

# Geographically weighted regression modelling of the spatial association between malaria cases and environmental factors in Cameroon.

**Marlvin Anemey Tewara**

Shandong University Cheeloo College of Medicine <https://orcid.org/0000-0003-3720-0084>

**Liu Yunxia**

Shandong University Cheeloo College of Medicine

**Weiqiang Lin**

Shandong University Cheeloo College of Medicine

**Binang Helen Barong**

Shandong University Cheeloo College of Medicine

**Prisca Ngetemalah Mbah-Fongkimeh**

The University of New Castle, School of Medicine and Public health

**Zheng Zhaolei**

Shandong University Cheeloo College of Medicine

**Liu Xinhui**

Shandong University Cheeloo College of Medicine

**Zhou Miao**

Shandong University Cheeloo College of Medicine

**Xiaojuan Liu**

Shandong University Cheeloo College of Medicine

**Fuzhong Xue** (✉ [xuefzh@sdu.edu.cn](mailto:xuefzh@sdu.edu.cn))

---

## Research article

**Keywords:** Geographically weighted regression, Ordinary least squares, malaria, spatial statistics, mapping, Geographical information systems.

**Posted Date:** August 18th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.13021/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background:** Studies have illustrated the association of malaria cases with environmental factors in Cameroon but limited in addressing how these factors vary in space for timely public health interventions. Thus, we want to find the spatial variability between malaria hotspot cases and environmental predictors using Geographically weighted regression (GWR) spatial modelling technique. **Methods:** The global Ordinary least squares (OLS) in the modelling spatial relationships tool in ArcGIS 10.3. was used to select candidate explanatory environmental variables for a properly specified GWR model. The local GWR model used the global OLS candidate variables to examine, predict and explore the spatial variability between environmental factors and malaria hotspot cases generated from Getis-Ord  $G_i^*$  statistical analysis. **Results:** The OLS candidate environmental variable coefficients were statistically significant (adjusted  $R^2 = 22.3\%$  and  $p < 0.01$ ) for a properly specified GWR model. The GWR model identified a strong spatial association between malaria cases and rainfall, vegetation index, population density, and drought episodes in most hotspot areas and a weak correlation with aridity and proximity to water with an overall model performance of 0.243 (adjusted  $R^2 = 24.3\%$ ). **Conclusion:** The generated GWR maps suggest that for policymakers to eliminate malaria in Cameroon, there should be the creation of malaria outreach programs and further investigations in areas where the environmental variables showed strong spatial associations with malaria hotspot cases. **Keywords:** Geographically weighted regression, Ordinary least squares, malaria, spatial statistics, mapping, Geographical information systems.

## Background

About 70% of the world's malaria burden is focused in just 11 countries; 10 in Sub-Saharan Africa (SSA) (Burkina Faso, Cameroon, the Democratic Republic of Congo, Ghana, Mali, Mozambique, Niger, Nigeria, Uganda and the United Republic of Tanzania) and India. These high-burden nations housed an estimated 151 million cases of malaria and 275 000 deaths. In 2017, all the high burden endemic countries in Africa saw an increase in malaria cases with Cameroon witnessing an increase of 131,000 malaria cases(1), from a total of 1,618,738 confirmed cases reported in health facilities and communities in 2016(2). The epidemiological transmission of malaria in Cameroon is high ( $> 1$  case per 1000 population) in about 71% (16.6 million people) and low (0–1 cases per 1000 population) in about 29% (6.8 million) in people of all sexes and age groups with children less than five at greater risk of the disease(2). In 2014, the morbidity of malaria in Cameroon was 30% in children and 18% in Adults(3). The government of Cameroon and partners have been combating malaria through the creation of national intervention programs including: the distribution of free insecticide-treated nets (ITN) that was established in 2011 to populations at high risk, provision of sulfadoxine-pyrimethamine drugs to pregnant woman, parasitological screening of suspected malaria cases, and the application of other WHO standard treatments(3,4). Because of the many socio-economic and environmental challenges posed by the malaria disease to Africa countries, investment programs and research towards the control and elimination of the disease is a global concern. The global technical strategy (GTS) report on malaria highlighted that, funding in the fight for malaria has remained relatively stable since 2010 and that to reach the GTS 2030 goals of: reduce malaria mortality rates globally, reduce malaria case incidence globally, eliminate malaria from countries in which malaria was transmitted in 2015 and prevent re-establishment of malaria in all countries that are malaria-free, an annual malaria funding will need to increase to at least US\$ 6.6 billion per year by 2020(5). Intensifying investments in malaria research and development by endemic SSA is a key to attaining the GTS targets and eradicating the disease from the SSA geolocations.

The application of spatial statistical methods to geolocational health data research has enabled complex scenarios of the malaria disease to be visualized through the creation of spatial maps within the Geographical information systems (GIS) technology(6–11). The study of the spatial variation between disease outcomes and associative socioeconomic or environmental factors with the GIS system has greatly improved our understanding of these factors with the health outcome in question. Malaria has been reported to be associated with environmental and climatic factors such as rainfall, humidity, temperature (12,13) and understanding the behavior of these factors in space with the application of spatial regression statistics(14) will further improve on timely control measures and resource allocations.

Regression analyses are statistical techniques that allow for the modelling, examining, and exploring of spatial relationships, to better understand the factors behind observed spatial patterns and hotspots, and to predict outcomes based on that understanding(14). Ordinary Least Squares regression (OLS) is a global regression method; It provides a global model of the variable or process to be predicted or studied. It creates a single regression equation to represent that process. Geographically Weighted Regression (GWR) is a local spatial regression method that allows the relationships to be modelled to vary across the study area by

fitting a regression equation to every feature in the dataset using candidate explanatory variables from the OLS. It is a local form of linear regression used to model spatially varying relationships. GWR statistical modelling technique has been applied to a range of malaria studies: Hasyim (15), used the GWR to find the spatial association between malaria cases and environmental factors in South Sumatra, Indonesia where altitude, distance from forest and rainfall were associated with malaria, Moise(16), in the seasonal and geographic variation of pediatric malaria in Burundi, identified the spatial variation between monthly rainfall and malaria prevalence. While the GWR spatial modelling technique has been a powerful tool in the understanding of spatial variability of malaria cases and environmental factors (rainfall, distance from forest, altitude, vegetation index, temperature, and proximity to water body), and other malaria prevention studies(15–18), its application has been valuable in the understanding of other health outcomes and social science studies including cancer events(19), dengue fever(20), mental depression(21), fire events(22), hospital accessibility study (23), alcohol and violence(24) and real estate housing crisis(25).

Massoda (26), compared malaria survey programs in different ecological zones in Cameroon and recommended on the needs of intervention programs during high transmission rainy seasons. Furthermore, Tewara(27), in a recent study on small area spatial statistical analysis of malaria clusters and hotspots in Cameroon, illustrated the linear association between malaria cases and environmental factors using the Pearson correlation statistics(28) but didn't demonstrate any spatial variability that would become the main aim of this study. Knowing the spatial variation between malaria cases and environmental factors will aid in targeted funding, resource allocations, and prevention programs. The specific objective of this study is to find the spatial variability between malaria hotspot cases and environmental predictors using the GWR spatial modelling technique.

## Methods

### Data source

Data for this study was obtained from the demographic and health survey (DHS) program website (<https://dhsprogram.com>) funded by the United States Agency for International Developments (USAIDS) following a written administrative clearance(29). The DHS program has been described elsewhere(27). Households are grouped as either an urban (city block or apartment building) or rural (village or group of villages) cluster points (figure 1) and displaced a distance up to 2 km for urban-city clusters and 5 km for rural clusters due to confidentiality. For this study, points and lattice data for the Cameroon DHS VI malaria survey was obtained from the DHS spatial data repository site(30). The malaria survey data for 2015 was linked with environmental covariates data; enhanced vegetation index (EVI), rainfall, drought episodes, population density, aridity, proximity to water, and analyzed using the ArcGIS 10.3 (ESRI, Redlands, California, USA) software.

### Malaria and environmental data description

The WHO recommends that all cases of suspected malaria should be confirmed using parasite-based diagnostic testing (either microscopy or rapid diagnostic test) before administering treatment. Thus, malaria cases were confirmed based on both rapid diagnostic tests and on laboratory analysis. A clinical case was defined as a malaria-attributable febrile episode (body temperature in excess of 37.5 °C), accompanied by headaches, nausea, excess sweating and/or fatigue censored by a 30-day window (6,27). Since the households are the variables to be analyzed, a malaria year as described in this study, is the average number of people per year who show clinical symptoms of *Plasmodium falciparum* malaria within the cells whose centroid falls within a radius of 10 km (for rural points) or 2 km (for urban points)(27). The environmental covariates data set used for this study are described in table 1.

Permission to use the data was obtained through a written request and subsequent approval from the DHS division of the USAID. During the DHS project, interviews and blood test analyses are conducted only if the respondent provides voluntary informed consent. Written informed consent was obtained from all participants.

### Statistical analysis

#### Getis-Ord $G_i^*$ statistics

The Getis-Ord  $G_i^*$  statistics is a local statistic that allows us to discover new locations with significant clusters of hot and coldspots. It assesses each malaria household cluster (or feature) within the context of neighboring malaria households and compares the local situation to the global situation. The 2015 Malaria indicator survey data was analyzed for malaria hotspots.

The Getis-Ord  $G_i^*$  local statistics is given as:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j - \bar{X} \sum_{j=1}^n w_{ij}}{s \sqrt{\left[ \frac{n \sum_{j=1}^n w_{ij}^2 - \left( \sum_{j=1}^n w_{ij} \right)^2}{n-1} \right]}} \quad (1)$$

Where  $x_j$  is the attribute value for feature  $j$ ;  $w_{ij}$  is the spatial weight between  $i$  and  $j$ ;  $n$  is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^n w_j}{n} \quad (2)$$

$$s = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (3)$$

### Spatial regression analysis

To investigate the spatial relationship between the distribution of malaria hotspots and environmental covariates, we used the regression analysis technique. The mathematical computation applied to the global OLS is given as;

$$y = \beta_0 + \sum_k \beta_k X_{ik} + \varepsilon_i \quad (a)$$

$y$  is the dependent or response variable for the observation  $i$  (the process to be predicted or understood), for example, the malaria hotspot cases,  $X_{ik}$  is the value of the independent or explanatory variable  $k$ th for  $i$  used to model or predict the dependent variables. The explanatory variables include: population density ( ), enhanced vegetation index ( ), rainfall ( ), aridity ( ) drought episodes ( ), and proximity to water ( ). The spatial model to be built will explain if the distribution or occurrence of malaria hotspots are due to the combination of these explanatory variables. This will help us create a prediction map that can be used for public health resource allocations due to the spatial relationship between the dependent (malaria hotspots) and explanatory variables.  $\beta_k$  is the regression coefficients for the variable  $k$ . They are values, one for each explanatory variable, that represent the strength and type of relationship the explanatory variable has to the dependent variable.  $\beta_0$  is the regression intercept. It represents the expected value for the dependent variable if all the independent variables are zero. The residuals  $\varepsilon_i$  represent the portion of the dependent variable that isn't explained by the model (14).

The global OLS method is based on the use of only one equation (a) to explore the relationship between variables. To account for the spatial component of the OLS variables for a specified model, the GWR local regression model creates an equation (b) for each element of the dependent variable data set, in order to capture geographic variations (23). The GWR model is computed as;

$$y = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \quad (b),$$

where  $u_i$  and  $v_i$  represent the point coordinates of  $i$ th in space. Thus, the GWR equation ( $\beta$ ) distinguishes that spatial variations in relationships might exist and provides a way in which they can be measured and this is based mainly in its spatial ability in a GIS environment with emphasis on the differences regarding space and local disaggregation of the local statistics as thoroughly illustrated by Fotheringham(39). Although the GWR model can be advantageous in order to distinguish heterogeneity of the space itself, it also makes it easier to go from a global OLS perspective to a local analysis, thus obtaining a better grade of details and precision using the GIS environment(39).

For this study, our regression models (OLS and GWR) was built using the modelling spatial relationship in ArcGIS. Multiple models were made using a set of candidate variables until selecting a model with high explanatory power, and that contained the six explanatory variables (to relevant for a properly specified model with respect to malaria transmission(23)). A positive coefficient of an OLS candidate variable means the explanatory  $X$  variables and the response or dependent  $Y$  variables changed in the same direction and if the environmental risk factor increases, then the number of confirmed malaria cases will increase. Similarly, a negative coefficient means  $X$  and  $Y$  changed in reverse directions (15). To have a properly specified GWR model, the generated OLS summary result of the model variables should pass a statistically significant six-test check (data pre-processing background analysis).

The GWR model was constructed using a default fixed distance as the kernel type and AICc (Akaike's Information Criterion) as the bandwidth method to find the optimal distance for better model performance. The GWR generates a map that was represented as raster surfaces for the model predictions (combined strength of the relationship amongst the variables used), residuals, local  $R$  squared (model significance), condition number (difficulties identifying spatial relationship) and coefficients (explaining the strength of the relationship between the dependent and explanatory variables) and an output table demonstrating the strength of adjusted  $R^2$  significance (from 0- 1).

## Results

### Hotspot analysis

The analysis depicted that, there exist high malaria hotspots distribution in most areas (rural, urban, and urban-city centers) of the Northwest, Southwest, Littoral, Yaoundé, and Center regions and low hotspots location elsewhere, as shown in figure 2.

### Ordinary Least Square (OLS)

The OLS global regression illustrates the statistical significance of the model variables and feasibility to be used in specifying the GWR model. The result demonstrated that two of the explanatory variable coefficients (aridity and proximity to water) had a negative relationship (negative sign) with the dependent variable, while all the other explanatory variables had a positive relationship with the dependent variable. The adjusted R-squared for the OLS had 22.3 % evaluation of the model performance (table 3).

### Geographically Weighted Regression (GWR)

In the current study, the coefficients ( $\beta$ ) of the population density( $X_1$ ), enhanced vegetation index ( $X_2$ ), rainfall ( $X_3$ ) and drought episodes ( $X_5$ ) exhibit high (strong) correlation with malaria hotspots in most areas in the western portion of the country and few elsewhere, while aridity ( $X_4$ ) and proximity to water ( $X_6$ ) showed a weak association as exemplified in figure 3

The GWR output also produced a predicted malaria map, local  $R$ -squared ( $R^2$ ) and residuals. The local  $R^2$  in figure 4A shows that the model performance was high in regions of the western part of the country and some areas in the Central and Yaoundé DHS regions with a model performance of 0.243 (adjusted  $R^2 = 24.3\%$ , table 3). The predicted map (the combined strength of the relationship amongst the variables used) showed high spatial association in areas of the Southwest, Northwest and the North regions; figure 4B, while the model residuals (areas not explained by the model) were seen in most parts in the South region; figure 4C.

The differences in model specification between OSL and GWR or how well the model has improved from the global OSL to the local GWR is summarized in table 3.

## Discussion

Studies on the association of environmental factors with malaria cases in Cameroon have been illustrated(27,40). In this study, we demonstrated the ability of the GWR technique as a facilitatory tool to understand the spatial interaction of environmental factors with malaria hotspots. Though malaria clusters and hotspots have been reported in Cameroon(27), our analysis confirmed rural and urban-city centers of the Western, Littoral, and Central part of the country as targeted areas for further research and intervention programs.

For the model specification, the OLS model (table 2) illustrated that aridity and proximity to water had negative associations. For example, the negative coefficient for aridity in our model means the malaria cases decrease with lack of water since aridity is a deficiency of moisture probably due to the lack of rainfall while an increase in rainfall in a given rainy season in Cameroon will promote malaria cases and these periods can be targeted for malaria prevention programs since rainfall creates breeding sites for female *Anopheline* mosquitoes(14). This understanding can promote malaria prevention campaign such as getting rid of stagnant waters around habitable household clusters or discarding water cans to prevent the growth of malaria-causing mosquitoes. Moreover, filling up of potholes during the dry season in high-risk areas will help diminish the mosquito breeding sites. The Koenker statistics (background check-test for non-stationarity) was statistically significant ( $P < 0.01$ ) and reflects that the relationships being modeled were consistent across the entire study area and thus nonstationary (except for drought episodes and EVI). Furthermore, the VIF values ( $< 7.5$ ; table2) indicates no redundancy among explanatory variables and hence no multicollinearity. The OLS model residuals were tested for clustering using the spatial autocorrelation (pre-analysis check) tool in ArcGIS as described elsewhere (41) and it indicated that the variables used were randomly distributed (no clustering). A statistically significant spatial autocorrelation in the model residuals would indicate that we neglected one or more key explanatory variables. This is a positive indicator of choosing a good model(42). The Jarque-Bera Statistics was statistically significant ( $p < 0.01$ ) indicating that our model predictions were biased (the residuals were not normally distributed); this may be due to the changing signs in some of the coefficients in the explanatory variable and thus causing variability. Though the test was biased, we proceeded to the GWR model because recent studies(15) have reported on spatial variations similar to our specified model variables and our main goal was to understand the behavior of these environmental variables with malaria cases for future research and intervention projects. The 22.3 % evaluation of the OLS model performance, indicates the explanatory variable coefficients were telling 22 % of the spatial relationship story between the malaria cases and the environmental factors we were trying to model. This may seem low per the  $R^2$  range (0-100 %) where higher  $R^2$  values depict good model performance. Hasyim (15), in the spatial modelling of malaria cases associated with environmental factors in South Sumatra, Indonesia, had low  $R^2$  (6.2%) variation of malaria incidences by environmental factors for the OLS model. Moise (16) also had low  $R^2$  ( $< 5\%$ ) in their OLS model variables in the seasonal and geographic variation of pediatric malaria in Burundi. This is to say that, the OLS model  $R^2$  varies with the explanatory variables under investigation and lower  $R^2$  does not always signify poor model performance(14).

The local GWR model was built based on the validated variables from the global OLS model. A validated OLS can promote a global policy for malaria control programs where similar statistically significant environmental or climatic predictors could be targeted across many malaria-endemic countries while a validated spatial relationship with GWR is an appropriate method to initiate prevention programs in local systems or village level within endemic countries(15). The GWR output coefficients maps (figure 3) indicated that population density, EVI, rainfall, and drought episodes had a strong spatial correlation or positive influence on malaria cases in our study locations. The strong correlations were seen in areas of the western part of Cameroon and few in the north. The population of these localities should be considered for malaria control programs in high transmission seasons in Cameroon. Whereas, aridity and proximity to water had a weak association in the above-cited locations meaning there was little or no spatial interdependence between these factors and malaria cases. The spatial variability ability of the GWR model is observed in figure 4A where the generated local  $R^2$  illustrated areas with 24.3% of the model performance between the environmental factors and malaria cases. The high spatial interdependence observed in the Northwest, Southwest, Littoral, Douala, Central and Yaoundé DHS regions calls for effective malaria surveillance. Likewise, active control measures in areas (East, Adamawa and North regions) with low spatial variability. The necessity for the creation of efficient malaria surveillance systems can further be demonstrated from our predicted map (figure 4B) highlighting strong spatial variability in some household clusters of the Northwest, Southwest, Adamawa and North regions for which public health interventions should be prominent.

The GWR residuals (figure 4C) indicated that areas, where the model did not work, were common in the southern part of the country. This implies that our model was unable to explain the spatial variability story between malaria cases and environmental factors in the locations depicted by the GWR residuals. Though our model finds it difficult explaining the spatial interdependence in these areas, the overall condition numbers from the attribute table output (background analysis verification) of the GWR indicates the model did not have a hard time solving; since the condition numbers from the explanatory environmental variables were  $< 30$  ( $>30$  would mean the model had difficulties solving the spatial relationships(42). Furthermore, table 3 illustrated that the GWR improved our understanding of the spatial relationship between malaria cases and environmental factors from the OLS as the AICc and  $R^2$  were lower and higher respectively. This signifies that the GWR is a better indicator for explaining spatial variability at the local level(15,23).

We had the following limitations: Firstly, we used only the malaria hotspot locations to specify our GWR model. This was because running the model on entire malaria- cases- location for the whole country would have missed key explanatory variables. Moreover, this will help cut down on resources allocation by targeting vulnerable hotspot communities. Secondly, our OLS model was biased and failed to pass the six tests check at the level of the Jarque-Bera statistics that was significant ( $p < 0.01$ ), indicating our explanatory predictors were not normally distributed in some locations. Thirdly, our model did not include other key environmental variables such as temperature and humidity and socioeconomic predictors for which data were limited or missing. Further investigations for practically available data for these environmental and socioeconomic predictors are required to provide detailed spatial variability coverage for the malaria disease in Cameroon. Though our study had some remits, it has demonstrated a rigorous understanding of the spatial interdependence between malaria cases and environmental risk factors and has provided new insights into the malaria disease at the local level in Cameroon by applying the GWR spatial modelling technique that was limited in most studies. Moreover, the methods in this study can be used to study other health outcomes in Cameroon that have been applied in available literature in other countries(17,21,23).

## Conclusions

This study demonstrated that rainfall, EVI, drought episodes, and population density had a strong spatial association with malaria distribution and could be targeted as important risk predictors for control programs at the local level in Cameroon. Given the greater availability of spatial data(30) and desktop GIS packages and statistical techniques, the challenges faced in the malaria disease investigation will be improved in the future. The generated GWR maps suggest that for policymakers to achieve the GTS targets for malaria by 2030, there should be the creation of outreach programs that will target malaria hotspot locations and carry out further investigations in areas where the environmental variables showed strong spatial associations with malaria cases.

## Abbreviations

SSA: Sub-Saharan Africa; WHO: World health organization; GTS: Global Technical Strategy; GIS: Geographic Information Systems; OLS: Ordinary Least Squares; GWR: Geographically Weighted Regression; ITN: Insecticide Treated Net; DHS: Demographic and Health Survey; USAIDS: United States Agency for International Development; EVI: Enhanced Vegetation index; AICc: Akaike's Information Criterion; VIF: Variance Inflation Factor

## Declarations

### Acknowledgments

We thank Shandong University PR China for funding support and DHS for data provision.

### Funding

Support and funding for this study were provided by the National Natural Science Foundation of China; Shandong University. Grant number 81573259.

### Availability of data and materials

Datasets supporting the conclusions of this article are available from the DHS program website <https://dhsprogram.com>. Public access to the data is closed and permission was obtained.

### Authors' contributions

MAT: data acquisition, data processing, management, design, and analysis. FZX, LY: Supervision, conceptualization, and interpretation of results. MAT, BHB, PNMF: drafting, writing, revisions. WL, ZZ, LX, ZM, XL: assisted with data analysis, writing, and revisions. All the authors read and approved the final manuscript.

### Ethics approval and consent to participate

DHS studies in Cameroon are approved by the Cameroon government. Written informed consent was obtained from all participants.

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests.

## References

1. WHO; High burden to high impact: A targeted malaria response. Available at: <https://apps.who.int/iris/bitstream/handle/10665/275868/WHO-CDS-GMP-2018.25-eng.pdf?ua=1>: Accessed 21 Mar 2019
2. WHO | Cameroon [Internet]. WHO. [cited 2019 Mar 24]. Available from: <http://www.who.int/countries/cmr/en/>
3. Ministry of Public Health. Sectorial health strategic plan 2001–2015. Ministry of Public Health; 2001. <http://www.nationalplanningcycles.org/file-repository/CMR>. [Internet]. [cited 2019 Apr 30]. Available from: <http://www.nationalplanningcycles.org/file-repository/CMR>
4. Institut National de la Statistique - INS/Cameroun and ICF. International. 2012. Enquête Démographique et de Santé et à Indicateurs Multiples du Cameroun 2011. Calverton, Maryland, USA: INS/Cameroun and ICF International. [Internet]. [cited 2019 Apr 30]. Available from: <https://dhsprogram.com/pubs/pdf/FR260/FR260.pdf>
5. World Health Organization, World Health Organization, Global Malaria Programme. Global technical strategy for malaria, 2016–2030 [Internet]. 2015 [cited 2019 Mar 19]. Available from: [http://apps.who.int/iris/bitstream/10665/176712/1/9789241564991\\_eng.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/176712/1/9789241564991_eng.pdf?ua=1)
6. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*. 2015 Sep 16;526(7572):207–11.
7. Gemperli A, Sogoba N, Fondjo E, Mabaso M, Bagayoko M, Briet OJT, et al. Mapping malaria transmission in West and Central Africa. *Tropical Medicine and International Health*. 2006 Jul;11(7):1032–46.
8. Gething PW, Casey DC, Weiss DJ, Bisanzio D, Bhatt S, Cameron E, et al. Mapping *Plasmodium falciparum* Mortality in Africa between 1990 and 2015. *New England Journal of Medicine*. 2016 Dec 22;375(25):2435–45.
9. Omumbo JA, Noor AM, Fall IS, Snow RW. How Well Are Malaria Maps Used to Design and Finance Malaria Control in Africa? Keating JA, editor. *PLoS ONE*. 2013 Jan 11;8(1):e53198.
10. Booman M, Durrheim DN, Grange KL, Martin C, Mabuza AM, Zitha A, et al. Using a geographical information system to plan a malaria control programme in South Africa. *Bulletin of the World Health Organization*. 2000;7.
11. Dalrymple U, Mappin B, Gething PW. Malaria mapping: understanding the global endemicity of *falciparum* and *vivax* malaria. *BMC Medicine* [Internet]. 2015 Dec [cited 2019 Mar 21];13(1). Available from: <http://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-015-0372-x>
12. Moukam Kakmeni FM, Guimapi RYA, Ndjomatchoua FT, Pedro SA, Mutunga J, Tonnang HEZ. Spatial panorama of malaria prevalence in Africa under climate change and interventions scenarios. *International Journal of Health Geographics* [Internet].

- 2018 Dec [cited 2018 Mar 15];17(1). Available from: <https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-018-0122-3>
13. Efe SI, Ojoh CO. Climate Variation and Malaria Prevalence in Warri Metropolis. *Atmospheric and Climate Sciences*. 2013;03(01):132–40.
  14. Regression analysis basics [Internet]. [cited 2019 Apr 13]. Available from: [http://resources.esri.com/help/9.3/ArcGISDesktop/com/Gp\\_ToolRef/Spatial\\_Statistics\\_toolbox/regression\\_analysis\\_basics.htm](http://resources.esri.com/help/9.3/ArcGISDesktop/com/Gp_ToolRef/Spatial_Statistics_toolbox/regression_analysis_basics.htm)
  15. Hasyim H, Nursafingi A, Haque U, Montag D, Groneberg DA, Dhimal M, et al. Spatial modelling of malaria cases associated with environmental factors in South Sumatra, Indonesia. *Malaria Journal* [Internet]. 2018 Dec [cited 2019 Apr 26];17(1). Available from: <https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2230-8>
  16. Moise I, Roy S, Nkengurutse D, Ndikubagenzi J. Seasonal and Geographic Variation of Pediatric Malaria in Burundi: 2011 to 2012. *International Journal of Environmental Research and Public Health*. 2016 Apr 15;13(4):425.
  17. Rusk A, Highfield L, Wilkerson JM, Harrell M, Obala A, Amick B. Geographically-weighted regression of knowledge and behaviour determinants to anti-malarial recommending and dispensing practice among medicine retailers in western Kenya: capacitating targeted interventions. *Malaria Journal* [Internet]. 2016 Dec [cited 2019 Apr 26];15(1). Available from: <http://malariajournal.biomedcentral.com/articles/10.1186/s12936-016-1599-5>
  18. Ehlkes L, Krefis A, Kreuels B, Krumkamp R, Adjei O, Ayim-Akonor M, et al. Geographically weighted regression of land cover determinants of Plasmodium falciparum transmission in the Ashanti Region of Ghana. *International Journal of Health Geographics*. 2014;13(1):35.
  19. Goovaerts P, Xiao H, Adunlin G, Ali A, Tan F, Gwede CK, et al. Geographically-weighted regression analysis of percentage of late-stage prostate cancer diagnosis in Florida. *Applied Geography*. 2015 Aug;62:191–200.
  20. Ren H, Zheng L, Li Q, Yuan W, Lu L. Exploring Determinants of Spatial Variations in the Dengue Fever Epidemic Using Geographically Weighted Regression Model: A Case Study in the Joint Guangzhou-Foshan Area, China, 2014. *International Journal of Environmental Research and Public Health*. 2017 Dec 6;14(12):1518.
  21. Choi H, Kim H. Analysis of the relationship between community characteristics and depression using geographically weighted regression. *Epidemiology and Health*. 2017 Jun 21;39:e2017025.
  22. Song C, Kwan M-P, Zhu J. Modeling Fire Occurrence at the City Scale: A Comparison between Geographically Weighted Regression and Global Linear Regression. *International Journal of Environmental Research and Public Health*. 2017 Apr 8;14(4):396.
  23. Martínez Bascuñán M, Rojas Quezada C. Geographically weighted regression for modelling the accessibility to the public hospital network in Concepción Metropolitan Area, Chile. *Geospatial Health* [Internet]. 2016 Nov 22 [cited 2019 Apr 26];11(3). Available from: <http://geospatialhealth.net/index.php/gh/article/view/451>
  24. Alcohol outlet density and violence: A geographically weighted regression approach - Cameron - 2016 - Drug and Alcohol Review - Wiley Online Library [Internet]. [cited 2019 Apr 25]. Available from: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/dar.12295>
  25. Legg R, Bowe T. Applying Geographically Weighted Regression to a Real Estate Problem. Available at: [https://www.esri.com/news/arcuser/0309/files/re\\_gwr.pdf](https://www.esri.com/news/arcuser/0309/files/re_gwr.pdf). Accessed 30 Apr 2019. :2.
  26. Massoda Tonye SG, Kouambeng C, Wounang R, Vounatsou P. Challenges of DHS and MIS to capture the entire pattern of malaria parasite risk and intervention effects in countries with different ecological zones: the case of Cameroon. *Malaria Journal* [Internet]. 2018 Dec [cited 2019 Apr 20];17(1). Available from: <https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2284-7>
  27. Tewara MA, Mbah-Fongkimeh PN, Dayimu A, Kang F, Xue F. Small-area spatial statistical analysis of malaria clusters and hotspots in Cameroon;2000–2015. *BMC Infectious Diseases* [Internet]. 2018 Dec [cited 2018 Dec 14];18(1). Available from: <https://bmcinfctdis.biomedcentral.com/articles/10.1186/s12879-018-3534-6>
  28. Puth M-T, Neuhäuser M, Ruxton GD. Effective use of Pearson's product–moment correlation coefficient. *Animal Behaviour*. 2014 Jul;93:183–9.
  29. The DHS Program - Quality information to plan, monitor and improve population, health, and nutrition programs [Internet]. [cited 2018 Feb 15]. Available from: <https://dhsprogram.com/>

30. Spatial Data Repository, The Demographic and Health Surveys Program. ICF International. Available from spatialdata.dhsprogram.com [Internet]. [cited 2019 Mar 25]. Available from: <http://spatialdata.dhsprogram.com/home/>
31. Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4SF2T42>. [Internet]. [cited 2019 Apr 30]. Available from: <https://beta.sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-adjusted-to-2015-unwpp-country-totals>
32. CGIAR-CSI. 2009. "Global Aridity – Annual." [Internet]. [cited 2019 Apr 30]. Available from: <http://www.cgiar-csi.org/data/global-aridity-and-pet-database>
33. Center for Hazards and Risk Research - Columbia University, Center for International Earth Science Information Network - Columbia University, and International Research Institute for Climate and Society - Columbia University. 2005. "Global Drought Hazard Frequency and Distribution." [Internet]. [cited 2019 Apr 30]. Available from: <http://dx.doi.org/10.7927/H4VX0DFT>
34. Kamel Didan. 2016. "NASA MEaSUREs Vegetation Index and Phenology (VIP) Phenology EVI2 Yearly Global 0.05Deg CMG" [Internet]. [cited 2019 Apr 30]. Available from: [https://doi.org/10.5067/measures/vip/vipphen\\_evi2.004](https://doi.org/10.5067/measures/vip/vipphen_evi2.004)
35. Wessel P, Smith WHF. A global, self-consistent, hierarchical, high-resolution shoreline database. *Journal of Geophysical Research: Solid Earth*. 1996 Apr 10;101(B4):8741–3.
36. Wessel, Paul, and Walter Smith. 2017. "A Global Self-consistent, Hierarchical, High-resolution Geography Database Version 2.3.7." [Internet]. [cited 2019 Apr 30]. Available from: <http://www.soest.hawaii.edu/pwessel/gshhg/>
37. Climate Hazards Group. 2017. "Climate Hazards Group InfraRed Precipitation with Station data 2.0." [Internet]. [cited 2019 Apr 30]. Available from: <http://chg.geog.ucsb.edu/data/chirps/index.html>
38. Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*. 2015 Dec 8;2:150066.
39. Fotheringham AS, Brunson C, Charlton M, 2002. *Geographically weighted regression: the analysis of spatially varying relationships*. Wiley, Hoboken, NJ, USA
40. Tatem AJ, Adamo S, Bharti N, Burgert CR, Castro M, Dorelien A, et al. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. *Population health metrics*. 2012;10(1):8.
41. How Spatial Autocorrelation (Global Moran's I) works—Help | ArcGIS Desktop [Internet]. [cited 2019 Mar 26]. Available from: <http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm>
42. Rosenshein L, Scott L, Pratt M. Finding a Meaningful Model. Available at: <https://www.esri.com/news/arcuser/0111/files/findmodel.pdf>. Accessed 2019 Apr 23

## Tables

Table 1: Covariates data used for this study

Variable	Units	Description	Resolution	Data source
<b>Population density</b>	Number of people	The average number of people in the cells whose centroid falls within a radius of 10 km <sup>#</sup> or 2 km*.	2.5 minutes	(31)
<b>Aridity</b>	Aridity Index between 0.01 (Hyper-Arid) and 0.99 (Humid)	The average aridity index of the cells whose centroid falls within a radius of 10 km <sup>#</sup> or 2 km*.	Unspecified	(32)
<b>Drought episodes</b>	Individual classes between 1 (Low Drought) and 10 (High Drought)	The average of the drought episodes indices of the cells whose centroid falls within a radius of 10 km <sup>#</sup> or 2 km*.	Unspecified	(33)
<b>EVI year</b>	EVI value between 0 (least vegetation) and 10000 (Most vegetation)	The EVI was calculated by measuring the density of green leaves in the near-infrared and visible bands and describe as the average EVI of the cells whose centroid falls within a radius of 10 km <sup>#</sup> or 2 km*.	Unspecified	(34)
<b>Proximity to water</b>	Meters	Straight-line distance to the nearest major water body. Based on the World Vector Shorelines, CIA World Data Bank II, and Atlas of the Cryosphere	1:1,000,000	(35,36)
<b>Rainfall</b>	Millimeters per year	The average rainfall of the cells whose centroid falls within a radius of 10 km <sup>#</sup> or 2 km*.	30 minutes	(37,38)

# rural points and \* urban points.

Table 2: Summary of OLS Model Variables

Variable	Coefficient	Std Error	t-Statistic	Probability	Robust SE	Robust t	Robust Pr	VIF
<b>Intercept</b>	-412.757664	126.980400	-3.250562	0.001234*	119.323296	-3.459154	0.000597*	-----
<b>Population density</b>	0.017716	0.006601	2.683856	0.007486*	0.007310	2.423448	0.015671*	1.406656
<b>Aridity</b>	-0.085866	0.009752	-8.804637	0.000000*	0.014935	-5.749357	0.000000*	2.970344
<b>Droughts episodes</b>	0.026051	0.008413	3.096489	0.002066*	0.016161	1.611959	0.107536	1.295128
<b>Enhanced vegetation index</b>	0.108842	0.029597	3.677506	0.000270*	0.067284	1.617653	0.106302	1.586277
<b>Proximity to water</b>	-0.001290	0.000575	-2.242510	0.025297*	0.000409	-3.152627	0.001716*	1.381426
<b>Rainfall</b>	0.746337	0.084883	8.792495	0.000000*	0.135624	5.502997	0.000000*	3.278200

**Note:** \* An asterisk next to a number indicates a statistically significant p-value ( $p < 0.01$ ) VIF (Variance Inflation Factor) checks for redundancy among explanatory variables.

Table 3: Comparison between OLS and GWR models

Value	OLS	GWR
AICc*	9400.76851	9386.566909
Multiple R <sup>2</sup>	0.231561	0.252845
R <sup>2</sup> adjusted	0.223486	0.242829

\* Akaike's Information Criterion (AICc), measures the model fit/performance; a lower value indicates a better model performance.

## Figures

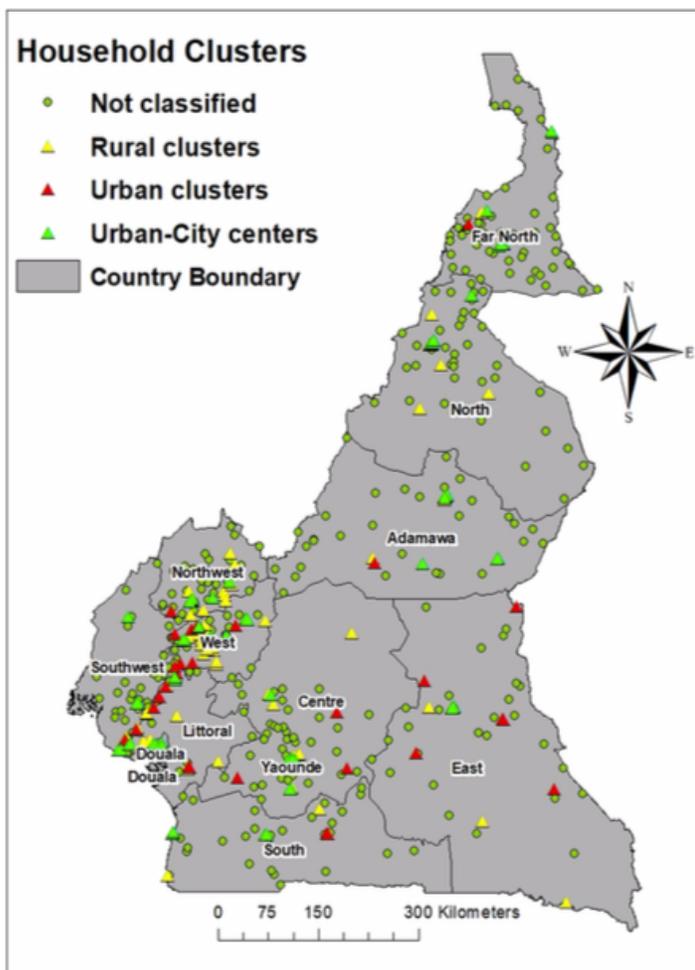


Figure 1

DHS household survey cluster points.

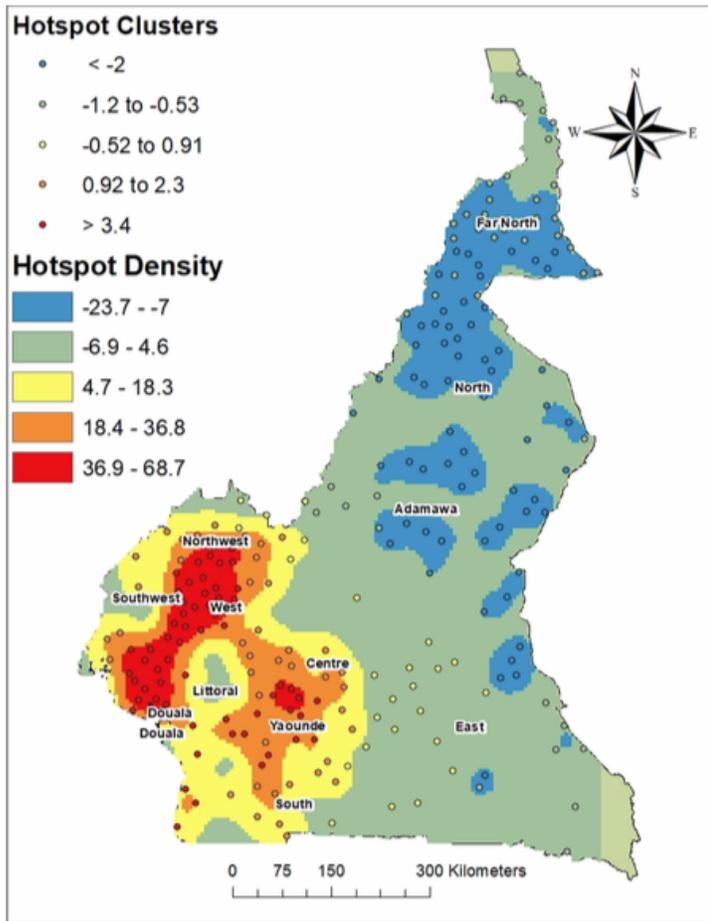
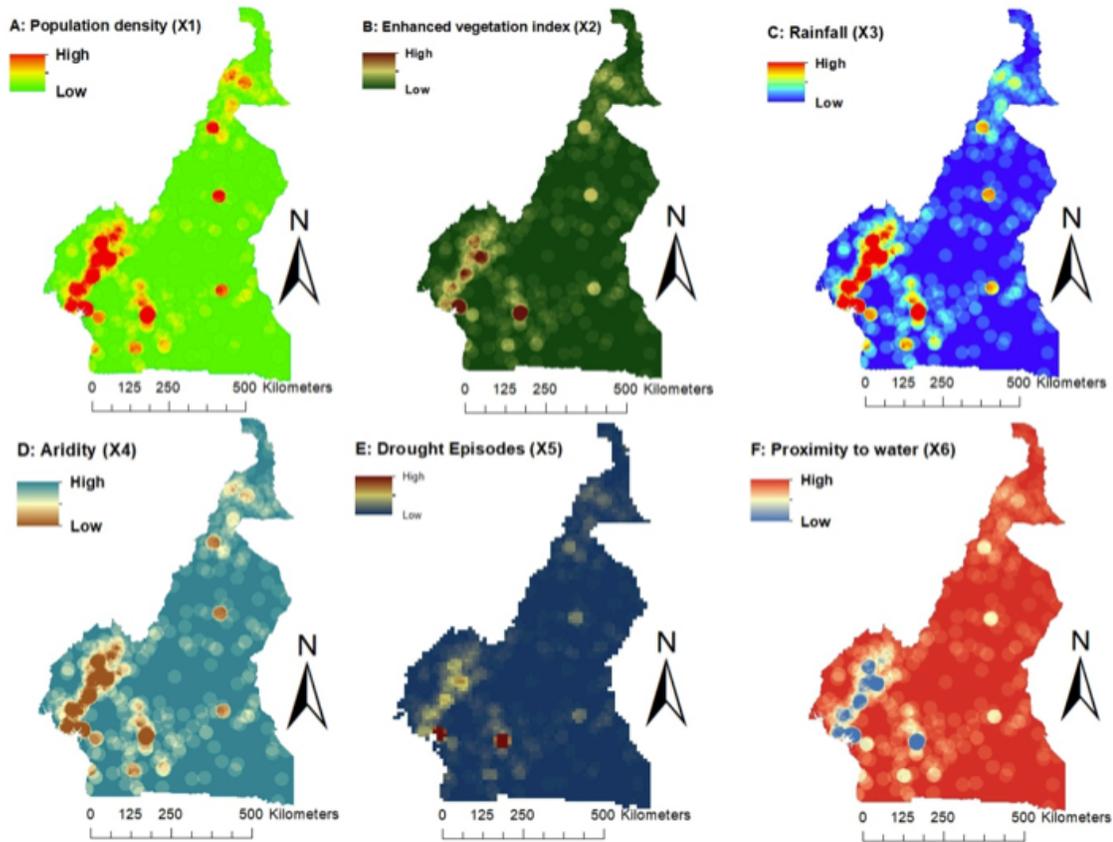


Figure 2

Map of malaria hotspot points and raster density.



**Figure 3**

Distribution of GWR model coefficients showing the spatial relationship between malaria hotspot cases and explanatory variables.

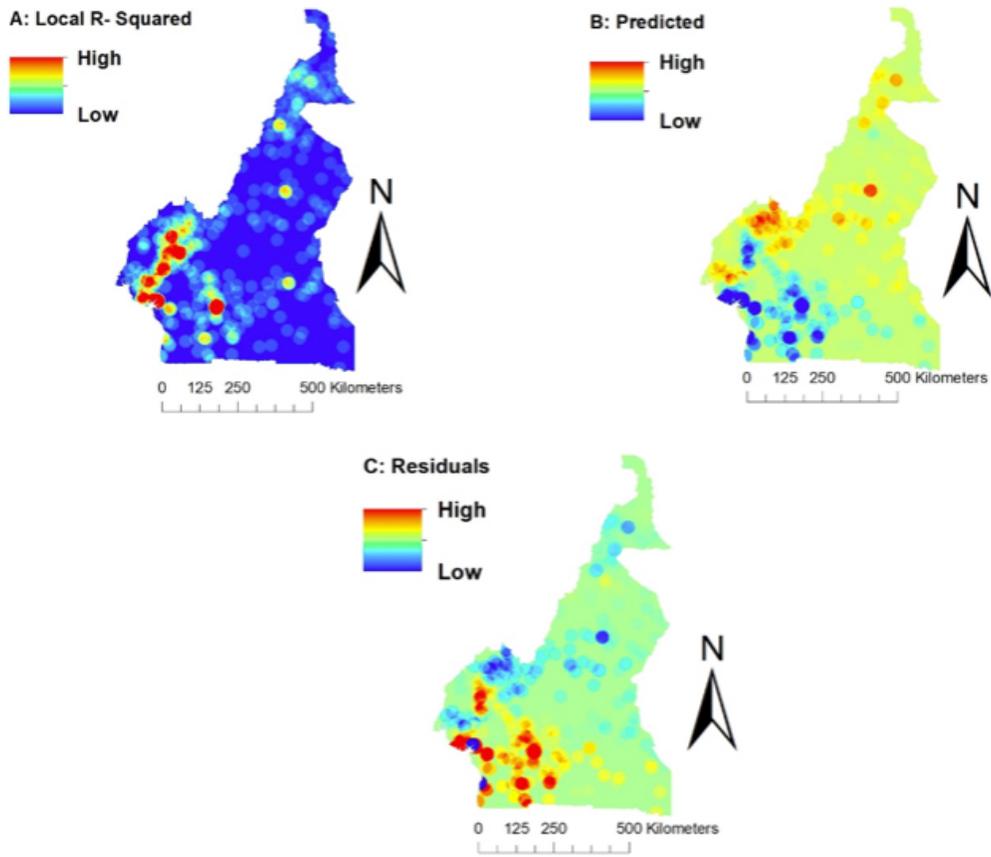


Figure 4

Map of GWR local  $R^2$ (A), predicted malaria areas (B), and model residuals (C)