



## 21 **Abstract**

22 Malaria remains a public health concern. Monitoring the fine-scale heterogeneity of the malaria  
23 burden enables more targeted control efforts. Although malaria indicator surveys (MIS) have been  
24 crucial in evaluating the progress of malaria control interventions, they are only designed to provide  
25 a cross-sectional national and regional malaria disease burden. Recent advances in geostatistical  
26 methods allow us to interpolate national survey data to describe subnational disease burden that is  
27 crucial in informing targeted control.

28 A binomial geostatistical model employing Markov chain Monte Carlo (MCMC) parameter estimation  
29 methods is used to understand the spatial drivers of malaria risk in Kenya and to predict malaria risk  
30 at a fine-scale resolution, including identifying hotspots.

31 A total of 11,549 children aged six months to 14 years from 207 clusters were sampled in this survey  
32 and used in the present analysis. The national malaria prevalence based on the data was 8.4%, with  
33 the highest in the lake endemic zone (18.1 %) and the lowest in the low-risk zone (<1 %). The  
34 analysis shows that elevation, proportion of ITN distributed, rainfall, temperature and urbanization  
35 covariates are all significant predictors of malaria transmission. The 5x5 Km resolution maps show  
36 that malaria is heterogeneous in Kenya, with hotspot areas in the lake endemic area, the coastal  
37 areas, and some parts of the shores of Lake Turkana and Kajiado.

38 The high-resolution malaria prevalence maps produced as part of the analysis have shown that  
39 Kenya has additional malaria hotspots, especially in areas least expected. These findings call for a  
40 rethinking of malaria burden classification in some regions for effective planning, implementation,  
41 resource mobilization, monitoring, and evaluation of malaria interventions in the country.

42

43

44

45

## 46 Background

47 Malaria remains a public health concern and continues to be one of the most important tropical  
48 diseases affecting human populations to date [1]. In 2020, an estimated 241 million malaria cases  
49 occurred worldwide, 90% of which were in sub-Saharan Africa [2]. The disease is caused by protozoa  
50 of the genus *Plasmodium* of which five known species, *Plasmodium falciparum*, *P. vivax*, *P. ovale*,  
51 and *P. malariae*, more recently, *P. knowlesi*, are responsible for human infection [3,4]. The vector  
52 responsible for human transmission is the female anopheles mosquito. In Sub-Saharan Africa,  
53 malaria is one of the leading causes of morbidity and mortality, especially in children under five.  
54 Other high-risk groups include pregnant women and immunologically naïve persons like travellers  
55 coming from non-endemic places [5].

56 Malaria remains a significant public health problem in Kenya, accounting for an estimated 13% to  
57 15% of outpatient cases, with nearly 70% of the population at risk for malaria [6]. Four out of the  
58 five species of malaria parasite that cause human infection are present in Kenya, but the  
59 *Plasmodium falciparum* parasite is the predominant cause of infection in the country. Over the past  
60 decade, Kenya has substantially scaled up available malaria control tools, such as insecticide-treated  
61 bed nets, indoor residual spraying and the use of artemisinin-based combination therapies [6].  
62 Evidence of this massive scale-up of interventions is the observed decline in prevalence. Kenya has  
63 experienced a decrease in the national prevalence of malaria among children ages six months to  
64 14 years, from 13 per cent in 2010 to 8 per cent in 2015, and % in 2020 [6].

65 One of the key objectives of the Kenya Health Policy 2014–2030 is the elimination of communicable  
66 diseases, including malaria. This is supported by the Kenya malaria strategy for 2019 to 2023, which  
67 sets a vision of a malaria-free Kenya and targets to reduce malaria incidence and mortality by  
68 seventy-five per cent by 2023, with 2016 as the baseline year[6].

69

70 Malaria transmission in Kenya varies geographically. This could be due to varied climatic conditions,  
71 vector and parasite resistance, differences in intervention uptake across populations and other  
72 unmeasured factors that are thought to be responsible for this increasing heterogeneity [6,7]. The  
73 country is administratively divided into five malaria epidemiological zones based on risk profiles.  
74 These zones include highland epidemic-prone areas, lake endemic areas, coast endemic, semi-arid  
75 seasonal, and low-risk malaria areas. The endemic areas lie in the lake and coastal regions with  
76 altitudes ranging from 0m to 1300 above sea level. These areas have perennial malaria transmission  
77 due to rainfall, temperature, humidity and other critical factors that drive malaria transmission. The  
78 semi-arid seasonal malaria transmission areas are in the country's northern, northeastern, and  
79 southeastern parts. These areas experience short periods of intense malaria. The highland epidemic-  
80 prone areas are located within the western highlands and have seasonal malaria transmission with  
81 some yearly variation. The altitude in these zones is relatively higher than the other zones, lying  
82 1500 meters above sea level. The malaria epidemics in the highland epidemic-prone zones are less  
83 predictable. Lastly, the low-risk malaria areas cover Nairobi and the central highland. Temperatures  
84 are usually too low to allow the completion of the sporogony cycle of the malaria parasite in the  
85 vector in the low-risk zones[6] .

86 Malaria indicator surveys (MIS) measure progress on key malaria indicators in Kenya. The country  
87 has conducted four MIS in 2007, 2010, 2015 and 2020. The MIS are nationally representative  
88 household surveys that provide estimates of national and regional malaria indicators to assist  
89 malaria control programs in tracking their progress and evaluating the impact of strategies and  
90 interventions. The MIS follow a standard methodology recommended by the Roll Back Malaria  
91 Monitoring and Evaluation working group guidelines [8]. Originally, MIS surveys were designed to  
92 measure the blanket scale up of interventions like bed nets, using a classic two-stage sample design  
93 and coverage indicators as the primary endpoints. Over time, as coverage increased, interest  
94 expanded to the impact of parasite prevalence. While it is still a norm that the MIS traditionally  
95 measures progress in these areas, the survey methodologies need to consider the underlying disease

96 heterogeneity [8]. Recent advances in statistical analyses, including geostatistics, have made it  
97 possible to make fine-scale inferences of malaria transmission from survey data like the MIS, that is  
98 not traditionally designed for such extrapolations. In this paper, we use the Kenya 2020 MIS data to  
99 i) understand the relationship between malaria prevalence and several factors including  
100 environmental factors and ii) understand the disease heterogeneity across the country's surface  
101 including identification of hotspots.

## 102 **Methods**

### 103 **Country Profile**

104 Kenya is an East African country that covers an area of 582,550 km<sup>2</sup>. It is bordered by Ethiopia to the  
105 north, Tanzania to the south, Uganda to the west, South Sudan to the northwest, and Somalia to the  
106 northeast. Approximately 80% of Kenya's land is arid and semi-arid, only 20% is arable, and only  
107 1.9% of the total surface area is occupied by standing water. The great East African Rift Valley  
108 extends from Lake Victoria to Lake Turkana and further southeast to the Indian Ocean [9]. The  
109 country has a number of large rivers including the Tana, Galana, Turkwel and Nzoia [10]. Figure 1  
110 below is a map of Kenya showing the five epidemiological zones as defined by the national malaria  
111 program [6].

112 *Figure 1: Kenya epidemiological malaria zones*

113

### 114 **Data**

115 This secondary analysis used data from the Kenya MIS [6]. Access to the dataset was given to the  
116 authors on Apr 28 2022. The datasets were de-identified. The IRB-approved procedures for DHS  
117 public-use datasets do not in any way allow respondents, households, or sample communities to be  
118 identified. Authors had no access to the names of individuals or household addresses in the data

119 files. Additionally, the geographic identifiers only go down to the regional level, which is hard to  
120 identify individuals.

121

122 The 2020 MIS, the fourth conducted by the country, followed a similar design and set-up as the  
123 former ones. It was conducted during the peak malaria season in November and December 2020. A  
124 two-stage stratified sampling design was used, powered to give malaria parasite prevalence  
125 estimates and other key malaria indicators at the national level (urban and rural areas) and for the  
126 five epidemiological zones. The first stage sampling unit was a cluster developed from enumeration  
127 areas (EAs). EAs are the smallest geographical areas created for purposes of census enumeration.  
128 The EAs used were based on the 2019 Kenya population census. In the Kenya MIS, a cluster was  
129 defined as either an EA or part of an EA. A total of 301 clusters (134 urban and 167 rural) were  
130 sampled in this first stage using the probability proportional to size approach. The second stage  
131 sampling unit was households. In each cluster, 30 households were selected from a line listing of the  
132 sampled clusters using a systematic random sampling approach. A total of 7,952 households were  
133 sampled. All women aged 15-49 in the selected households were eligible for individual interviews.  
134 They were asked questions about preventing malaria during pregnancy and treating childhood  
135 fevers. In addition, the survey included testing for anaemia and malaria among children aged six  
136 months to 14 years using a finger- or heel-prick blood sample.

137 Permission to use the dataset was obtained from The Demographic and Health Surveys (DHS)  
138 Program through the archiving office. The original study received ethical clearance from the  
139 Kenyatta National Hospital/University of Nairobi Scientific and Ethics Review Committee. All  
140 participants provided oral informed consent.

141

## 142 **Variables**

### 143 **Outcome variable**

144 In this analysis, the outcome variable was a binary outcome derived from the total number of  
145 children tested and the total number testing positive. This was extrapolated to estimate the cluster-  
146 level plasmodium falciparum malaria prevalence (*PfPR*).

147

### 148 **Explanatory variables**

149 The explanatory variables included cluster-level factors such as rainfall, temperature, elevation, and  
150 urbanization and individual-level characteristics such as gender and age.

### 151 **Data cleaning**

152 The data cleaning and analysis were done in R[11]. Maps produced in the analysis were further  
153 processed for better visualization in QGIS (Version 3.2). Relevant variables were extracted and  
154 renamed to shorter names for ease of coding. Coordinates in the initial dataset were given the  
155 longitude and latitude system, which were transformed into the universal coordinate system (UTM).  
156 All distances were scaled to kilometres.

### 157 **Exploratory analysis**

158 The initial exploratory analysis was descriptive to understand the data and to explore the initial  
159 relationships between the outcome variable of prevalence with the covariates in the data set.  
160 Scatter plots with fitted linear regression lines were used for this step to observe the relationship  
161 between prevalence and the explanatory variables. To further understand the variables, correlation  
162 plots were used to understand the relationships between the variables to guide the decisions of  
163 which covariates to include in the Model. The additional exploratory analysis involved plotting the  
164 clusters on the Kenyan map's surface, showing the sampled cluster's distribution and the crude  
165 malaria prevalence.

### 166 **Model fitting**

167 The first objective of the analysis was to understand the relationships between malaria prevalence  
168 and several factors, including environmental factors. Several steps were followed:

- 169 1. Fitting a generalized linear model
- 170 2. Assessing evidence of residual correlation
- 171 3. Fitting a generalized linear mixed model
- 172 4. Reassessing evidence of residual correlations
- 173 5. Fitting a binomial geostatistical model and parameter estimation
- 174 6. Model validation

175 The model description for the generalized linear Model and the generalized linear mixed Model are  
176 described in Supplementary file 1.

177

## 178 **Model description for the binomial geostatistical Model**

179 Let  $Y_i$  denote the number of individuals that test positive for plasmodium falciparum at survey  
180 cluster location  $x_i$

181 And that the survey team went to the sampled clusters given by  $x_i$  and sampled  $m_i; i = 1 \dots n$   
182 individuals at risk in the cluster and recoded the outcome of every person that tests positive and  
183 negative for plasmodium falciparum malaria.

184 Then standard geostatistical Model assumes that:

$$185 \quad Y_i \sim \text{Binomial} ( m_i, P( x_i) )$$

186  $Y_i$  is a Binomial distribution with  $m_i$  trials and probability of a positive test  $P( x_i)$  specified in the  
187 binomial geostatistical Model below:

188

$$189 \quad \log \left\{ \frac{P(x)}{1 - P(x)} \right\} = \alpha + d( x_i)^T \beta + S(x) + Z_i$$



190 Where  $\alpha$  is the intercept parameter,  $S(x)$  are the spatial random effects, representing the spatial  
191 variation between the sampled clusters.  $Z_i$  are mutually independent zero-mean Gaussian random  
192 variables with variance  $r$  and in this analysis represent the spatial variation within cluster variation,  
193 measurement error or small-scale spatial variation.

194  $d(x_i)^T$  is a vector of observed spatially referenced explanatory variables associated with the  
195 response  $Y_i$ , and  $\beta$  is a vector of spatial regression coefficients for the covariates.

196 The Matérn correlation function for the stationary Gaussian processes  $S(x)$  used in this analysis, a  
197 two-parameter family is given by:

198

$$199 \quad p(u, \varphi, k) = 2^{k-1} (u/\varphi)^k K_k + (u/\varphi)$$

200 Where:

- 201 •  $u$  denotes the distance between two locations  $x$  and  $x'$ ,
- 202 •  $\varphi > 0$  is a scale parameter that determines the rate at which correlation decays to 0 as the  
203 distance increases, and
- 204 •  $k > 0$ , is a smoothness parameter which determines the analytic smoothness of the underlying  
205 process  $S(x)$ .

206 In the binomial geostatistical regression for this analysis, the Matérn shape parameter  $k$  was set to  
207 0.5 variance parameters  $\tau^2$  to 0.

208 The covariates  $d(x_i)^T$  used in the binomial geostatistical Model for prediction were obtained from  
209 an exploratory analysis set to understand the relationship of the variables with the outcome variable  
210 of malaria prevalence. This Model included the covariates: elevation, ITN usage, mean temperature,  
211 rainfall, and cluster urbanization (urban vs rural). The Markov chain Monte Carlo (MCMC) methods  
212 were used for parameter estimation in this Model. Confidence intervals of the estimates are  
213 calculated on the log scale then transformed back to the non-log scale that is used to report the  
214 results.

215 To test whether there was any evidence against spatial correlation in the data, empirical variogram  
216 methods are used. A simulation of 1000 empirical variograms around the fitted Model is ran and  
217 these are used to compute 95% confidence intervals at any given spatial distance of the variogram. A  
218 conclusion is reached that that there is a spatial correlation in the data if the empirical variogram  
219 obtained from the data falls outside the 95% tolerance bandwidth.

220 The second objective is to understand the disease heterogeneity across the surface of the country,  
221 including identification of hotspots and the uncertainty attached to these hot spots. For this  
222 purpose, a binomial geostastical model was used as described above but with covariates that were  
223 available as raster. These included urbanization, temperature, and precipitation. The target for the  
224 predictions was a prevalence of malaria over the 5 x 5 km regular grid surface covering the whole  
225 surface of Kenya. A map of malaria prevalence was generated. Uncertainty of the prevalence was  
226 addressed using *Exceedance Probabilities*, an approach that is more relevant to policy makers, than  
227 the traditional approach of using confidence intervals. *Exceedance Probabilities (EP)* method sets  
228 policy relevant thresholds. The *EP* can be formally expressed as:

$$229 \quad EP = \text{Probability} \{, P(x_i) > t \mid \text{data} \}$$

230 where  $t$  is the prevalence threshold, set to 10% in this analysis.

## 231 **Results**

232 A total of 11,549 children aged six months to 14 years were sampled. The analysis used 297 clusters.  
233 The number of clusters per transmission zone is shown in Table 1 below. The lake endemic area had  
234 the greatest number of clusters (98), while the coastal endemic area had the lowest number of  
235 clusters (29). The national malaria prevalence based on the data was 8.4%, with the highest in the  
236 lake endemic zone (18.1 %) and the lowest in the low-risk zone (<1 %).

237 *Table 1 Malaria Prevalence in Kenya across five epidemiological zones*

<b>Epidemiological Zone</b>	<b>Total clusters</b>	<b>Number tested(N)</b>	<b>Number positive(n)</b>	<b>Prevalence (n/N)%</b>
Coastal Endemic	29	1088	59	4.95
Highland Epidemic Prone	55	2122	33	1.56
Lake Endemic	97	4621	836	17.93
Low Risk	54	1307	1	0.08
Seasonal	56	2210	33	1.49

238

239 The map in Figure one below shows the sampled locations on the left and the cluster-level malaria  
240 prevalence on the right. The lake endemic area has the highest number of clusters sampled and is  
241 also the zone with the highest prevalence estimates at the cluster level.

242 *Figure 2 Map of Sampled locations(left) and Malaria Prevalence (Right)*

243 The weather pattern varies across the surface of Kenya. The maps in Figure 2 below show the  
244 variation in temperature across space on the top, and the variation in annual precipitation for the  
245 year 2020 on the bottom.

246 *Figure 3 Mean temperature (degrees Celsius) (top) and Annual precipitation (mm)in Kenya (2020)*  
247 *(Bottom)*

248

249

## 250 **Binomial geostatistical model results**

251 The binomial geostatistical Model results indicate that the elevation, proportion of ITN distributed,  
252 rainfall, temperature and urbanization covariates are all significant predictors of malaria  
253 transmission (Table 2).

254 The odds of malaria transmission are less in the urban clusters compared to the rural ones. Urban  
255 clusters have nearly 68% less malaria prevalence than rural ones (OR 0.32 CI: 0.26-0.39, P value  
256 <0.0001). The higher the rainfall, the higher the risk of malaria transmission. Every mm increase in

257 the average rain increases malaria prevalence by 1.9 times (OR 1.91 CI 1.69- 2.15, P value <0.0001).  
 258 Rise in mean temperature also increases the risk of malaria prevalence. Every degree increase in  
 259 temperature increases the odds of malaria prevalence by 1.4 times (OR 1.37CI 1.28-1.47, P value  
 260 <0.0001).

261 *Table 2: Binomial geostatistical Model*

262

	Estimate	Standard error	Lower Bound	Upper Bound	p-value
<b>Intercept</b>	0.00	3.54	0.00	0.00	<0.0001
<b>elevation</b>	1.86	1.05	1.70	2.02	<0.0001
<b>INT coverage</b>	3.15	1.07	2.75	3.61	<0.0001
<b>Urban vs Rural</b>	0.32	1.11	0.26	0.39	<0.0001
<b>Rainfall(mm)</b>	1.91	1.06	1.69	2.15	<0.0001
<b>Temperature</b>	1.37	1.04	1.28	1.47	<0.0001
<b>Age in months (ref &lt;12)</b>					
<b>12-23</b>	1.93	1.54	0.83	4.48	<0.001
<b>24-35</b>	2.72	1.52	1.19	6.21	<0.01
<b>36-47</b>	3.56	1.51	1.59	7.96	<0.001
<b>48-59</b>	7.84	1.45	3.76	16.35	<0.001
<b>Female vs male</b>	1.20	1.08	1.03	1.40	<0.001
<b>Sigma^2*</b>	0.56	1.44	0.27	1.14	NA
<b>Phi**</b>	59.57	1.85	17.87	198.50	NA
<b>Tau^2***</b>	0.95	2.24	0.20	4.58	NA

263 **sigma2 is the variance of the Gaussian process, phi is the scale parameter of the spatial correlation**  
 264 **and tau2 is the variance of the nugget effect**

## 265 **Model validation**

266 Using variogram-based techniques described above, the Model above was tested for evidence of  
 267 spatial correlation. The results of this process are shown in figure 4 below. Since the empirical semi-  
 268 variogram (solid line) falls within the 95% confidence interval (grey envelope), this shows that the  
 269 Model is valid; the Model for malaria prevalence is compatible with the data.

270 *Figure 4 Model Validation*

271

## 272 **Prediction**

273 To understand the disease heterogeneity across the country's surface, including the identification of  
274 hotspots, a 5 × 5 km resolution map for malaria prevalence in children six months to 14 years is  
275 presented in Figure 5 below. Overall, malaria prevalence is low in most parts of the country.

276 Hotspots were notable in Western Kenya in the lake endemic areas around Lake Victoria, in the  
277 endemic coastal regions along the Indian Ocean and three hotspot areas within the seasonal  
278 epidemiological zone, one around the Lake Turkana region, one around the humid and sub-humid  
279 belt in Meru County and the other in the semi-arid belt of Kajiado County.

280 *Figure 5 Malaria prevalence predictions among children six months- 14 years in Kenya*

281 Figure 6 below presents a map of malaria exceedance and probabilities, showing areas where  $p(x) \geq$   
282 10% with certainty on the colour gradient.

283 *Figure 6 Exceedance Probabilities.*

284

## 285 **Discussion**

286 Understanding the spatial distribution of malaria and the factors that drive its transmission are key  
287 in malaria control. Given the heterogeneity of malaria transmission in Kenya, defining the malaria  
288 burden at more localized locations is important to allow for targeted control activities. The national  
289 malaria indicator surveys performed in the country are not designed to provide malaria prevalence  
290 estimates at localized levels. This paper uses Model-Based geostatistical methods to understand  
291 malaria transmission drivers in Kenya and map out malaria prevalence at a very high resolution (5 x 5  
292 Km square grid).

293 In our analysis, we have found that several factors influence malaria transmission. These include  
294 gender, age, temperature, rainfall, bed net coverage, elevation, and urbanization. This is consistent  
295 with the well-known predictors of malaria transmission. Studies in the same area have previously  
296 found a higher risk of malaria among males and an increasing risk with age compared to the first  
297 year of life [11]. Both the natural environment and the artificial environment are known to affect  
298 malaria transmission. Temperature, humidity, and rainfall all have interactions with mosquitos at  
299 specific points in their life cycle. Temperature regulates the development of mosquitos at each  
300 stage. The laying of eggs by mosquitos is reduced in temperature extremes, either too cold or too  
301 hot [12]. Temperature also affects the mosquito stage transition, with the optimal temperature  
302 being between 22 and 26 degrees Celsius[11].

303 Rainfall has been shown to be positively correlated with high malaria transmission. During the rainy  
304 season, there is usually water logging in the ground, creating mosquito breeding grounds. This  
305 analysis observed that prevalence nearly doubles for every mm increase in the annual rainfall. The  
306 areas observed to have a higher prevalence of malaria in Kenya are known to have prolonged rainy  
307 seasons[6].

308 The analysis also identified that malaria transmission is higher in rural areas compared to urban  
309 areas. This finding is consistent with other studies in the same region. Urban areas may have better  
310 housing and improved health services that are easier to access. These factors contribute to the lower  
311 risk of malaria. Conversely, rural areas are primarily associated with favourable conditions for  
312 malaria, including stagnant water, poor housing, inaccessible health services and agricultural  
313 activities [13].

314 The finding of increasing malaria prevalence with higher bed net coverage can be explained through  
315 reverse causality, which is often observed due to the higher distribution of bed nets in areas heavily  
316 affected by malaria.

317 Malaria hotspot areas identified in the analysis include the entire lake and coastal regions classified  
318 as malaria endemic [6]. This finding is in keeping with other previous analyses done for past time  
319 points [18]. The climatic condition in these areas is known to support malaria transmission. We do  
320 find additional hotspots, which highlights the strength in our analysis approach. Localized malaria  
321 hotspots are identified in the county of Turkana. Though this area is classified as a seasonal malaria  
322 transmission zone, a reactive case detection in the area conducted from 2018 to 2019 also detected  
323 high malaria transmission with a prevalence as high as 33.6% [15] . Another study in a refugee camp  
324 in the same region identified a malaria prevalence of 64.2% [15]. Evidence from a recent study  
325 examining the contribution of *P. falciparum* parasite importation to local malaria transmission in  
326 Central Turkana confirms that malaria in the area is rather endemic, with intense local transmission  
327 as opposed to the importation of malaria [16]. Due to its malaria risk classification status, Turkana is  
328 often left out of malaria control activities. This is an important finding where an area's transmission  
329 is misclassified. The recent WHO malaria surveillance guide calls for countries to view malaria  
330 transmission as a continuum in space and stratify the malaria burden for better targeting and  
331 improved efficiency of malaria interventions[8]. As malaria transmission declines, it becomes  
332 increasingly focal and prone to outbreaks. Understanding and predicting patterns of transmission  
333 risk becomes an essential component of an effective elimination campaign, allowing limited  
334 resources for control and elimination to be targeted cost-effectively. In this study, we also find  
335 additional hotspots in the counties of Meru and Kajiado, areas with humid and arid weather  
336 conditions, respectively. There is a need for more local surveillance in the area. These areas are also  
337 characterized by low implementation of malaria control measures.

338

339 There are several strengths and limitations of the data used in the analysis. To the best of our  
340 knowledge, this is the latest nationally representative data on malaria prevalence. With this, the  
341 results of this study are generalizable to the entire population of Kenya. Use of the geostatistical  
342 Model as opposed to the traditional non-spatial Model, is a key strength. It allows us to borrow

343 information from the sampled cluster to infer for the unsampled ones and at the same time, account  
344 for predictors that influence malaria transmission. The major limitation in the analysis is the lack of  
345 adequate environmental covariates to improve spatial predictions.

## 346 **Conclusion**

347 This analysis has shown that rainfall, urbanization, temperature, and bed net coverage are important  
348 factors that affect malaria transmission. The high-resolution malaria prevalence maps produced as  
349 part of the analysis are important in identifying hotspots which is an essential element in planning,  
350 implementation, resource mobilization, monitoring, and evaluation of malaria interventions in the  
351 country. We have also identified malaria hotspots in areas not traditionally classified as endemic,  
352 highlighting the need to rethink the classification of malaria transmission epidemiology in Kenya.

## 353 **Acknowledgements**

354 Many thanks to the DHS programme team for allowing the authors to use the dataset and to the  
355 participants that provided the data.

## 356 **References**

- 357 1. Murray CJL, Rosenfeld LC, Lim SS, Andrews KG, Foreman KJ, Haring D, et al. Global malaria  
358 mortality between 1980 and 2010: a systematic analysis. *Lancet*. 2012 Feb 4;379(9814):413–  
359 31.
- 360 2. World Health Organization. World malaria report 2021 [Internet]. 2021 [cited 2022 Jun 12].  
361 Available from: [https://www.who.int/teams/global-malaria-programme/reports/world-](https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021)  
362 [malaria-report-2021](https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021)
- 363 3. Millar SB, Cox-Singh J. Human infections with *Plasmodium knowlesi*—zoonotic malaria. *Clin*  
364 *Microbiol Infect*. 2015 Jul 1;21(7):640–8.
- 365 4. White NJ. *Plasmodium knowlesi*: The Fifth Human Malaria Parasite. *Clin Infect Dis*. 2008 Jan  
366 15;46(2):172–3.
- 367 5. Lalloo DG, Hill DR. Preventing malaria in travellers. *BMJ*. 2008 Jun;336(7657):1362–6.



- 368 6. Division of National Malaria Programme (DNMP) [Kenya], ICF. Kenya Malaria Indicator Survey  
369 2020. 2021;158.
- 370 7. Mogeni P, Omedo I, Nyundo C, Kamau A, Noor A, Bejon P, et al. Effect of transmission  
371 intensity on hotspots and micro-epidemiology of malaria in sub-Saharan Africa. *BMC Med.*  
372 2017 Jun;15(1).
- 373 8. World Health Organization. Malaria surveillance, monitoring & evaluation: A reference  
374 manual. Geneva; 2018.
- 375 9. Ingham K. Kenya | History, Map, Flag, Climate, Capital, & Facts | Britannica [Internet].  
376 Britannica. 2019 [cited 2022 Jun 18]. Available from:  
377 <https://www.britannica.com/place/Kenya>
- 378 10. KNBS. Population and housing census of Kenya, 2009 - Kenya Data Portal. 2010.
- 379 11. R Development Core Team. R: A Language and Environment for Statistical Computing.  
380 Vienna, Austria; 2020.
- 381 12. Zhao X, Chen F, Feng Z, Li X, Zhou XH. Characterizing the effect of temperature fluctuation on  
382 the incidence of malaria: An epidemiological study in south-west China using the varying  
383 coefficient distributed lag non-linear model. *Malar J.* 2014 May 27;13(1):1–10.
- 384 13. Larson PS, Eisenberg JNS, Berrocal VJ, Mathanga DP, Wilson ML. An urban-to-rural continuum  
385 of malaria risk: new analytic approaches characterize patterns in Malawi. *Malar J.* 2021  
386 Dec;20(1):1–14.
- 387 14. Gopal S, Ma Y, Xin C, Pitts J, Were L. Characterizing the Spatial Determinants and Prevention  
388 of Malaria in Kenya. *Int J Environ Res Public Heal* 2019, Vol 16, Page 5078. 2019  
389 Dec;16(24):5078.
- 390 15. Meredith HR, Wesolowski A, Menya D, Esimit D, Lokoel G, Kipkoech J, et al. Epidemiology of  
391 Plasmodium falciparum Infections in a Semi-Arid Rural African Setting: Evidence from  
392 Reactive Case Detection in Northwestern Kenya. *Am J Trop Med Hyg.* 2021 Oct;105(4):1076–  
393 84.
- 394 16. Markwalter CF, Menya D, Wesolowski A, Esimit D, Lokoel G, Kipkoech J, et al. Plasmodium  
395 falciparum importation does not sustain malaria transmission in a semi-arid region of Kenya.  
396 *PLOS Glob Public Heal.* 2022;2(8):e0000807.

397

398

399

400

401

402

403



Figure 1

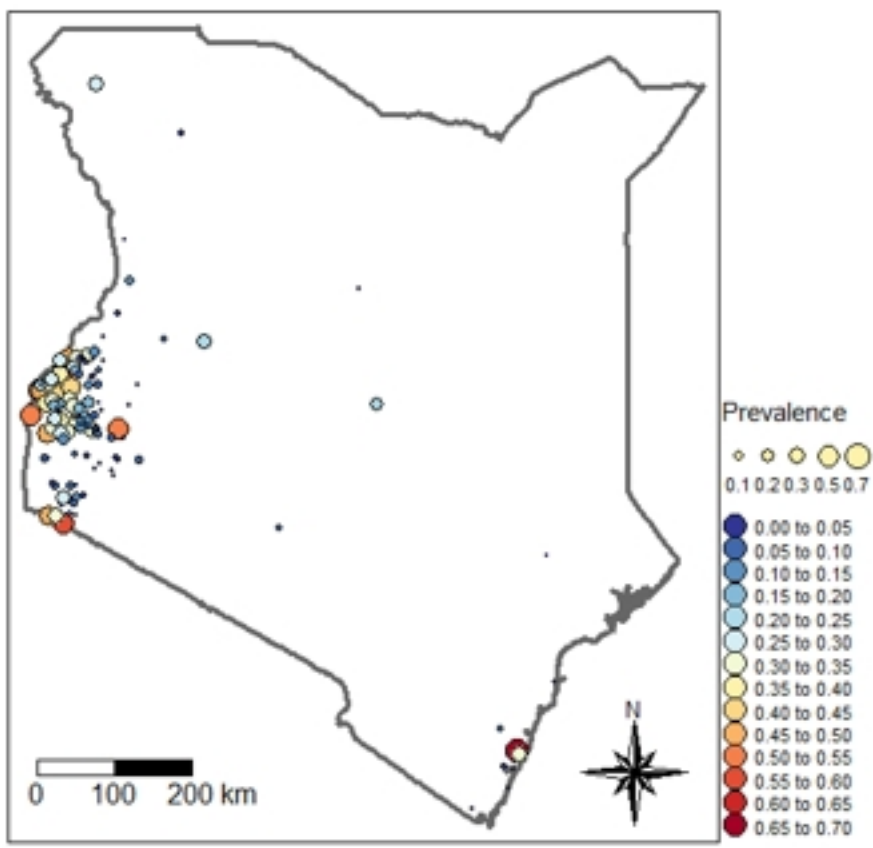
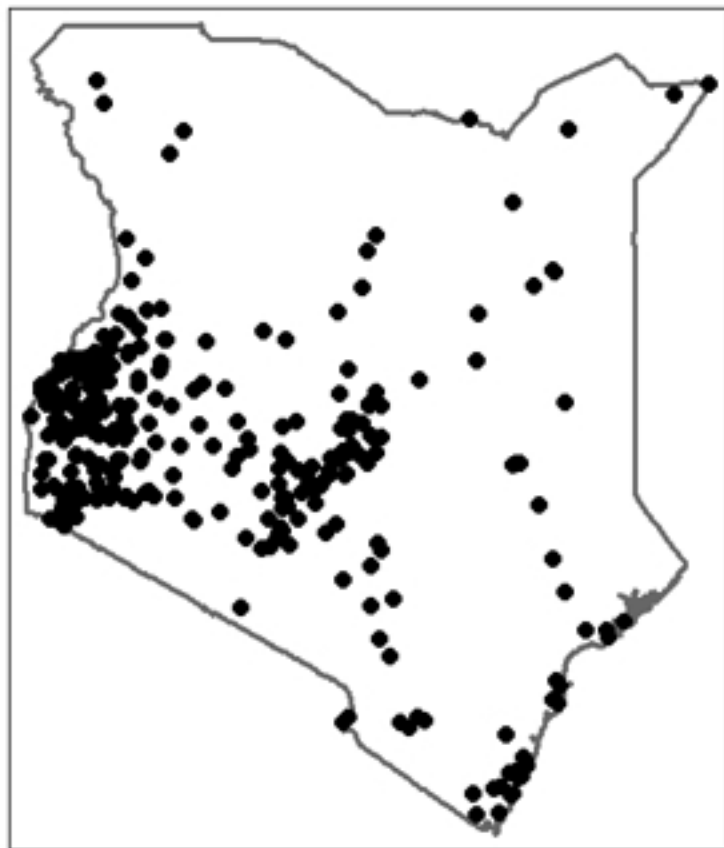
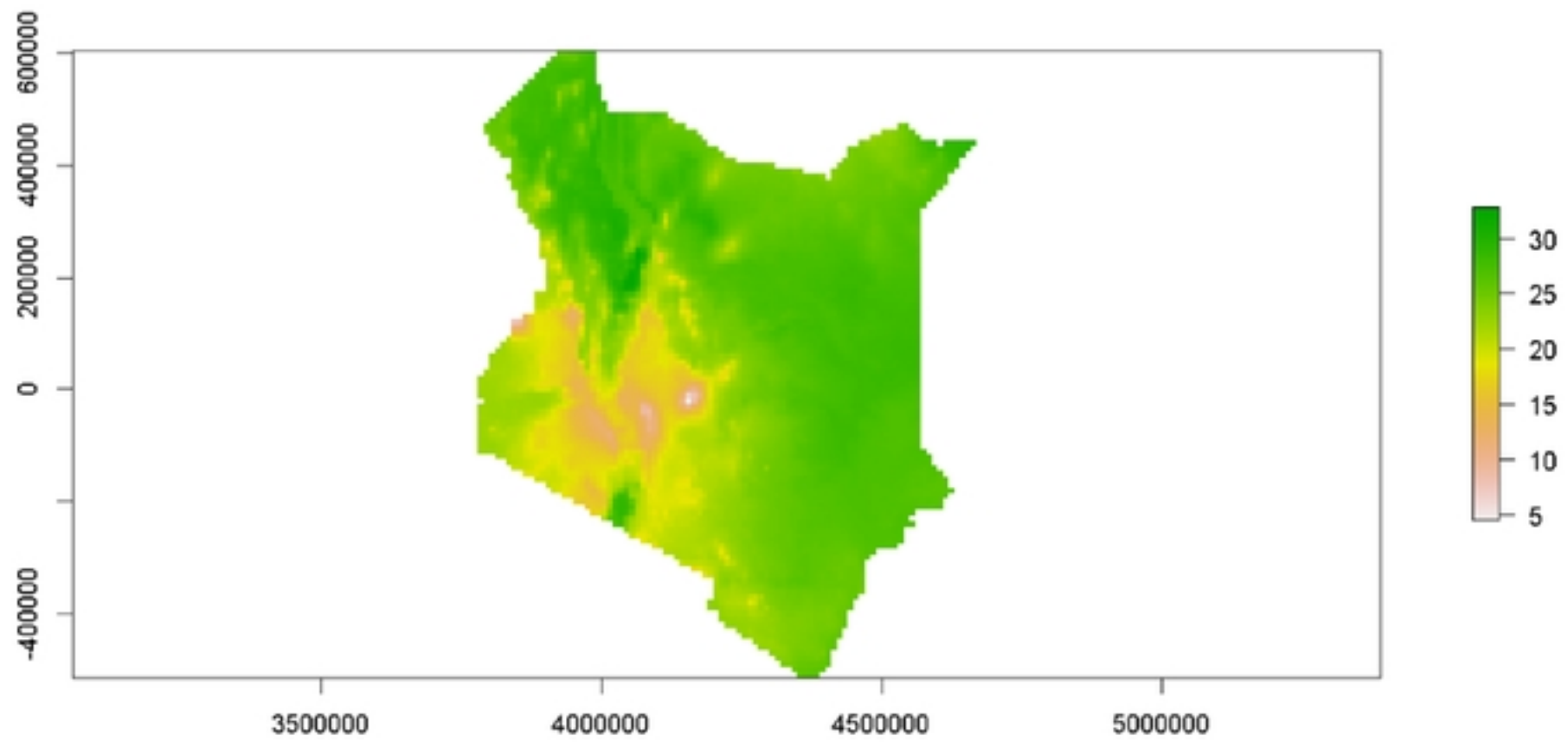
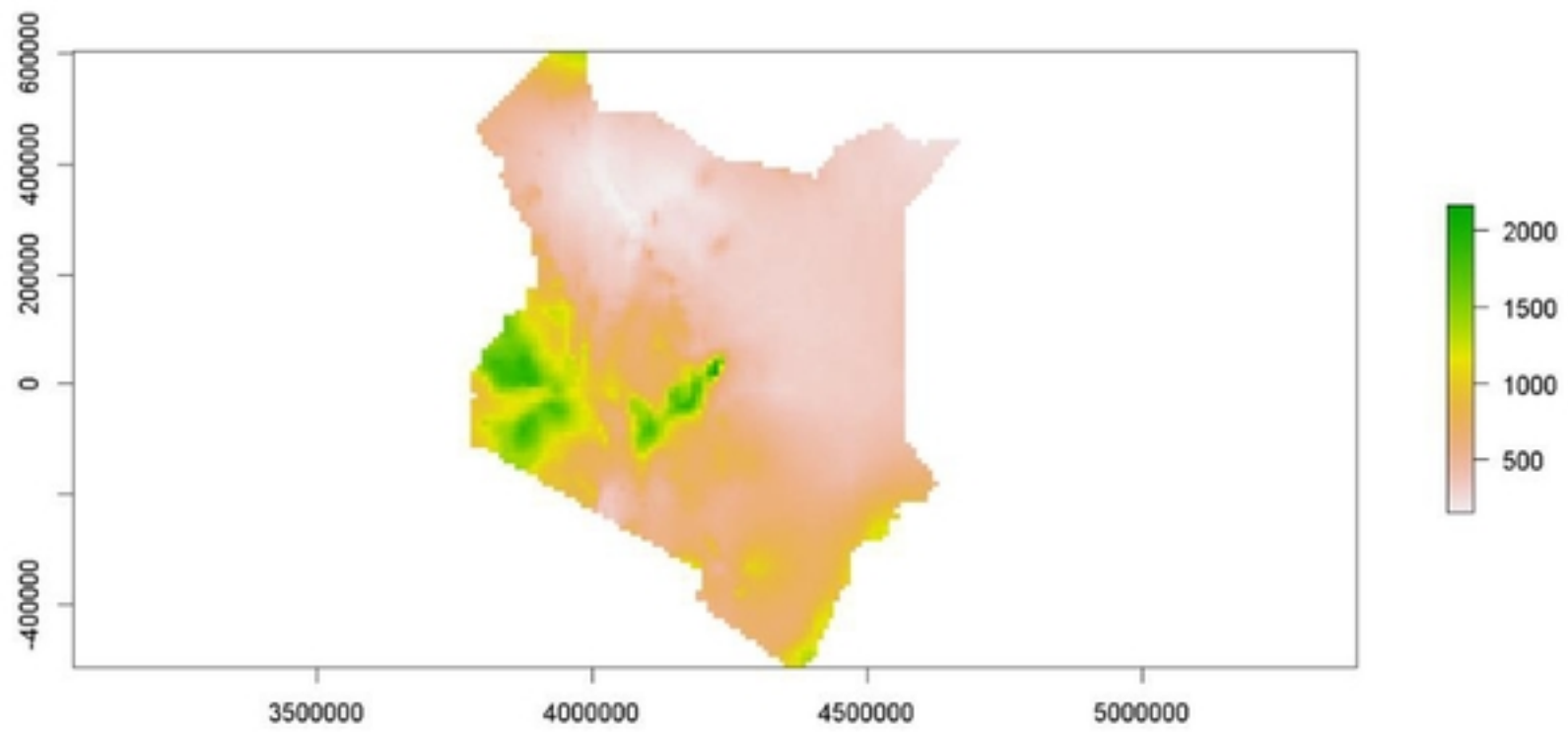


Figure 2





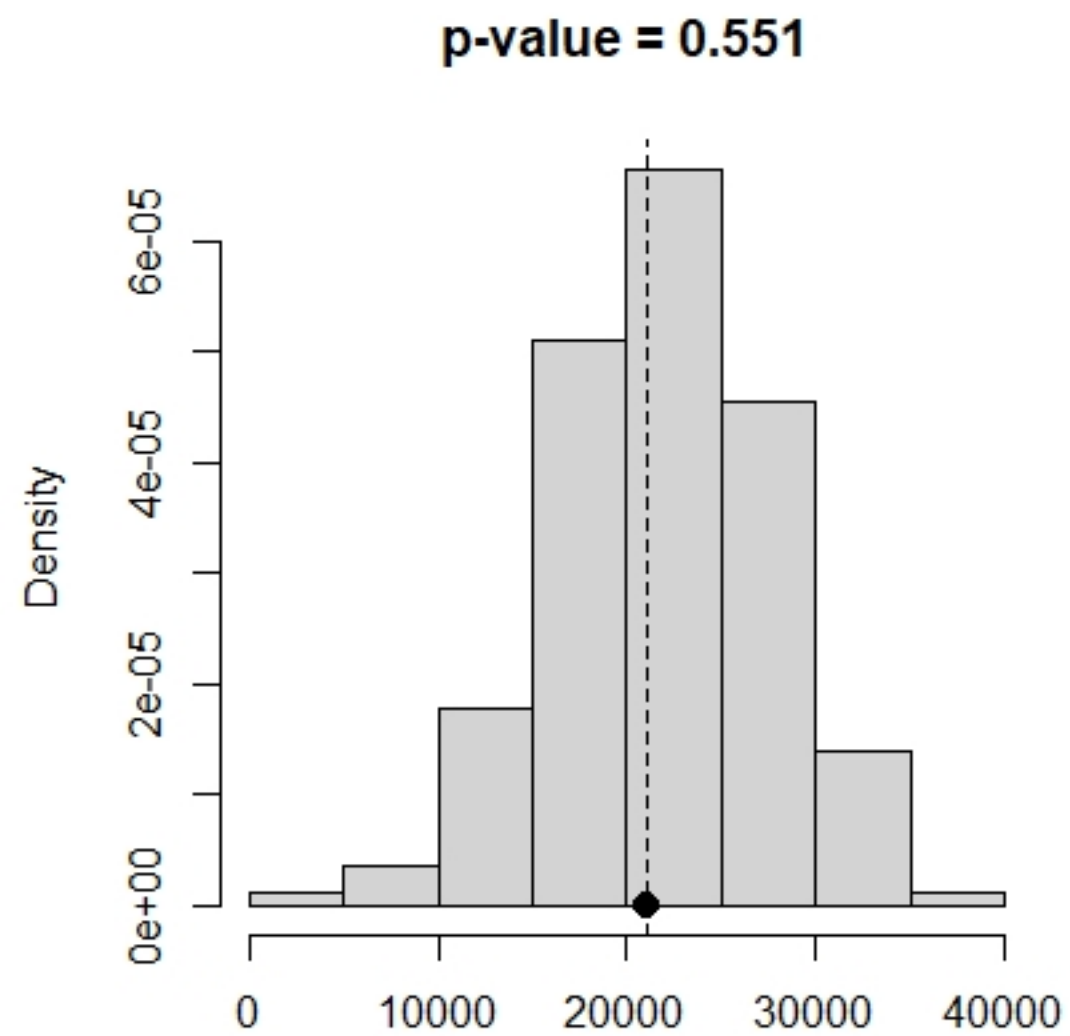
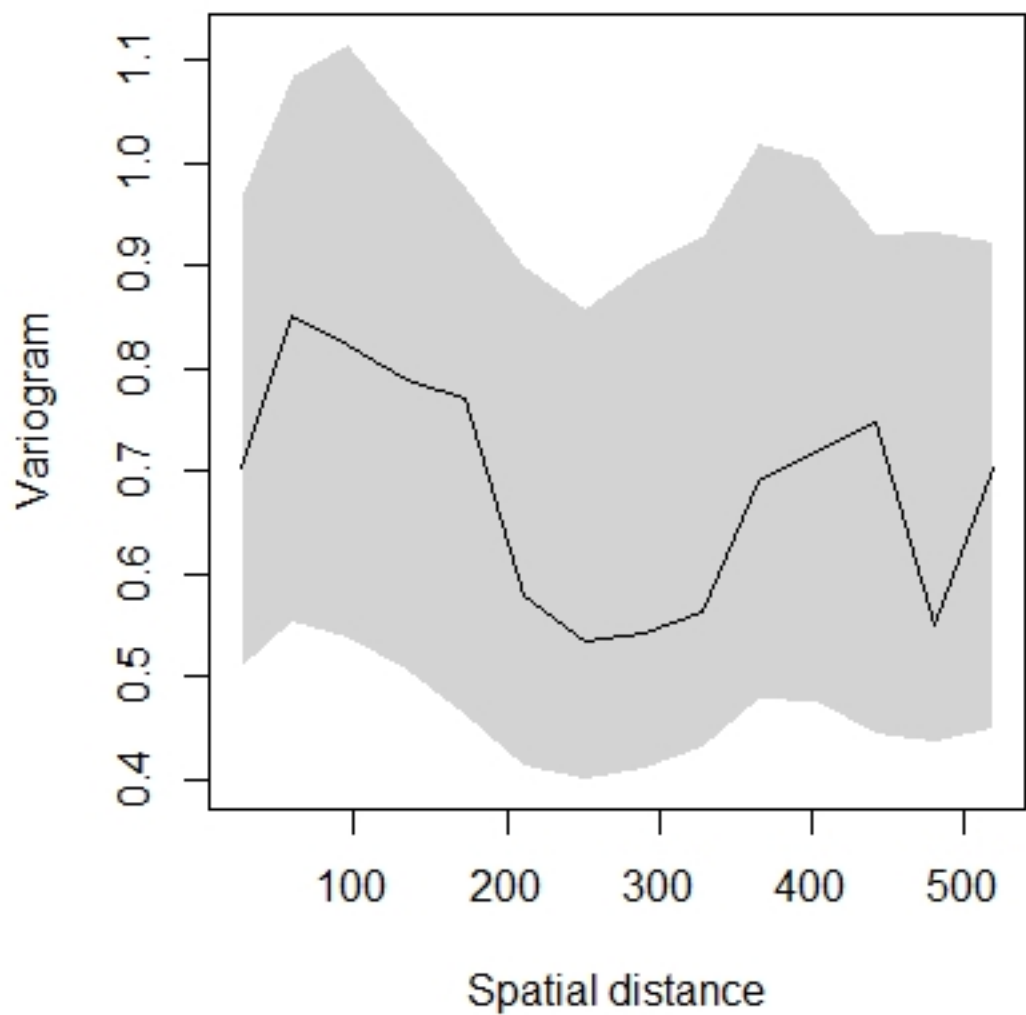


Figure 4

Malaria Prevalence in Kenya

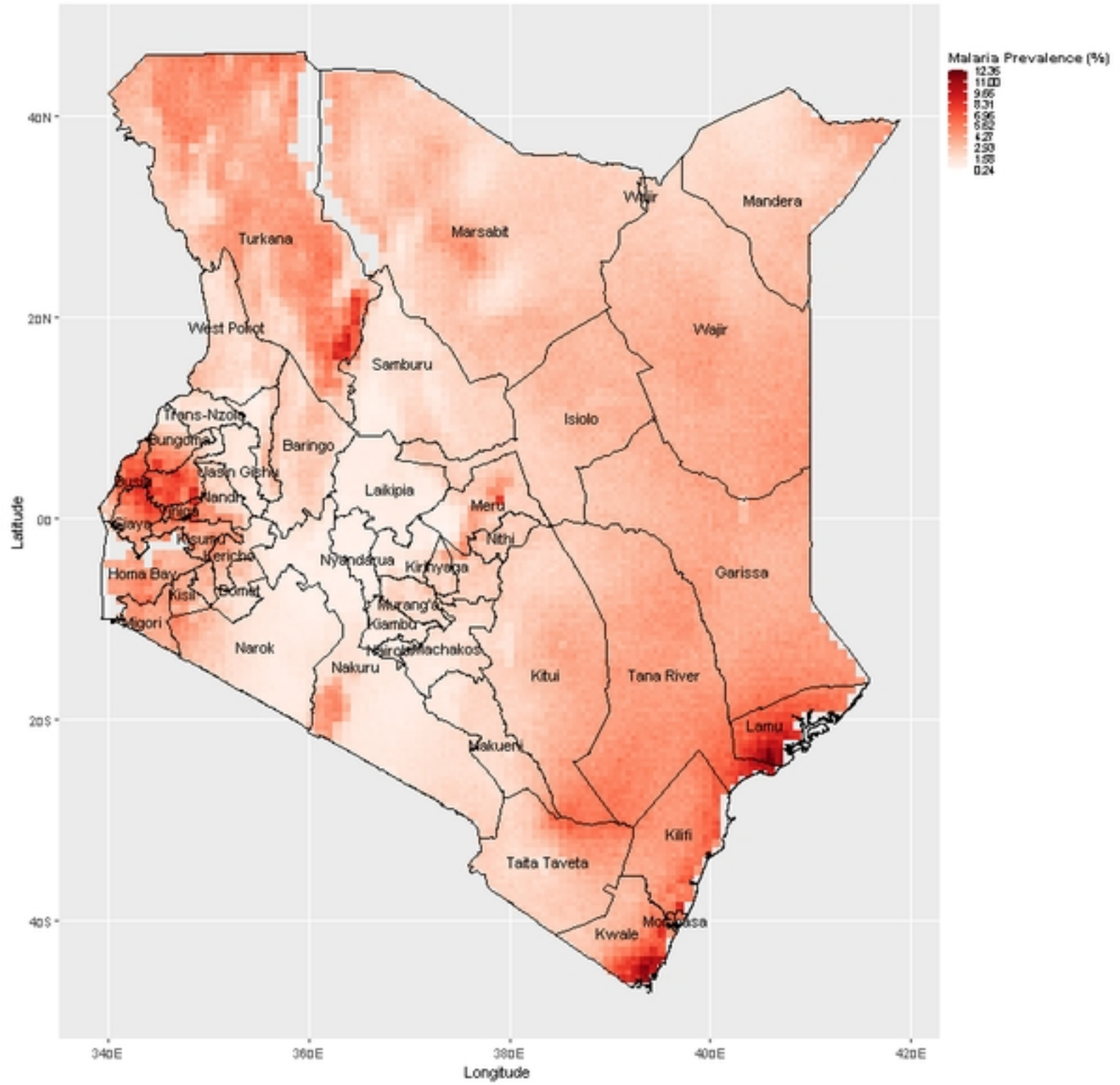


Figure 5



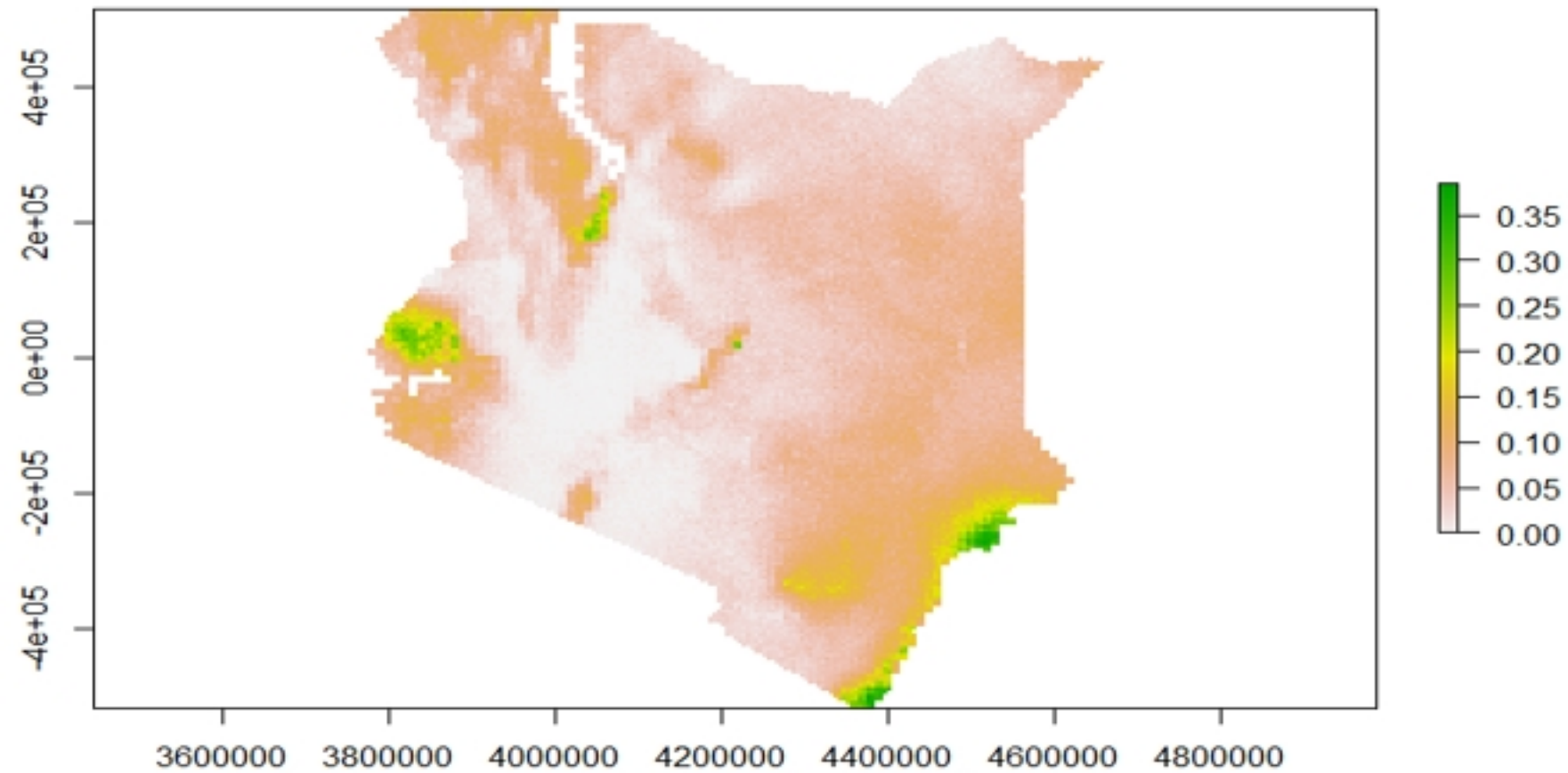


Figure 6