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 land 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 edge effects. Our results illustrate the role of human land use and fine-scale movements in infectious disease transmission by intensifying interactions between pathogens, insect vectors and people around habitat edges.

Keywords (min 3):

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

45 **Introduction:**

46

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

57

58 The emergence of the zoonotic malaria *Plasmodium knowlesi* has been positively associated with
59 both forest cover and historical deforestation (8, 9). However, out of necessity, statistical
60 approaches to assess environmental risk factors for *P. knowlesi* and other infectious diseases
61 typically evaluate relationships between disease metrics and local land cover surrounding houses or
62 villages. While an individual may spend most of their time within the vicinity of their residence, this
63 area does not necessarily represent where they are most likely to be exposed to a disease. This is
64 supported by varying associations between *P. knowlesi* occurrence and landscape variables at
65 different spatial scales, likely partially due to human movement into different surrounding habitats
66 (8, 10). Although land cover variables describing physical terrestrial surfaces are frequently
67 incorporated into disease models, land use, the purposes for which people use specific areas, is
68 rarely quantified. Places with the similar types of land cover may be used very differently, with the
69 activities and frequencies with which people visit these places determining the spatial distribution of
70 disease (1).

Mathematical modelling studies have revealed the importance of spatial variation in contact rates due to the movement of individuals through heterogeneous environments with varying transmission intensity (11). A multi-species transmission model of *P. knowlesi* highlighted the role of mixing patterns between populations in different ecological settings in determining the basic reproductive rate and subsequent modelling studies illustrate the sensitivity of this disease system to population densities of both people and wildlife hosts (7, 12). However, although mechanistic models have been extended to explore the potential importance of these heterogeneities in disease dynamics, there are inherent constraints on model complexity and most models make simplistic assumptions about the habitat uses of different populations.

Empirical data on human population movement is increasingly available, allowing assessment of the impact of mobility on infectious disease dispersion and risks (5). On larger spatial scales, mobile phone data has revealed the role of human migration in the transmission of infectious diseases such as malaria, dengue and rubella (13-15). Analysis of these data can provide insight into the creation of sources and sinks of infection following long range movements of infected or susceptible populations (16). However, spatial resolution of this data is limited, particularly in areas with poor or no mobile coverage, such as forested areas, limiting the utility to assess movements across different habitats (17). Alternatively, the advent of low-cost GPS tracking devices allows quantification of movements on finer spatial scales, demonstrating marked heterogeneity in individual movement and risk behaviours (3, 18). Combining these data with detailed data on land cover and vector dynamics can provide new insights into how landscapes affect *P. knowlesi* transmission.

Previous studies of *P. knowlesi* have relied on questionnaire surveys, identifying self-reported travel to nearby plantations and forest areas as a risk factor for *P. knowlesi* and other malaria infections (e.g. (19-21)). However, the resultant spatial range and frequency of these movements remain

97 unknown and the definition of different habitat types is entirely subjective. Further, little is known
98 about differences in local movement patterns in different demographic groups. While infections in
99 male adults have been linked to forest and plantation work, it is unknown whether infections
100 reported in women and young children are likely to arise from exposure to similar environments
101 (22). The main mosquito vector in this area, *An. balabacensis*, is primarily exophagic and has been
102 identified in farm, forest and village areas near houses (23, 24). Macaque populations are reported
103 in close proximity to human settlements and molecular and modelling studies suggest transmission
104 remains primarily zoonotic in this area (7, 25, 26).
105
106 Key questions remain about where individuals are likely to be exposed to *P. knowlesi* and how
107 landscape determines risk. Reported presence of both vectors and macaques around households
108 suggests the possibility of peridomestic transmission and understanding the importance of these
109 habitats is essential to effectively design intervention strategies and predict impacts of future
110 environmental changes. Functional ecology approaches allow the modelling of distribution of
111 different populations based on biological resources and relate transmission to landscape and
112 environmental factors (27). By integrating methods to collect detailed GPS tracking data on human
113 movements with approaches developed to model resource utilisation in wildlife populations, fine-
114 scale patterns of human land use can be quantified and overlaid with spatiotemporal models of
115 mosquito distribution. This allows exploration how landscape composition, as well as configuration
116 and connectivity between habitats, impacts human exposure to *P. knowlesi* and other vector-borne
117 and zoonotic diseases.
118
119 To quantify the role of heterogeneity in land use on disease exposure, we conducted rolling cross-
120 sectional surveys in two study areas with on-going *P. knowlesi* transmission in Northern Sabah,
121 Malaysia (Matunggong and Limbuak (28)). Here, randomly selected people (age 8 and over) were
122 asked to carry a GPS tracking device for two-week periods. We aimed to characterise local

123 movement patterns and identify individuals and locations associated with increased *P. knowlesi*
124 exposure risks by: 1. analysing individual movement patterns and developing predictive maps of
125 human space use relative to spatial and environmental factors, 2. modelling biting rates of the main
126 vector *An. balabacensis*, and 3. assessing exposure risks for *P. knowlesi* based on predicted mosquito
127 and human densities. Integrating these three approaches allowed a uniquely spatially explicit
128 examination of disease risk.
129
130 **Methods**
131
132 *Study site*
133 This study was conducted in two rural communities in Northern Sabah, Malaysia: Matunggong,
134 Kudat (6°47N, 116°48E, population: 1260) and Limbuak, Pulau Banggi (7°09N, 117°05E, population:
135 1009). These areas were the focus for integrated entomology, primatology and social science studies
136 for risk factors for *P. knowlesi* (http://malaria.lshtm.ac.uk/MONKEYBAR), with clinical cases and
137 submicroscopic infections reported from both sites and *P. knowlesi* sero-prevalence estimated as
138 6.8% and 11.7% in Matunggong and Limbuak respectively (28). Demographic data and GPS locations
139 of primary residences were collected for all individuals residing in these areas.
140
141 *GPS tracking survey*
142 A minimum of 50 participants per site were targeted in a rolling cross-sectional survey (29). During
143 pre-defined two week intervals, randomly selected participants were asked to carry a QStarz BT-
144 QT13000XT GPS tracking device (QStarz, Taipei, Taiwan) programmed to record coordinates
145 continuously at one minute intervals for at least 14 days. Individuals were excluded if they were not
146 primarily residing in the study area, under 8 years old or did not consent. Trained fieldworkers
147 visited the participant every two days to confirm the device was functioning, replace batteries and
148 administer questionnaires on locations visited and GPS use. Fieldworkers recorded whether the

149 device was working and if the individual was observed carrying the GPS device to assess compliance.
150 Individuals were excluded from analysis if insufficient GPS data were collected (less than 33% of
151 sampling period) or individuals were observed not using the device for two or more visits.
152
153 *Human space use*
154 Biased random bridges (BRBs) were used to calculate individual utilisation distributions (UDs), the
155 probability of an individual being in a location in space within the sampled time period (30). UDs are
156 typically calculated using kernel smoothing methods, smoothing observed locations of individuals or
157 species to estimate average probabilities of use per grid cell to create a probability density function
158 (31). BRBs improve on these methods by estimating the UD as a time-ordered series of points, taking
159 advantage of the autocorrelated nature of GPS tracks to bias movement predictions towards
160 subsequent locations in a time series (30). Within this study, large proportions of GPS fixes were
161 missed due to technical issues with batteries and GPS tracking; BRBs were used to interpolate
162 between known locations and adjust for missing data. UDs were calculated separately for each
163 individual for all movement and night-time only movements (6pm – 6am) (Supplementary
164 Information). Estimates of the core utilisation area (home range) were based on the 99th percentile,
165 representing the area with a 99% cumulative probability distribution of use by the sampled
166 individual.
167
168 To assess relationships between space use and environmental factors and develop predictive maps
169 of community space use, we fit resource utilisation functions, regression models in which the UDs
170 are used as the response variable, improving on models using raw GPS count points as the response
171 when there is location uncertainty and missing data (32). The probability density function (UD) per
172 individual was rasterised to $30m^2$ grid cells and environmental and spatial covariates extracted for
173 each grid cell. Potential environmental covariates included distance to the individual's own house,
174 distance to closest house, distance to roads, land use class (forest, agriculture, cleared or water),

distance to forest edge, elevation and slope (Supplementary Information, Table S1). Resource utilisation was modelled as a Bayesian semi-continuous (hurdle) model with two functionally independent components, a Bernoulli distribution for presence/absence of individual *j* in a specific grid cell *i* (z_i) and a gamma distribution for the UD in grid cells visited (y_i) (33, 34). We defined absences as grid cells with a UD of less than 0.00001 and randomly sampled equal numbers of absences within the study site for each presence point for each individual (35). Occurrence was defined as:

$$z_{ij} = \begin{Bmatrix} 1 & \text{if } y_{ij} > 0 \\ 0 & \text{otherwise} \end{Bmatrix}$$

The utilisation distribution for grid cells visited was defined as:

$$y_{ij} = \begin{Bmatrix} \text{NA} & \text{if } z_{ij} = 0 \\ y_{ij} & \text{otherwise} \end{Bmatrix}$$

The full model was specified as:

$$z_{ij} \sim \text{Bernoulli}(\phi_{ij})$$

$$y_{ij} \sim \text{Gamma}\left(\frac{\mu_{ij}^2}{\sigma^2}, \frac{\sigma^2}{\mu_{ij}}\right)$$

With the linear predictor for the Bernoulli model specified as:

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

193 Where β_0 represents the intercept, $X_{ij}^T\beta_i$ represents a vector of covariate effects and γ_j represents

194 the additive terms of random effects for individual. For the Gamma component, σ^2 is the variance

195 and the linear predictor μ_{ij} is specified as:

196

$$\log(\mu_{ij}) = \alpha_0 + X_{ij}^T\alpha + \gamma_j$$

197

198 With α_0 representing the intercept, $X_{ij}^T\alpha$ representing a vector of coordinates and γ_j representing

199 the random effects. Weakly informative normal priors specified as *Normal* (0,0.01) were used for all

200 intercepts and coefficients. Bayesian inference was implemented using integrated nested Laplace

201 approximation (INLA) (36). This approach uses a deterministic algorithm for Bayesian inference,

202 increasing computational efficiency relative to Markov chain Monte Carlo and other simulation-

203 based approaches (33). We did not explicitly include spatial autocorrelation as several distance-

204 based covariates were included (e.g. distance from own house) and multicollinearity between

205 covariates and spatial structure has been shown to decrease performance of resource utilisation

206 functions (32). Models were limited to land areas within 5km of households included in the study

207 site and separate models were fit for each site.

208

209 *Exposure to infected vectors*

210 To estimate vector biting rates, we assembled data from 328 nights of entomological surveillance

211 activities conducted with 5km of the Matunggong study site while GPS tracking was on-going,

212 including: monthly longitudinal surveillance (23), investigations surrounding households of cases and

213 controls (24), and environmentally stratified outdoor catches (37). We limited this data to counts of

214 *An. balabacensis*, the primary *knowlesi* vector which comprises over 95% of *Anopheles* caught in this

215 region. Plausible environmental covariates were assembled, including land use type, slope, aspect,

216 elevation, topographic wetness index, enhanced vegetation index (EVI), population density and

217 average monthly temperature and rainfall. A Bayesian hierarchical spatiotemporal model was fit

with *An. balabacensis* counts as the outcome and the log number of person nights per catch as an offset, including month as a random effect (Supplementary Information).

As these vectors are rarely reported indoors (24) and human landing catches were primarily conducted outside, we excluded areas within houses for calculations of exposure risks. Spatially explicit exposure risks were calculated as derived quantity from human resource utilisation and mosquito biting rate models, using a beta distribution for *P. knowlesi* sporozoite rates parameterised using available infection data from this area. Individual risk was explored using a simple exposure assessment model where an individual's risk of exposure is the sum of the probability of exposure to an infected vector across all locations visited, with the risk of individual *i* being exposed to an infected vector in month *t* as:

Eq. 1

$$r_{it} = \sum_{j=1}^{J} y_{ij} a_{jt} Z_j$$

Where *j* indexes the grid cells visited, y_{ij} is the utilisation distribution, a_{jt} is the number of bites per individual in that cell and month, and Z_j is the proportion of infectious mosquitoes in that cell (4). As these vectors are rarely reported indoors and human landing catch data was not available for inside houses, we excluded areas inside houses. To evaluate places associated with exposure for the entire community, we calculated the number of infectious bites per grid cell each month as:

Eq. 2

$$r_{jt} = \sum_{i=1}^{I} Y_{ij} a_{jt} Z_j$$

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) (36). Model fit was assessed using deviance information

criteria (DIC) and area under the receiver operating curve (AUC) or root mean square error (RMSE) (Supplementary Information).

Ethics approval

This study was approved by the Medical Research Sub-Committee of the Malaysian Ministry of Health and the Research Ethics Committee of the London School of Hygiene and Tropical Medicine. 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 (109 in Limbuak and 134 in Matunggong). The most commonly reported occupation was farm or plantation work (n=74), 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, varied by gender and occupation. Individuals at the more rural Limbuak site covered larger distances (Supplementary Information, Table S2). 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).

264 For both study areas, we developed a model of community space use during peak mosquito biting
265 hours (6pm – 6am), in the form of resource utilisation functions, i.e. predictions of time- and space-
266 specific presence (UDs) on the basis of spatial and environmental variables (31). During these times,
267 human space use (UDs) was mostly predictable and negatively correlated with distance from the
268 individual's house, other houses, roads and slope. The AUC for presence/ absence models was 0.936
269 for Matunggong and 0.938 for Limbuak and RMSE for the overall model was 0.0073 and 0.0043 for
270 Matunggong and Limbuak respectively. While individuals were more likely to use areas further away
271 from forests in the Matunggong site, human space use was positively correlated with proximity to
272 forests in the Limbuak site (Table 1). Despite marked differences between different demographic
273 groups and seasons observed during 24 hour movements, these factors did not improve the
274 predictive power of the model for movements between 6pm and 6am.
275
276 **Table 1.** Estimated coefficients for fixed effects of resource utilisation functions

	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

277
278 Data on mosquito biting rates were assembled from contemporaneous human landing catches
279 conducted in the Matunggong site (Supplementary Information, Table S4). Between August 2013

280 and December 2015, 4814 *An. balabacensis* were caught from 328 sampling nights in 155 unique
281 locations. The median biting rate was 2.1 bites per night per person, ranging from 0 – 28 bites per
282 person per night. Although no associations were identified between land classification and vector
283 density in this site, models identified positive relationships with enhanced vegetation indices (EVI)
284 and negative associations with distance to forest and human population density (Supplementary
285 Information, Table S7). Despite monthly variation, no significant time lags or seasonal patterns in
286 mosquito biting rates were identified (Supplementary Information, Figure S4). Of 1524 mosquitoes
287 tested for *Plasmodium* sporozoites, 4 were positive for *P. knowlesi*, giving a median sporozoite rate
288 of 0.24%.
289
290 **Table S6.** Posterior rate ratio estimates and 95% Bayesian credible interval (BCI) for model 4

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

291
292
293 For individuals included in the GPS tracking study in Matunggong, where both human movement
294 and entomology data was available, we calculated exposure risks as a derived quantity from
295 utilisation distributions and mosquito biting rate models (Methods, Eq. 1). Exposure varied markedly
296 between individuals, with an overall 150-fold difference in predicted mean probabilities of infected
297 bites per night (range: 0.00005-0.0078). No clear differences were observed between genders, age
298 groups or occupations of individuals sampled (Supplementary Information, Figure S5) and there was
299 no association between risk and distance travelled (R^2= 0.0032, p = 0.24).
300
301 Using the resource utilisation function with demographic and spatial data for all individuals in
302 Matunggong, we predicted community-wide space use (Figure 3) and estimated exposure to

303 infected mosquitoes (Methods, Eq. 2) (Figure 1). The predicted number of person nights per grid cell
304 for the entire community ranged from 0 to 12.79 (median: 0.01, IQR: 0.0004 – 0.99), with the mean
305 probability of a community member exposed to an infected bite per grid cell of 0.00082 (IQR:
306 0.00001, 0.00050). Although over 43% of the study site is forest and relatively high biting rates were
307 predicted in forests during the study period (mean: 1.94, range: 0.04 – 12.59), this habitat was rarely
308 used by people in the evenings, with less than 8% of predicted person-nights in forests. Due to these
309 space use patterns, over 91% of predicted infected bites were predicted within 500m of houses.
310 Highest exposure risks were consistently found near forest edges and in close proximity to
311 households, despite spatial and temporal heterogeneity and model uncertainty (Figure 2).
312
313 **Discussion**
314
315 This study highlights the importance of human movement and behaviour patterns and land use in
316 determining exposure to zoonotic and vector-borne diseases such as *P. knowlesi*. Although *P.*
317 *knowlesi* has previously been associated with forest exposure (e.g. (19)) and higher biting rates have
318 been reported in forest interiors (23), this novel approach incorporating both mosquito and human
319 space use data provides a new perspective on peri-domestic transmission, with more than 90% of
320 infectious bites predicted in areas surrounding households at forest edges. This study additionally
321 demonstrates the utility of ecological methods to understand human movement and identify
322 geographical areas associated with higher contact with disease vectors.
323
324 Within these communities, local movement patterns during peak vector times were largely
325 predictable and could be explained by spatial and environmental factors. However, despite this
326 finding, there was substantial variation in predicted exposure between individuals as a result of
327 heterogeneity in habitats used. No significant differences in exposure were predicted between men
328 and women, with individuals with high exposure risks identified across occupational and age groups.

329 Although this finding differs from clinical reports, a comprehensive survey within this community
330 identified equal proportions of men and women exposed to *P. knowlesi* as evidenced by specific
331 antibody responses and data on asymptomatic infections suggests higher numbers of non-clinical
332 infections in women (28, 38). While infrequent events (such as hunting trips) may contribute to
333 these differences in clinical cases, this analysis highlights the importance of routine movements into
334 local environments in shaping exposure risks.
335
336 This improved understanding of how local human land use is related to exposure risk has important
337 implications for surveillance and control programmes. Malaria control programmes often rely on
338 interventions within the house, such as insecticide treated bednets and indoor residual spraying;
339 however, movements outside during peak biting times highlight the importance of also targeting
340 outdoor transmission. Although initial *P. knowlesi* cases were primarily identified in adult men living
341 and working in forests (20), this study illustrates the potential importance of peri-domestic habitats
342 in transmission and provides quantitative insight on mixing between people and infected
343 mosquitoes in forest fringe areas. As Malaysia moves towards malaria elimination, surveillance
344 systems are incorporating novel focal investigation methods, including monitoring changes in local
345 land use and populations at risk likely to be affected by these movement patterns (39).
346
347 Even with the large and highly detailed movement dataset analysed, this study was limited by the
348 availability of mosquito data; as human landing catch data was assembled from other studies, there
349 was not uniform spatial and temporal coverage of the study site increasing uncertainty. The limited
350 mosquito data availability precluded development of mosquito biting rate models for Limbuak and
351 other outlying islands. An additional limitation to estimating mosquito biting rates was the difficulty
352 obtaining spatially and temporally resolute remote sensing data for predictors due to high cloud
353 cover (40). As few positive mosquitoes were identified, uniform estimates of sporozoite rates based
354 on available data were used across the Matunggong site; if further data was available, these models

355 could be refined to incorporate estimates of human and macaque density, mosquito biting
356 preferences in different habitats and infection levels in all hosts (41).
357
358 Despite these limitations, this is the first large-scale study to utilise GPS tracking data and ecological
359 methods to create fine-scale maps of exposure risk. This study highlights the importance of
360 incorporating heterogenous patterns of human space use into disease models, as the majority of
361 human exposure may occur in areas with lower vector biting rates but greater probabilities of
362 human use. Further, results quantitatively illustrate the importance of forest edges and local habitat
363 in *P. knowlesi* transmission and can inform understanding of other zoonotic and vector-borne
364 diseases.
365
366 **Acknowledgements**
367 We would like to thank the MONKEYBAR project team and the participants in Sabah, Malaysia for
368 their help with these studies. We acknowledge the Medical Research Council, Natural Environmental
369 Research Council, Economic and Social Research Council and Biotechnology and Biosciences
370 Research Council for funding received for this project through the Environmental and Social Ecology
371 of Human Infectious Diseases Initiative, grant no. G1100796. This work was additionally supported
372 by the National Socio-Environmental Synthesis Center (SESYNC).

373 **References:**

374 1. Lambin EF, Tran A, Vanwambeke SO, Linard C, Soti V. Pathogenic landscapes: interactions
375 between land, people, disease vectors, and their animal hosts. Int J Health Geogr. 2010;9:54.
376 2. Patz JA, Daszak P, Tabor GM, Aguirre AA, Pearl M, Epstein J, et al. Unhealthy landscapes:
377 Policy recommendations on land use change and infectious disease emergence. Environ Health
378 Perspect. 2004;112(10):1092-8.
379 3. Stoddard ST, Forshey BM, Morrison AC, Paz-Soldan VA, Vazquez-Prokopec GM, Astete H, et
380 al. House-to-house human movement drives dengue virus transmission. Proc Natl Acad Sci U S A.
381 2013;110(3):994-9.
382 4. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Paz Soldan V, Kochel TJ, Kitron U, et al.
383 The role of human movement in the transmission of vector-borne pathogens. PLoS Negl Trop Dis.
384 2009;3(7):e481.
385 5. Pindolia DK, Garcia AJ, Wesolowski A, Smith DL, Buckee CO, Noor AM, et al. Human
386 movement data for malaria control and elimination strategic planning. Malar J. 2012;11(205).
387 6. Moyes CL, Shearer FM, Huang Z, Wiebe A, Gibson HS, Nijman V, et al. Predicting the
388 geographical distributions of the macaque hosts and mosquito vectors of Plasmodium knowlesi
389 malaria in forested and non-forested areas. Parasit Vectors. 2016;9(1):242.
390 7. Imai N, White MT, Ghani AC, Drakeley CJ. Transmission and control of Plasmodium knowlesi:
391 a mathematical modelling study. PLoS Negl Trop Dis. 2014;8(7):e2978.
392 8. Fornace KM, Abidin TR, Alexander N, Brock P, Grigg MJ, Murphy A, et al. Association
393 between Landscape Factors and Spatial Patterns of Plasmodium knowlesi Infections in Sabah,
394 Malaysia. Emerg Infect Dis. 2016;22(2):201-8.
395 9. Shearer FM, Huang Z, Weiss DJ, Wiebe A, Gibson HS, Battle KE, et al. Estimating
396 Geographical Variation in the Risk of Zoonotic Plasmodium knowlesi Infection in Countries
397 Eliminating Malaria. PLoS Negl Trop Dis. 2016;10(8):e0004915.
398 10. Brock PM, Fornace KM, Grigg MJ, Anstey NM, William T, Cox J, et al. Predictive analysis
399 across spatial scales links zoonotic malaria to deforestation. Proceedings of the Royal Society B.
400 2019;286(1894).
401 11. Acevedo MA, Prosper O, Lopiano K, Ruktanonchai N, Caughlin TT, Martcheva M, et al. Spatial
402 heterogeneity, host movement and mosquito-borne disease transmission. PLoS One.
403 2015;10(6):e0127552.
404 12. Yakob L, Lloyd AL, Kao RR, Ferguson HM, Brock PM, Drakeley C, et al. Plasmodium knowlesi
405 invasion following spread by infected mosquitoes, macaques and humans. Parasitology.
406 2018;145(1):101-10.
407 13. Chang HH, Wesolowski A, Sinha I, Jacob CG, Mahmud A, Uddin D, et al. Mapping imported
408 malaria in Bangladesh using parasite genetic and human mobility data. Elife. 2019;8.
409 14. Wesolowski A, Qureshi T, Boni MF, Sundsoy PR, Johansson MA, Rasheed SB, et al. Impact of
410 human mobility on the emergence of dengue epidemics in Pakistan. Proc Natl Acad Sci U S A.
411 2015;112(38):11887-92.
412 15. Wesolowski A, Metcalf CJ, Eagle N, Kombich J, Grenfell BT, Bjornstad ON, et al. Quantifying
413 seasonal population fluxes driving rubella transmission dynamics using mobile phone data. Proc Natl
414 Acad Sci U S A. 2015;112(35):11114-9.
415 16. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, et al. Quantifying the
416 impact of human mobility on malaria. Science. 2012;338(6104):267-70.
417 17. Wesolowski A, Buckee CO, Engo-Monsen K, Metcalf CJE. Connecting Mobility to Infectious
418 Diseases: The Promise and Limits of Mobile Phone Data. J Infect Dis. 2016;214(suppl_4):S414-S20.
419 18. Vazquez-Prokopec GM, Bisanzio D, Stoddard ST, Paz-Soldan V, Morrison AC, Elder JP, et al.
420 Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics
421 in a resource-poor urban environment. PLoS One. 2013;8(4):e58802.

422 19. Grigg MJ, Cox J, William T, Jelip J, Fornace KM, Brock PM, et al. Individual-level factors
423 associated with the risk of acquiring human Plasmodium knowlesi malaria in Malaysia: a case control
424 study. Lancet Planetary Health. 2017;1:e97-104.
425 20. Singh B, Kim Sung L, Matusop A, Radhakrishnan A, Shamsul SS, Cox-Singh J, et al. A large
426 focus of naturally acquired Plasmodium knowlesi infections in human beings. Lancet.
427 2004;363(9414):1017-24.
428 21. Yasuoka J, Levins R. Impact of deforestation and agricultural development on anopheline
429 ecology and malaria epidemiology. Am J Trop Med Hyg. 2007;76(3):450-60.
430 22. Barber BE, William T, Dhararaj P, Anderios F, Grigg MJ, Yeo TW, et al. Epidemiology of
431 Plasmodium knowlesi malaria in north-east Sabah, Malaysia: family clusters and wide age
432 distribution. Malar J. 2012;11:401.
433 23. Wong ML, Chua TH, Leong CS, Khaw LT, Fornace K, Wan-Sulaiman WY, et al. Seasonal and
434 Spatial Dynamics of the Primary Vector of Plasmodium knowlesi within a Major Transmission Focus
435 in Sabah, Malaysia. PLoS Negl Trop Dis. 2015;9(10):e0004135.
436 24. Manin BO, Ferguson HM, Vythilingam I, Fornace K, William T, Torr SJ, et al. Investigating the
437 Contribution of Peri-domestic Transmission to Risk of Zoonotic Malaria Infection in Humans. PLoS
438 Negl Trop Dis. 2016;10(10):e0005064.
439 25. Lee KS, Divis PC, Zakaria SK, Matusop A, Julin RA, Conway DJ, et al. Plasmodium knowlesi:
440 reservoir hosts and tracking the emergence in humans and macaques. PLoS Pathog.
441 2011;7(4):e1002015.
442 26. Chua TH, Manin BO, Daim S, Vythilingam I, Drakeley C. Phylogenetic analysis of simian
443 Plasmodium spp. infecting Anopheles balabacensis Baisas in Sabah, Malaysia. PLoS Negl Trop Dis.
444 2017;11(10):e0005991.
445 27. Hartemink N, Vanwambeke SO, Purse BV, Gilbert M, Van Dyck H. Towards a resource-based
446 habitat approach for spatial modelling of vector-borne disease risks. Biol Rev Camb Philos Soc.
447 2015;90(4):1151-62.
448 28. Fornace KM, Herman LS, Abidin TR, Chua TH, Daim S, Lorenzo PJ, et al. Exposure and
449 infection to Plasmodium knowlesi in case study communities in Northern Sabah, Malaysia and
450 Palawan, The Philippines. PLoS Negl Trop Dis. 2018;12(6):e0006432.
451 29. Johnston R, Brady HE. The rolling cross-section design. Electoral Studies. 2002;21(2):283-95.
452 30. Benhamou S. Dynamic approach to space and habitat use based on biased random bridges.
453 PLoS One. 2011;6(1):e14592.
454 31. Papworth SK, Bunnefeld N, Slocombe K, Milner-Gulland EJ. Movement ecology of human
455 resource users: using net squared displacement, biased random bridges and resource utilization
456 functions to quantify hunter and gatherer behaviour. Methods in Ecology and Evolution.
457 2012;3(3):584-94.
458 32. Hooten MB, Hanks EM, Johnson DS, Alldredge MW. Reconciling resource utilization and
459 resource selection functions. J Anim Ecol. 2013;82(6):1146-54.
460 33. Blangiardo M, Cameletti M. Spatial and Spatio-temporal Bayesian Models with R-INLA.
461 Chichester, United Kingdom: John Wiley & Sons, Ltd; 2015.
462 34. Sadykova D, Scott BE, De Dominicis M, Wakelin SL, Sadykov A, Wolf J. Bayesian joint models
463 with INLA exploring marine mobile predator-prey and competitor species habitat overlap. Ecology
464 and Evolution. 2017;7:5212-26.
465 35. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species
466 distribution models: how, where and how many? Methods in Ecology and Evolution. 2012;3:327-38.
467 36. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by
468 using integrated nested Laplace approximations. Journal of the Royal Statistical Society.
469 2009;71(Part 2):319-92.
470 37. Ng SH, Homathevi R, Chua TH. Mosquitoes of Kudat: species composition and their medical
471 importance (Diptera: Culicidae). Serangga. 2016;21(2):149-62.

472 38. Fornace KM, Nuin NA, Betson M, Grigg MJ, William T, Anstey NM, et al. Asymptomatic and
473 Submicroscopic Carriage of Plasmodium knowlesi Malaria in Household and Community Members of
474 Clinical Cases in Sabah, Malaysia. J Infect Dis. 2015.
475 39. Sektor Penyakit Bawaan Vektor Bahagian Kawalan Penyakit Kementarian Kesihatan
476 Malaysia. Garis Panduan Pencegahan Malaria Re-Introduction di Malaysia. Putra Jaya: Kementerian
477 Kesihatan Malaysia; 2016. Contract No.: KKM/TM/620/2016.
478 40. Weiss DJ, Mappin B, Dalrymple U, Bhatt S, Cameron E, Hay SI, et al. Re-examining
479 environmental correlates of Plasmodium falciparum malaria endemicity: a data-intensive variable
480 selection approach. Malar J. 2015;14:68.
481 41. Yakob L, Bonsall MB, Yan G. Modelling knowlesi malaria transmission in humans: vector
482 preference and host competence. Malar J. 2010;9:329.

483

484

485 **Figure legends**

486 **Figure 1.** Human movement relative to habitat. A. Example of GPS tracks from a 22-year-old male
487 plantation worker in Matunggong over aerial imagery, B. Individual utilisation distribution calculated
488 from GPS tracks

489

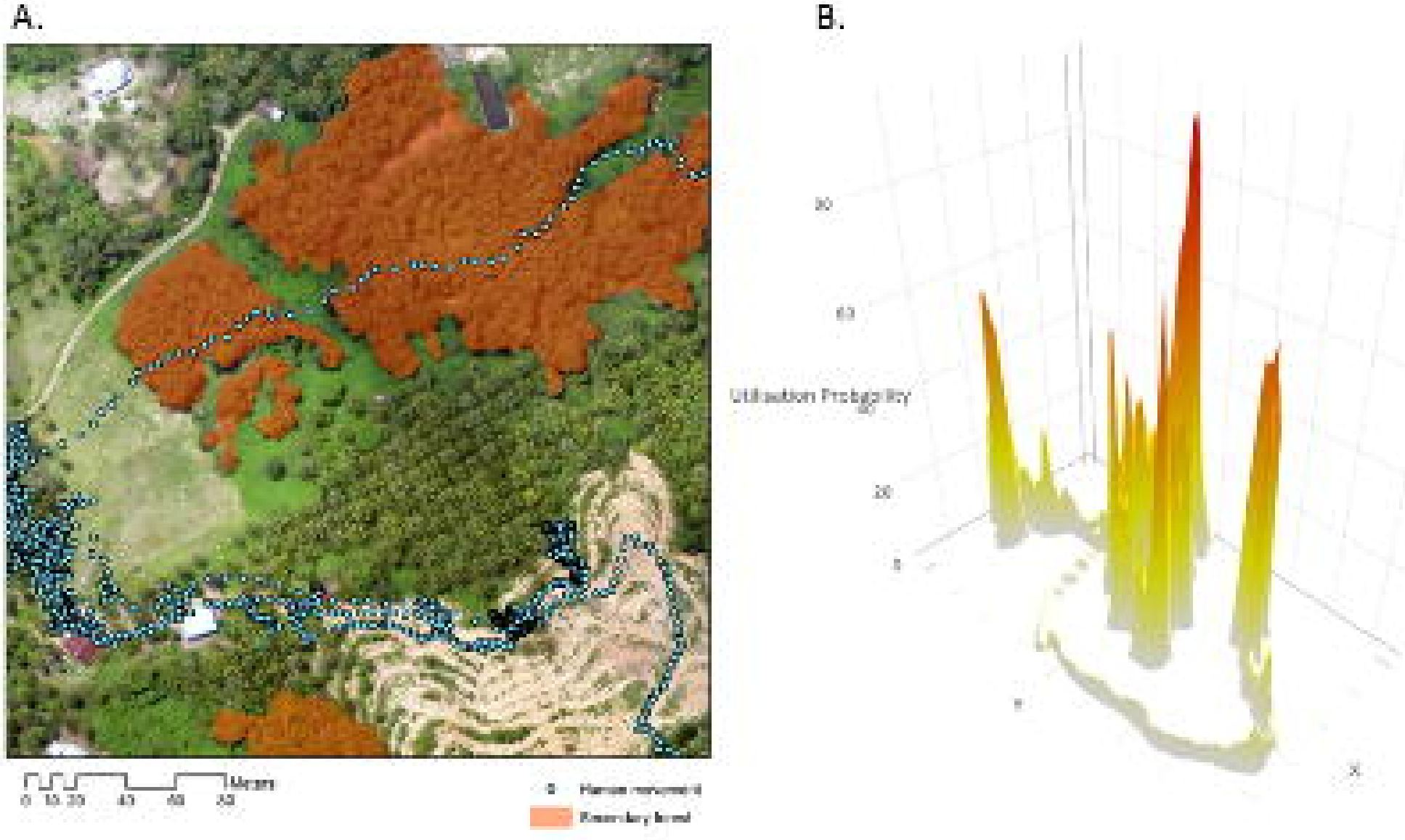
490 **Figure 2.** Mosquito biting rates. A. *An. balabacensis* biting rate per person-night from data collected
491 in Matunggong, B. Predicted mean *An. balabacensis* biting rates per month from spatiotemporal
492 models (Supplementary Information), C. Predicted number of bites for all individuals residing in
493 Matunggong by distance from secondary forest, and by D. Distance from households

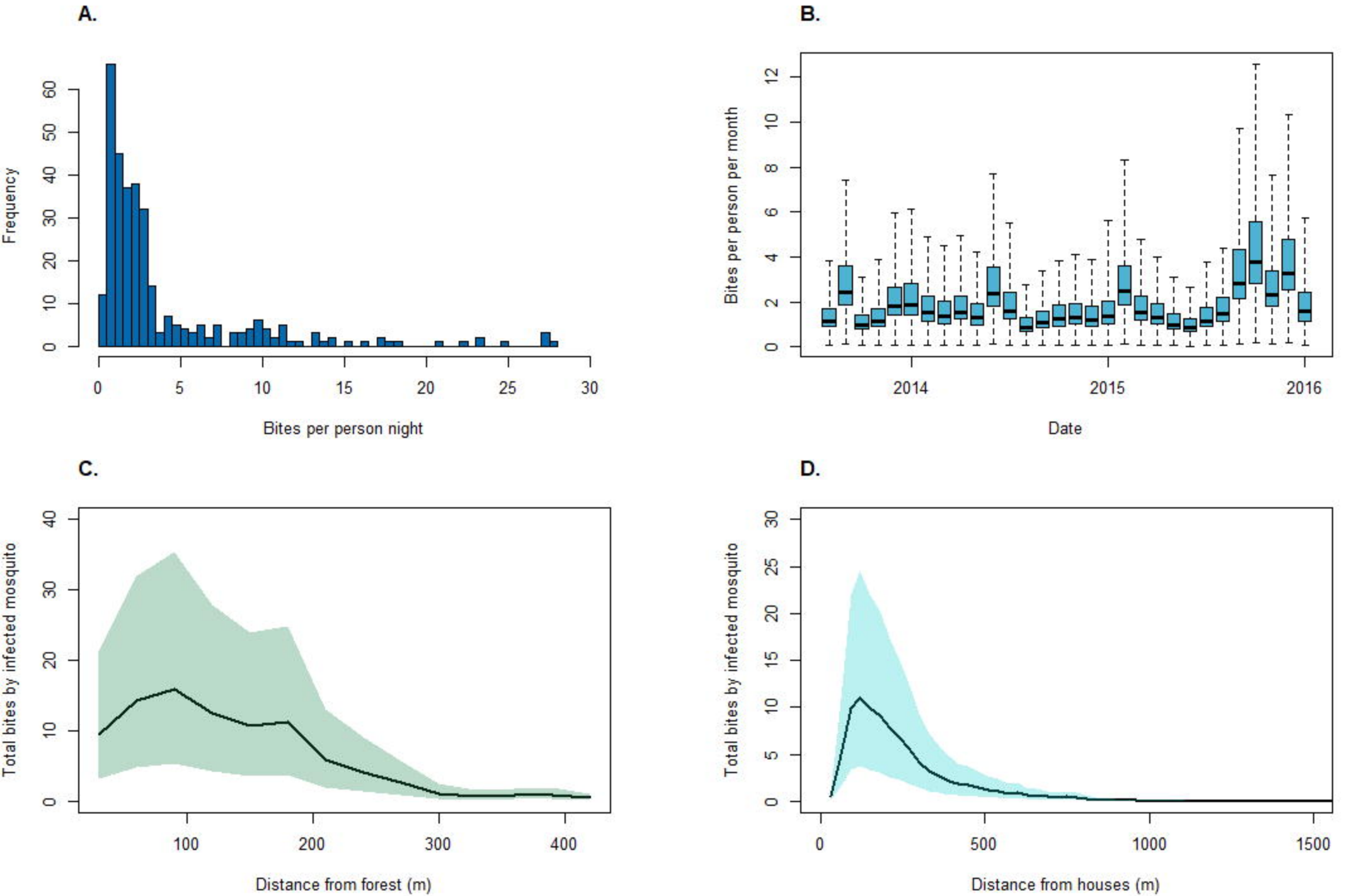
494

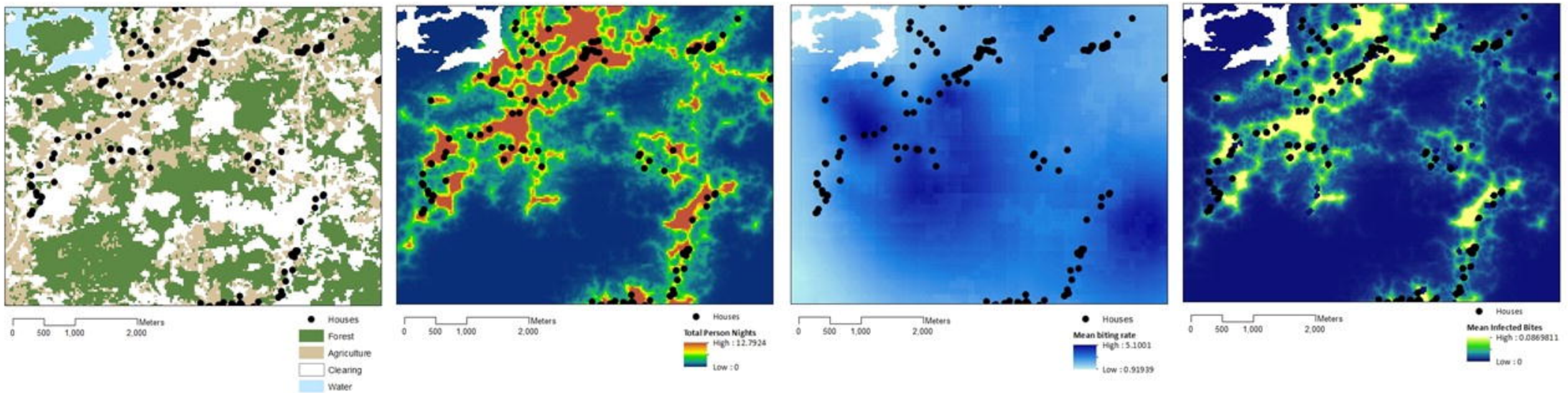
495 **Figure 3.** A. Land use in Matunggong site, B. Predicted number of person- nights for entire
496 community per grid cell, C. Predicted mosquito biting rates, D. Predicted infected bites per grid cell

497

498


A.
B.
Utilisation Probability
Meters
0 10 20 40 60 80
Human movement
Boundary level


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)


0 500 1,000 2,000 Meters
Houses
Forest
Agriculture
Clearing
Water
0 500 1,000 2,000 Meters
Houses
Total Person Nights
High : 12.7924
Low : 0
0 500 1,000 2,000 Meters
Houses
Mean biting rate
High : 5.1001
Low : 0.91939
0 500 1,000 2,000 Meters
Houses
Mean Infected Bites
High : 0.0869811
Low : 0