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

Marlvin Anemey Tewara

Department of Biostatistics, School of Public Health, Shandong University Cheeloo College of Medicine, Jinan 250012

<https://orcid.org/0000-0003-3720-0084>

Liu Yunxia

Department of Biostatistics, School of Public Health, Shandong University Cheeloo College of Medicine, Jinan 250012

Prisca Ngetemalah Mbah-Fongkimeh

School of Medicine and Public Health, Faculty of Health and Medicine, The University of New Castle Callaghan

Zheng Zhaolei

Department of Biostatistics, School of Public Health, Shandong University Cheeloo College of Medicine, Jinan 250012

Helen Barong Binang

Department of Clinical Laboratory Diagnosis, Shandong University Cheeloo College of Medicine, Jinan 250012

Liu Xinhui

Department of Biostatistics, School of Public Health, Shandong University Cheeloo College of Medicine Jinan 250012

Zhou Miao

Department of Biostatistics, School of Public Health, Shandong University Cheeloo College of Medicine, Jinan 250012

Xiaojuan Liu

Department of Biostatistics, School of Public Health Shandong University Cheeloo College of Medicine, Jinan 250012

Fuzhong Xue (✉ xuefzh@sdu.edu.cn)

Research article

Keywords: Geographically weighted regression, Ordinary least squares, malaria, spatial statistics, mapping, insecticidal net, households

Posted Date: May 26th, 2019

DOI: <https://doi.org/10.21203/rs.2.9820/v1>

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

Abstract

Background Cameroon has witnessed a 131,000 increase in malaria cases, according to a recent report addressing the malaria burden and control strategies in endemic regions. Studies have illustrated the association between malaria cases and environmental factors in Cameroon but limited in addressing how these factors vary in space for timely 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) tool 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 OLS candidate variables to examine, predict and explore the spatial variability between environmental factors and malaria hotspot cases generated from Getis-Ord Gi* statistical analysis. Spatial maps of mosquito bed net ownership and GWR outputs were also created for public health surveillance. **Results** The OLS candidate environmental variable coefficients were statistically significant for a properly specified GWR model (adjusted R² = 22.3% and p < 0.01). The GWR model identified a strong 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 (adjusted R²= 24.3%). The mosquito bed nets analysis maps demonstrated an overall low coverage(<50%) of household ownership. **Conclusion** The generated GWR maps suggest that for policymakers to eliminate malaria in Cameroon by 2030, there should be the creation of outreach programs that will target malaria hotspots locations, intensify free insecticidal net distribution, allocate specific funding, establish vaccination campaigns and carry out further investigations in areas where the environmental variables showed strong spatial associations with malaria hotspot cases.

Background

In a recent “high burden to high impact; a targeted malaria response” call report to address malaria burden and control strategies in endemic regions, Cameroon witnessed an increase in malaria cases of 131000 and 1.3 million additional cases in neighboring Nigeria. 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. Only India marked progress in reducing its disease burden, registering a 24% decrease compared to 2016(1). 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(2,3). Because of the many socio-economic 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(4). 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(5–10). 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 (11,12) and understanding the behavior of these factors in space with the application of regression statistics(13) 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(13). 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 (14), 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 was associated with malaria, Moise(15), 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(14–17), its application has been valuable in the understanding of other health outcomes and social science studies including cancer events(18), dengue fever(19), mental depression(20), fire events(21), hospital accessibility study (22), alcohol and violence(23) and real estate housing crisis(24).

Massoda (25), 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(26), in a recent study on small area spatial statistical analysis of malaria clusters and hotspots in Cameroon, found the association between malaria cases and environmental factors using the Pearson correlation statistics(27) 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 and resource allocations, prevention programs such as insecticide-treated nets (ITN) distribution, clinical diagnosis, and treatments. The specific objective of this study is to identify ITN coverage in Cameroon and to find the spatial variability between malaria hotspot cases(26) 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(28). The DHS program has been described elsewhere(26). Households are grouped as either an urban(city block or apartment building) or rural (village or group of villages) cluster points 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 2011, DHS VI malaria survey year was obtained from the DHS spatial data repository site(29). The recent 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 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)(26). The environmental covariates data set used for this study are described in table 1.

Ethical approval

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 are conducted only if the respondent provides voluntary informed consent. Written informed consent was obtained from all participants.

Mapping ITN usage.

To understand the distribution of malaria with respect to the given household clusters and how this has been controlled in the past, we mapped selected behavioral categories of household ownership of mosquito bed nets. The characteristics mapped include: household with at least one mosquito bed net, household with at least one ITNs, household with at least one long-lasting insecticidal net (LLITNs) ,

population who slept under long-lasting insecticidal net (LLITNs) the last night, population who slept under any net the previous night and the number of people living in a house with at least one ITNs. These characteristics were mapped using ArcGIS and data values ranged from light (lower data value) to dark (high data value).

Statistical analysis

Getis-Ord G_i^* statistics

A recent study by Tewara (26), identified malaria hotspots locations for the 2015 malaria year that was used as our response variable for the regression analysis model specification. 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. We represented the malaria hotspots map by (26) as raster surfaces with high and low data values.

The Getis-Ord local statistics is given as:

$$G_i^* = \frac{z_i - \bar{z}}{s} \frac{\sum_{j=1}^n w_{ij} z_j}{\sum_{j=1}^n w_{ij}} \quad (1)$$

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:

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

$$s^2 = \frac{\sum_{j=1}^n w_j z_j^2 - \bar{z}^2}{\sum_{j=1}^n w_j} \quad (3)$$

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 both the dependent and explanatory variables in the regression statistics used in our model is given as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 \dots \beta_n + \varepsilon \quad (4)$$

Y is the dependent or response variable representing the process to be predicted or understood; for example, the malaria hotspot cases, X is the independent or explanatory variables used to model or predict the dependent variables. The explanatory variables include: population density (X_1), enhanced vegetation index (X_2), rainfall (X_3), aridity (X_4), drought episodes (X_5), and proximity to water (X_6). 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 or interdependence between the dependent (malaria hotspots) and explanatory variables. β is the regression coefficients computed by the regression tool in the ArcGIS software. They are values, one for each explanatory variable, that represent the strength and type of relationship the explanatory variable has to the dependent variable. β_0 is the regression intercept. It represents the expected value for the dependent variable if all the independent variables are zero. The residuals ε represent the portion of the dependent variable that isn't explained by the model; the model under and over predictions (13).

For this study our regression model was built using the exploratory regression tool in ArcGIS; to find the candidate variables and their spatial significance for a properly specified model. These variables include all the explanatory variables (X_1 to X_6). The candidate variables were then used to build a global regression model using the OLS tool in the modelling spatial relationships tool in ArcGIS. To have a properly specified GWR model, the generated OLS summary result of the model variables should pass a statistically significant six-test check and the candidate variables from the OLS model will then be used to specify our GWR model.

The GWR tool in the spatial modelling toolset was then used to build our prediction model 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

ITNs usage coverage

For regions with at least one mosquito net regardless of whether it was a treated net or not, households in the North and Far North regions had high usage coverage ($> 50\%$) while households in the East and West regions of the country had low coverage of homes having at least one treated or not treated mosquito net as depicted in figure 1A. For household clusters with at least one ITN, the North and the Northwest regions had high coverage of bed net distribution while the Center and the Littoral regions had moderate coverage and low coverage in the Far North region as shown in figure 1B. For the category of household clusters with at least one LLITNs, the North and the Northwest regions had high percentages of household ownership of LLITNs while the Far North and South regions had low LLITNs distribution; figure 1C.

For the number of persons living in a household with at least one ITN, the North region alone had high ITN ownership (>2500 people) while the South region saw a low ITN ownership < 409 people as demonstrated in figure 2A. For the population who slept under any net the night before the survey was conducted, Littoral, Centre, South, and Adamawa regions had a high percentage of the population while the Far North had the least percentage ($<9.6\%$) of the population in this category; figure 2B. Finally, for the population who slept under LLITNs the night before the survey, the Adamawa, Northwest, Littoral, and Douala regions had higher percentages of people than the rest of the other regions. The Far North and the West regions had the lowest percentages of people who slept under LLITNs the night before the survey was conducted as seen in figure 2C below.

Hotspot analysis

The hotspot map was analyzed from the most recent 2015 malaria year data to find out whether ex malaria hotspots show a spatial variability with environmental covariates. The analysis depicted that, there was high malaria hotspots distribution in some areas of the Far North, North, Northwest, Southwest, Littoral, Yaoundé, East, and Adamawa regions shown as high data values in red; figure 3. Some locations in the Littoral, Center, Far North, South and Northwest regions had low malaria hotspot distribution.

Ordinary Least Square (OLS)

The OLS global regression summary result illustrates the statistical significance of the model variables and feasibility to be used in 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 the other explanatory variables had a positive relationship with the dependent variable. The OLS standard deviation residuals were randomly distributed after a global spatial autocorrelation check for randomness. The adjusted R-squared [d] had a 22.3 % evaluation of the model performance as illustrated in table 2.

Geographically Weighted Regression (GWR)

The local GWR model fits one equation (equation a^*) for each feature or explanatory variable from the global OLS model to explore the correlations between the dependent (Y) and the explanatory (X) variables. In the current study, the coefficients (β) 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 cases in most areas in the western portion of the country and few elsewhere (areas in red), while aridity (X_4) and proximity to water (X_6) showed a weak association (areas in blue) as exemplified in figure 4 below.

The GWR output produced a predicted malaria map, local R- squared (R^2) and residuals. The local R^2 in figure 5A shows that the model performance was high or strong in regions of the western part of the country and some areas in the Central and Yaoundé DHS regions. The value of R^2 ranges from 0 to 100 percent and the adjusted R^2 for this model was 0.24 (24%) as seen in table 3. The predicted map

(the combined strength of the relationship amongst the variables used) was high in areas of the Southwest, Northwest and the North regions; figure 5B, while the model residuals (areas not explained by the model) were seen in most parts in the South region; figure 5C.

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. The result established that the GWR improved our OSL for the model specification and performance.

Discussion

One of the most effective methods of the malaria vector control in Cameroon is the use of mosquito bed nets and depending on whether it is a long-lasting insecticidal one, further increases the vector control chances. Several studies have highlighted the importance of increased mosquito net usage and LLITN and their role in reducing the incidence of malaria in SSA (38–44). In this study, most urban-city centers and rural clusters household settlements in the North region had at least one mosquito net whether or not it was a treated (> 50%), while the West and East regions recorded a low coverage for the above ITNs coverage category. For households that own at least one ITN and one LLITN, the North and the Northwest regions had high coverage of the mosquito net ownership. The consistent high coverage of bed nets in the North region tells us of the prevention awareness in that region. The characteristics maps are indications of how well malaria maps can be used to initiate and monitor malaria programs(8). The effectiveness of ITNs has been studied in the semi-urban and rural communities in the Southwest region of Cameroon (45); our generated maps can be used to further understand the effectiveness of the mosquito net coverage in the rural, urban and urban-city clusters in other regions for effective control programs. The number of persons living in a household with at least one ITN was highest in the North region correlating with the high coverage of mosquito bed net coverage in that region. This interesting characteristic could be supporting the reasons why this region recorded low malaria cases in the past as reported by Gemperli (6), though Massoda (25), demonstrated a drawback with the DHS data set as compared to Malaria indicator survey (MIS) in the underreporting of high malaria parasite risk in the North and Far North regions(46). This calls for cautions and continues promotion of mosquito net usage and other preventive campaigns. Population who slept under LLITNs last night was highest in the Adamawa, Northwest, Littoral and Douala DHS regions while for population who slept under any net last night was high in Adamawa, Center, Littoral and South regions; this apparent high coverage in these regions could be associated with the increased population density demonstrated in recent studies(26), due to increase urbanizations (47) and administrative geolocational importance of Yaoundé(Center) and Douala(Littoral) receiving the highest number of city dwellers.

Malaria hotspots were identified in rural and urban-city centers of the Western, Central, East and Northern part of the country. These hotspot locations are the baseline for further research and intervention programs. We used the hotspot locations as the subset of the population for our GWR model.

For the model specification, the OLS model (table 2) illustrated that, all the explanatory variables from the candidate variable exploratory regression analysis were statistically significant ($p < 0.01$); meaning the coefficients were statistically significant at the 95 percent confidence level and that the variables explain our model, though the coefficients of the aridity and proximity to water had negative associations (table5 [a]). A positive coefficient 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. For example, an increase in rainfall in a 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. Similarly, a negative coefficient means X and Y changed in reverse directions(14). For example, the negative coefficient for aridity in our model means the malaria cases decreases with lack of water since aridity is a deficiency of moisture probably due to the lack of rainfall. 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 (test for non-stationarity) was statistically significant ($P < 0.01$) and reflects that the relationships being modeled are consistent across the entire study area and thus nonstationary (except for drought episodes and EVI) as seen from the robust probability (table2 [b]). Furthermore, the Variance Inflation Factor (VIF) values (< 7.5; table2 [c]) indicates no redundancy among explanatory variables. The OLS model also produces an output residual. These residuals were tested for clustering using the spatial autocorrelation tool in ArcGIS(48) and it indicated that the variables used were randomly distributed and did not have a statistically significant z-score (p -value = 0.34 and z-score: 0.73), because a statistically significant spatial autocorrelation ($p < 0.05$ and z-score ≥ 1.96) in the model residuals would indicate that we were missing one or more key explanatory variables. This is a positive indicator of choosing a good model(49). Joint F and Wald Statistics (table 2 [e]): Asterisk (*) indicates overall model significance ($p < 0.01$); and because the Koenker (BP) Statistic [f] was statistically significant, we used the Wald Statistic to determine overall model significance(49).The Jarque-Bera Statistics[g] 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 GWR model because recent studies(14) have reported on spatial variations similar to our specified model variables and our main goal was to understand the behavior of these variables with malaria cases for future research and intervention projects. The R-squared and Akaike's Information Criterion (AICc) measures the model fit/performance. The adjusted R-squared (table2 [d]) had a 22.3 % evaluation of the model performance, indicating that the combined effect of the explanatory variables in the OLS model was telling 22 % of the relationship story between the malaria cases and the environmental factors we are trying to model. This may seem low per the R^2 range (0-100 %) where higher R^2 values depict good model performance. However, environmental variables tend to be very complex since they most often are natural processes and will be very cumbersome to test for all higher R^2 variables due to time constraint. Hasyim (14), 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 (15) 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(13).

The local GWR model was built based on the variables from the global OSL model. A validated OLS can lead to a global policy for malaria control programs while a validated spatial relationship with GWR is an appropriate method to initiate prevention programs in local systems(14). The GWR output coefficients maps indicated that population density, EVI, rainfall, and drought episodes had a strong 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; figure 4. Whereas, aridity and proximity to water had a negative or weak association in the above-cited locations. The GWR adjusted R^2 value was higher (24.3 %) than the OLS model as expected to indicate improvement with local GWR model. This signifies that the GWR is a better indicator for explaining spatial variability at the local level(14,22). The map of the generated R^2 (figure 5A) indicates that, the GWR explains 24.3% of the spatial variability between malaria cases and the environmental factors in the Northwest, Southwest, Littoral, Douala, Central and Yaoundé DHS regions and a weak or low spatial variability in some areas in the East, Adamawa and North regions. This can further be demonstrated from our predicted map(figure 5B) highlighting some household clusters in the Northwest, Southwest, Adamawa and North regions as the predicted locations for malaria control and public health interventions.

The GWR residuals indicated that areas with over and under predictions 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 of the GWR indicates that our 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). Furthermore, table 3 illustrated that the GWR improved our understanding of the spatial relationship between malaria cases and environmental factors since the AICc and R^2 were higher than that of OLS.

We had the following limitations: Firstly, we did not include all the categories that were used for the assessment of mosquito bed net ownership and usage, though we had a wider presentation of the analysis that was limited in previous studies(26) in Cameroon. Secondly, 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. Thirdly, our OLS model was biased and failed to pass the six tests check at the level of the Jarque-Bera statistics (table 2 [g]) that was significant ($p < 0.01$) and indicating our explanatory predictors were not normally distributed in some locations. We proceeded because our main goal was to understand the spatial variability between malaria cases and environmental factors in Cameroon that has been demonstrated in other countries(14,15) for intervention programs. Applying a log function to transform the variables would have improved the model biased problem, but will need re-running and removing possible outliers and testing for a new properly specified model which was challenging(49). Fourthly, the timing and the displacement nature of the DHS data points for confidentiality may have falsified the results as the analyzed points were displaced prior to release though maintained in their respective DHS regions. Future investigations are required to test for other environmental and socioeconomic variables to provide a detailed spatial interdependence with the malaria disease. Though our study had some remits, it has demonstrated a rigorous understanding of the spatial interdependence between malaria cases and environmental predictors. Moreover, the methods in this study can be used to study other health outcomes in Cameroon that have been applied in available literature(16,20,22). To the best of our search, this study has provided new insights into the malaria disease at the local level in Cameroon by applying the GWR spatial modelling technique that has never been done before.

Conclusions

The application of a local GWR spatial modelling to malaria research is of immense help in answering the pressing scientific questions faced by policymakers in controlling malaria. Given the greater availability of spatial data⁽²⁹⁾ and desktop GIS packages and statistical techniques, the challenges faced in the malaria disease investigation will be improved in the future. Our analyses demonstrated low coverage of mosquito net ownership in Cameroon, especially in public health resource-limited areas. The generated GWR maps suggest that for policymakers to archive the GTS targets for malaria by 2030, there should be the creation of outreach programs that will target malaria hotspots locations, intensify ITNs distribution, allocate specific funding, establish vaccination campaigns and carry out future investigations in areas where the environmental variables showed strong spatial associations with malaria hotspot 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; LLITN: Lon-Lasting Insecticide Net; AICc: Akaike's Information Criterion; VIF: Variance Inflation Factor; MIS: Malaria Indicator Survey.

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. 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. Institute 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. Available from:
<https://dhsprogram.com/pubs/pdf/FR260/FR260.pdf> Accessed 30 Apr 2019
3. Ministry of Public Health. Sectorial health strategic plan 2001–2015. Available from: <http://www.nationalplanningcycles.org/file-repository/CMR> Accessed 30 Apr 2019
4. World Health Organization, World Health Organization, Global Malaria Programme. Global technical strategy for malaria, 2016-2030. Available from: http://apps.who.int/iris/bitstream/10665/176712/1/9789241564991_eng.pdf?ua=1 Accessed 19 Mar 2019
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. Dalrymple U, Mappin B, Gething PW. Malaria mapping: understanding the global endemicity of *falciparum* and *vivax* malaria. *BMC Medicine* 2015 Dec;13(1).
11. 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*. 2018 Dec;17(1). Available from:
<https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-018-0122-3>
12. Efe SI, Ojoh CO. Climate Variation and Malaria Prevalence in Warri Metropolis. *Atmospheric and Climate Sciences*. 2013;03(01):132–40.
13. Regression analysis basics [Internet]. Available from:
http://resources.esri.com/help/9.3/ArcGISDesktop/com/Gp_ToolRef/Spatial_Statistics_toolbox/regression_analysis_basics.htm. Accessed 13 Apr 2019
14. 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;17(1). Available from:
<https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2230-8>
15. 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.
16. 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;15(1). Available from:
<http://malariajournal.biomedcentral.com/articles/10.1186/s12936-016-1599-5>
17. 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.

18. 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.
19. 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.
20. 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.
21. 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.
22. 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>
23. 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>
24. 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. Accessed 30 Apr 2019.
25. 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;17(1). Available from: <https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2284-7>
26. 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;18(1). Available from: <https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3534-6>
27. Puth M-T, Neuhäuser M, Ruxton GD. Effective use of Pearson's product–moment correlation coefficient. *Animal Behaviour*. 2014 Jul;93:183–9.
28. 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/>
29. Spatial Data Repository, The Demographic and Health Surveys Program. ICF International. Available from spatialdata.dhsprogram.com. Available from: <http://spatialdata.dhsprogram.com/home/> Accessed 25 Mar 2019
30. 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>
31. CGIAR-CSI. 2009. "Global Aridity – Annual." [Internet]. Available from: <http://www.cgiar-csi.org/data/global-aridity-and-pet-database> Accessed 30 Apr 2019
32. 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." Available from: <http://dx.doi.org/10.7927/H4VX0DFT> Accessed 30 Apr 2019
33. 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
34. 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.

35. 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/>
36. 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>
37. 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.
38. Etang J, Nwane P, Piameu M, Manga B, Souop D, Awono-Ambene P. Evaluation of New Tools for Malaria Vector Control in Cameroon: Focus on Long Lasting Insecticidal Nets. Eisele T, editor. *PLoS ONE*. 2013 Sep 23;8(9):e74929.
39. Bhatt S, Weiss DJ, Mappin B, Dalrymple U, Cameron E, Bisanzio D, et al. Coverage and system efficiencies of insecticide-treated nets in Africa from 2000 to 2017. *eLife* [Internet]. 2015 Dec 29 [cited 2018 Mar 24];4. Available from: <https://elifesciences.org/articles/09672>
40. Noor AM, Mutheu JJ, Tatem AJ, Hay SI, Snow RW. Insecticide-treated net coverage in Africa: mapping progress in 2000–07. *The Lancet*. 2009;373(9657):58–67.
41. Levitz L, Janko M, Mwandagaliwa K, Thwai KL, Likwela JL, Tshetu AK, et al. Effect of individual and community-level bed net usage on malaria prevalence among under-fives in the Democratic Republic of Congo. *Malaria Journal* [Internet]. 2018 Dec [cited 2019 Apr 26];17(1). Available from: <https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2183-y>
42. Dolan CB, BenYishay A, Grépin KA, Tanner JC, Kimmel AD, Wheeler DC, et al. The impact of an insecticide treated bed net campaign on all-cause child mortality: A geospatial impact evaluation from the Democratic Republic of Congo. Carvalho LH, editor. *PLOS ONE*. 2019 Feb 22;14(2):e0212890.
43. Allcock SH, Young EH, Sandhu MS. A cross-sectional analysis of ITN and IRS coverage in Namibia in 2013. *Malaria Journal* [Internet]. 2018 Dec [cited 2019 Apr 26];17(1). Available from: <https://malariajournal.biomedcentral.com/articles/10.1186/s12936-018-2417-z>
44. Saleh J-EA, Saddiq A, Uchenna AA. LLIN Ownership, Utilization, and Malaria Prevalence: An Outlook at the 2015 Nigeria Malaria Indicator Survey. *OALib*. 2018;05(01):1–12.
45. Apinjoh TO, Anchang-Kimbi JK, Mugri RN, Tangoh DA, Nyingchu RV, Chi HF, et al. The Effect of Insecticide Treated Nets (ITNs) on Plasmodium falciparum Infection in Rural and Semi-Urban Communities in the South West Region of Cameroon. Culleton R, editor. *PLOS ONE*. 2015 Feb 25;10(2):e0116300.
46. WHO, Recrudescence de cas de paludisme dans la région de l'Extrême-Nord du Cameroun | Regional Office for Africa [Internet]. [cited 2019 Apr 27]. Available from: <https://www.afro.who.int/fr/news/recrudescence-de-cas-de-paludisme-dans-la-region-de-lextreme-nord-du-cameroun>
47. Kabaria CW, Gilbert M, Noor AM, Snow RW, Linard C. The impact of urbanization and population density on childhood Plasmodium falciparum parasite prevalence rates in Africa. *Malaria Journal* [Internet]. 2017 Dec [cited 2018 Mar 15];16(1). Available from: <http://malariajournal.biomedcentral.com/articles/10.1186/s12936-017-1694-2>
48. 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>
49. 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	Data source
Population density	Number of people	The average number of people in the cells whose centroid falls within a radius of 10 km# or 2 km*.	(30)
Aridity	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# or 2 km*.	(31)
Drought episodes	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# or 2 km*.	(32)
EVI year	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# or 2 km*.	(33)
Proximity to water	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	(34,35)
Rainfall	Millimeters per year	The average rainfall of the cells whose centroid falls within a radius of 10 km# or 2 km*.	(36,37)

rural points and * urban points.

Table 2: Summary of OLS Results - Model Variables

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

OLS Diagnostics		
Number of Observations:	578	Akaike's Information Criterion (AICc) [d]: 9400.76851
Multiple R-Squared [d]:		Adjusted R-Squared [d]:
Joint F-Statistic [e]:		Prob(>F), (6,571) degrees of freedom:
Joint Wald Statistic [e]:		Prob(>chi-squared), (6) degrees of freedom:
Koenker (BP) Statistic [f]:		Prob(>chi-squared), (6) degrees of freedom:
		Prob(>chi-squared), (2) degrees of freedom:
	0.231561	0.223486
	28.677491	0.000000*
	47.830227	0.000000*
	39.159382	0.000001*
Jarque-Bera Statistic [g]:	299512.123672	0.000000*

Note: * An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$). The Letters [a] to [g] highlights the steps in the interpretation of the OLS results. VIF (Variance Inflation Factor) values (> 7.5) indicates redundancy among explanatory variables.

Table 3: Comparison between OLS and GWR models

Value	OLS	GWR
AICc*	9400.76851	9386.566909
Multiple R2	0.231561	0.252845
R2 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

Household categories of mosquito bed net ownership and usage.

Figure 2

Population characteristics of mosquito bed net usage.

Figure 3

Malaria hotspot raster map from the Getis-Ord Gi* analysis.

Figure 4

Distribution of GWR model coefficients (A-F) showing the spatial relationship between malaria hotspot cases and explanatory variables (X1 to X6).

Figure 5

Map of GWR local R2(A), predicted(B), and model residuals (C).