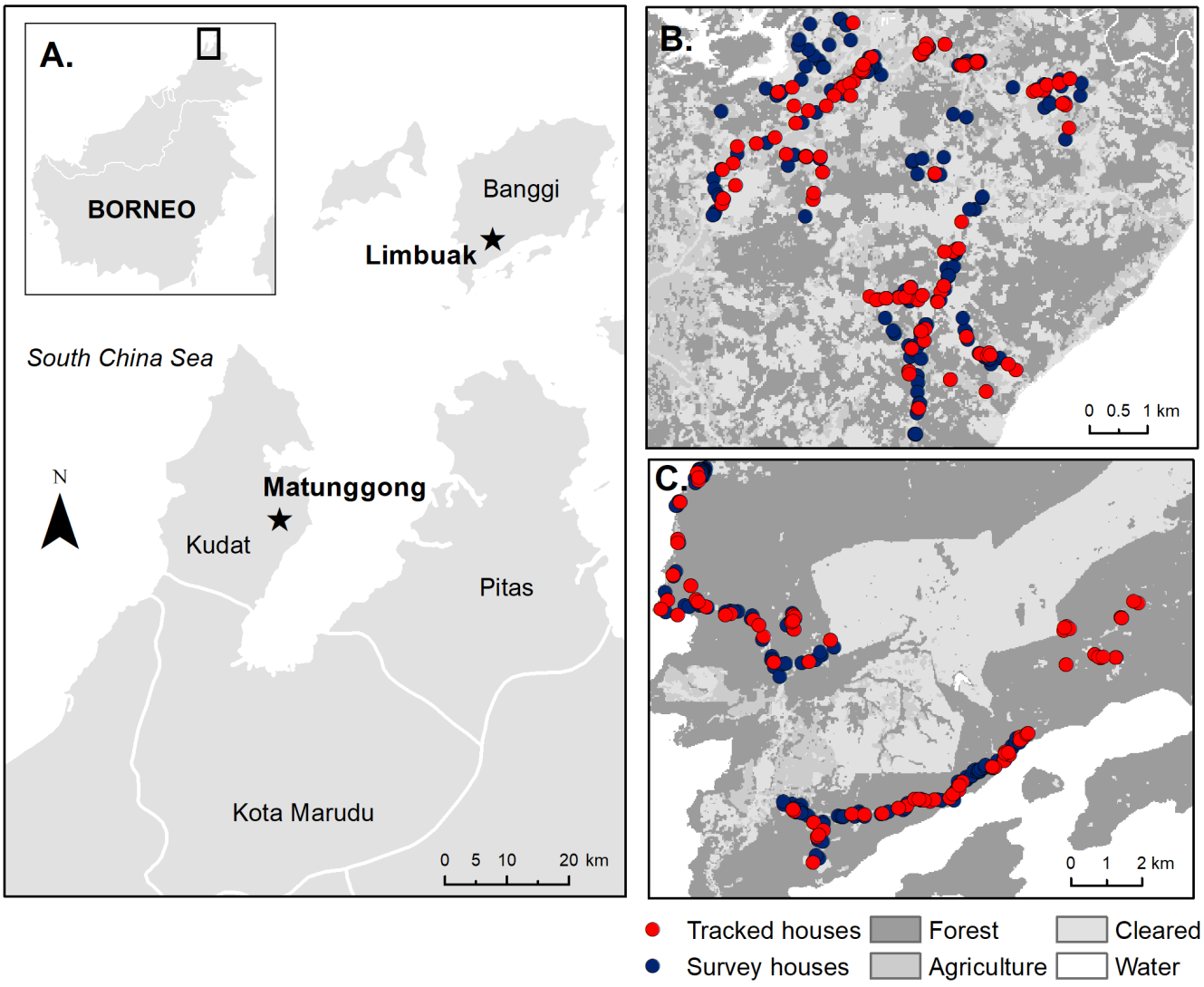
- Identify spatial and environmental factors associated with increased mosquito density
- Check residual spatial and temporal correlation
- Develop predictive model of biting rates

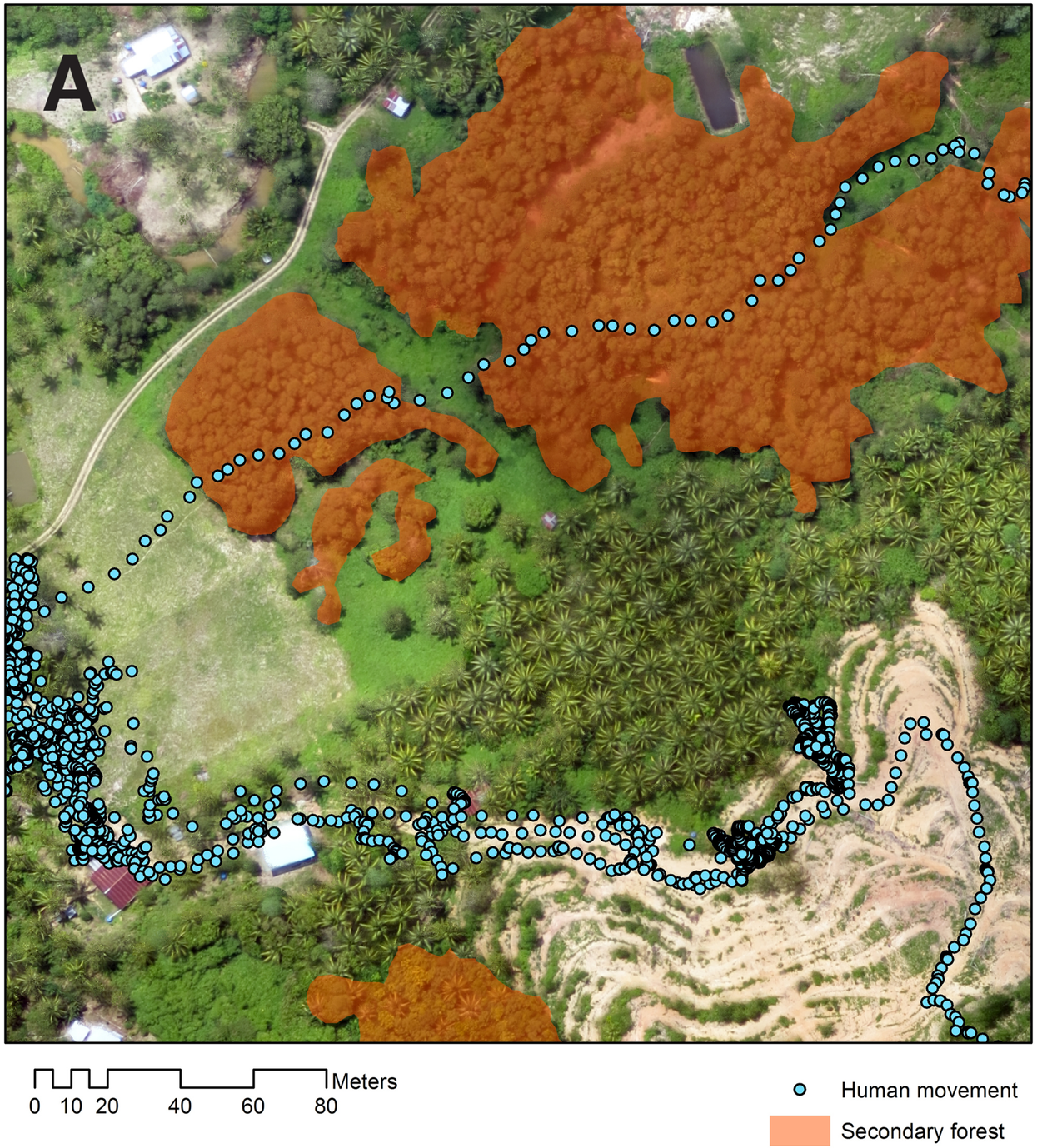
Probability of human space use during mosquito biting hours

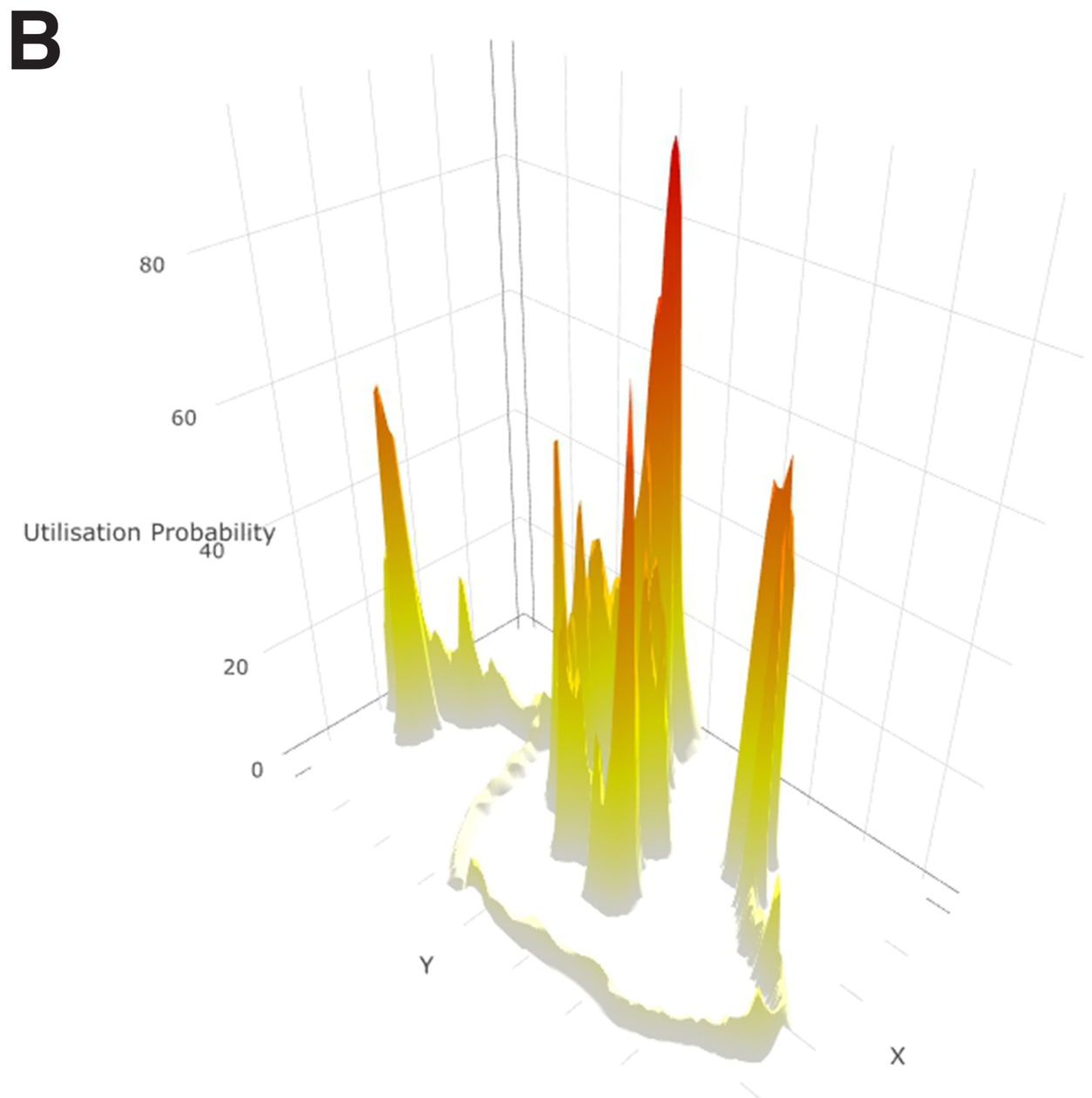
Predicted *An. balabacensis* biting rate

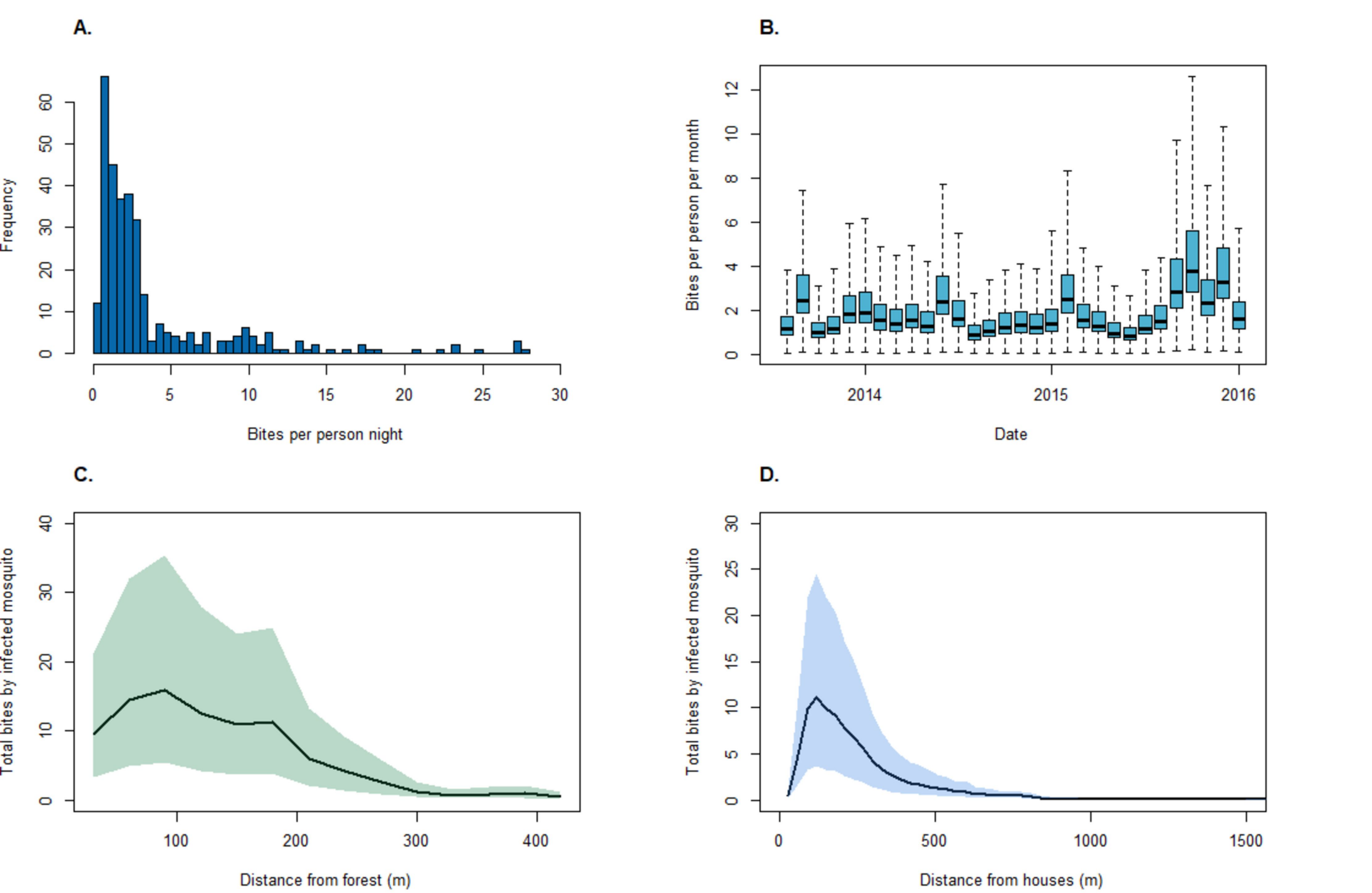
Probability of infection with *P. knowlesi*

Exposure to *P. knowlesi*
Number of infectious bites on humans


A.
BORNEO
Banggi
Limbuak
South China Sea
N
Matunggong
Kudat
Pitas
Kota Marudu
0 5 10 20 km
B.
0 0.5 1 km
C.
0 1 2 km
Tracked houses
Survey houses
Forest
Agriculture
Cleared
Water


A
0 10 20 40 60 80
Meters
Human movement
Secondary forest


B
80
60
Utilisation Probability
40
20
0
Y
X


A.
Frequency
0
10
20
30
40
50
60
0
5
10
15
20
25
30
Bites per person night
B.
Bites per person per month
0
2
4
6
8
10
12
2014
2015
2016
Date
C.
Total bites by infected mosquito
0
10
20
30
40
100
200
300
400
Distance from forest (m)
D.
Total bites by infected mosquito
0
5
10
15
20
25
30
0
500
1000
1500
Distance from houses (m)

A

0 500 1,000 2,000 Meters

Houses
Forest
Agriculture
Clearing
Water

B

0 500 1,000 2,000 Meters

Houses
Total Person Nights
High : 12.7924
Low : 0

C

0 500 1,000 2,000 Meters

Houses
Mean biting rate
High : 5.1001
Low : 0.91939

D

0 500 1,000 2,000 Meters

Houses
Mean Infected Bites
High : 0.0869811
Low : 0