UNIVERSITY OF THE WITWATERSRAND

FACULTY OF HEALTH SCIENCES

SCHOOL OF PUBLIC HEALTH

Bayesian spatio-temporal analysis of malaria prevalence in children aged 2-10 years between 2000-2015 in Gabon

Fabrice Lotola Mougeni

Student number: 2485529

Research report Submitted to the faculty of Health Sciences, University of the Witwatersrand, in partial fulfilment of the requirement for the Degree of MSc Biostatistics

May 2023

Declaration

University of the Witwatersrand

Faculty of Health Science

Private Bag 3 Wits, 2050

Johannesburg

RE: AGREEMENT BY CO-AUTHORS

Fabrice Mougeni, student number 2485529, wrote the protocol, acquired ethical approval, cleaned the data, performed data analysis, wrote the first draft and received critical comment from all co-authors. By signing this declaration, the co-authors listed below agree to the use of the article by the student as part of her Research Report.

Fabrice Mougeni __ _____22/05/2023_____

Bertrand Lell _____

Ngianga-Bakwin Kandala _____ _____

Tobias Chirwa _____ _____

Dedication

À Josiane Itoumba et Kathy Mbutu mes mamans d’amour.

À maman Sylvie Makarighi, Gloria Ndoulou, Marya Anne-Marie en souvenir de mon cher onlce, Ndambola Joseph qui nous a été arraché si brutalement.

À toute ma famille et toutes les personnes qui me sont chères, qui m’ont toujours encouragées, et qui ont cru en moi pour le début de mes aventures depuis le premier temps.

Acknowledgement

I would first like to express my gratitude to Professor Tobias Chirwa, Head of the school, who honored me by accepting to direct this work, when at the beginning I did not have a clear idea of what I should work on, and even what data I could use, until we was able to obtain the data from the DHS program. My gratitude also goes to my co-supervisors, Prof. Kandala Bakwin and Prof. Bertrand Lell for their expertise, feedback, and clear direction to improve the quality of the work. I am very happy to have worked with you, and I would appreciate to continue working with you and learning more. I also find a place here to be grateful for USAID's initiative through the DHS program to make public data available. These have played a major role in this research.

I also believe that the success of this work was deeply dependent on the quality of the teaching in biostatistics provided in the Epidemiology and Biostatistics Division of the University of the Witwatersrand. I still remember how amazed I was by the way Prof Levin explained the course on hypothesis testing or Prof Musengue introduced us to the use of REDcap or STATA and Prof Maposa to the analysis of spatial data with R. I thank you very much. I can't mention all the names here because there are so many of them working in quality, but please find my gratitude for what you brought to me from the epidemiology course to the specific courses for biostatisticians including the comments while we were preparing our protocols our first year. I also think of my classmates Alberto and Esnath. I remember our encouragement during the year. Thanks a lot.

I want to thank the PAMAfrica project for the scholarship during these two years in the framework of capacity building. I am very grateful for everything, for the time you took to be informed about the progress during the master and the interest to see young researchers to be trained in malaria research. It was simply not about finance only but also about commitment to really help in this field of public health. I also thank the Directors of CERMEL Prof Akim Adegnika, Prof Bertrand Lell and Prof Maxime for their guidance during the years of work. The knowledge acquired during this time has also largely contributed to the quality of this work which I can say is on a standard scale.

This work is also the contribution of my family including friends. Thanks for your efficient support. I am thinking very strongly as I write these few lines about Papa Joseph. I wish he would have read these words, that he would have been there with me, so that I could have told him in detail about my days during this master's program as we liked to do. He was a source of motivation, a hand that was missing somewhere in my heart for my journey, Mais helas! I thank my dear friend Dr. Larissa Ngombi Mavoungou for the time she was able to dedicate to read this work, aligning yourself to the time requirements. Thank you also for your encouragements which are also a support to continue this work. At that point, I did not forget you, Ms Honoraldie Nzadieni for your multi-form support in my life. Thank you a lot.

Finally, I thank God for his grace upon me, for the results of this master's degree, and for all that he has certainly placed before me among the points mentioned above. The example of excellence comes to us properly from him. I want to say thank you to him for the people I met in church, especially Pastor Josué Kukimunu, Mama Eugénie Kukimunu and Papa Dieu-donné, with wonderful preachings, adding a stone to my faith, motivating me to do more through, to give the best of myself in what I do. I

thank the Lord for having met you and the whole family. There are many people who have given me good advice from far and near. Excuse me for not mentioning you here but you are in my heart.

Table of content

List of figures

List of tables

List of abbreviations

ACT	Artemisine Combination Therapy
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
AR	Auto-Regressive
BART	Bayesian Additive Regression Trees
BYM	Besage York Mollié
CAR	Conditional Autoregressive
CDC	Disease Control and Prevention
CI	Credible Interval
DALYs	Disability Adjusted Life Years
DGM	Direction Générale de la Météorologie
DGS	Direction Générale de la Statistique
DHS	Demographic and Health Survey
DIC	Deviance Information Criterion

EVI	Enhanced Vegetation Index
EWS	Early Warning System
GAM	Geo-Additive Model
GBM	Gradient Boosted Trees
GF	Gaussian field
GLM	Generalized linear model
GMRF	Gaussian Markov Random Field
GPS	Global Positioning System
GWR	Geographically Weighted Regression
iid	Idependent Identically Distributed
INLA	Integrated Nested Laplace approximation
ITNs	Insecticide Treated Bed Nets
KLD	Kullback-Leibler Divergence
LISA	Local Indicator of Spatial Autocorrelation
LM	Lagrange Multiplier
MCMC	Markov Chain Monte Carlo
NMCP	National Malaria Control program
OLS	Ordinary Least Square, Ordinary Least squared
PC	Penalized Complexity
PfPR	Plasmodium falciparum Parasite Rate
RW	Random Walk
SAE	Small Area Estimation

SDEM	Spatial Durbin Error Model
SDGs	Sustainable Development Goals
SEM	Structural Equation Model, Spatial Error Model
SLM	Spatial Lag Model
SPDE	Stochastic Partial Differential Equation approach
SPTOLS	Spatio-Temporal Model for OLS
SVC	Spatial Varying Coefficient
TCL	Theorem of Central Limit
USAID	United State Agency for International Development
VIF	Variance Inflator Factor
WHO	World Health Organization

CHAPTER 1 - INTRODUCTION

1.1 Background

This study focused on understanding malaria burden by using the space and time effects, and the effect of some factors. This introduction part is divided as follows: a brief overview of malaria in Gabon for the past 20 years is presented. This includes the benefit of interventions, the epidemiological change in the most at-risk age group, the importance of using the *Plasmodium falciparum* parasite rate (PfPR) for children with age in the range 2-10 years, the trend in prevalence over time, and the factors accounted for in the variation of the prevalence. Then, in the literature review section, a slight insight on the impact of malaria, its transmission or associated risk factors was provided. Also the section reviewed Bayesian modeling in malaria research and provided discussion on some results. Then next sections focus on the problem statement, justification, research question, and end up with aims and objectives of the study.

1.2 Background

Malaria is an infection caused by the *Plasmodium* parasite. The transmission to human is possible by an infected female *Anopheles* (1). About 241 million cases and 627 000 deaths were related to malaria in 2020 in large scale (2). In Africa, the Sub-Saharan region accounts for approximately 90% of the malaria cases and mortality caused by the *Plasmodium falciparum.* It is the one that most often causes fatal or severe malaria, of the other four species, that is *Plasmodium vivax, knowlesi, ovale and malariae* (1).

Gabon is a Sub-saharian country with a high prevalence of malaria (3). Malaria transmission is almost present during all seasons of the year, in particular due to the position on the equatorial line, and making the area hyper-endemic to the *P. falciparum.* On average, the prevalence can exceed 50% for children under five in some specific regions (4), and may hamper the economic development from the district to the national level. Following the sustainable development goals (SDGs) of Gabon for good health and well-being, there is need to decrease importantly the rate of mortality for children under-five from 41% to 2.6%. To contribute in such a goal, reducing the burden of malaria may be an important point to target, since malaria is the first cause of consulting, hospitalization, and deaths in Gabon. To contribute in reducing this clinical burden, Gabon through the national program is committed in the initiative of taking action against malaria for its eradication by 2030.

Focusing on children, previous studies up until 2005 to 2008, have shown the most affected age group is that of children with age under 5, since they can rapidly develop severe disease due to the low immunity (5). In fact, to acquire a natural immunity against an infection such as malaria, it is necessary to contract the infection several times. While this may generate a natural protection, it does not confer to the immune system the ability to stop the multiplication of the pathogen (sterile immunity) nor the ability to remember the identity of the pathogen for a very long time. This is due to the ability of the parasite to be able of avoiding to be destroyed by the immune system of the body, and to reduce its ability to fight the infection and the development of immunity.

However, although this can well explaine the reason why those children can be the most at risk, previous studies have shown that children over 5 years old have became the most affected age group (3). This shift, as highlighted in some studies suggests higher exposure of this group due to the focus of the new strategies on those under five years (5). Also, possibly because they were not exposed and uninfected at a young age, they may have delay in acquiring immunity (6). However, to use a standardized measure for endemicity, *Plasmodium falciparum* parasite rate (PfPR) for children in the range of 2 - 10 years old is considered. Their prevalence can also exceed 46% (4,7,8). In fact, three reasons have driven this choice when considering malaria intensity transmission (9): (i) the constant PfPR in this age group, (ii) the PfPR may be least influenced by drug treatment and not affected by immunity and it's reflecting stationary state predicted from mathematical model in terms of clinical malaria in this group. (iii) they are aligned with the historical approaches of endemicity statement. It is well known that PfPR is the best indicator of the intensity of the transmission of malaria infection (9). Additionally, Its relationship with clinical incidence has been defined to help in the estimation of clinical burden of *P. falciparum* when PfPR is available (10).

In Gabon, after the improvement of the antimalarial policy, through interventions including impregnated bed nets, and artemisine combination therapy (ACT), some studies observed a decline in the prevalence of malaria since 2000. A cross-sectional survey conducted in Libreville, Port-gentil, Melen and Oyem, based on health facilities, showed that between 2005 and 2008, the prevalence fell from 31% to 18% for children under 11 years of age (3). Another study conducted in Libreville in 2009, using a large number of febrile children, showed a decrease of malaria between 2000 and 2008. After 2008, several studies conducted separately suggested either that there was no evidence of decline of the prevalence, up to 2012 or 2013, or some suggested an increase in malaria prevalence to 40% in children (3,7,8). However, for studies conducted in Gabon, the conclusion has remained in some way ambiguous if there is truly a decrease or an increase of the prevalence after certain time in the whole country, may be because of the methodology used, the availability or the quality of the data.

Focusing now on factors related to this prevalence, among studies estimating the prevalence in Gabon, few of them have investigated on the factors related to the infection or driving the change in the prevalence, especially ecological factors (5,8,11). They often use the level of education, the type of house, the open water body, the source of drinking water as the factors predicting the infection, and most of them are conducted on a local scale.

This study described the spatial distribution of malaria prevalence over time for children between 2-10 years based on cluster of households. This attempted to understand the spatio-temporal distribution of malaria prevalence and how some ecological variables were influencing this prevalence, while taking into account the spatial effect. This was done using a national representative survey covering all the provinces based on the cluster of households. This can provide to public health program managers and policy makers information on areas where there is need of improving malaria interventions all over the country and how to better improve the strategies to be applied.

1.3 Literature review

1.3.1 Introduction

Since many years it has been established that malaria is a great issue for public health in the world. It is estimated that the global burden is over 40 million disability adjusted life years (DALYs) (12). Although research pertaining to malaria has made progress, it still poses a real problem involving not only on health but also on the economy, and on social development. In terms of socio-economic status, malaria treatment can have a great impact on the health budget of the households, therefore threatening the lives of those living in poverty.

Hence, understanding factors related to malaria distribution may contribute to reduce the impact on the economy, and improve the life quality of the population. Among the factors that can influence the distribution of the prevalence, the variation of certain environmental factors such as temperature, rainfall, precipitation, altitude, crowding index for household implies that from country level to area level inside each country, the way malaria is affecting lives is quite different over time. In fact, these factors can create spatio-temporal variation, so that we can observe some regions to have recorded a high variation in the occurrence of malaria, while in some other areas, low variations have been observed. However close regions have similar values changing over time (13).

The most important factors, among ecological factors affecting malaria are temperature and rainfall (14). In Gabon, temperature is almost constant throughout the country but varies slightly in the central and north parts annually. On average, the highest temperature is around 30°C (along coast and Libreville) and almost constant around 25°C and 27°C, with the mean annual temperature being around 25°C or 28°C (15,16). The temperature decreases in the south with a long dry season (16,17). It can also drop to 23°C in the center and the south. On average, the rainfall in Gabon falls between 1500 and 3050 mm/year, with 2600 mm/year in Libreville (16). Among the other ecological factors there are wet days, and aridity. Wet days are known to change widely over the year with the high value reaching up to 23 days in a month usually in the capital city (18). The aridity, defined as the ratio between rainfall and evapotranspiration, is high in Gabon and therefore the country is considered very humid.

1.3.2 Demographic and Health Survey

The data used in this work were obtained from the Demographic and Health Survey (DHS) which is funded by the United State Agency for International Development (USAID), and implemented by Macro International Incorporation. It is a program aimed to provide reliable data at national and regional level for many health outcomes and factors widely used in government, academia and private sector (19). Data are publicly available but the access is only given for research purpose upon request. Many analyses have been conducted in Africa and all over the world using these data (20,21).

Some of the DHS survey include also malaria indicator for children under or above five years old as well as some ecological factors related to the cluster of households. Although individuals' data are available, sometimes DHS program does not release the infection status of each individual, hence it becomes difficult to link status of infection with some variables such as sex, or education level. Some other variables can be extracted and summarized so that the analysis can be done on that level.

1.3.3 Malaria transmission and risk factors for infection

Malaria is a vector-born disease. Its life cycle is complex (figure below) (22). The parasite *Plasmodium* is the one that causes malaria in humans. The life cycle of *P. falciparum*, involving human or *Anopheles* as host or intermediate host, is divided into three (3) parts or stages, that is, sporogonic cycle, erythrocytic cycle and exo-erythrocyte cycle.

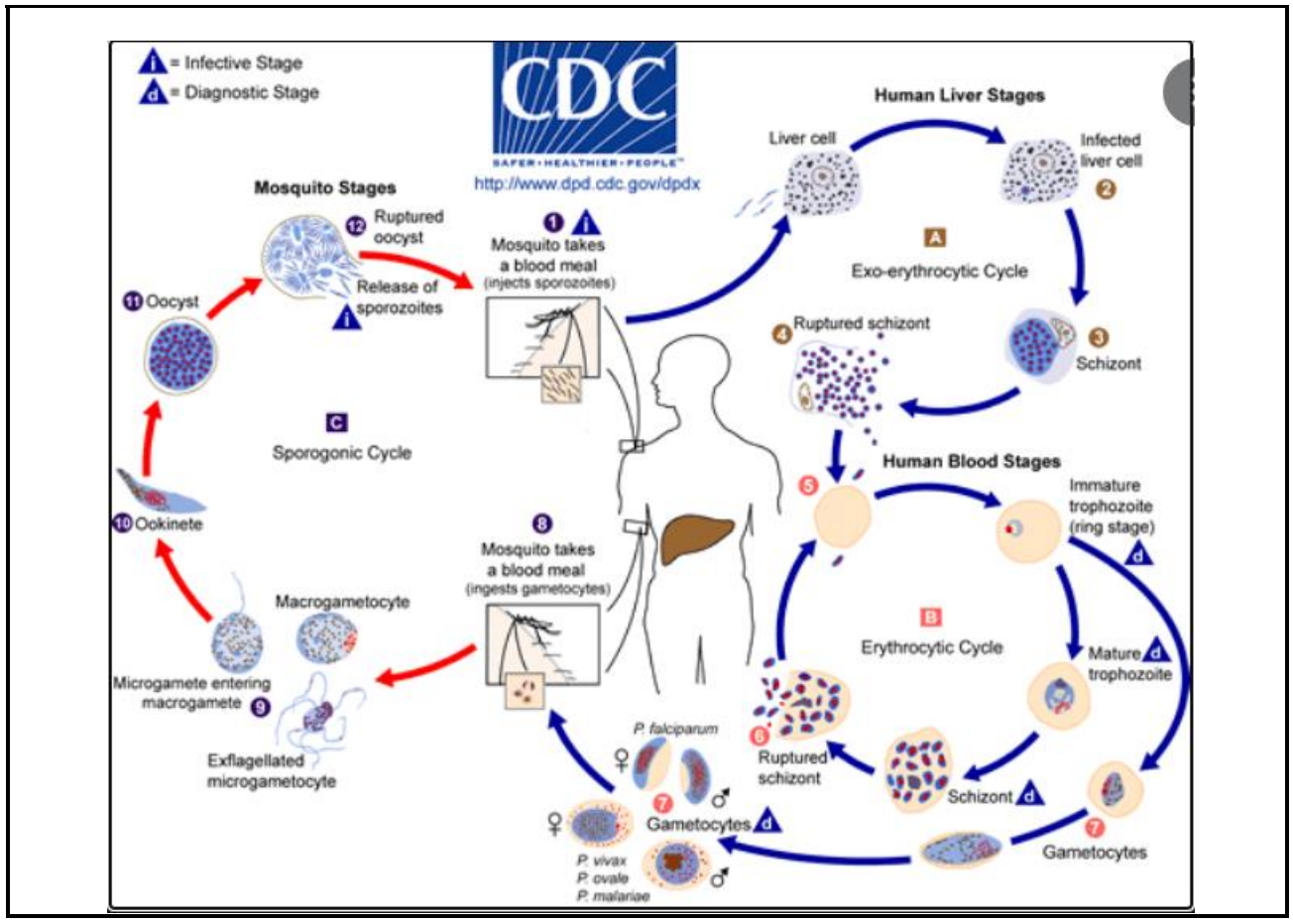


Figure 1.1*:Life cycle of* Plasmodium *– source:*
https://upload.wikimedia.org/wikipedia/commons/4/47/Malaria_LifeCycle%28French_version%29.GIF

In malaria transmission, factors are defined to act differently on: **(1)** the host-vector, **(2)** the parasite and (**3)** the vector. In fact, the temperature is very important for both vector and parasite, with the optimal temperature being between 25 and 27°C (23,24). For the parasite, it affects the sporogonic duration which is influenced by environment (1). Below 16°C, parasites stop growing and the cycle in the mosquito is interrupted. Consequently, the transmission cannot take place (25). Above 28°C or 32°C, there is also a delay in the erythrocyte cycle. In the vector, it affects the population dynamics or the abundance of adult which is an important factor in transmission. Temperature and water availability play a key role for breeding of larval of the vector, as it can help in the duration of development for larvae. When the temperature is above 28°C or 30°C the growth of *Anopheles mosquito* and its distribution can decrease very fast, hence implies a reduction in the risk of being bitten (24,25). Usually for the analysis of the factors related to the prevalence or the incidence of malaria, the temperature of the night land surface and day land surface, mean temperature, minimal and maximal temperature are widely used. Among other factors, there are arid conditions, rainfall, humidity. In fact, arid conditions also act on the development of adult *Anopheles*. When the water surface is limited, it

involves a reduction in the suitable sites and the survival for all stages of development. Rainfall and humidity are important for the living condition, egg laying and multiplication, thereby contributing to the population increase. In fact, rainfall increases the humidity, hence extending adult *Anopheles mosquitoes'* life and helps in their behavior. But it can also wash-off the breeding sites and reduces the ability to detect the relationship with malaria prevalence. According to some studies, the average rainfall around which mosquitoes can be found is between 1100 mm and 7400 mm, and when humidity above 60%. For the vector-host, the land coverage can play an important role as it can locate mosquito abundance (26).

To help reduce the intensity of the transmission, the WHO recommended two primary vector control measures, that is, indoor residual spraying and ITNs. ITNs are the most effective, even in the area where there is high transmission. It can help in reducing malaria mortality for under-five years of age by 20% or more (27–29). In Gabon, the Ministry of Health greatly involved many control programs, and despite the well-known knowledge on the importance of ITNs, the ITNs coverage seems decreasing and low (30).

1.3.4 Bayesian spatio-temporal modeling and malaria research

In the last decades, spatial and spatio-temporal analysis has been widely used to understand the dynamic of many diseases such as malaria. It can help to produce risk maps to locate the most affected area. Moreover, it helps in predicting areas where cases are likely to occur. It can also help to quantify the contribution of the space in the estimation of either the prevalence or the incidence over time, and to evaluate variables such as the climatic variables and some environmental variables that are involved in the spatio-temporal variations of the outcome of interest (31,32).

To do this, many methods were developed and applied in the last decades. Some of them used likelihood-based estimation and some others used Bayesian-based estimation (33,34). However, Bayesian methods are the most used due to their flexibility in the estimation or in providing risk maps (35,36). For example, to obtain an estimate of the excess probability when evaluating the decrease of the incidence, a study in Rwanda used Bayesian spatio-temporal model. They found an increase of the incidence between 2013 and 2018, with 47% of sectors have failed in the reduction of malaria. But they did not include any covariate to assess if this effect can be influenced if some covariates were introduced. In Mozambique, study investigating on the geospatial analysis of malaria risk, used also Bayesian method. They found high prevalence of malaria all over the country, and significant relationship between malaria prevalence and age, place of residence, education level of mother, and sleeping under bednet. However, they did not assess in the model the effect of the climatic variables.

In Burundi, the relationship between meteorological variables and malaria incidence was evaluated using Bayesian modelling. They found that given a month, the incidence of malaria is positively associated with the minimum temperature of the month before. In the same configuration, maximum temperature and rainfall had a negative effect on the incidence. However, they did not adjust with other variables like socio-economic factors or demographic factors to see if the same effect remains unchanged. In contrast to the above studies, study conducted in Uganda investigated on the effect of interventions with variables such as the use of ITNs, ACT, proportion of malaria seeking behaviour and wealth score, and the climatic condition with variables such as rainfall, night land surface, day land surface, altitude, land cover, distance to water bodies, and

normalized difference vegetation index, on the spatial and temporal distribution of malaria including age. They found a fast decrease of malaria in children under 5 compared to children above 5 years old of ages. The climatic variables were associated with the decrease of the incidence from 2013 to 2017. The interventions have a positive effect on malaria prevalence.

In the same way, in a study conducted in Tanzania, in order to describe the relationship of environmental/climatic, socio-economic and intervention factors with parasitemia risk, using almost the same variables used in the above study plus type of residence, Bayesian geostatistical modeling was used. They found a prevalence ranging from 0.3% to 19%. In contrats to the results in Uganda, they did not found any evidence of the relationship of environmental variables and malaria. The same framework was used in a study conducted in Angola investigating on the modeling of the spatio-temporal distribution. They provided a risk map of malaria, and found that the demographic and the socio-economic variables were related to malaria prevalence negatively and positively, respectively. In contrats to the finding in Uganda or Tanzania for the effect of intervention, using Bayesian modelling, a study in Nigeria, using almost the same variables but adjusted with population density, found that the intervention was not significant in malaria risk.

1.4 Problem statement

Often when collecting or analyzing data, the assumption made fundamentally is the independence of the observations, or the lack of influence of observations at different locations. However, sometimes this assumption may not hold because of the involvement of either the structure of the place where the data are collected, or some variables like environmental/ecological factors.

In fact, **(i)** the environmental exposure can also be defined by ecological factors such as temperature, precipitation, annual rainfall or aridity. It is known that the climatic conditions and environment are important in the survival of Anopheles mosquitoes and the development of their eggs. These specific conditions of climate and environment may differ by geographic area (37). Therefore, the distribution of Anopheles densities can be heterogeneous in different locations and similar in some other locations. **(ii)** Due to socio-economic factors or other variables, the localization of the population can be very different in terms of environmental exposures like stagnant pools of water, swamps or garbage heaps and construction type and quality of houses. Some people are probably more exposed than others, as for example, those with high incomes may have high chance of having a good environment for education or for living. They tend to be close to each other, creating similarity in the environment, therefore imply a dependency in the value of the outcome in this area which will affect the distribution of the health outcomes. Hence, considering these facts, to understand the distribution of the prevalence of malaria, it will be inadequate to ignore this dependency induced by either ecological variables or socio-economic factors.

Notwithstanding this fact, few studies have looked at malaria ecological factors in Gabon and the importance of spatial component in malaria prevalence throughout the country. There is a need to understand the spatial patterns in malaria prevalence over time, in order to understand the progress made in some regions and which regions are still lagging behind and which ecological factors are strongly associated with malaria in Gabon.

Thus, in this work, the spatial method in order to remove variability due to the small area was used. In fact, when conducting spatial analysis in the small-area: 1) the population at risk is reduced and there is a small number of events leading to unstable risk estimates, 2) local variation in data quality implying biases in the findings, and 3) difficulty in interpreting the disease maps (38). Because of the random variation, another source of variability that can be handled using the Bayesian approach to analysis, is introduced.

1.5 Justification

Looking back over the past few years, although there has been some progress in the intervention like the development of malaria vaccine, many strategies have failed in reducing malaria incidence or prevalence. Hence, it is always important to better understand malaria and its risk factors using different approaches. The spatial distribution possibly created either by the Anopheles' distribution or by socio-economic factors can have an impact on malaria control and outcomes, and potentially increase the transmission. This creates the need for spatial component and environmental or ecological factors to be investigated as both spatial context and environment define social ecological models to understand disease prevalence strategies and help in formulation of research hypothesis for future research and appropriate interventions. Malaria burden is almost carried by children. Hence estimating the clinical burden and understand factors influencing the transmission or the prevalence will help in, which strategies to use, and how to plan it, in order to decrease the risk of morbidity and mortality.

1.6 Research question

What is the spatial-temporal distribution of malaria prevalence among children from 2 - 10 years of age in Gabon from 2000 to 2015 and the ecological risk factors influencing this distribution?

1.7 Aims

The aims of this study are to describe the spatial-temporal distribution of malaria prevalence in children between 2 and 10 years of age as well as to assess the effect of ecological risk factors on this distribution in Gabon from 2000 to 2015.

1.8 Objectives

- To determine the prevalence of malaria in Gabon between 2000 to 2015 among children between 2 - 10 years of age by province and type of residence.

- To determine if the value of the prevalence in a cluster is influenced by the nearby clusters, and identify areas where high/low malaria prevalence cluster together.

- To assess factors associated with the malaria prevalence in Gabon while accounting for spatial effects.

CHAPTER 2 METHODS AND MATERIALS

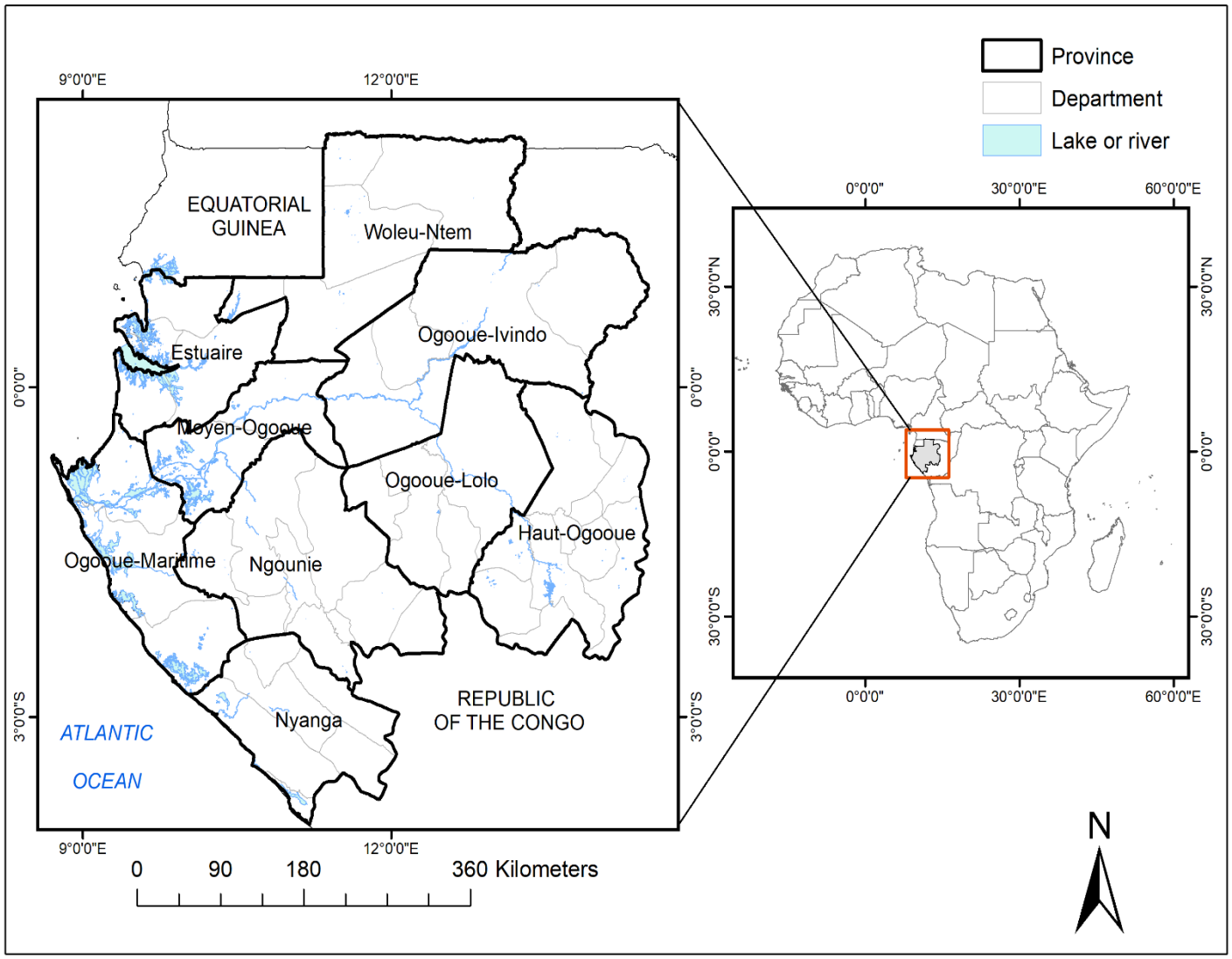
2.1 Study design

This study used secondary data from the DHS program. The data consisted of a series of cross-sectional surveys carried out in 4 time points (2000, 2005, 2010 and 2015) to provide representative estimate at the national or regional level (39). Four time points were used as Gabon does not have recent data. There were about 30 clusters in each province which were further separated into rural and urban type of residence. The malaria measure in this data were the prevalence of malaria *P. falciparum* in children between 2-10 years of age per cluster as measured by rapid diagnostic tests. Data between 2000 and 2015 were used as there were no other source of data after 2015. This will give better insights on the dynamic of malaria in Gabon for futur research

2.2 Study site

The study was conducted in Gabon in both urban and rural areas using Gabon DHS data. It is situated in West coast of Central Africa bordering the Atlantic Ocean and crossed by the equator. The rainy season extends from October to May with a short dry season from June to August (40). The area is almost forested and generally hot and very humid throughout the year. Gabon has a total of 9 provinces with Libreville as its capital city. The whole area is subdivided in 49 departments. with an estimated total population of 2.1 million (2019). Almost all provinces are endemic for *P. falciparum*.

Map of Gabon



232 **Figure 2**: Map of the Republic of Gabon

233

234 ## 2.3 Study population

235 The population in this study was constituted of children between 2 to 10 years of age living in Gabon in the years 2000, 2005,
236 2010 and 2015 as presented in the DHS survey.

237

238 ## 2.4 Data sources and sampling

239 Information on the sampling frame was obtained from the 2000 report of a health survey of the Gabonese DGS (41), namely
240 the number of households in each cluster and number of clusters in each residence and resident area. This sampling frame
241 was used for the 4 DHS surveys between 2000 and 2015. All surveys used this sampling frame and there was no resampling
242 between surveys. Data from these DHS surveys, funded by USAID, are made available to researchers upon request.

Information about the date when the four DHS surveys have started and ended was not available. To conduct the survey, DHS used stratified multi-stage sampling (42). The country was divided into region consisting of provinces and two major towns (Libreville and Port-Gentil). Within each region the primary sampling units were clusters of households stratified by area (urban or rural), totaling 20 strata. Then, a random number of households was selected within each chosen cluster, their sampling weight and their centroid calculated. The number of households was in the range of 100 – 300 for each cluster. For all the surveys, 336 clusters were sampled with 5 to 30 households per cluster. Spatial data was available for each cluster centroid.

2.4.1 Outcome

The main outcome variable was malaria prevalence (PfPR) which is a continuous variable, named 'Malaria Prevalence YEAR' in the DHS database. The PfPR is the proportion of the population in a cluster found carrying asexual blood-stage parasites (43). While the exact method of determining parasitemia was not available from the data, a report of a UNAID DHS surveys carried out in 29 African countries used both rapid diagnostic test and malaria microcopy (64). It is likely that at least RDT was used in the Gabon survey, given its ease of use compared to microscopy.

2.4.2 Covariates

Data of ecological variables and the malaria outcome were downloaded as csv file and contained a specific ID for each cluster. Hence these variables can therefore be linked to the DHS data set. They were selected based on the investigations of several studies showing the possible relation with either malaria prevalence or incidence. Details on the variable dictionary can be found in the Geospatial dataset manual (42). They are summarized in table below.

264 ***Table 1**: Description of the variables used*

Covariate	Unit	Description
Enhanced Vegetation Index (EVI)	0-10000	The average vegetation index value within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS cluster at the time of measurement (year)
Proximity to waters (Coast/Large Lakes)	Meters	Straight-line distance to the nearest major water body
Population count	Number of people	The count of individuals living within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster at the time of measurement (year)
Precipitation	Millimeters (mm)	The average precipitation measured within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster in a given year
Travel time to nearest settlement >50,000 inhabitants	Minutes	The average time (min) required to reach a high-density urban center from the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS cluster location, based on year 2015 infrastructure data
Minimum temperature	Degrees Celsius (°C)	The average annual maximum temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS cluster location
Maximum temperature	Degrees Celsius (°C)	The average annual minimum temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS cluster location
Wet days	Number of days (days/month/year)	The average number of days receiving rainfall within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS cluster location per month for a year
Urban–rural settlement	Categorical	This is urban–rural population classification of the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location
Rainfall	Millimeters per year (mm/year)	The average number of drought episodes (categorized between 1 (low) and 10 (high)) for the areas within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location
Aridity	0 (most arid) and 300 (most wet)	The average yearly precipitation divided by average yearly potential evapotranspiration, an aridity index defined by the United Nations Environmental Program (UNEP)
Day land surface temperature	Degrees Celsius (°C)	The mean annual daytime land surface temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location
Night land surface temperature	Degrees Celsius (°C)	The average nighttime land surface temperature within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location.
Land surface temperature	Degrees Celsius (°C)	estimates the temperature of the land surface (skin temperature), which is detected by satellites by looking through the atmosphere to the ground
ITN coverage	Number of people	The average number of people within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location who slept under an insecticide treated net the night before they were surveyed
Growing season length	classes between 1 and 16	The length of the growing season in days (reported in one of 16 categories) for the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS survey cluster location

265

266 **2.5 Data management**

267 Geospatial data was downloaded as a shapefile format. The covariates associated with these households were provided in a
268 csv file. These were imported into R software in order to extract data from the shapefile and to combine with the covariates

data. A private Gitlab repository was set up for raw and converted data as well as for the analysis scripts. The data available had been anonymized and delocalized, to avoid making an exact identification of household location. Delocalization was performed by randomly placing the centroid within a buffer zone of 2 km or 10 km around the true centroid for urban and rural areas, respectively. Data manipulation including data cleaning, combining data, coding, re-coding, selecting variables, and handling NA (Not Available) observations were performed using the tidyverse package. The final analysis dataset contained data from the DHS final report, from the individual level data and the covariates data containing also the outcome. The analysis from descriptive statistics to multi-level modelling were performed using R software (version 4.0.5). The analysis used the INLA package for Bayesian computation and the SUMMER package for analyzing the surveys. ARCGIS software version 10.7.1 was used for the mapping of hotspots and cold spots.

2.6 Statistical analysis

(A) Bayesian inference with INLA and different approches of the spatio-temporal models

2.6.1 Bayesian inference and computation

In the Bayesian framework, if y is the observed data, the given prevalence of malaria here in the data, and θ denotes all the parameters, or the mean prevalence, then the estimation of θ , the average prevalence, its standard deviation, will be based on the information from the data and the previous knowledge on the parameter. Bayesian inference helps to provide a distribution to the parameters given the data. This distribution is called posterior distribution ($p(\theta|y)$). It is obtained as approximatively the product of the distribution given to this parameter before seeing the data called prior ($p(\theta)$), and the likelihood (product of the individual probability of $y_i|\theta$)) using Bayes theorem, and can be briefly expressed as:

$$p(\theta|y) \propto p(y|\theta) \times p(\theta) \quad (1)$$

The computation of the marginal posterior distribution is done via Markov Chain Monte Carlo (MCMC) method which can be expensive in terms of computation time, due to the complexity of the algorithm. To deal with this, in the last decades thanks to the work of Rue et al (44), a new approach has been suggested called INLA. This approach is well implemented in R software, and widely used in many studies such as a study to identify risk factors of malaria prevalence in Mozambique (45), and to characterize environmental variables in the transmission intensity (46). It is based on the latent Gaussian model, the Gaussian Markov Random Field model (GMRF) and the Laplace approximation for approximating the normal distribution. The idea behind this approach is to reduce the computation time by using a sparse matrix. To obtain such a matrix, one assumption is added: the conditional dependence of the parameters. The results compared with MCMC computation are closely related in many cases (44,47).

The model can be briefly explained as follows: by denoting that $x = (\omega, \alpha, \beta)$ is the latent fields, and $\phi = (\sigma_e^2, k, \sigma^2)$ is the hyper-parameters, to find the marginal posterior, $p(x_k|y)$, an approximation of $p(\phi|y)$ and $p(x_k|\phi, y)$ is found, as one can write:

$$p(x_k|y) = \int p(x_i|\phi, y)p(\phi|y)d\phi \quad (2)$$

Then the Laplace approximation and the nested integration are used to obtain an approximation $\tilde{p}(\phi|y)$ of $p(\phi|y)$, and $\tilde{p}(x_i|\phi, y)$ of $p(x_i|\phi, y)$. Lastly, a numerical integration is used to obtain the approximation $\tilde{p}(x_k|y)$ of $p(x_k|y)$ which takes the following form (47,48):

$$\tilde{p}(x_k|y) = \sum_k \tilde{p}\,(x_i|\phi_k, y)\tilde{p}(\phi_k|y)\Delta_k \quad (3)$$

2.6.2 Structured and unstructured spatial effects

In the spatial model, two random effects can be introduced: spatial unstructured and spatial structured effects. These account for region and neighborhood effects, and these are known as BYM or BYM2 model or convolution model. Additionally, as in many models, in the spatio-temporal model, each variable or random effect can vary by time plus an interaction effect for counting simultaneously the variation by time and region. Many studies have used such a model for the estimation of the incidence or the prevalence, and for evaluating the risk factors associated with malaria transmission. Using this model, a study conducted in Rwanda estimated the excess probability of the incidence within small areas (49). In another study conducted in Senegal, it was used to identify malaria hot spots varying over time (50).

These random effects help to deal with the small area estimation using a Bayesian framework. In fact, these remove the variation introduced when estimating the outcome of interest, and hence help to obtain an estimate which is more reliable. To obtain an unbiased estimate, it is suggested that to obtain the space-time estimate of the prevalence through the Horvitz estimation conditional auto-regressive model (CAR) by borrowing strength from neighbors and using the logit of the obtained prevalence. In the original model of Fay-Herriot, only one term was introduced. This helps to obtain precision of the transformed weighted estimates. It is done using the logit function, to ensure it will fall between 0 and 1 (51). Later, Banerji introduced a spatial component (52,53). Although the estimation can be on the areal-level or the unit-level, when the unit-level model is used the spatial effect is always on the area level based on large administrative boundaries such as provinces.

2.6.3 Estimation in areal and geo-statistical data

In the spatial analysis two types of data can be used: the areal data and the geo-statistical data. In the areal data, the unit is a polygon representing an area like a province or a district. If the number of regions is too small, the estimation or the assessment of the effect of the covariates becomes unreliable. When the spatial data are on a small level such as the districts or the

aggregations of the households, or individuals the analysis can be conducted on that level (54). Particularly, in the case of health outcome, the units of the analysis can be the aggregation of, for example, households. In this case, each location can be considered as a partial realization of a stochastic process.

In contrast to areal data where the neighbors separated by boundaries are used, for geo-statistical data, the spatial effect is estimated using the Euclidean distance alone. This is refered to as an isotropic process. In this case the idea is to measure the spatial dependence by using only the distance, as to say, nearby points have the same value. This can be modeled using a covariance function. The most widely used covariance function in the literature that models well in many situations is the Matérn covariance function (55). Using this isotropic process, the variance-covariance matrix, carrying the spatial correlation, is not easily computable, because it is very large. This issue is known as the 'big n problem'. A suggested method to deal with this issue is the use of a stochastic partial differential equation (SPDE). This approach allows us to find a differential operator on a SPDE which has the solution described by such a covariance function. This solution, a Gaussian field (GF), is then approximated using triangulation or meshes, hence the approximation is not anymore a GF but a GMRF and therefore resulting in a sparse precision matrix (48). In practice, this can be implemented using the integrated nested Laplace approximation and the results are very similar to those obtained by MCMC method. (56).

2.6.4 Geo-additive model and spatial econometrics in INLA

Often, the linearity of the covariates with the outcome is assumed, while in fact some variables such as the environmental variables are known to have a non-linear relationship with the outcome. Some authors have extended the conventional Generalized linear model (GLM) by using the GAM, in order to allow the shape of the relationship to be closely related to the observed data. In such a GAM, the relationship between the outcome and the covariates is no longer conveyed by one linear function for all the covariates, but assumes that the relation is described for each continuous variable by a function which may be linear or not (34). Many studies have used this model. For example, a study in Burundi investigated the association of malaria incidence with climatic variables like temperature and rainfall (57), and another study in Parà (Brazil) investigated the effect of climate and environmental variables for children under five years of age (35). In this approach, the relation between the outcome and the covariates are the sum of different functions, linear and non-linear (58,59).

The methods mentioned above do not allow to quantify clearly the spillover effect, that is, the way the change of a covariate in a region can affect the prevalence in the same region or in another region. Such models are widely used in econometrics. These include the SLM, where the lag is on the dependent variable only, the SEM, where the lag is on the error, the SDM where the lag is only on the covariates, and the SDEM with the lag on both covariates and the error (58,59). Their implementation is possible within INLA by conditioning on a parameter to obtain the GMRF (56,60). In the case of malaria,

they were used in many studies: association between incidence of malaria and environmental variables in Nigeria (14), and risk factors associated with prevalence of malaria in Ethiopia (61).

While these models are widely used in econometrics, they are not directly implemented in INLA. To be INLA-fitable, it is suggested to condition on the spatial autocorrelation parameter. Briefly, instead of the likelihood $p(y|\phi)$, we consider $p(y|\phi, \rho_0)$ to obtain the conditioned $p(x_i|y, \rho_0)$, with ρ_0 a fixed value. Then the marginal distribution of ρ is given by:

$$p(\rho|y) \propto p(y|\rho)p(\rho) \tag{4}$$

Where $p(\rho)$ is the prior for ρ

Consequently, the marginal distribution $p(x_i|y)$ is obtained as:

$$p(x_i|y) \simeq \sum_{k=1}^{m} p\,(x_k|\rho_k, y)\lambda_k \Delta_k, \qquad \lambda_k = \frac{p(y|\rho_k)p(\rho_k)}{\sum_{i=1}^{m} p\,(y|\rho_i)p(\rho_i)\Delta_i} \tag{5}$$

The spatial autocorrelation parameter is bound by the inverse of minimum and maximum eigenvalues of the adjancy matrix W. When W is row standardized, ρ is between 0 and 1.

2.6.5 Geographically weighted regression and spatial varying coefficient

Another method used in the context of spatial analysis is the GWR. It is considered as an exploratory tool for the spatial analysis. In fact, linear regression uses the assumption of a stationary relationship and provides a global estimate of the coefficients in the regression. However, this assumption is often not met. To solve this problem, the GWR method allows to break down this assumption by providing an estimate of the coefficient for each location, using neighborhood windows chosen with an adaptive or a fixed bandwidth method. The number of regression equations equals the number of locations in the data. The adaptive bandwidth method allows to choose the proportion of neighbors to use in each regression, in contrast to the fixed bandwidth. In the GWR, sometimes multicollinearity in global and local level can be an issue for the correct interpretation of the results. In fact, local multicollinearity can appear even after the global multicollinearity was assessed and avoided. Many studies have highlighted this fact including one that was on characterizing the association between malaria cases and environmental variables conducted in South Sumatra (Indonesia) (23). An alternative to this case of local multicollinearity can be a Bayesian model called here spatial varying coefficient implemented in INLA (62).

(B) Estimation and related equations in spatio-temporal analysis

2.6.6 Estimation of the prevalence

To estimate the prevalence, the weight on the cluster level was obtained as suggested by DHS and detailed below (41).

(1) Approximation of the Weight in cluster level (level-2 weight)

The sampling weight used in the DHS data was calculated for each stage of the sampling. The stage 1 was the cluster of households, and the stage 2 was the household. But the final weight provided in the data was the product of the weight on these two levels. This weight is not helpful when conducting the multilevel model. Indeed, methods such as pseudo-likelihood require using the weight in both levels. However, the DHS program has not provided this information to ensure confidentiality. To help in the approximation, DHS program advises to use the available information in the data and the DHS final report, to obtain an estimate of the weight in the required level (41). The formula suggested to approximate the weight, including the correction for non-response between the clusters was as follows:

$$W_{hi} = \frac{A_h f_{hi}^{\alpha}}{a_h^c} \quad (6)$$

With the variation factor $f^{\alpha} = \frac{d_{hi}^{HH}}{\frac{A_h M b_h}{a_h^c S_h}}$, $\alpha \in [0,1]$ and $d_{hi}^{HH} = \frac{M}{m^c}$ the denormalized type of the household weight provided in the DHS data as HV005 variable. h is the stratum from 1 to 20, which represents the number of strata by rural and urban area for all the regions used as stratification, and i represents the cluster. A_h is the total number of clusters from the census in the stratum h, a_h^c is the number of clusters interviewed in stratum h, Mb_h the average number of households by sampling strata for each cluster from the census data found in Table A.3 of the Gabon DHS final report, m^c the number of households completed, and M is the number of total households according to the census frame. For this work, α was set to 0.5. The sensitivity analysis could be performed for the other values (41).

(2) Space-time estimation of the prevalence

(a) Areal model

To estimate the prevalence by province using clusters of continuous points, y_j was defined to be the cluster's outcome and w_j, if available, the weight associated with the cluster j. The estimator for each region i at time t using the Horvitz and Thompson formula is (51):

$$p_{it} = \frac{\sum_{k=1}^{n_{it}} w_k\, y_k}{\sum_{k=1}^{n_{it}} w_k} \; i \in I \tag{7}$$

, I is the partition of the region, and n_{it} is the number of clusters at time t in the area i. Using the logit of the estimated prevalence, $logit(p_{it})$, and its variance from linearization V_{it}^{L}, the true prevalence was estimated using the following space-time smoothing model:

$$logit(p_{it})|\lambda_{it} \sim N\left(\lambda_{it}, V_{it}^{L}\right) \tag{8}$$

,

$$\lambda_{it} = \beta + \beta_{it}(rural) + (\lambda + \gamma_{it})(urban) + \alpha_{it} + \epsilon_{it} + S_i + e_i + \delta_{it} \tag{9}$$

β is the intercept for a cluster identified as in rural area, $\beta + \lambda$ intercept for urban area. β_{it} and γ_{it} are temporal main time-varying effects. They are introduced to account for the stratification effect as in the sample design. α_{it} is the structured time trends, ϵ_{it} is the temporal term (iid), S_i is the structured spatial trends, e_i the unstructured spatial terms and δ_{it} the interaction between space and time. In this model, the structured and the unstructured spatial effects were considered, often referred to as BYM model. This model allowed to combine the effect of the region and the effect of neighborhood, both modeled using either the intrinsic conditional auto-regressive random effects for S, and $i.i.d$ errors following a Gaussian distribution with mean 0 and a precision τ_e for the unstructured spatial effect e. In order to consider the dependency between the two components, instead of the BYM model, its reparameterization known as the BYM2 was used. For the spatio-time interaction effect, four types of models of interaction were compared (66):**Type 1**, only the interaction of the unstructured term, that is, specific area time trends interact with random effect of the region without being influenced by neighbors , **Type2**, the temporal trends are not spatially correlated, that is, the temporal trend of a specific area is not influenced by its neighboor, **Type 3**, the spatial dependence or spatial pattern is not changing according to time, and **Type 4** : spatial patterns correlated in time or temporal trends spatially correlated specific to each cluster. For temporal model, a random walk of order 1 or 2 was used. The spatial model chosen was the ICAR model. The details are provided below. For time effect α, the following was used:

- $\alpha_{it} \sim N(\alpha_{t-1}, \sigma^2)$ (random walk order 1) or
- $N(2\alpha_{t-1} + \alpha_{t-2}, \sigma^2)$ (random walk order 2)

The auto-regressive model AR(1) is defined as:

$$\alpha_t = \phi\alpha_{t-1} + \omega_t \quad \omega_t \sim N\left(0, k^{(-1)}\right), t \geq 2 \text{ and} \tag{10}$$

$\alpha_1 \sim N(0, \tau_1)$, $\tau_1 = k(1 - \phi^2)$,

- The spatial structure of δ_{it} is modeled by the ICAR model:

$$\delta_{it}|\delta_{jt}, \tau_\delta \sim N\left(\frac{1}{n_i}\sum_{i\in n_{\partial i}} \delta_{it}, \frac{t_\delta}{n_i}\right),\ i \neq j \quad (11)$$

Where ∂i= the set of neighbors for the area i, $n_{\partial i}$ number of these neighbors, and $b\delta_{it}$ the mean of the neighbors at time t.

When a spatio-temporal model is used without an interaction, it is assumed that each cluster has no specific temporal evolution. This assumption does not always hold. Hence, introducing the interaction term allowed to relax this assumption, hence to have a specific temporal evolution for each cluster.

(b) Unit-level model

In the above estimation, since the unit of the analysis was the province, the area-level model was used. In fact, in this model, admin 1, which corresponds to the province, was firstly used as a random variable. Then in th unit-level model, the cluster was used as the new geographic unit of the analysis. Hence, the estimation in admin 1 level was obtained by moving from the cluster level to the province level. The estimation by type of residence was also provided. While the unit was the cluster, the spatial effect was from the provinces. Therefore, y_c was defined to be the cluster's outcome, and the following space-smoothing model was used:

$$logit(p_{ct})|\lambda_c \sim N(\lambda_c, V_c^u) \quad (12)$$

β is the intercept for rural clusters, $\beta + \lambda$ intercept for rural clusters, and p_{ct} the prevalence associated with cluster c at time t.

(3) Penalized Complexity (PC) Priors model

In the estimation of the prevalence explained above, the penalized complexity priors (PC) was used (63). This prior is less sensitive to the change of parameters. The idea of using this prior is based on the difficulty of setting a prior for a hierarchical model as noted in the study of Simpson (63). It stands on the idea of a weakly informative prior and the use of the base model, which is a model controlled by a flexible parameter ξ. This parameter can be the variance or a function of the variance like the precision, but set to the null value or 0. It is based on the principle of Occam's razor. It is better to prefer a simpler model than a complex model, unless there is sufficient support for the second. In fact, this prior was first constructed to be a function depending on the distance. This distance is used to quantify the complexity between the flexible model and the base model,

by using the square root of the Kullback-Leibler divergence (KLD), so that the prior does not lead to much loss of information by putting significant mass where the flexible model has been replaced by the base model through constant-rate penalization. Then, based on penalizing the deviation from the base model, parameterized with distance d, the prior on the flexible parameter takes the following form:

$$\pi(\xi) = \pi_d\big(d(\xi)\big)\left|\frac{\partial d(\xi)}{\partial \xi}\right| = \lambda e^{-\lambda d(\xi)}\left|\frac{\partial d(\xi)}{\partial \xi}\right| \tag{13}$$

Where λ, defining the magnitude of the penalty or the deviation from the base, is selected when controlling the prior mass in the tail using the following expression: $P(Q(\xi) > U) = \alpha$, where U is defined to be the upper bound. This bound is the tail event for the interpretable parameter $Q(\xi)$, which is the form of the parameter in the distribution used, for example, the standard deviation $\sqrt{variance}$. The parameter α is its probability or its mass. These parameters represent the information of the PC prior specified by the users according to prior information on the parameter of interest.

- BYM2 model and PC priors

In the case of our model, which uses smoothing components, a reparameterization of the structured and unstructured component was used to account for the dependency between the two components S_i and e_i. Therefore, instead of $S_i + e_i$, the model used $\frac{1}{\sqrt{\tau}}\big(\sqrt{(1-\phi)}S_i + \sqrt{\phi}e_i\big)$. Here, $\frac{1}{\tau}$ is the marginal precision of S and e. The fraction of the variance explained by the spatial component e_i was the mixing parameter ϕ which is between 0 and 1. Hence, to deal with these two components, the priors were defined on the hyperparameters ϕ and τ. In this case, the base model assumed no effect of these components. Fixing the marginal precision, for example using ϕ, the base model was defined with ϕ as 0 because no spatial effect was assumed. Hence the base model using the new parameterization was only with S. This is detailed in Simpson et al (63). The prior of ϕ is not on a close form. It was deduced using the defined base model and the flexible model, which was $\big(\sqrt{(1-\phi)}S_i + \sqrt{\phi}e_i\big)$. The KLD for measuring the distance between them was used, defining the prior on the distance. The prior on τ in the same way as follows:

$$\pi(\tau) = \frac{\lambda}{2}\tau^{-3/2}e^{-\lambda\tau^{-1/2}}, \tau > 0, \lambda > 0 \tag{14}$$

- PC priors for time effect

In the case of AR(1), the PC prior for ϕ, was not on a closed form as well, and was derived defining the base model as the model with $\phi = 1$, that is no change over time. The PC prior for τ_1 had the same distribution with the PC prior of τ. The parameter λ was defined using the tail as explained above (64).

2.6.7 Spatial autocorrelation -hotspot and cold spot

Before conducting the spatial analysis, Moran's index I was used for detecting the spatial autocorrelation on the outcome at the cluster level. A positive Moran's statistics close to 1 or Geary c statistic close to 0 indicated a strong autocorrelation of the prevalence. For identifying hot spots and cold spots, the local indicator of spatial autocorrelation (LISA) was used.

2.6.8 Factors associated with the prevalence – Spatial Regression with INLA

The unit of the analysis was the cluster of the households. A Gaussian random effect model to select relevant ecological variables and testing for multicollinearity using the VIF was used, with the cut-off set at 10. The selection of the variables was also based on the knowledge of the variables commonly used in the literature. Linear regression was used to see whether the variation observed in the spatial scale can be sufficiently explained by the predictors, and to obtain estimates of the effect of the main variables after adjusting for the other covariates. To do this, the spatial autocorrelation was measured in the residuals using I Moran index. Spatial analysis was conducted considering each point as a realization of a Gaussian process, thus defining a stochastic process. This process has a covariance function which represents the distance between two points. This covariance function is called Màtern covariance as explained in the section 2.6.3, page 14.

This covariance Màtern function was defined as follows:

$$cov\big(f(0), f(x)\big) = \frac{\sigma^2}{2^{v-1}\Gamma(v)}(k \parallel x \parallel)^v K_v(k \parallel x\parallel) \quad (15)$$

and is related to a process which is a weak solution of the SPDE given by:

$$(k^2 - \Delta)^{\alpha/2}\big(\tau f(x)\big) = W(x) \quad (16)$$

where W(x) is a Gaussian white nose process, Δ is the Laplacian, k is a scale parameter related to the range r, τ affects the variance of f, α is related to the smoothness of f. $\parallel x \parallel$ represents the Euclidean distance for two locations x_i and x_j $(j \neq i)$, $f(x)$ is the exact solution which is a stationary process having the Matérn covariance function given above. This solution is

approximated by a GMRF through triangulation. The range is the distance after which the spatial correlation is small, i.e. almost null or less than 0.1.

The default value for α was 2. Therefore, the range and σ^2 are given by:

$$r = \frac{\sqrt{8\lambda}}{k}, \quad \sigma^2 = \frac{1}{4\pi k^2 \tau^2}, \quad \lambda = \alpha - d/2 = 2 - 2/2 = 1$$

For the SPDE, implemented in INLA, the default values were used. The transformation of τ and k defined by $\theta_1 = log(\tau)$ and $\theta_2 = log(k)$ was used with an independent normal prior distribution with precision equal 1 and mean 0 (29). The model defined for spatial regression based on SPDE approach is as follows:

$$y_{it} \sim \mathcal{N}(\eta_{it}, \sigma_e^2) \tag{17}$$

with

$$\eta_{it} = \sum_{j=1}^{p} \beta_j \, x_{ji} + \omega_{it}, \quad \omega_{it} = a\omega_{i(t-1)} + \xi_{it}, \quad |a| < 1$$

x_{mi} is the *mth* variable values at cluster i, σ_e^2 the variance for the measurement error modeled through a Gaussian process spatially and serially uncorrelated. In contrast to the linear random mixed model where the term ω_{it} is not allowed to be correlated, here ω_{it} is from the latent process changing in time using AR(1) with coefficient a which is spatially correlated. The values of ω_{it} depend on the values of $\omega_{i(t-1)}$. If a, the coefficient of AR(1), is close to 1, the value at time t is similar to that of $t-1$. ξ_{it} is a GF (Gaussian Field) independent in time, with mean 0 such that for $t_1 = t_2$, $Cov(\xi_{it_1}, \xi_{it_2})$ is non-null and is the Matèrn covariance function defined as above.

The vector of the parameters was given by $\theta = (\omega, \alpha, \beta)$, and for the hyper-parameters by $\phi = (\sigma_e^2, k, \sigma^2)$ for the implementation in INLA. A Gaussian Markov Random Field (GMRF) prior was assumed for θ, with the precision matrix Q, and mean 0. The parameters of this precision matrix were k and σ^2. The PC prior was used for the autocorrelation parameter a such that the probability of the standard deviation to be greater than 0 is 0.9. For the remaining parameters, loggamma distributions were used with mean 1, and precision of 0.00001 or 0.001 as also used by Musenge et al (48).

For the flexibility on the relationship between the outcome and the covariates, the GAM was used and presented as follows:

$$y_{it} \sim \mathcal{N}(\eta_{it}, \sigma_e^2)$$

with

$$\eta_{it} = \sum_{j=1}^{p} f_j\left(x_{ji}\right) + v_{it}, \quad v_{it} = \rho v_{i(t-1)} + u_{it} \tag{18}$$

The prior specification for the functions is f_j for all j is a random walk process of order 2 (RW(2)). This prior is chosen to avoid deviation from the linear trend. It is used when smoothing with the spline, for its computational efficiency and Markov property. It was defined as follows:

$f(t) = 2f(t-1) - f(t-2) + u(t)$, $f(t) = f\left(x_{(t)}\right)$, $x_{(t)}$ is an ordered statistic, with $u(t) \sim N(0, \tau^2)$ and diffuse prior set on $f(1)$ and $f(2)$.

The methods above did not allow to quantify the spillover effect, or the impact of a covariate on the outcome when the values are changing somewhere. To do this, spatial econometrics models including the SAR model was used. Among them the following were computed:

- SEM, lag on the error

$$y = X\beta + u; \text{ with } u = \rho Mu + e \text{ with } e \sim MVN\left(0, \sigma^2 I_n\right) \tag{19}$$

The average total impact (direct) is the posterior marginal of the coefficient β_r and the average indirect impact is 0.

- The SLM, lag on the outcome,

$$y = \rho Wy + X\beta + WX\gamma + e;\ e \sim MVN(0, \sigma^2 I_n) \tag{20}$$

- The SDM, lag only on the covariates,

$$y = \rho Wy + X\beta + WX\gamma + e; \text{ with } e \sim MVN\left(0, \sigma^2 I_n\right) \tag{21}$$

The average impact is $n^{-1}tr((I_n + \rho W)^{-1})\beta_r + n^{-1}tr((I_n + \rho W)^{-1}W)\gamma_r)$, where the tr is the sum of diagonal elements of a matrix.

- SDEM

$$y = X\beta + WX\gamma + u;\ u = \rho Mu + e \text{ with } e \sim MVN(0, \sigma^2 I_n) \quad (22)$$

583

584

585 The average total impact is the sum of respectively direct impact and indirect impact: $\beta_r + \gamma_r$

586

587 These are implemented in INLA as spatial latent model by conditioning on ρ to allow the model to take the suitable form (53).
588 In this estimation, a Gaussian prior was used for $log\left(\frac{\rho}{1-\rho}\right)$, and a normal prior for β with mean 0 and very large variance. In
589 the implementation, the variance of the likelihood was set to e^{-15} (60). Therefore, it is not possible to compare models fitted
590 within INLA, but the obtained DIC can perhaps be compared with the one obtained with the other models such as the GAM.

591

592 Lastly, GWR or SVC were performed to assess if the relationship between the prevalence and the covariates was stationary
593 or not all over the country. Spatial varying coefficient using Bayesian framework was used as a method to supply GWR because
594 of the local multicollinearity observed by looking at the conditional number (65).

595 The model suggested can be written as follows:

596 • For the GWR,

$$y_k = \beta_{i0} + \sum_{j=1}^{p} \beta_{kj}\, x_{jk} + \epsilon_i \quad (23)$$

597

598 x_{jk} is the value for the *jth* covariates at location k, p number of covariates, and β_{i0} the intercept for location k, and β_{kj} the
599 coefficient associated with the covariate J at location k, ϵ_i and the error at location i.

600 • For the SVC model, the equation is given by:
601

$$\eta_i = \alpha + (\beta_0 + \beta_i)x_i \quad (24)$$

602

603 with $\beta_0 \sim \mathcal{N}(0,1)$; $\beta|\xi \sim \pi(\beta|\xi)$, and β is a vector of all random effects to account for variation by cluster of the coefficient (66).

604

CHAPTER 3 - RESULTS

3.1 Descriptive analysis

3.1.1 Distribution of the prevalence and the covariates

This study included 336 clusters, with 153 (46%) clusters in rural areas and 183 (54%) clusters in urban areas. As shown in *Table 2*, the number of the clusters was approximately the same in each province. According to province, the prevalence of malaria was high in Estuaire province, followed by Moyen-Ogooue and Woleu-Ntem. The lowest prevalence was found in Haut-Ogooue province. In general, prevalence was similar between the provinces and have large standard deviation.

***Table 2**: Distribution of the clusters and malaria prevalence by province from 2000-2015*

Province	Number of clusters	Percentage (%)	Mean prevalence (SD)
Estuaire	31	11	0.29 (0.16)
Moyen-Ogooue	31	11	0.22 (0.13)
Woleu-Ntem	32	11	0.21 (0.1)
Ngounie	32	11	0.19 (0.11)
Ogooue-Ivindo	32	11	0.19 (0.12)
Nyanga	31	11	0.18 (0.08)
Ogooue-Lolo	32	11	0.18 (0.11)
Ogooue Maritime	32	11	0.18 (0.11)
Haut-Ogooue	29	10	0.17 (0.09)

Observing the distribution of the clusters by type of residence, it was found that 78% of the provinces had a high number of clusters sampled in rural areas (Table 3). Over time, in each province, the prevalence appeared to be on average high in rural areas (see Table 3,Table 4).

Table 3: *Distribution of clusters and malaria prevalence by province and type of residence from 2000-2015*

Province	Type of residence	Number of clusters	Percentage (%)	Mean prevalence (SD)
Estuaire	R	17	55%	0.3 (0.16)
	U	14	45%	0.25 (0.15)
Haut-Ogooue	R	9	31%	0.22 (0.11)
	U	20	69%	0.15 (0.07)
Moyen-Ogooue	R	18	58%	0.25 (0.14)
	U	13	42%	0.19 (0.11)
Ngounie	R	17	53%	0.22 (0.12)
	U	15	47%	0.16 (0.1)
Nyanga	R	14	45%	0.21 (0.07)
	U	17	55%	0.16 (0.07)
Ogooue-Ivindo	R	19	59%	0.22 (0.13)
	U	13	41%	0.15 (0.1)
Ogooue-Lolo	R	20	62%	0.21 (0.11)
	U	12	38%	0.13 (0.09)
Ogooue Maritime	R	20	62%	0.21 (0.12)
	U	12	38%	0.14 (0.07)
Woleu-Ntem	R	19	59%	0.25 (0.09)
	U	13	41%	0.15 (0.07)

Table 4, showed that over time, the prevalence was different by type of residence in each province. It was also found that the prevalence was very low in the year 2010 compared to the other years. The highest prevalence was found in the urban area of the province of Estuaire, with 54% (±3%) and 32% (±2%) during 2000 and 2015, respectively. The lowest prevalence was found in Ogooue-Lolo, 5% (±1%) and 7% (±1%) for the 2005 and 2010 respectively.

628

629 ***Table 4**: Distribution of malaria prevalence by type of residence and year from 2000-2015*

Province	Type of residence	Mean prevalence (SD) %			
		2000	**2005**	**2010**	**2015**
Estuaire	R	0.54 (0.03)	0.18 (0.05)	0.17 (0.03)	0.32 (0.02)
	U	0.48 (0.07)	0.12 (0.03)	0.13 (0.03)	0.27 (0.04)
Haut-Ogooue	R	0.35 (0.07)	0.22 (0.1)	0.14 (0.06)	0.19 (0.05)
	U	0.26 (0.04)	0.12 (0.04)	0.08 (0.02)	0.13 (0.02)
Moyen-Ogooue	R	0.45 (0.06)	0.16 (0.06)	0.12 (0.03)	0.27 (0.03)
	U	0.36 (0.04)	0.11 (0.02)	0.08 (0.01)	0.2 (0.03)
Ngounie	R	0.4 (0.03)	0.11 (0.04)	0.13 (0.02)	0.24 (0.03)
	U	0.31 (0.04)	0.07 (0.02)	0.09 (0.01)	0.18 (0.03)
Nyanga	R	0.3 (0.04)	0.2 (0.06)	0.14 (0.02)	0.2 (0.03)
	U	0.25 (0.04)	0.13 (0.04)	0.1 (0.02)	0.15 (0.03)
Ogooue-Ivindo	R	0.4 (0.05)	0.12 (0.04)	0.12 (0.02)	0.26 (0.04)
	U	0.29 (0.04)	0.07 (0.03)	0.08 (0.01)	0.17 (0.04)
Ogooue-Lolo	R	0.39 (0.02)	0.11 (0.04)	0.12 (0.01)	0.22 (0.02)
	U	0.27 (0.03)	0.05 (0.01)	0.07 (0.01)	0.14 (0.02)
Ogooue Maritime	R	0.39 (0.04)	0.14 (0.03)	0.07 (0.01)	0.23 (0.02)
	U	0.25 (0.04)	0.11 (0)	0.05 (0)	0.14 (0.01)
Woleu-Ntem	R	0.36 (0.06)	0.2 (0.09)	0.19 (0.05)	0.27 (0.03)
	U	0.23 (0.08)	0.08 (0.03)	0.12 (0.02)	0.19 (0.04)

630

631

In Table 5, the mean prevalence all over the country was 20% (±12%). It was also seen that the standard deviation was showing a small variation for almost all variables, excepted for proximity to water, ITNs coverage, all population count, and aridity.

Table 5: *Distribution of prevalence and covariates from 2000-2015*

Covariate	Mean (SD)
Annual Precipitation (mm/year)	14.39 (2.46)
Aridity	55.95 (8.7)
Day Land Surface Temp (°C)	26 (1)
Diurnal Temperature Range (°C)	8 (1)
Enhanced Vegetation Index	39 (6)
Frost Days	0 (0)
Growing Season Length	12 (1)
ITN Coverage (%)	24 (15)
Land Surface Temperature (°C)	23 (1)
Maximum Temperature (°C)	30 (1)
Mean Temperature (°C)	26 (1)
Minimum Temperature (°C)	22 (1)
Night Land Surface Temp (°C)	20 (1)
Prevalence (%)	20 (12)
Proximity to Water (meters)	1453 (1208)
Rainfall (mm/year)	1754 (349)
Travel Times (min)	453 (271)
Wet Days (days/month/year)	11 (1)
All Population Count (number of people)	16600 (75300)

In Table 6, it was observed that the mean value of almost all the variables is not changing over time. Only rainfall, all population count, ITNs coverage, and aridity are slightly changing. ITNs coverage was the most varying, showing a decrease over time.

642 ***Table 6**: Distribution of covariates and outcome by year*

Covariate	Mean prevalence (SD) %			
	2005	**2010**	**2015**	**2000**
All Population Count	14800 (66800)	16600 (74800)	18600 (83700)	
Annual Precipitation (mm/year)	13.18 (1.33)	15.59 (1.67)	12.49 (1.37)	16.3 (2.74)
Aridity	51.2 (4.06)	59.89 (5.47)	48.69 (4.83)	64.03 (8.77)
Day Land Surface Temp (°C)	25 (1)	26 (1)	26 (1)	26 (1)
Diurnal Temperature Range (°C)	8. (1)	8. (1)	8. (1)	8. (1)
Enhanced Vegetation Index	390.4 (61.2)	394.2 (61.9)	394.9 (61.8)	365.4 (57.1)
Frost Days	0 (0)	0 (0)	0 (0)	0 (0)
Growing Season Length	12. (1)	12. (1)	12. (1)	12. (1)
ITN Coverage	0.2 (0.05)	0.42 (0.07)	0.09 (0.05)	
Land Surface Temperature (°C)	23 (0.88)	23 (1)	23 (1)	23 (1)
Maximum Temperature (°C)	30 (1)	30 (1)	30 (1)	30 (1)
Mean Temperature (°C)	26 (1)	26 (1)	26 (1)	26 (1)
Minimum Temperature (°C)	22 (1)	22 (1)	22 (1)	22 (1)
Night Land Surface Temp (°C)	20 (1)	20 (1)	20 (1)	19 (1)
Prevalence	0.13 (0.06)	0.11 (0.04)	0.21 (0.06)	0.35 (0.1)
Proximity to Water (meters)	1453 (1210)	1453 (1210)	1453 (1210)	1453 (1210)
Rainfall (mm/year)	1666 (2500)	1872 (301)	1478 (239)	2000 (344)
Travel Times (min)			363 (253)	542.21 (258)
Wet Days (days/month / year)	11 (1)	12 (1)	10 (1)	12 (1)

643

644

Table **7** shows that the mean of malaria prevalence was decreasing drastically from the year 2000 up to 2005. However, the standard deviation was almost the same over time, with a small variation in the prevalence observed by province. The mean of aridity, rainfall, ITNs coverage and wet days was slightly varying by year. The ITNs coverage was varying greatly compared to the other values, starting to increase from 2005 to 2010, then to decrease until 6% in the provinces of Estuaire and Haut-Ogooue. The high coverage of the ITNs was observed in the province of Ogooue-Ivindo during the year 2010, 52% (±6%). Also, it seems like the province of Estuaire recorded high values of some ecological variables such as aridity, wet day, proximity to water, and rainfall at Estuaire province. For each region, Enhanced vegetation index did not change over time but by area, with high value at Ogooue-Lolo in 2015. Aridity had the same time trend in all the provinces, decreasing from 2000 to 2005, increasing from 2005 to 2010 then decreasing from 2010 to 2015. No particular time trend for wet days. For all the remaining variables, it was like the mean was not varying by year and by region, almost the same.

***Table 7**: Covariates distribution by year and type of residence*

Province	Year	Prevalence Mean (SD)	Population count Mean (SD)	ITN coverage Mean (SD)	Aridity Mean (SD)	Rainfall (mm) Mean (SD)	Wet days (day(s)) Mean (SD)	EVI Mean (SD)	Day Land Surface Temp (°C) Mean (SD)
Estuaire	2000	0.52 (0.06)			76.03 (1.39)	2673 (2610)	14.96 (0.43)	320.6 (48.7)	26.74 (1.33)
	2005	0.16 (0.05)	114700 (186000)	0.16 (0.03)	55.68 (1.07)	2209 (1260)	12.22 (0.36)	343.7 (54.9)	26.86 (1.2)
	2010	0.16 (0.03)	128400 (208100)	0.34 (0.06)	63.41 (0.99)	2457 (1780)	13.09 (0.41)	344.4 (54.6)	27.54 (1.51)
	2015	0.3 (0.04)	143600 (232900)	0.06 (0.01)	48.35 (0.77)	1985 (1410)	11.61 (0.36)	347.6 (55.8)	27.24 (1.12)
Haut-Ogooue	2000	0.29 (0.07)			56.21 (2.68)	2189 (1490)	11.97 (0.36)	386.8 (34.3)	28.07 (1.73)
	2005	0.15 (0.08)	11800 (13300)	0.17 (0.02)	50.98 (1.99)	1715 (85)	11.59 (0.34)	410 (42.6)	27.66 (1.75)
	2010	0.1 (0.04)	13200 (14900)	0.38 (0.04)	58.48 (2.38)	1894 (120)	12.15 (0.33)	409.7 (42.7)	27.91 (1.61)
	2015	0.15 (0.04)	14800 (16700)	0.06 (0.02)	51.88 (2.16)	1620 (72)	11.12 (0.31)	409.9 (42.2)	28.16 (1.68)
Moyen-Ogooue	2000	0.41 (0.07)			72.13 (1.94)	1929 (130)	13.15 (0.23)	370.2 (25.5)	25.4 (0.37)
	2005	0.14 (0.05)	4200 (5400)	0.18 (0.03)	51.88 (1.03)	1719 (103)	10.73 (0.22)	391.5 (25.3)	25.2 (0.43)
	2010	0.11 (0.03)	4700 (6000)	0.41 (0.04)	61.57 (1.33)	1889 (134)	11.54 (0.19)	401.5 (31.3)	25.36 (0.71)
	2015	0.24 (0.04)	5300 (6800)	0.07 (0.04)	46.46 (1.17)	1465 (92)	10.13 (0.21)	398.2 (29.6)	25.32 (0.53)
Ngounie	2000	0.36 (0.06)			74 (5.08)	2326 (160)	13.2 (0.49)	365.4 (22.6)	26.41 (1.4)
	2005	0.09 (0.04)	1900 (2200)	0.17 (0.03)	57.5 (3.14)	1837 (87)	11.35 (0.33)	386.7 (30.6)	25.91 (1.23)
	2010	0.11 (0.03)	2100 (2500)	0.4 (0.04)	68.91 (3.78)	2174 (96)	12.33 (0.42)	386.1 (29.3)	26.69 (1.5)
	2015	0.21 (0.04)	2300 (2800)	0.07 (0.03)	55.85 (2.4)	1561 (109)	10.76 (0.3)	387.6 (31.6)	26.7 (1.56)
Nyanga	2000	0.28 (0.05)			60.96 (5.44)	1753 (228)	11.64 (0.59)	355.7 (49.7)	27.07 (1.28)
	2005	0.16 (0.07)	1200 (1000)	0.19 (0.02)	48.53 (3.41)	1397 (164)	10.25 (0.38)	369.4 (45.7)	26.73 (1.37)
	2010	0.12 (0.03)	1300 (1200)	0.41 (0.03)	59.58 (3.64)	1638 (214)	11.13 (0.43)	381.2 (47.9)	26.91 (1.27)
	2015	0.18 (0.04)	1500 (1300)	0.09 (0.03)	50.64 (2.16)	1169 (107)	9.67 (0.49)	382.7 (47)	26.65 (1.16)
Ogooue-Ivindo	2000	0.36 (0.07)			54.34 (4.57)	1701 (116)	11.14 (0.45)	400.1 (34.2)	25.59 (0.73)
	2005	0.1 (0.04)	2400 (3600)	0.24 (0.04)	47.74 (1.31)	1459 (75)	10.23 (0.13)	431.4 (43.9)	25.32 (0.6)
	2010	0.1 (0.02)	2600 (4000)	0.52 (0.02)	53.74 (1.4)	1503 (52)	11.07 (0.13)	430.7 (37.7)	25.56 (0.71)
	2015	0.22 (0.06)	3000 (4500)	0.12 (0.03)	43.98 (1.69)	1350 (90)	9.56 (0.19)	432.4 (40.6)	25.67 (0.66)
Ogooue-Lolo	2000	0.34 (0.07)			61.41 (5.55)	1921 (188)	11.75 (0.71)	400.7 (17.3)	25.65 (0.75)
	2005	0.09 (0.04)	3100 (3700)	0.18 (0.05)	52.64 (2.91)	1503 (165)	10.77 (0.5)	432.1 (14.8)	25.41 (0.67)
	2010	0.1 (0.03)	3400 (4100)	0.42 (0.06)	61.19 (3.43)	1750 (169)	11.54 (0.54)	445.8 (16.4)	26.01 (0.93)
	2015	0.19 (0.05)	3900 (4600)	0.07 (0.05)	50.48 (2.75)	1411 (136)	10.18 (0.48)	446.9 (16.1)	25.94 (0.93)
Ogooue Maritime	2000	0.34 (0.08)			67.75 (1.52)	2018 (200)	12.31 (0.33)	279.8 (86.4)	26.07 (0.65)
	2005	0.13 (0.03)	500 (1600)	0.27 (0.04)	49.41 (2.75)	1668 (192)	10.25 (0.17)	311.1 (96.6)	25.98 (0.69)
	2010	0.06 (0.02)	600 (1800)	0.51 (0.07)	60.33 (3.68)	1971 (189)	10.89 (0.27)	309.1 (92.1)	26.17 (0.66)
	2015	0.19 (0.05)	600 (2000)	0.18 (0.05)	49.99 (1.99)	1526 (243)	9.82 (0.34)	309.7 (91.7)	25.97 (0.62)
Woleu-Ntem	2000	0.31 (0.09)			56.1 (5.3)	1654 (135)	12.33 (0.65)	398.6 (21.3)	25.03 (0.39)

	2005	0.15 (0.09)	5900 (7400)	0.19 (0.03)	47.48 (2.3)	2673 (2610)	11.14 (0.39)	425.2 (22.9)	24.84 (0.51)
	2010	0.16 (0.05)	6600 (8200)	0.42 (0.05)	52.63 (2.4)	2209 (1260)	11.97 (0.52)	427.3 (29.1)	25.04 (0.45)
	2015	0.24 (0.05)	7300 (9200)	0.08 (0.07)	40.74 (2.97)	2457 (1780)	9.97 (0.39)	427.6 (27)	24.92 (0.54)

657

3.1.2 Visualization of malaria prevalence

As shown in Figure 3 (see appendix), the distribution of the prevalence was right skewed. To reduce the skewness and to approximate by the normal distribution, the prevalence was transformed using the log or the square root.

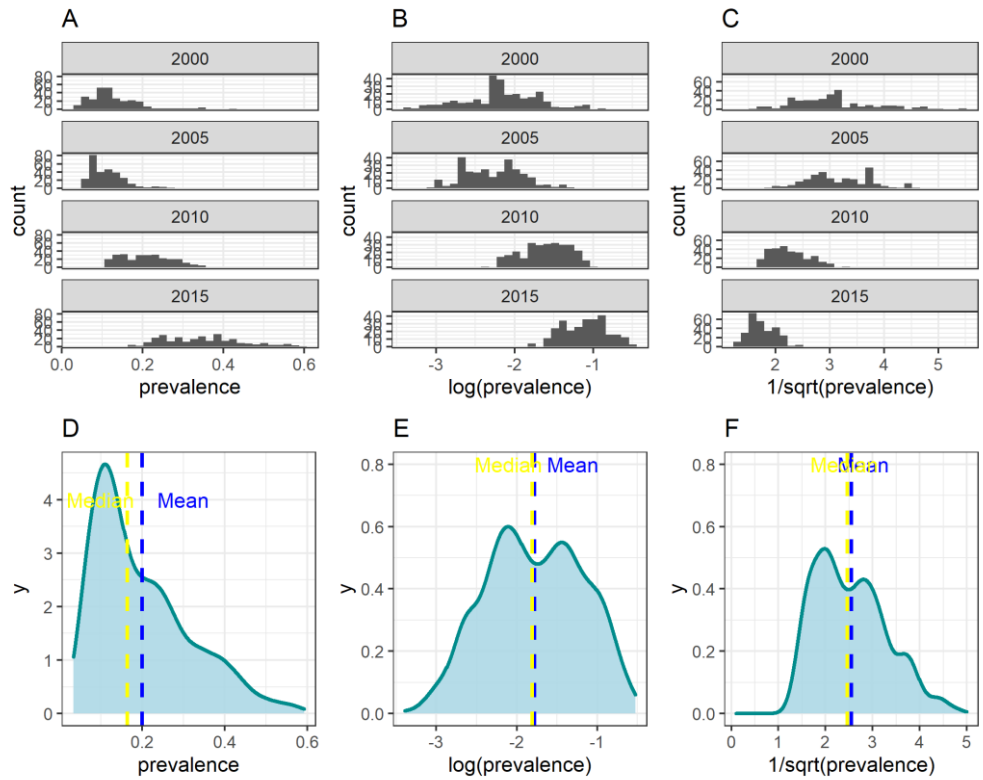


Figure 3: Panel (A), (B) and (C) show the distribution of the prevalence using histogram before and after transforming with log and square root for each year. Panel (D), (E) and (F) show the density in overall for each transformation

After the transformations, no one of the other transformations (Figure 4) was following the normal distribution even by type of residence.

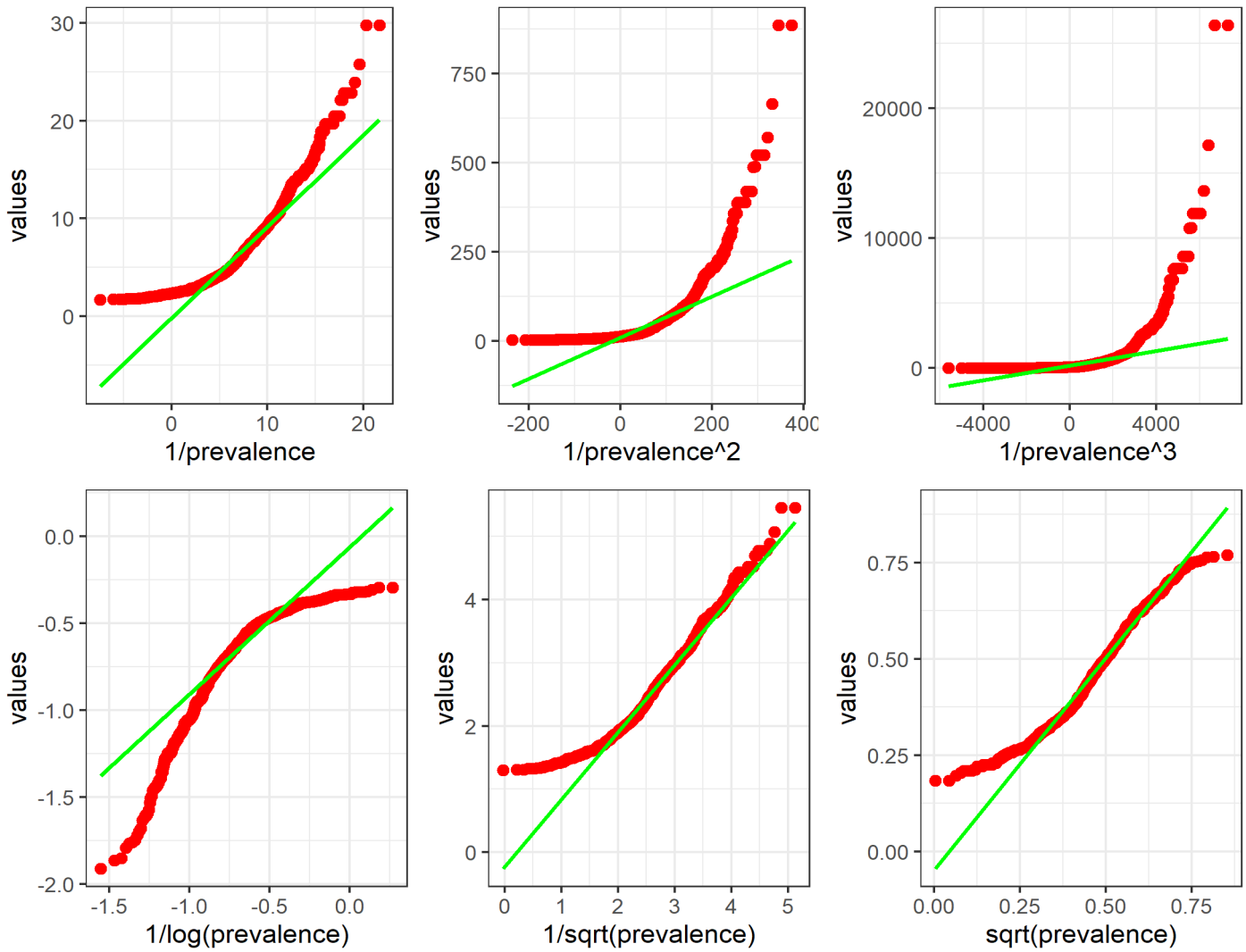


***Figure 4**: Overview of transformation of the prevalence using normal plot probability to approximate normal distribution*

However, as shown in Figure 5, the log transformation helped to obtain a more symmetric distribution but bimodal. From the normal probability plot, some departures from the ideal line at the start and the end was observed. For the inverse square root transformation, the distribution was also bimodal, somehow symmetric but with tail. From the normal probability plot, there was some departures from the ideal line only at the starting point. In contrast to the log transformation, the mean and median were slightly different for the inverse transformation (0.08 vs 0.02). Therefore, more symmetric with the log transformation than the inverse square root transformation. Hence, normality was assumed or even the theorem of central limit (TCL) was used to assume normality even for this bimodal distribution.

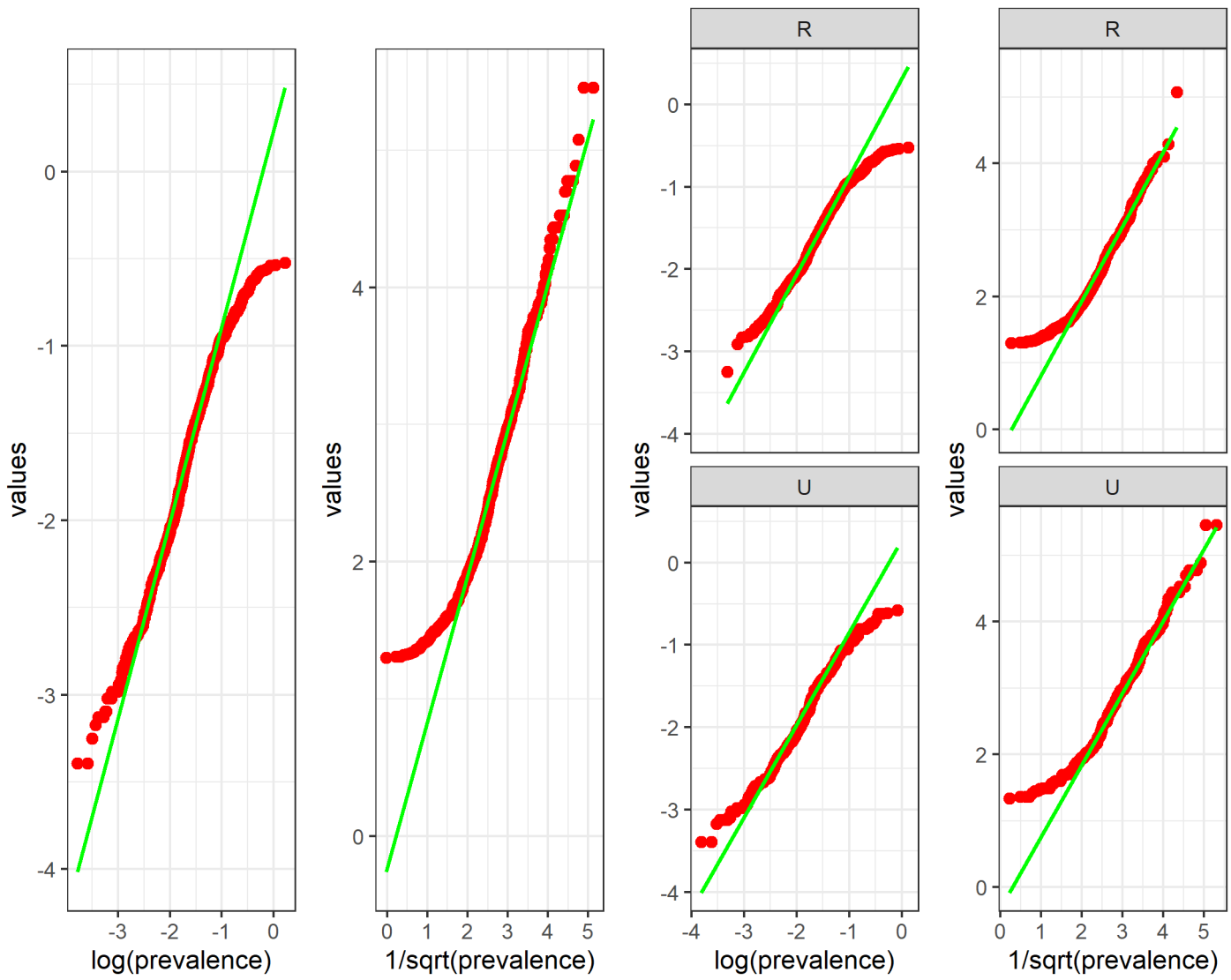


Figure 5*: Distribution of the prevalence using normal plot probability to compare log and inverse square root prevalence in overall (left) and by type of residence (Urban-Rural - right)*

The same was found for the distribution by type of residence (Figure 6).

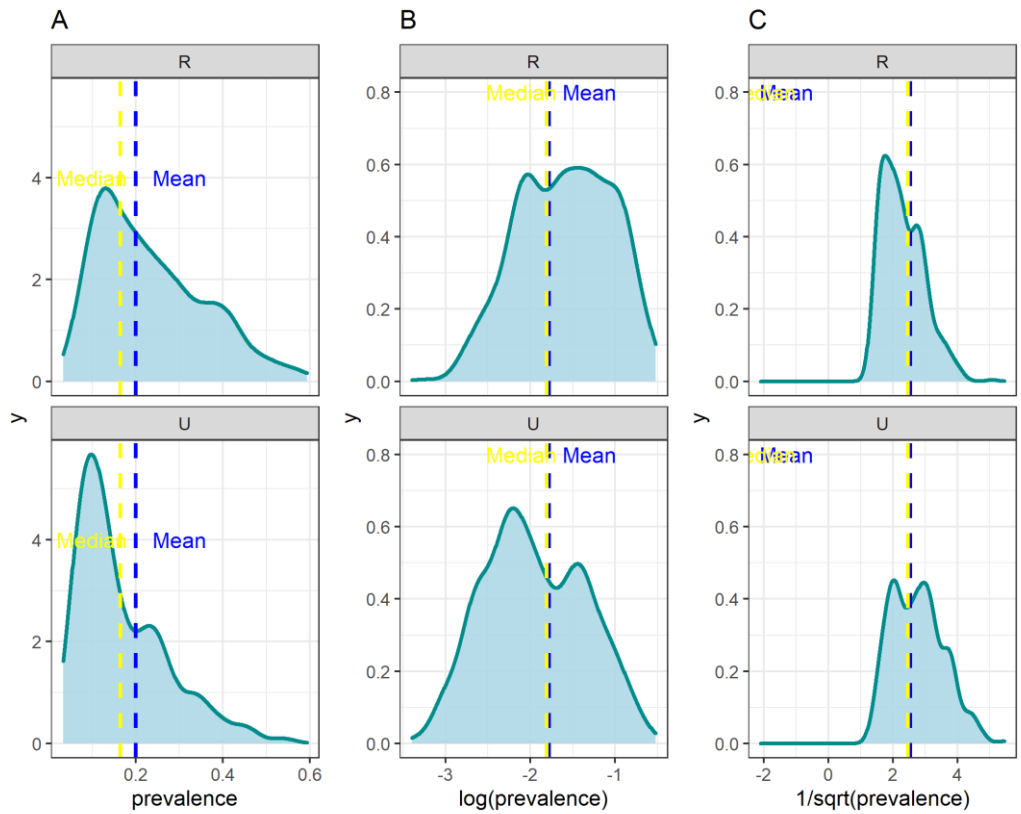


***Figure 6**: Panel (A), (B) and (C) show the distribution of the prevalence using histogram before transforming and after transforming with log and square root Prevalence in overall by type of residence (Urban-Rural)*

Observing the series of the map (Figure 7), on overage, it was observed that there was a decrease in the prevalence from the year 2000 to the year 2010. After the year 2010, there was a trend of increase until 2015.

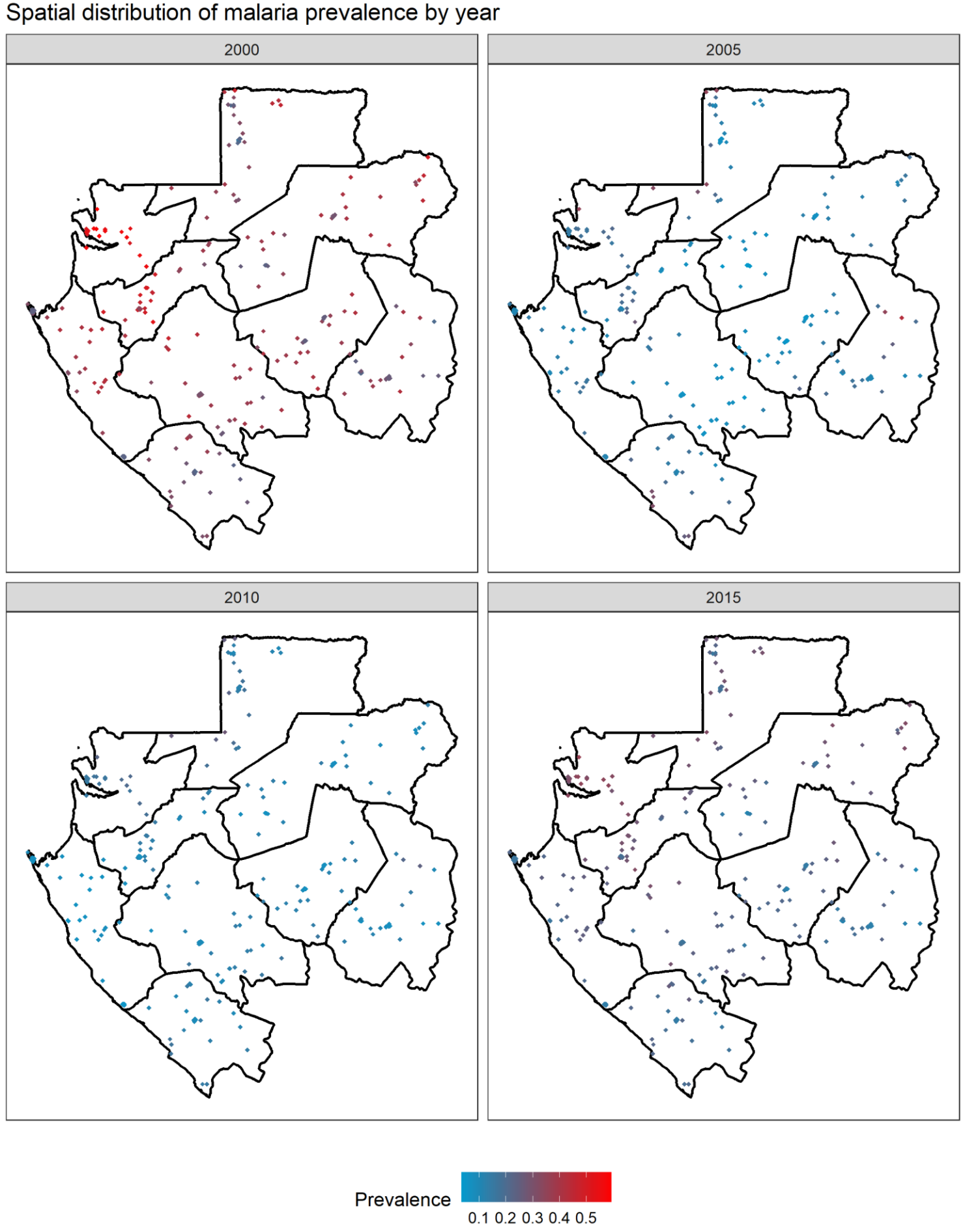


***Figure 7**: Map of prevalence by cluster and over time in all Gabon before estimating with smooth and complexity of the*

survey

Figure 8 shows that for almost every cluster the prevalence decreased over time and showed a small trend of increase. On average, as showed with the green line, the prevalence showed a U-shape over time.

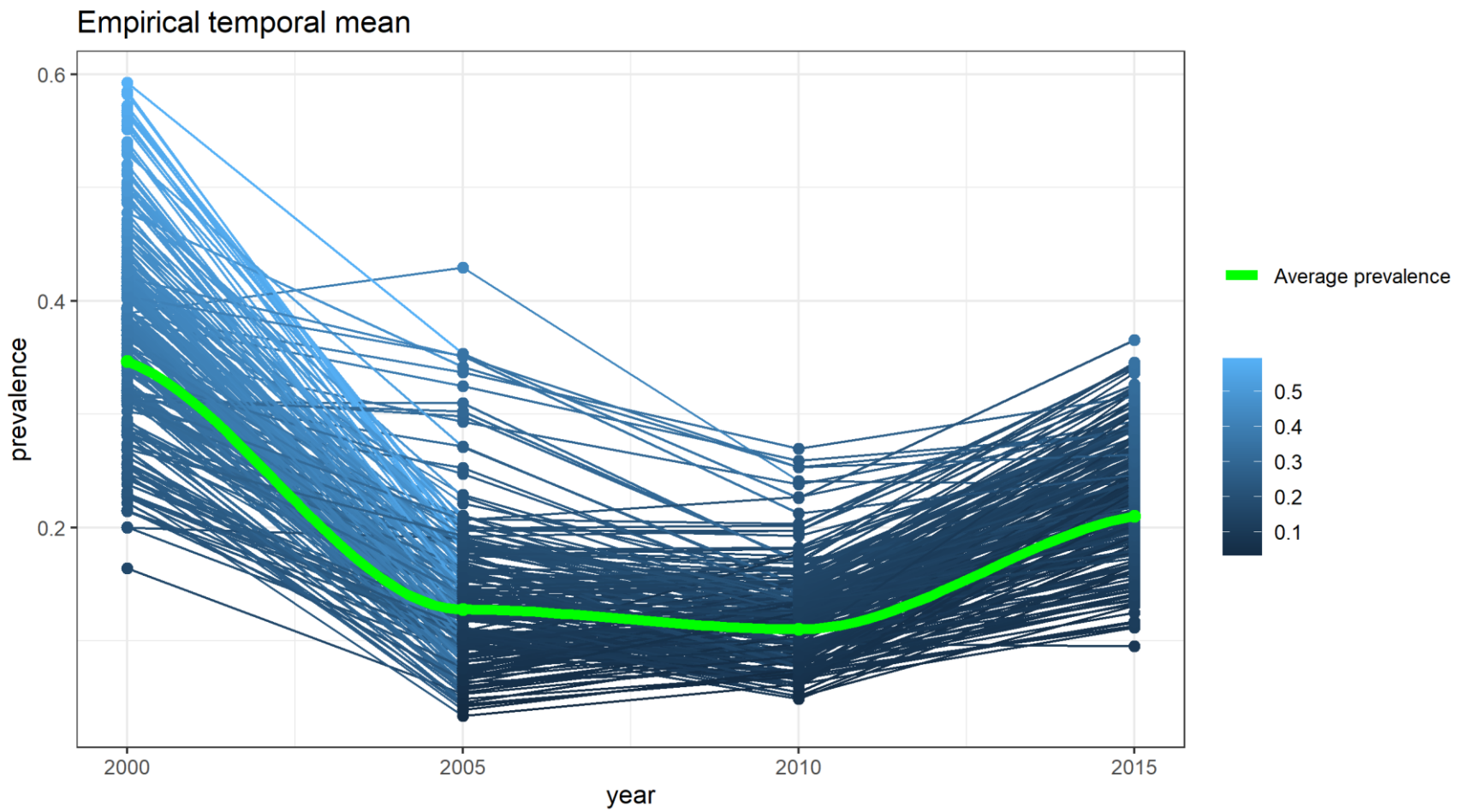


***Figure 8**: Prevalence trend for all clusters. The green line shows the average prevalence. The line was smoothed using polynomial function.*

3.1.3 Relationship between malaria prevalence and covariates

Observing Figure 9, some variables, such as rainfall had a linear trend after using either the log transformation or the inverse square root transformation compared to the untransformed data

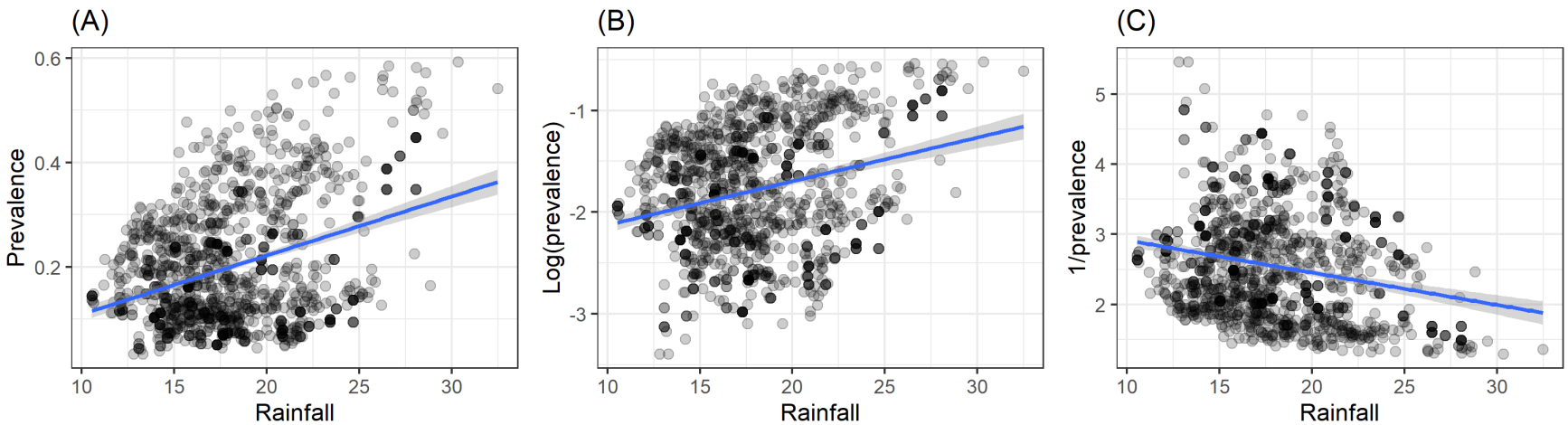


***Figure 9**: Scatterplot of malaria prevalence and Rainfall to visualize the trend of the relationship with different transformation. Panels (A), (B), (C) are showing the untransformed prevalence, the log transformed prevalence and the inverse square root transformation respectively. Blue line, through linear regression fitting the data is used to show the tendency of the data.*

Figure 10, shows that in contrast to some variables such us rainfall, for some other variables such us ITNs coverage, it may be better to approach the relationship with the outcome by a non-linear function as shown.

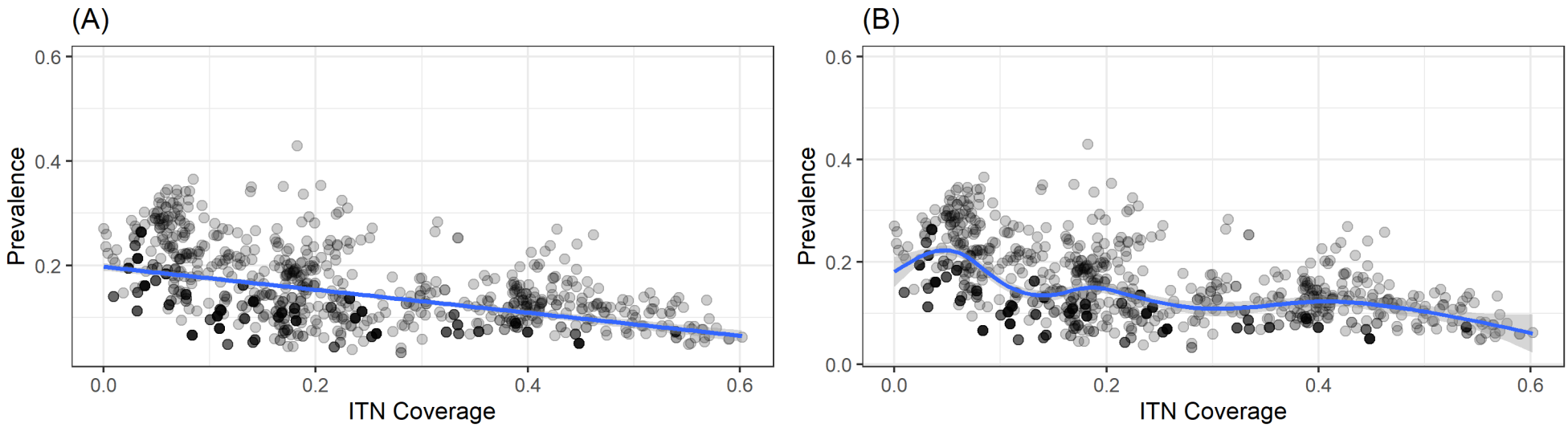


***Figure 10**: Scatterplot of malaria prevalence and ITNs coverage to visualize the trend of the relationship. Panels (A) and (B) show the shape of the relationship by using linear model and parametric model, respectively.*

Observing Figure 11, the high values for population density and some low values for enhanced vegetation index was observed.

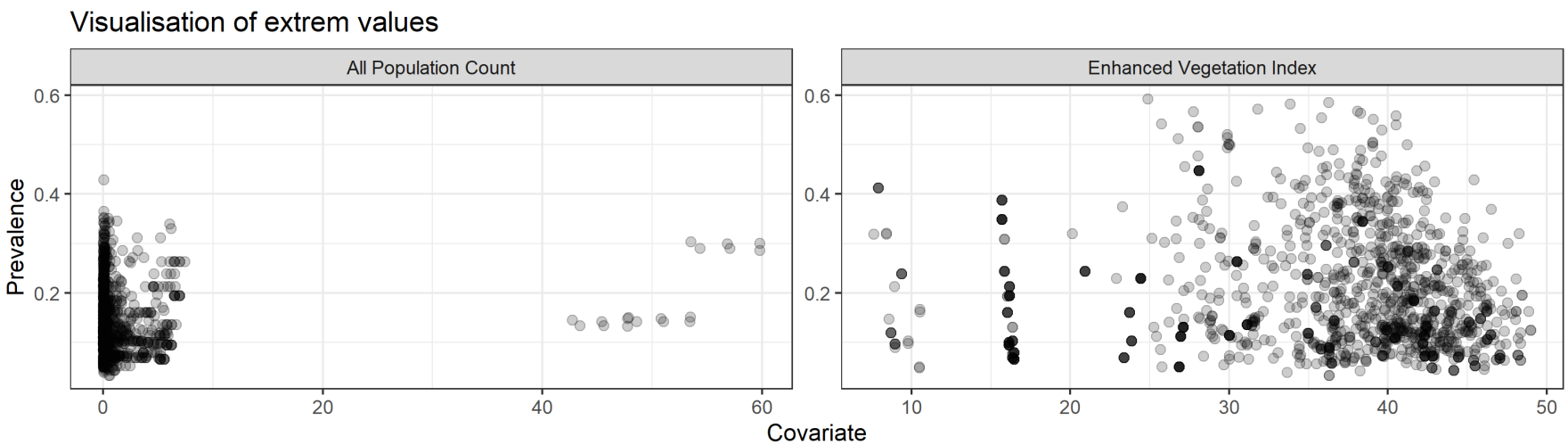


***Figure 11**: Scatterplot of malaria prevalence and covariates to visualize the extreme values.*

Therefore, two approaches were suggested. Firstly, a linear relationship with the log or the square root of the prevalence was assumed, hence, the model assuming linearity was run. Secondly, another model of variables with linear relationship and variables with non-linear relationship was run. Then the models were compared based on the DIC. As our aim was only to evaluate the effect of predictors, the parsimony model in term of interpretation or complexity were used (67).

3.2 Areal level model space-time - Estimation

Before the estimation of the prevalence, 64 clusters were removed because of missing values, 20 clusters for the variable malaria prevalence and 44 clusters in Libreville and Port-Gentil. These were removed because many values for these towns were missing. These were identified in DHS database as -9.999 (showing Not Available [NA]), including those for which weight was not found. Only 272 remained for areal data analysis. These 272 were only for the estimation of the prevalence. For the other analysis, 316 clusters were considered.

3.2.1 Space-time estimation of the prevalence

For each unit or province and type of residence, the values too close to 0, and too close to 1 were not observed. Because if these were observed, the logit would not be defined. Comparing different types of models of interactions, the interaction of type 2, based on the DIC was the best (see Table 8).

***Table 8**: Comparison of models based DIC for interaction effect*

Random Walk type	Space-time type	DIC	Interaction
rw1	0	-143.4477	Model without interaction
rw2	0	-141.2145	Model without interaction
rw1	1	-144.8954	Time and space iid
rw2	1	-145.1627	Time and space iid
rw1	2	-145.6341	Temporal trends in each cluster are not spatially correlated
rw2	2	-145.8665	Temporal trends in each cluster are not spatially correlated
rw1	3	-140.3611	Time-specific spatial patterns are not temporally correlated
rw2	3	-142.4178	Time-specific spatial patterns are not temporally correlated
rw1	4	-143.4477	Cluster-specific temporal trends spatially correlated - or spatial patterns temporally correlated
rw2	4	-141.2145	Cluster-specific temporal trends spatially correlated - or spatial patterns temporally correlated

rw = random walk, DIC = Deviance Information Criterion

Comparing the model with and without interaction, the model introducing interaction was slightly better (-145 vs 141). This model had also the highest spatial fraction, which is the mixing parameter (see Table 9).

***Table 9**: DIC model interaction vs without interaction*

Model	DIC	Spatial fraction (SD)[a]	95% CI[b]
Model 0	-141.2145	0.35 (0.27)	0.02 - 0.91
Model 1	-145.8665	0.41 (0.29)	0.02 - 0.95

[a]Spatial fraction is the mixing parameter in the model;
[b] CI = Credible Interval

As shown in Figure 12, the comparison between the estimate obtained using the smoothing method with the sample design, and the smoothing method without the sample design showed a reduction of the variance. Hence, the spatial effect used as smoothing reduced the variability introduced by the small area, when taking into account the sample design.

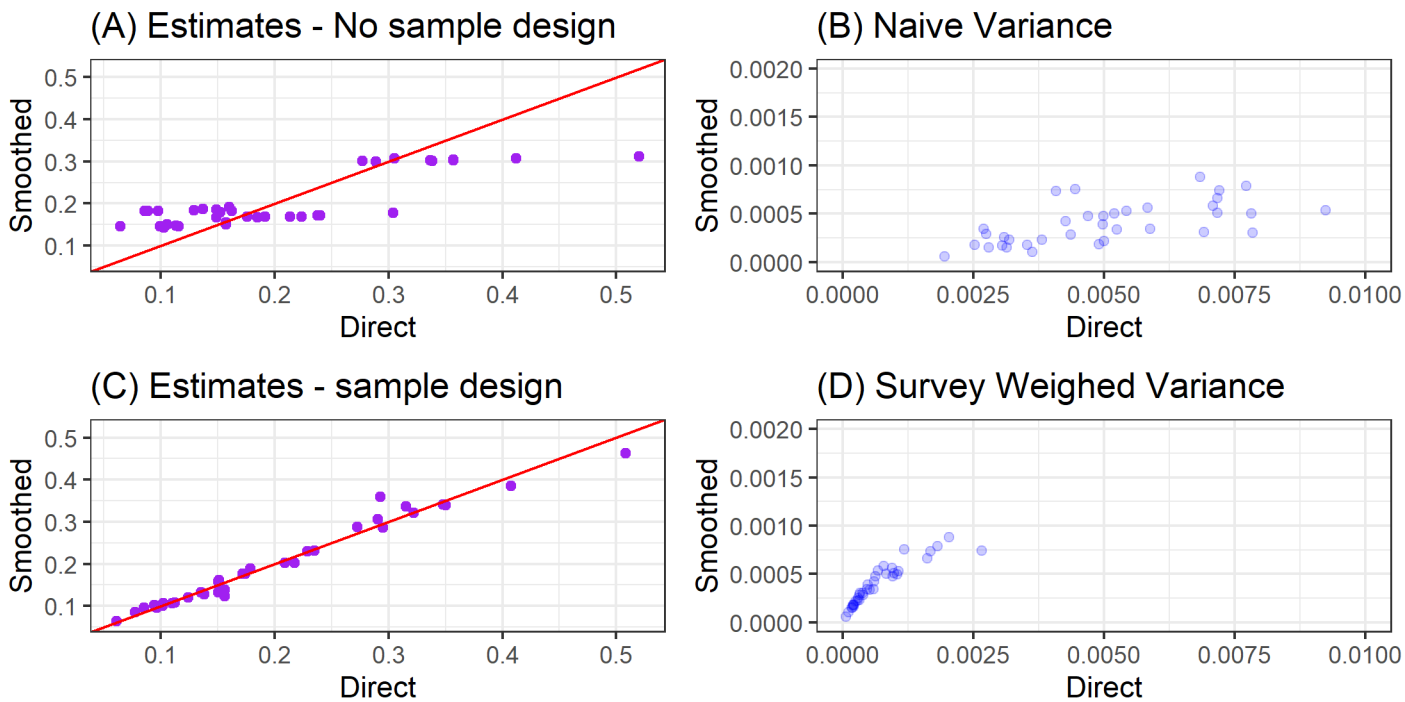


***Figure 12**: Comparison of variance with estimate obtained using smoothing of the sample design. Panel (A) and Panel (B) show the estimation obtained without considering the design, and panel (C) and Panel (D) are the estimation obtained when the survey design included. For each plot the direct model is plotted against the smoothed model.*

Hence, using the model with the sample design and the smooth method, the highest prevalence was found in Estuaire province for all the years, followed by Moyen-Ogooue Province (Table 10). In this table, the prevalence was high in 2000, decreased over time and increased slightly again in 2015, but for each year the values of the prevalence were roughly the same in each province.

***Table 10**: Space-time smoothing estimation of malaria prevalence*

Region	Prevalence 2000 - 95% CI	Prevalence 2005 - 95% CI	Prevalence 2010 - 95% CI	Prevalence 2015 - 95% CI
Estuaire	0.46 (0.42 - 0.51)	0.16 (0.12 - 0.20)	0.15 (0.12 - 0.18)	0.29 (0.25 - 0.32)
Moyen-Ogooue	0.38 (0.34 - 0.43)	0.13 (0.10 - 0.16)	0.11 (0.08 - 0.13)	0.23 (0.19 - 0.27)
Woleu-Ntem	0.36 (0.30 - 0.42)	0.13 (0.08 - 0.18)	0.13 (0.09 - 0.17)	0.23 (0.18 - 0.28)
Ngounie	0.34 (0.30 - 0.38)	0.10 (0.07 - 0.12)	0.11 (0.08 - 0.13)	0.20 (0.17 - 0.24)
Ogooue-Ivindo	0.34 (0.29 - 0.39)	0.10 (0.07 - 0.13)	0.10 (0.07 - 0.12)	0.20 (0.16 - 0.25)
Ogooue-Maritime	0.34 (0.28 - 0.39)	0.12 (0.10 - 0.14)	0.06 (0.05 - 0.08)	0.19 (0.15 - 0.23)
Ogooue-Lolo	0.32 (0.27 - 0.38)	0.09 (0.05 - 0.12)	0.10 (0.07 - 0.13)	0.18 (0.13 - 0.22)
Haut-Ogooue	0.31 (0.26 - 0.35)	0.12 (0.07 - 0.18)	0.10 (0.07 - 0.13)	0.16 (0.13 - 0.19)
Nyanga	0.29 (0.25 - 0.32)	0.14 (0.10 - 0.18)	0.11 (0.08 - 0.13)	0.18 (0.15 - 0.20)

In Figure 13, Ogooue-Maritime province was characterized by a very low transmission during the year 2010 (prevalence less 10% - hypo-endemic area). In general, the whole country remained a mesoendemic area over time with prevalences between 10% and 50%.

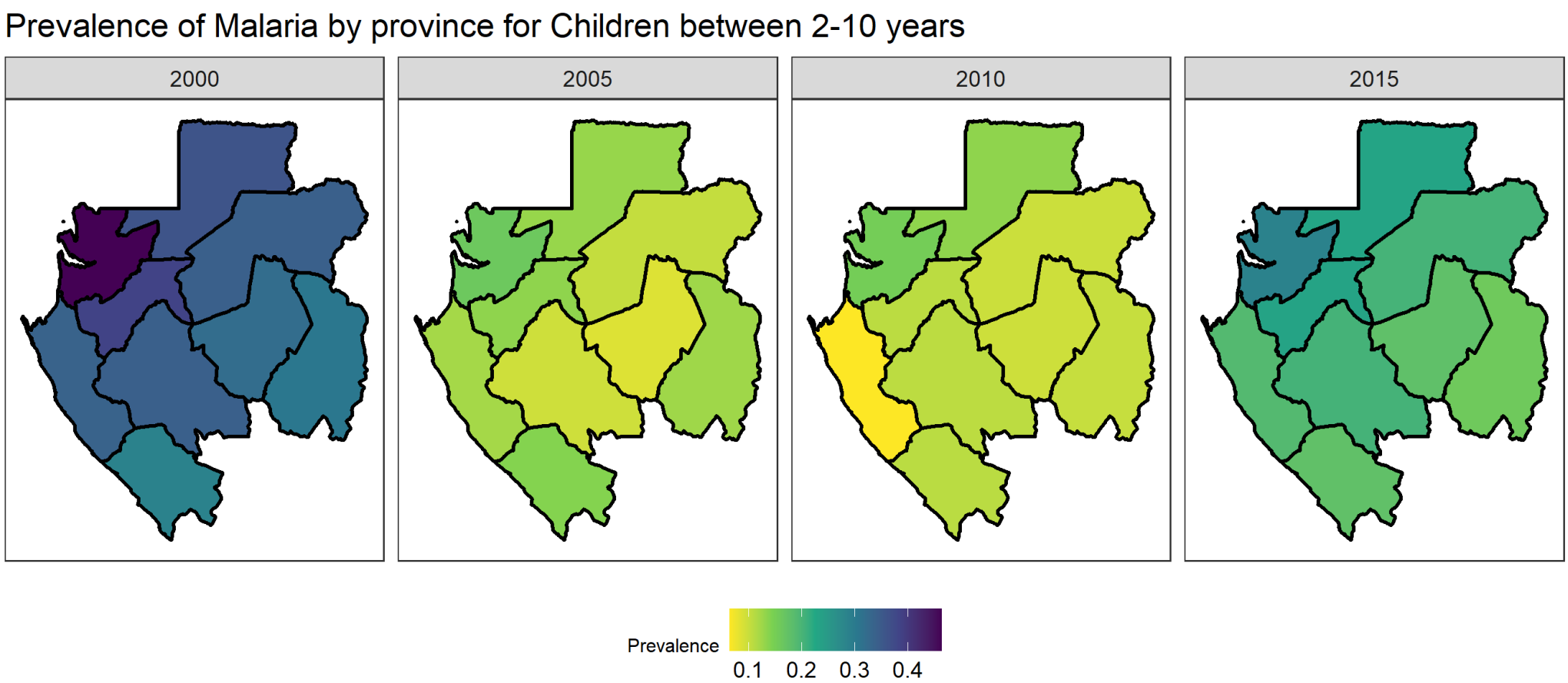


***Figure 13**: Map of the prevalence by year. This plot presents the estimate obtained with the smoothing model and survey complexity.*

3.3 Unit-level model

In Table 11, after using the unit-level model with stratification by type of residence, the length of the credible interval was small compared to the areal-level model. The maximum length of the credible interval went from 12% to 4%, and the minimal length went from 4% to 1% from area-level model to unit-level model, respectively. The highest prevalence was also found in Estuaire province for all the years, followed by Moyen-Ogooue province by type of residence. It was observed from the model averaging the prevalence and the model ran for each year, the prevalence of malaria was different between type of residence with higher prevalence in Rural area, with a difference of roughly 7%. Using this model, the rural areas of the Estuaire province were found to be hyperendemic areas during the year 2000, while it was found that almost all urban areas have low transmission intensity during the year 2010. In general, all the country was mesoendemic until 2015. All the models showed a decrease from 2000 to 2010 for both types of residence.

***Table 11**: Space-time smoothing estimation prevalence by year, province and type of residence from the unit level model*

Province	Type of residence	Prevalence 2000 - 95% CI	Prevalence 2005 - 95% CI	Prevalence 2010 - 95% CI	Prevalence 2015 - 95% CI
Estuaire	U	0.45 (0.43 - 0.47)	0.12 (0.10 - 0.14)	0.13 (0.12 - 0.14)	0.26 (0.25 - 0.27)
	R	0.55 (0.54 - 0.57)	0.18 (0.16 - 0.20)	0.17 (0.16 - 0.18)	0.33 (0.32 - 0.34)
Haut-Ogooue	U	0.26 (0.24 - 0.28)	0.13 (0.11 - 0.15)	0.09 (0.08 - 0.10)	0.13 (0.12 - 0.14)
	R	0.36 (0.34 - 0.38)	0.19 (0.17 - 0.21)	0.13 (0.12 - 0.14)	0.20 (0.19 - 0.21)
Moyen-Ogooue	U	0.35 (0.34 - 0.37)	0.10 (0.08 - 0.12)	0.08 (0.07 - 0.09)	0.20 (0.19 - 0.21)
	R	0.45 (0.44 - 0.47)	0.16 (0.15 - 0.18)	0.12 (0.11 - 0.13)	0.27 (0.26 - 0.28)
Ngounie	U	0.30 (0.29 - 0.32)	0.06 (0.04 - 0.08)	0.09 (0.08 - 0.10)	0.18 (0.16 - 0.19)
	R	0.40 (0.39 - 0.42)	0.12 (0.10 - 0.14)	0.13 (0.12 - 0.14)	0.25 (0.24 - 0.26)
Nyanga	U	0.23 (0.21 - 0.25)	0.13 (0.11 - 0.15)	0.09 (0.09 - 0.10)	0.14 (0.13 - 0.15)
	R	0.33 (0.31 - 0.35)	0.19 (0.17 - 0.21)	0.14 (0.13 - 0.15)	0.21 (0.20 - 0.23)
Ogooue-Ivindo	U	0.30 (0.28 - 0.32)	0.06 (0.04 - 0.08)	0.07 (0.07 - 0.08)	0.18 (0.17 - 0.19)
	R	0.40 (0.38 - 0.42)	0.13 (0.11 - 0.14)	0.12 (0.11 - 0.13)	0.25 (0.24 - 0.26)
Ogooue-Lolo	U	0.28 (0.26 - 0.30)	0.05 (0.03 - 0.07)	0.08 (0.07 - 0.09)	0.14 (0.13 - 0.16)
	R	0.38 (0.36 - 0.40)	0.11 (0.10 - 0.13)	0.12 (0.11 - 0.13)	0.21 (0.20 - 0.23)
Ogooue-Maritime	U	0.28 (0.26 - 0.29)	0.09 (0.07 - 0.11)	0.04 (0.03 - 0.05)	0.15 (0.14 - 0.16)
	R	0.38 (0.36 - 0.39)	0.15 (0.14 - 0.17)	0.08 (0.07 - 0.09)	0.22 (0.21 - 0.23)
Woleu-Ntem	U	0.25 (0.23 - 0.27)	0.11 (0.09 - 0.13)	0.13 (0.12 - 0.14)	0.20 (0.18 - 0.21)
	R	0.35 (0.33 - 0.37)	0.17 (0.16 - 0.19)	0.17 (0.16 - 0.18)	0.27 (0.26 - 0.28)

As shown in Figure 14, using the unit-level model, the prevalence had a U-shape for each province and for each type of residence over time. For each province the prevalence over time was high for cluster in rural areas over time

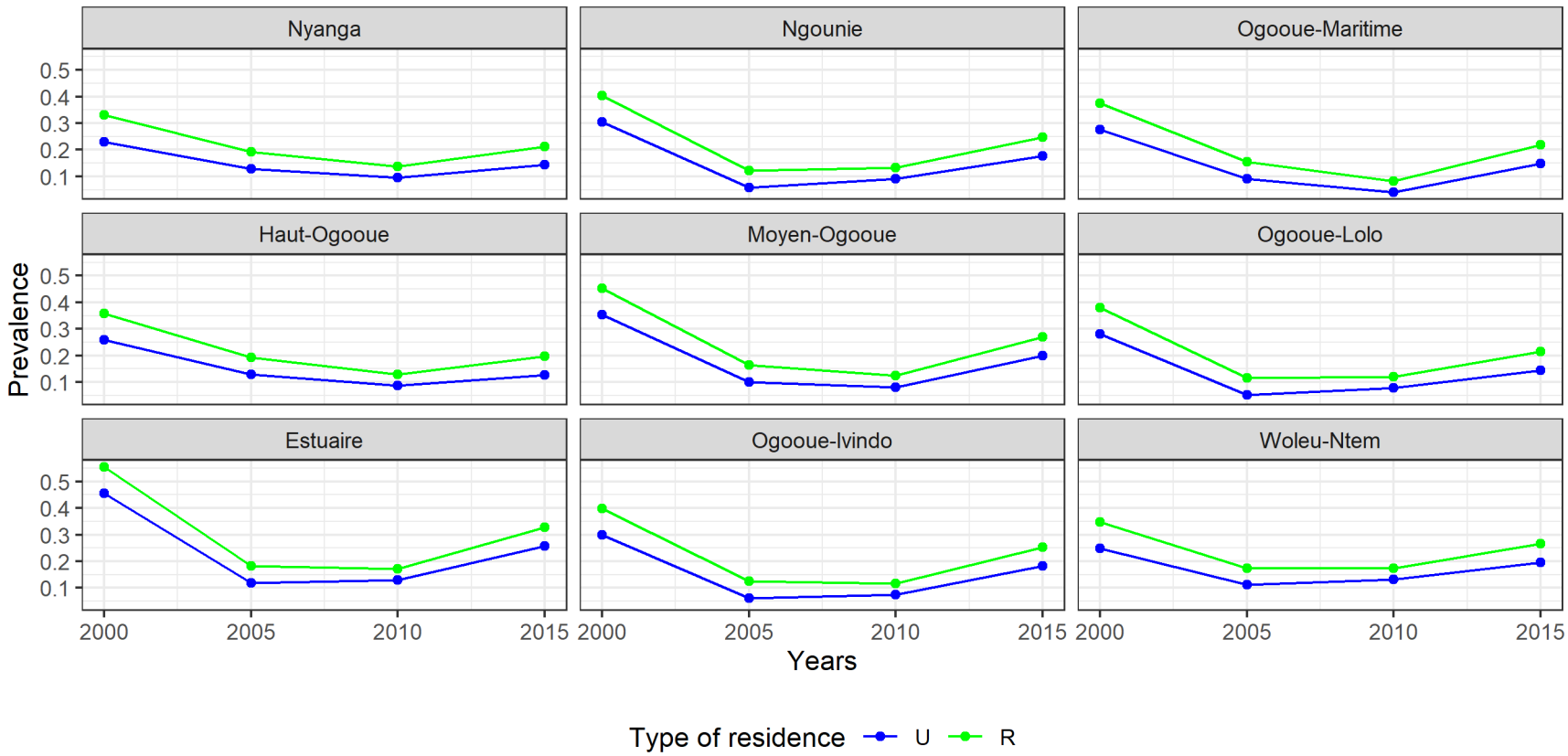


***Figure 14**: Smoothed prevalence by type of residence obtained from the unit-level model estimated for each year. CIs were omitted in this plot*

The spatial fraction or the mixing parameter was observed to vary slightly by year (Table 12), following the trend of the prevalence estimated over time.

***Table 12**: Distribution of mixing parameters over time*

Parameters	Year	Mean (SD)	95% CI
Phi for region struct	2000	0.35 (0.27)	(0.02, 0.91)
Phi for region struct	2005	0.32 (0.30)	(0.01, 0.96)
Phi for region struct	2010	0.30 (0.24)	(0.01, 0.86)
Phi for region struct	2015	0.46 (0.29)	(0.03, 0.96)
Phi for region struct	Overall	0.38 (0.27)	(0.02, 0.93)

3.4 Spatial autocorrelation of malaria prevalence

Table 13 showed that there was overwhelming evidence ($p < 0.001$) of a strong autocorrelation of malaria prevalence in the cluster of the households as the Moran's index I was positive and close to 1. This was almost constant over time.

Table 13: *Spatial autocorrelation malaria prevalence over time*

Year	Moran Index - outcome	Statistic 1	P-value
2000	0.88	11.81	<0.001
2005	0.87	11.62	<0.001
2010	0.83	11.04	<0.001
2015	0.83	11.12	<0.001

3.5 Hot spot and cold spot analysis

Although the spatial autocorrelation of the prevalence was strong and almost constant, the location of hot spots and cold spots identified appeared to change over time and only slightly in space. As shown in ***Figure 15***, significant hot spots were identified and the magnitude was varying over time. In the year 2000, a significant cluster of high prevalence was observed at the western part of Gabon at 99% confidence. In 2005, this cluster was reduced at the western part but afterwards appeared in the southern, and partly in the northern and the south-east part of Gabon. During the year 2010, a cluster of high prevalence was found at the northwest part of Gabon, and a small cluster of high prevalence was found at the east part of Gabon. In the year 2015, a great cluster of high prevalence at 99% confidence was found at the western part of Gabon, and a small cluster at 99% confidence was found at north-east part. Significant cold spots were found with magnitude changing over time. During the year 2000, the greatest cluster of low prevalence was found in the southern part, followed by the southeast, and the northwest. In 2005, they were found mainly at the center part of the country at 99% confidence. During the year 2010, a cluster with small prevalence at 99% confidence was found at the southwest, and the southeast part. In 2015, a great cluster of low prevalence was found at the southeast and the south part of the country.

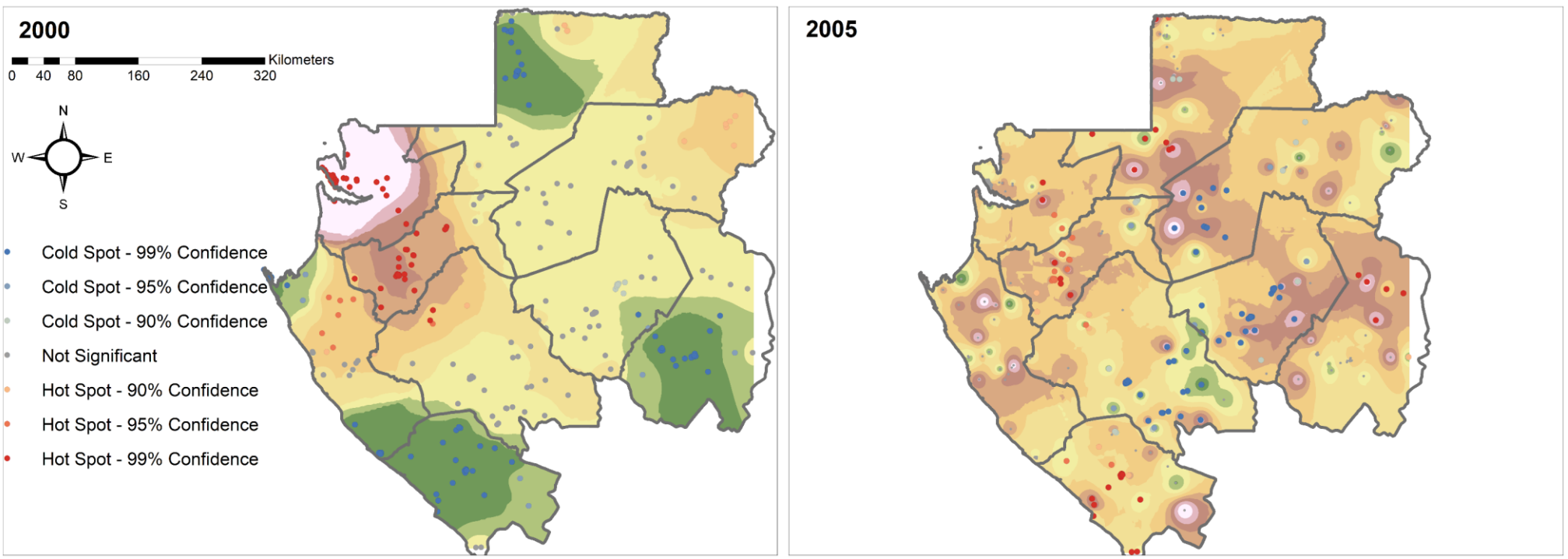


***Figure 15**: Malaria hot spots and cold spots identification over year in Gabon.*

3.6 Linear mixed model and variable selection

3.6.1 Pairwise correlation

Looking at the correlation plot (Figure 16), malaria prevalence was negatively associated with proximity to water, wealth index, annual precipitation, aridity, day land surface temperature, land surface temperature, ITN coverage, night land surface temperature. It was positively associated with maximum temperature, mean temperature, wet days, rainfall, and population count. The association between growing length, diurnal temperature, enhanced vegetation, minimum temperature and travel time did not provide any evidence of correlation. A very strong correlation was found between annual precipitation and aridity, mean temperature and minimum temperature, land surface and day land surface temperature, diurnal temperature and proximity to water, minimum temperature, diurnal temperature and proximity to water, and finally between annual precipitation and growing season length. Minimum and maximum temperature were removed from the model as the prevalence did not have a significant correlation with minimum temperature, but there was a strong correlation of minimum temperature and mean temperature. Therefore, the mean temperature was used in the model. On average, temperature was not really varying by year in the data as described previously.

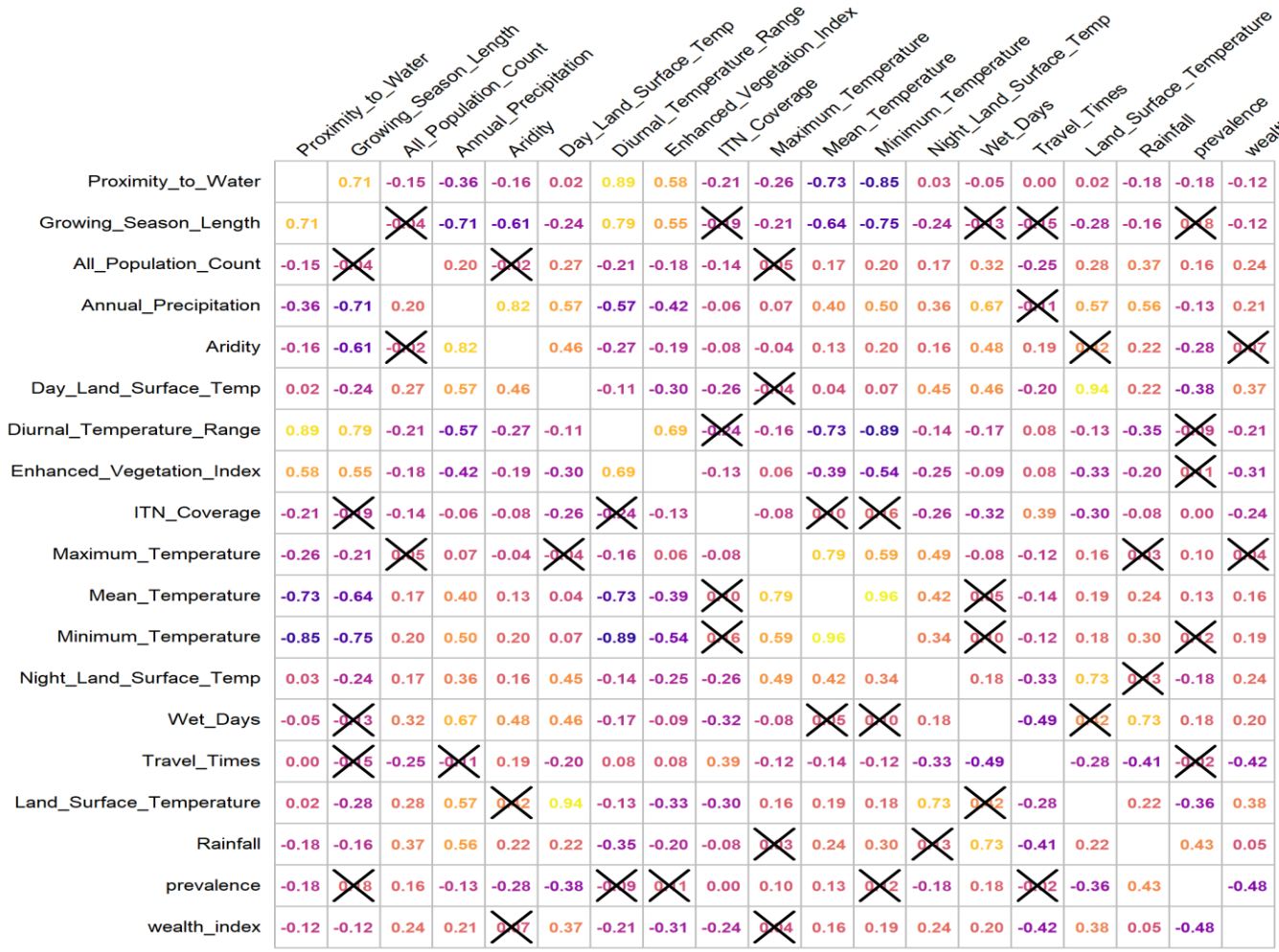


Figure 16: *Correlation matrix estimating the correlation between the covariates and malaria prevalence. The black cross means no evidence of association between two variables*.

3.6.2 Model building

In this part, a model that can best describe the relationship between covariates and the prevalence was built. Firstly, the model including the random slope on the year with the model including only the random intercept were compared and the VIF calculated. Secondly, after detecting the multicollinearity in the obtained model, variables with high VIF were excluded (table 15). Thirdly, from the selected model, model comparison between day land surface temperature and land surface temperature was made again, as previously a very strong association between these two variables was found. Then, the contribution of diurnal temperature in the model was evaluated. Finally, the comparison between the selected model and the model including the random slope on both ITN and year was done.

As shown in Table 14, the results suggested that using the model with the random slope on time improved the fit of the model significantly. The results need to be interpreted with caution, due to collinearity of the parameters and the assumptions.

***Table 14**: Random intercept vs random slope year*

Term	Npar	AIC	statistic	df	p.value
Random intercept	16	-2,654.082			
Random intercept + slope ITN	18	-2,846.571	196.4885	2	< 0.001

(1) Multi-collinearity

As shown in Table 15, collinearity was detected for the variables annual precipitation and aridity. The VIF was above 10, therefore annual precipitation was removed from the model. Also, the problem of multicollinearity was detected in the association of precipitation variable with diurnal day land temperature, land surface temperature, wet days and proximity to water. Therefore, precipitation was dropped from the model. The growing season was removed as there was some evidence of correlation with aridity as shown in the plot. However, aridity was used in the model because some studies have suggested an association with the prevalence of malaria (14). After the exclusion, the VIF was re-calculated for the remaining variables. Thus, in Table 15, the VIF was less than 10 for all the variables after removing some variables.

***Table 15**: Multicollinearity after removing variables*

Covariates	VIF	VIF (After removing variables)
Proximity to Water	7.44	6.87
Growing Season Length	5.44	-
All Population Count	1.65	1.72
Annual Precipitation	50.68	-
Aridity	27.10	7.23
Day Land Surface Temp	7.60	7.79
Diurnal Temperature Range	11.96	8.17
Enhanced Vegetation Index	2.47	2.80
ITN Coverage	5.53	4.15
Mean Temperature	3.63	3.33
Wet Days	7.43	6.40
Land Surface Temperature	8.16	8.71
wealth index	1.54	1.60
Rainfall	6.28	4.49

VIF = Variance Inflation Factor

(2) Comparison of the models: day land surface and land surface temperature

After evaluating the multicollinearity, the following models were compared: **1)** model with covariates + day land temperature and **2)** model with covariates + land surface temperature. The idea was to reduce as much as possible the multicollinearity by removing potential variables which were not improving the fit of the model. As a result, from the Table 16, the model with day land surface temperature was the best as the AIC was the lowest.

***Table 16**: Model + Day land surface vs Model + Land surface temperature - AIC*

Model	AIC
Covariates + Day Land surface temperature	-2836.79
Covariates + Land surface temperature	-2825.85

(3) comparing the models: including and excluding diurnal temperature

In Table 17, the AIC for the model without diurnal temperature was the lowest. Hence, diurnal temperature was excluded from the model. The following variables were the final variables selected: *Proximity to Water, all population count, aridity, day land surface temperature, enhanced vegetation index, ITN coverage, mean temperature, wealth index, wet days and rainfall*. For this analysis the main variables were ITNs coverage, rainfall, aridity and wet days.

***Table 17**: model with diurnall vs without diurnal*

Model	AIC
Covariates + Diurnal temp	-2836.79
Covariates - Diurnal temp	-2838.78

(4) Varying the intercept and slope on ITN coverage and year

After selecting this model, to assess if the fit of the model was better when the random slope is introduced on the variable year and ITNs coverage, the two models were compared. In Table 18, the model including slopes on both ITNs coverage and year was better than the model including the random slope only on year (ANOVA; p-value < 0.05). The random slope was considered only for the variables ITN and year. In fact, if rainfall, aridity and the other variables were used with a random slope, the number of observations in the data would not allow to fit a good model. Also, according to our description, ITNs coverage appeared to be the most varying in space and time.

***Table 18**: Random slope year vs Random slopes ITN + year*

Term	npar	AIC	Statistic	df	P-value
Random slope ITN	15	-2838.78	NA		
Random slopes ITN + year	17	-2867.35	32.57	2	< 0.001

3.6.3 Diagnostics of the selected model

As shown in Figure 17, the variance was not constant as the vertical spread was not the same for different values on the x-axis. The probability normal plot either for the random effect or the residuals showed that there was a slight departure from normal distribution as points were more deviating from the ideal line. However, the log transformation was better in terms of reducing the variance to become more constant, and stretching the points better than the untransformed data. The departure from the normal distribution either for the random effect or the residuals was reduced.

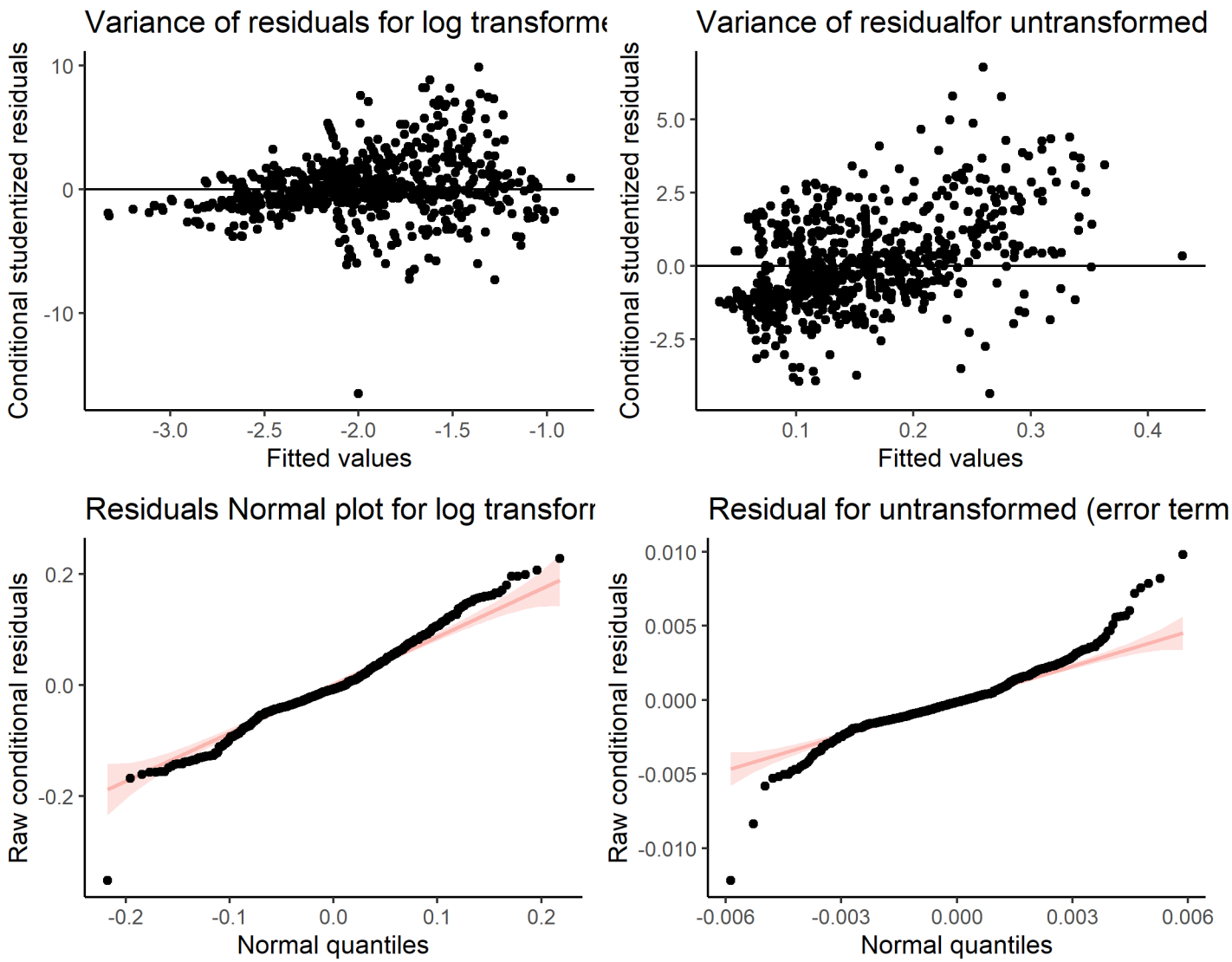


Figure 17*: Comparison of residuals from the log transformed and the untransformed prevalence. The plot at the left and right represents the residual variance of log transformed and for untransformed malaria prevalence respectively*

As shown in Figure 18, after using the square root transformation, the residuals and the random effects became more normal, and the variance became more constant than observed with the log transformation. Using the inverse square root transformation, the random effect for year and ITN did not approximate the normal distribution more than the square root transformation (see appendix **Figure 27** ().

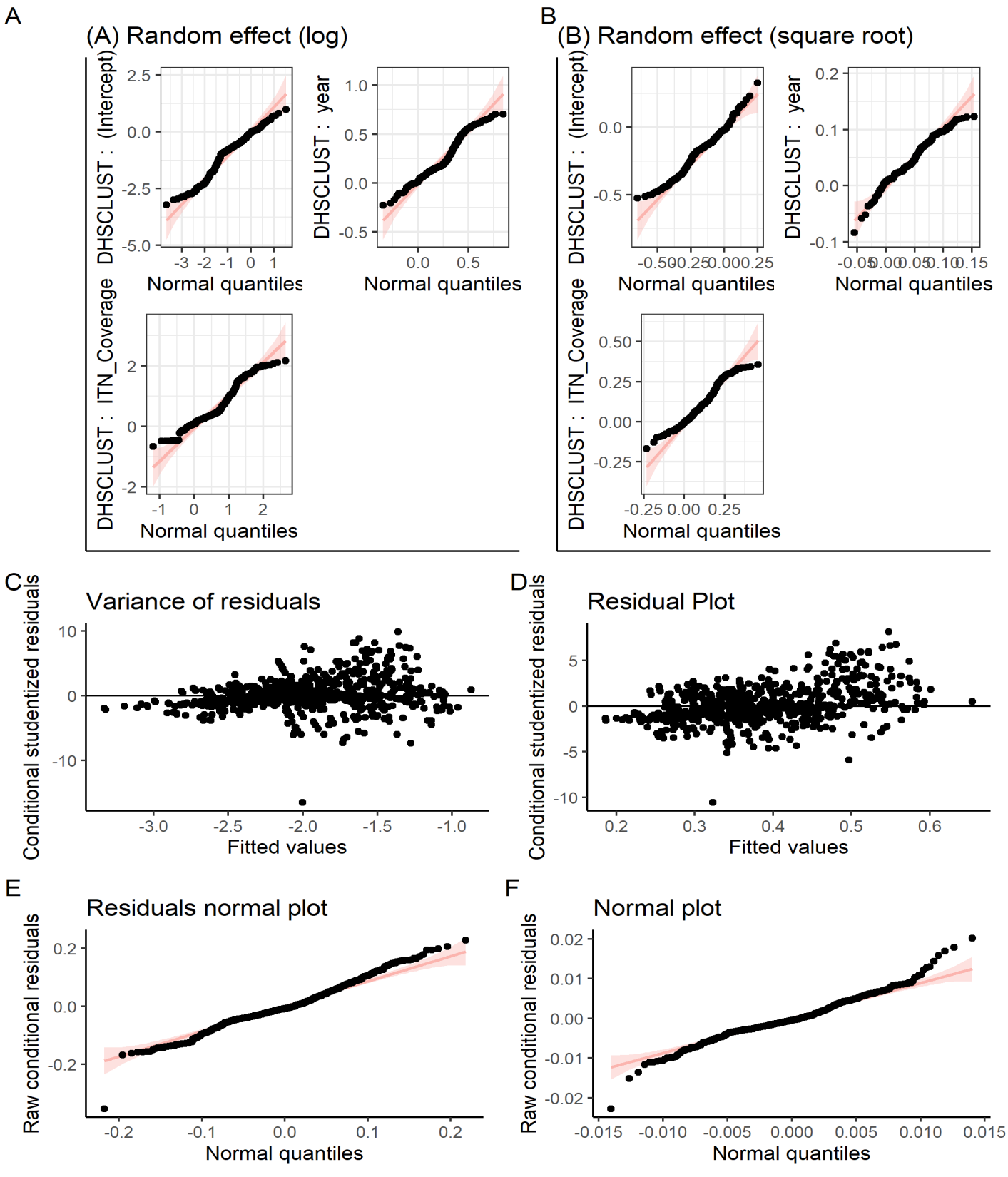


Figure 18: *Comparison of residual's variance and distribution for the diagnostic of the model between log and square root transformation. Panel A and B represent the random effects, C to E the residuals of model using the log, D and F the residual of the model using square root transformation on the outcome.*

Final Conclusion:

Despite little deviation from the assumption, the final model was interpreted using either the log or the square root transformation, because the linear model is robust for small deviations. The interest here was in the random effect, hence the model with square root transformation could be used. Indeed, with this model, the assumptions were less violated for the random effect (68). According to this model (Figure 19), the average effect of the prevalence was different from cluster to cluster when not adjusting for the other variables. This result was significant because the credible interval for the intercept did not contan 0. The effect of year and ITNs coverage on the prevalence by cluster was significant. It was represented by those black points who did not cross the red line with their credible interval. Using the model with log transformation, the same result was also observed. However, no significant effect on the year was obtained for the model using the inverse square root transformation (see appendix).

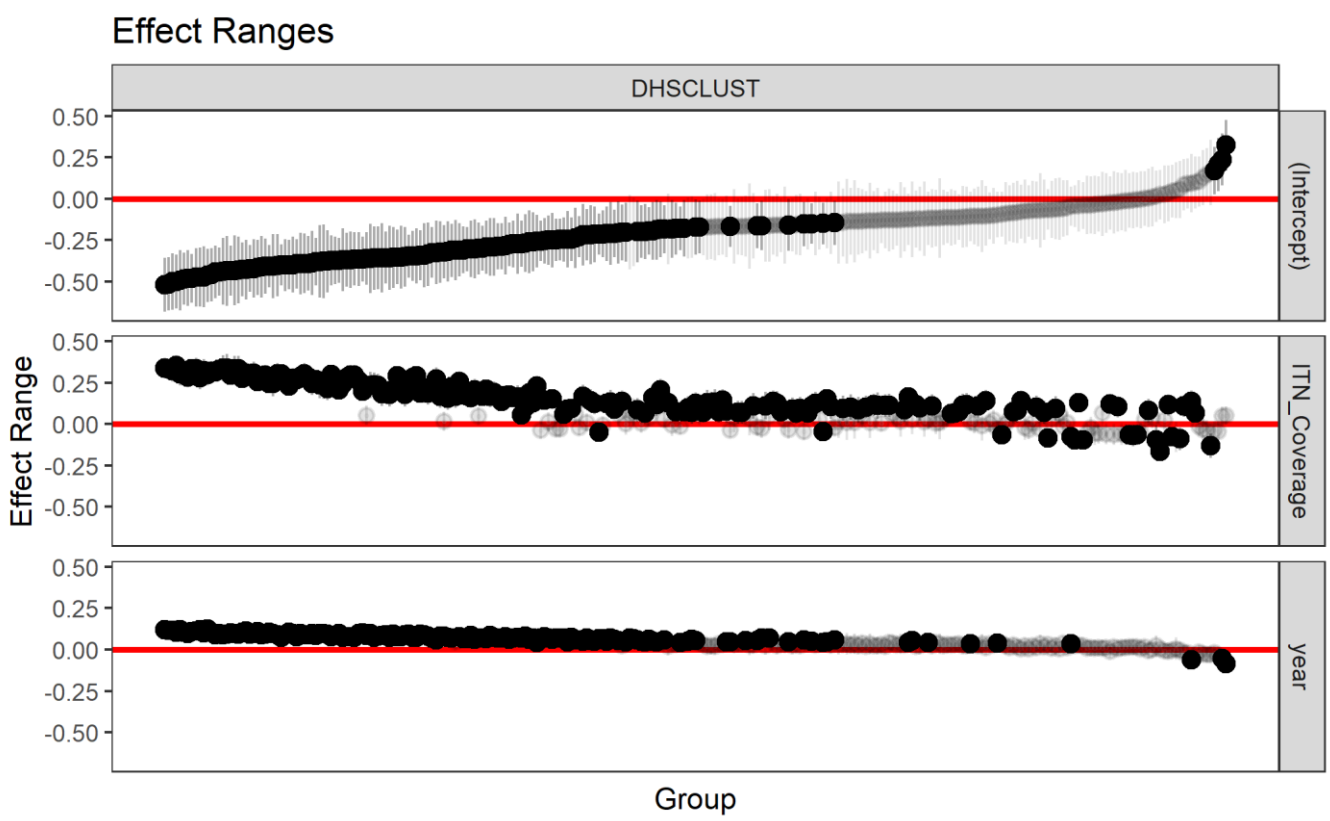


***Figure 19**: Effect of ITNs coverage and year on the prevalence from each cluster*

3.7 Multiple linear regression of malaria prevalence

Multiple linear regression for different years was performed using variables selected previously. However, as there was no data available for the variables ITNs coverage and all population count for the year 2000, the linear regression was conducted for this year without these two variables. The results with and without extreme values found on the variable population count were compared. The results of the regression were not modified except the coefficient which became significant after removing the extreme values. Hence, all the analysis was performed without these values. For all the years, the F-test showed with overwhelming evidence that at least one variable was associated with the outcome among all the variables selected. The assumption of linearity was assumed looking to the scatterplot presented in Figure 9.

After using either the log or square root transformation, the assumption of normality and homoscedasticity were met (Figure 20).

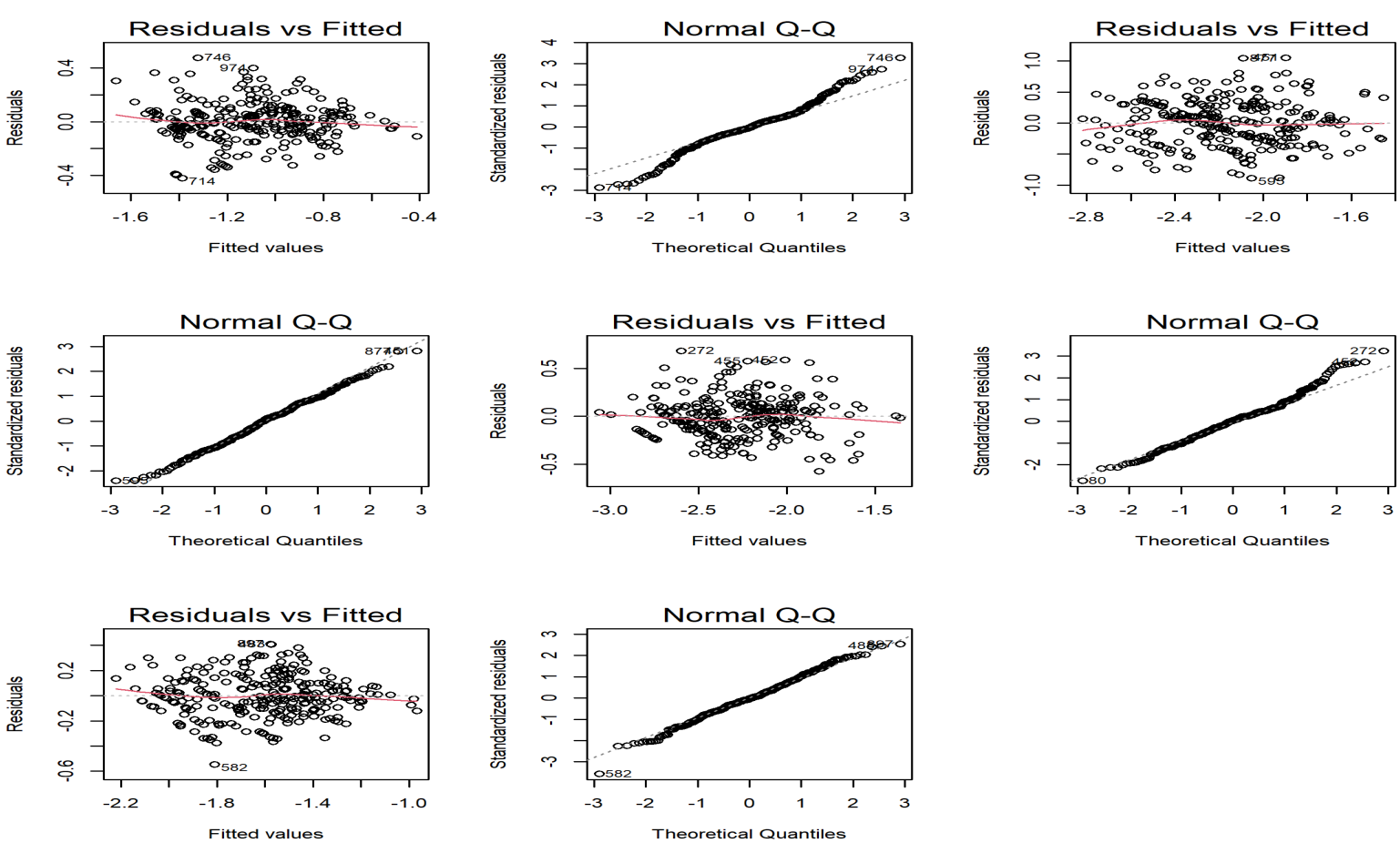


Figure 20: *Residuals inspection: The pair of plots represent the diagnostic for each year respectively 2000, 2005, 2010 and 2015*

For the assumption of independence on the residuals, overwhelming evidence of strong spatial autocorrelation in the residuals for each year was found in Table 19: Spatial autocorrelation on residuals and malaria prevalence over time. After including variables to explain the variation of the prevalence, the spatial autocorrelation observed on the prevalence was slightly reduced but not enough. A small effect of time was observed from the year 2000 to 2005, 2005 to 2010, 2010 to 2015, that is, a difference of spatial autocorrelation of 0.12, 0.26, and 0.06 respectively. This difference was almost constant when considering only the dependent variable.

***Table 19**: Spatial autocorrelation on residuals and malaria prevalence over time*

year	Moran Index – outcome	statistic 1	P-value	Moran Index - residuals	statistic 2	P-value
2000	0.88	11.81	<0.001	0.71	9.66	<0.001
2005	0.87	11.62	<0.001	0.84	11.18	<0.001
2010	0.83	11.04	<0.001	0.57	7.71	<0.001
2015	0.83	11.12	<0.001	0.62	8.31	<0.001

This OLS regression showed that ITNs coverage was significantly associated with the prevalence, in such a way that one-unit increase was associated with 82% decrease of the prevalence in a linear fashion. Its magnitude was varying slightly by year. The prevalence was increasing by 12% with a one unit increased of wet days. Wealth index was also increasing with the prevalence. In fact, almost all the variables were significantly related with the prevalence of malaria. It was also noted that the coefficients were varying slightly by year (Table 20).

***Table 20**: Ordinary least square regression for each year with logtransformation*

Variables	Estimate 1 - 2000		Estimate 2 – 2005		Estimate 3 - 2010		Estimate 4 – 2015	
	Beta	P-value	Beta	P-value	Beta	P-value	Beta	P-value
(Intercept)	-4.9	< 0.001	-0.70	0.64	-0.92	0.31	-0.04	0.95
ITN coverage			-1.74	< 0.001	-1.87	< 0.001	-1.02	< 0.001
Mean temperature	0.13	< 0.001	-0.09	0.10	-0.06	0.06	-0.01	0.71
Wet days	0.12	< 0.001	0.28	< 0.001	0.26	< 0.001	0.08	< 0.001
Night land surface temp			0.06	0.17	0.03	0.19	0.00	0.90
Day land surface temp	-0.044	< 0.001	-0.01	0.52	-0.05	< 0.001	-0.07	< 0.001
Wealth index	-0.0000007	0.0018	0.00	< 0.001	0.00	< 0.001	0.00	< 0.001
Rainfall	0.019	< 0.001	0.02	0.16	0.01	0.31	0.05	< 0.001
Proximity to water	0.0078	< 0.001	-0.02	< 0.001	-0.01	< 0.001	-0.01	< 0.001
Enhanced vegetation index	-0.0011	0.56	0.01	0.31	0.01	< 0.001	0.00	0.57
All population count			-0.10	< 0.001	-0.06	< 0.001	-0.04	< 0.001
Aridity	-0.0047	0.069	-0.06	< 0.001	-0.03	< 0.001	-0.02	< 0.001

3.8 Geographical Weighted Regression

Using the GWR model, the multicollinearity problem was not detected in the global and the local level using the VIF. However, by using the condition number, as advised in the literature, the multicollinearity problem was detected at the local level, because the CN was larger than the cut off value of 30. In fact, interpreting the results by considering only the VIF would lead to the misinterpretation of the model. In addition, the R^2 was very high (Table 21), and the predicted variable values were very close to the true values (see supplementary), arising the overfitting problem. Although the Moran index ran on the residuals was very small, and have shown no significant spatial autocorrelation, the reliability of the results was still uncertain. To deal with this, the spatial varying coefficient method in Bayesian framework, as advised in some work, was used (62).

***Table 21**: Diagnostic of GWR method*

Diagnostics for local multicollinearity				Diagnostics on spatial autocorrelation	
model	Rsquared (Min - Max)	local CN (Min - Max)	VIF (Min - Max)	Moran's I index	P-value
All values	0.8-0.97	295.67-297.89	1.53-5.17	0.10	0.08

3.9 Spatial varying coefficients

No evidence was found to support that the relationship for some variables was varying by location. However, only day land surface temperature and ITNs coverage were plotted to represent the variation showed in Figure 21. All the other variables were not showing any pattern (see supplementary). The patterns observed assumed that the increase of the ITNs coverage was associated with the decrease of the prevalence in the center of the country, and a slight increase around. An increase of the day land surface temperature had was associated with the decrease of the prevalence in the center part. But they were not significant.

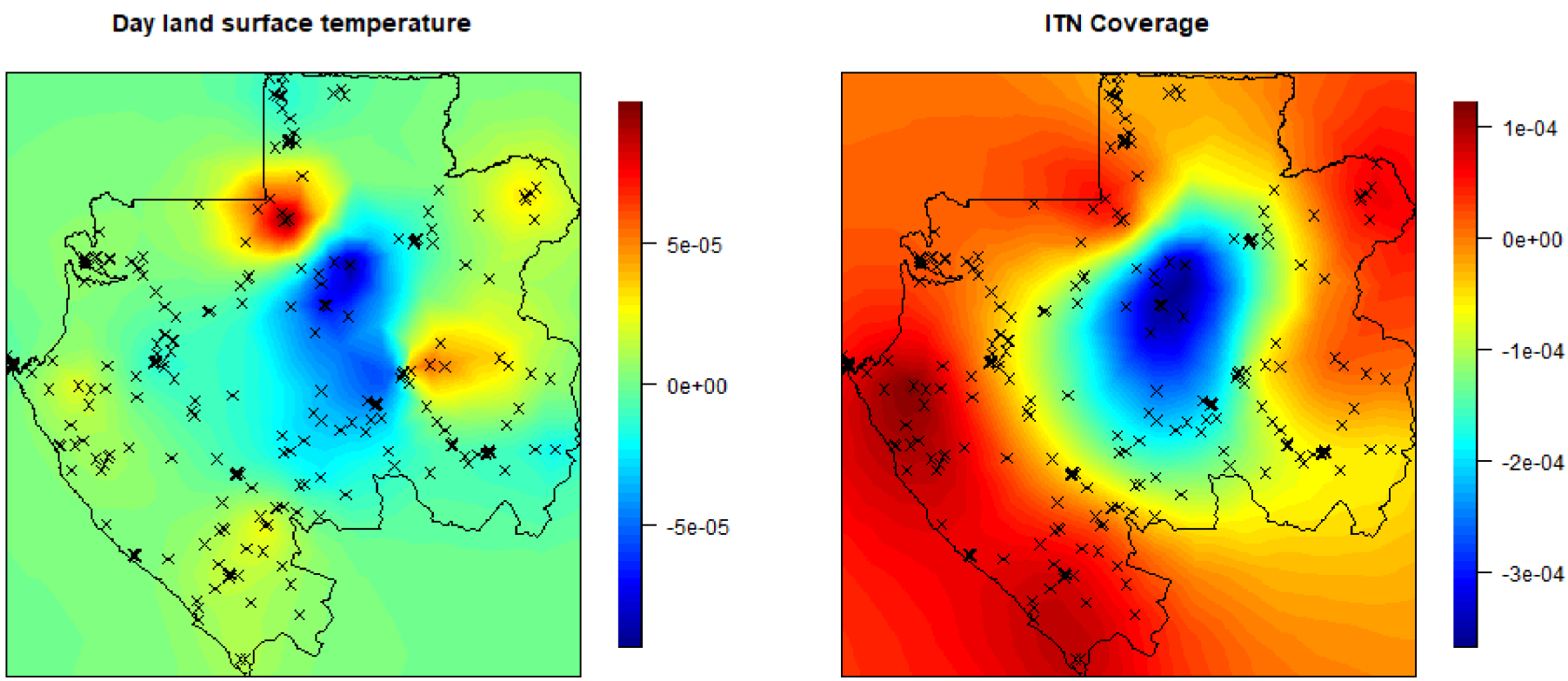


***Figure 21**: Spatial varying coefficient - visualization of the prevalence of malaria with Day land surface temperature (left) and ITN (right)*

3.10 Spatio-temporal model

3.10.1 Linear models

In this part, spatio-temporal models were run as the appropriate method to deal with spatial effect observed. The model was split into four sub-models: M1 (model without covariates), M2 (model with only ITNs coverage, population count and wealth index), M3 (model with only ecological variables including type of residence) and M4 (the full model). The variation of the spatial parameters was observed in each model as suggested in building model to explain the disease prevalence (69). In all the sub-models, the variable year was included. From M1, after adjusting with all the covariates, the spatial variation was reduced by 76% (Table 22). In fact, when introducing only other variables with ITNs coverage, the spatial variation was reduced by 68%, while when introducing only the environmental variables, the spatial variation was reduced by 88%. The best model with small DIC was the full model (Table 24). The model was used based on logtransformation, but already backtransformed as presented in all the remained tables.

The value of the range showed that the data had a strong spatial correlation which was decreasing slowly. This was present up to 142 km. For all the models, the spatial variance was obviously greater than the measure error by more than 95%. The coefficients for time model (ρ) were almost the same for all the three sub models. While strong evidence of the spatial effect in the variance of the prevalence was found, no evidence for the time effect was found, because the credible interval contains 0, even if one-year increase would yield in 12% decrease of the prevalence. Also, the coefficient for time was too high and the credible interval too tight, suggesting that the value for the next year was almost the same as the previous time.

Using the full model, increasing the ITN by 1%, decreased significantly malaria prevalence by about 52%; increasing the night land temperature by 1-degree increased malaria prevalence by 3%. As the population increased, the prevalence also decreased slightly and significantly by 2% for one percent increase. Wealth index was also significantly associated with the prevalence, but had a very small effect. As year was increasing the prevalence was also decreasing, but it was not significant (Table 22).

***Table 22**: Coefficients of the spatial model*

Type of covariate	Parameter	Mean (95% CI)
M1	(Intercept)	-32 (-86, 227)
	Year	-16 (-36, 9.6)
M2	Year	-15 (-30, 4.1)
	ITN coverage	-53 (-63, -40)
	Wealth index	-0.0000037 (-0.000024, 0.000016)
	All population count	-1.4 (-2.2, -0.63)
M3	Year	-15 (-36, 12)
	Mean temperature	-4.9 (-16, 7.8)
	Wet days	-6.2 (-16, 5)
	Day land surface temp	-1 (-2.3, 0.19)
	Enhanced vegetation index	0.23 (-0.16, 0.62)
	Rainfall	-0.52 (-2.1, 1.1)
	Proximity to water	1.4 (-9.4, 13)
	Aridity	1 (-0.45, 2.5)
	Rural	6.6 (0.17, 11)
	Urban	-6.2 (-9.8, -0.18)
M4	ITN coverage	-52 (-63, -39)
	Mean temperature	-22 (-40, 2.7)
	Wet days	-4.6 (-15, 6.8)
	Year	-12 (-29, 9.7)
	Day land surface temp	-0.77 (-2, 0.43)
	Night land surface temp	2.7 (1.1, 4.3)
	Enhanced vegetation index	0.25 (-0.15, 0.64)
	Wealth index	0.0000023 (-0.000019, 0.000024)
	Rainfall	-0.61 (-2.2, 0.97)
	Proximity to water	-0.33 (-11, 11)
	Aridity	1.1 (-0.4, 2.6)
	All population count	-1.7 (-2.5, -0.85)
	Rural	5.2 (0.3, 11)
	Urban	-5 (-10, -0.29)

*M1 = model without covariates but only year;
M2 = model with only ITN, population count and wealth index;
M3 = model with only ecological variables including type of residence;
M4 = the full model

3.10.2 Geo-additive model

In this geo-additive model, two approaches were developed. The first one assumed that the coefficient of AR(1), which is ρ, is zero. This approach is named replicated model. The second one assumed that ρ is not 0, and is named space-time correlation model. In this approach ρ is estimated. For each method, four models were run, and compared based on the DIC, in order to assess the effect of the ecological variables in the model on the malaria prevalence variance and the spatial distribution, before and after adjusting for the other variables. For variables with non-linear relationship, the smoothers were presented.

(1) Replicated model (rho = 0)

As shown in Table 24, the model with the lowest DIC was the full model (M4), followed by M3. This was when the spatial model was used with all the environmental variables assumed to have a non-linear relationship with the prevalence. In Figure 22 and Figure 23, the prevalence was decreasing as the ITNs coverage, population count, aridity, and day land surface was increasing, increasing slowly with increasing of rainfall, increasing strongly with the increase of wet days. The prevalence was decreasing with low values and increasing with high values of the proximity to water. It was almost constant with the increase of EVI with small variation, decreasing with the mean temperature for values less than 24°C, and became constant for values between 24°C and 27°C, then decreasing again.

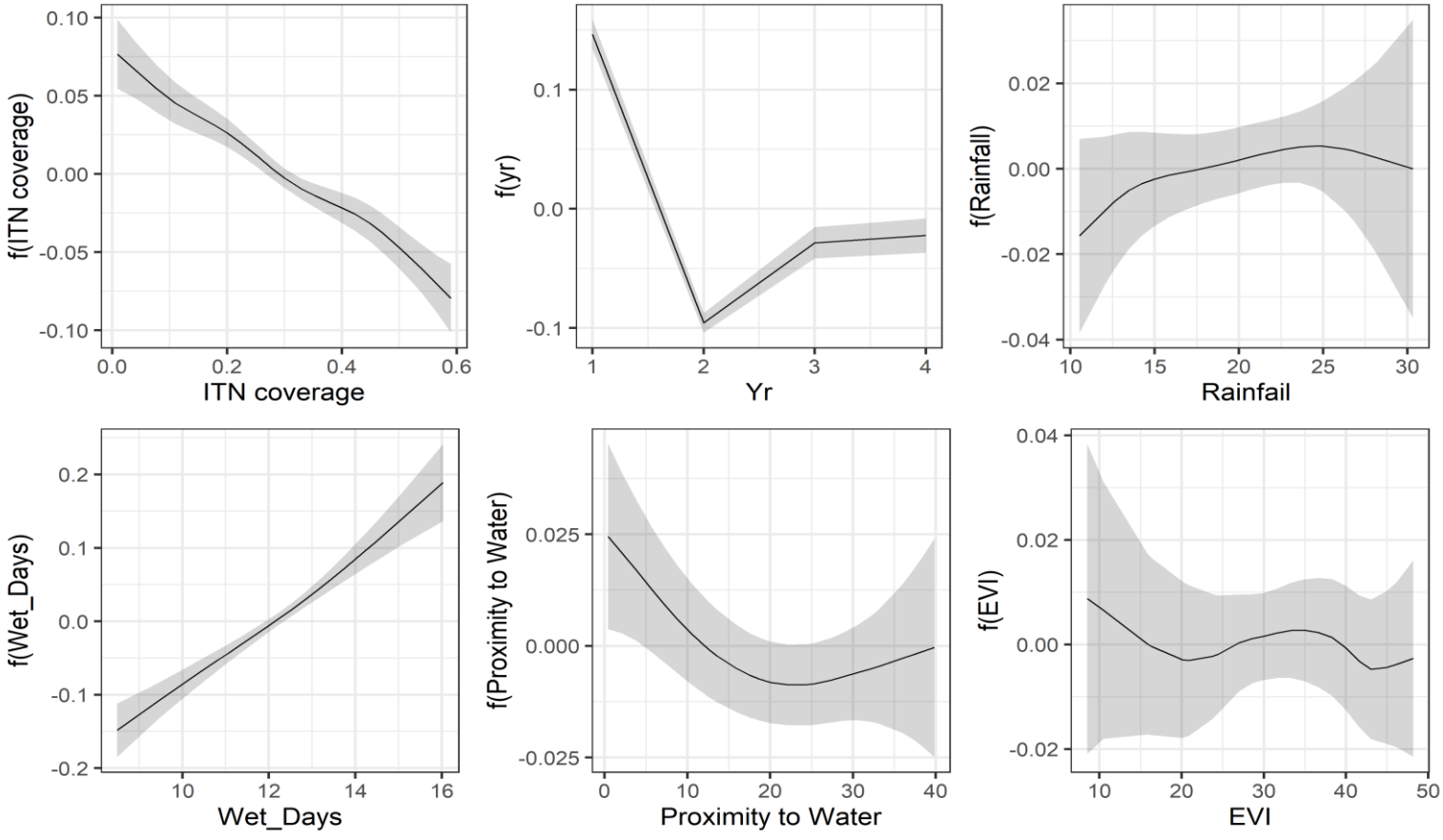


***Figure 22**: Smoother of malaria prevalence and covariates estimated from GAM with replicated method*

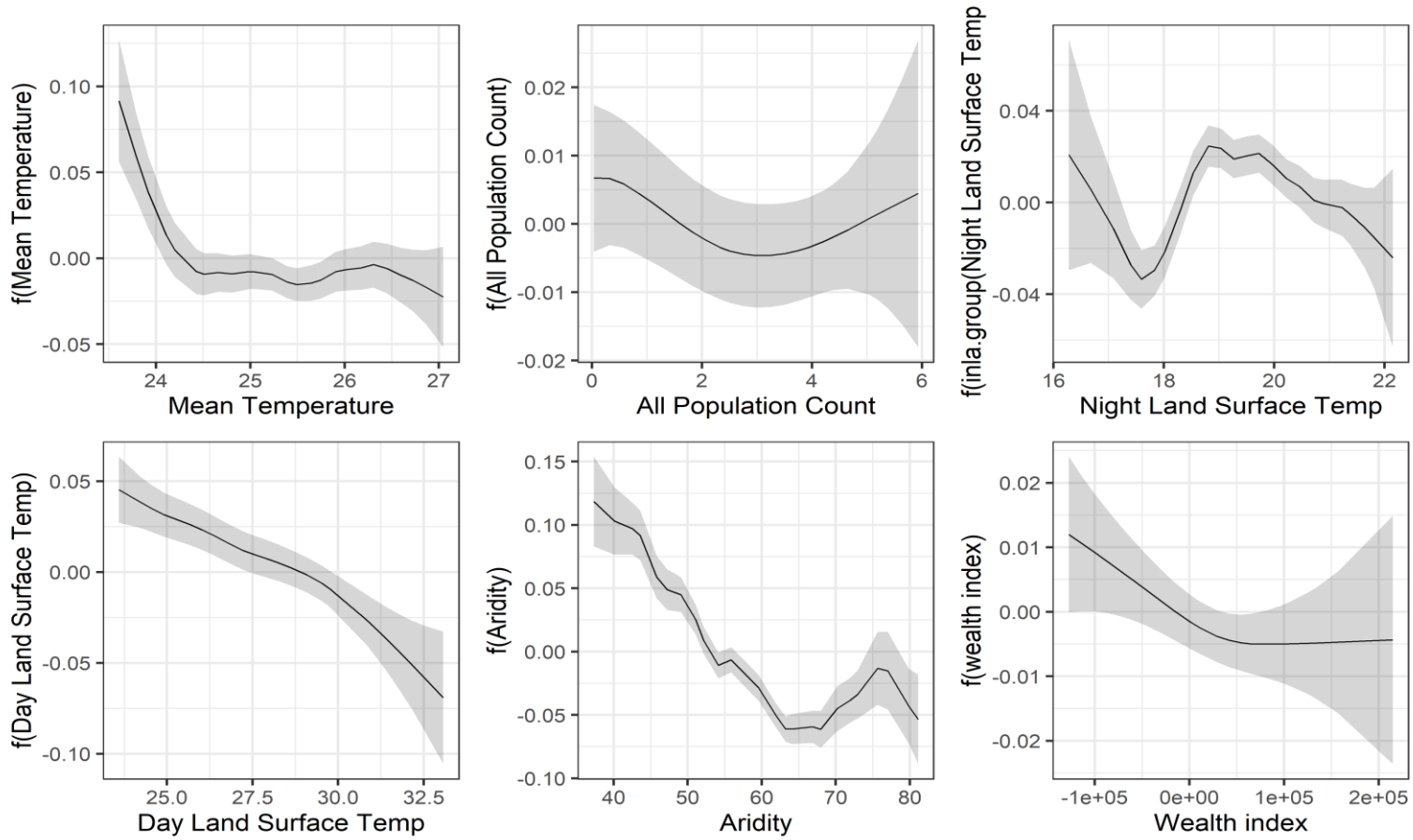


***Figure 23**: Smoother of malaria prevalence and covariates estimated from GAM with replicated method (cont)*

(2) Space-time correlation with year

The model considering that the spatial effect was correlated with year was computed, and ρ, the coefficient in AR(1)) was estimated. Using this GAM, the best model in overall was the full model, followed by M2. The spatial variance reduction was 9%, 26% with M3 and M2 respectively, with the lowest range found in M2 which was 278 km. The spatial reduction for the full model was 27% (Table 23).

***Table 23**: Spatial parameters GAM with all variables considered as non-linear*

		Spatio-temporal OLS model		GAM with non-linear variables	
Model	**Parameter**	**Mean (SD)**	**95% CI**	**Mean (SD)**	**95% CI**
M1	Measurement error	0.000064 (0.0000077)	(0.00005,0.00008)	0.00053 (0.000064)	0.00042;0.00067
	Spatial variance	0.17 (0.036)	(0.11,0.25)	11 (2.2)	7;16
	Range	333 (35)	(270,406)	332 (31)	276;396
	Time coefficient (a)	0.97 (0.0044)	(0.96,0.98)	0.99 (0.0016)	0.98;0.99
M2	Measurement error	0.000026 (0.0000045)	(0.000018,0.000036)	0.00047 (0.000062)	0.00036;0.0006
	Spatial variance	0.055 (0.01)	(0.038,0.077)	8.1 (1.7)	5.2;12
	Range	154 (15)	(128,185)	278 (27)	229;336
	Time coefficient (a)	0.97 (0.0044)	(0.96,0.98)	0.99 (0.0015)	0.99;0.99
M3	Measurement error	0.000071 (0.000012)	(0.00005,0.000095)	0.0006 (0.000082)	0.00045;0.00077
	Spatial variance	0.021 (0.0053)	(0.012,0.033)	10 (2.3)	6.4;15
	Range	153 (21)	(114,197)	342 (36)	279;422
	Time coefficient (a)	0.93 (0.014)	(0.89,0.95)	0.99 (0.0018)	0.98;0.99
M4	Measurement error	0.000023 (0.0000043)	(0.000016,0.000032)	0.00053 (0.000073)	0.0004;0.00068
	Spatial variance	0.044 (0.011)	(0.026,0.068)	8 (1.9)	4.9;12
	Range	142 (17)	(111,176)	290 (33)	230;360
	Time coefficient (a)	0.96 (0.0062)	(0.95,0.97)	0.99 (0.0016)	0.98;0.99

*GAM = Geo-additive Model; OLS = Ordinary Least Squared

3.10.3 GAM with linear and non-linear variables

The previous graphs showed that wet days and ITNs coverage had linear tendency. Hence, the model including them as having a linear relationship with the outcome was run, and compared with the previous model to see whether it would improve the fit of the model. In Table 24, comparing all those models from linear model to additive non-linear model, it was found that model with some variables as linear and some others as non-linear improved only the fit of the model with environmental variables and where not the best. Therefore, the model including only non-linear variables was the best fitting the data with the lowest DIC (-7888.7) on the full model, followed by M3. Using the model setting ρ to 0, M3, the model where the covariates were the environment variables, was found to be better than M2. A very small variation of the time was found. Since no evidence of the effect of the time on the prevalence was found, the model without time by taking the mean value for all the years was run.

***Table 24**: DIC for all GAMs*

Model	Linear relation variables with rho estimated	SGAM non-linear all environment variables with rho = 0	SGAM non-linear all environment variables with rho estimated	SGAM mixed linear non linear with rho estimated
M1	-4435.66	-3668.78	-6878.09	-6878.10
M2	-4565.98	-3680.14	-7746.42	-7544.43
M3	-4325.32	-4320.39	-6748.46	-6811.30
M4	-4497.76	-4385.93	-7888.71	-7801.80

- **GAM on summarized data over year**

Before interpreting this model, the spatial econometric model was run on the same data and the models were evaluated. The spatial econometrics models were introduced to understand the way the change of a covariate in a nearby clusters or in a cluster can affect the prevalence in another cluster or in the same cluster.

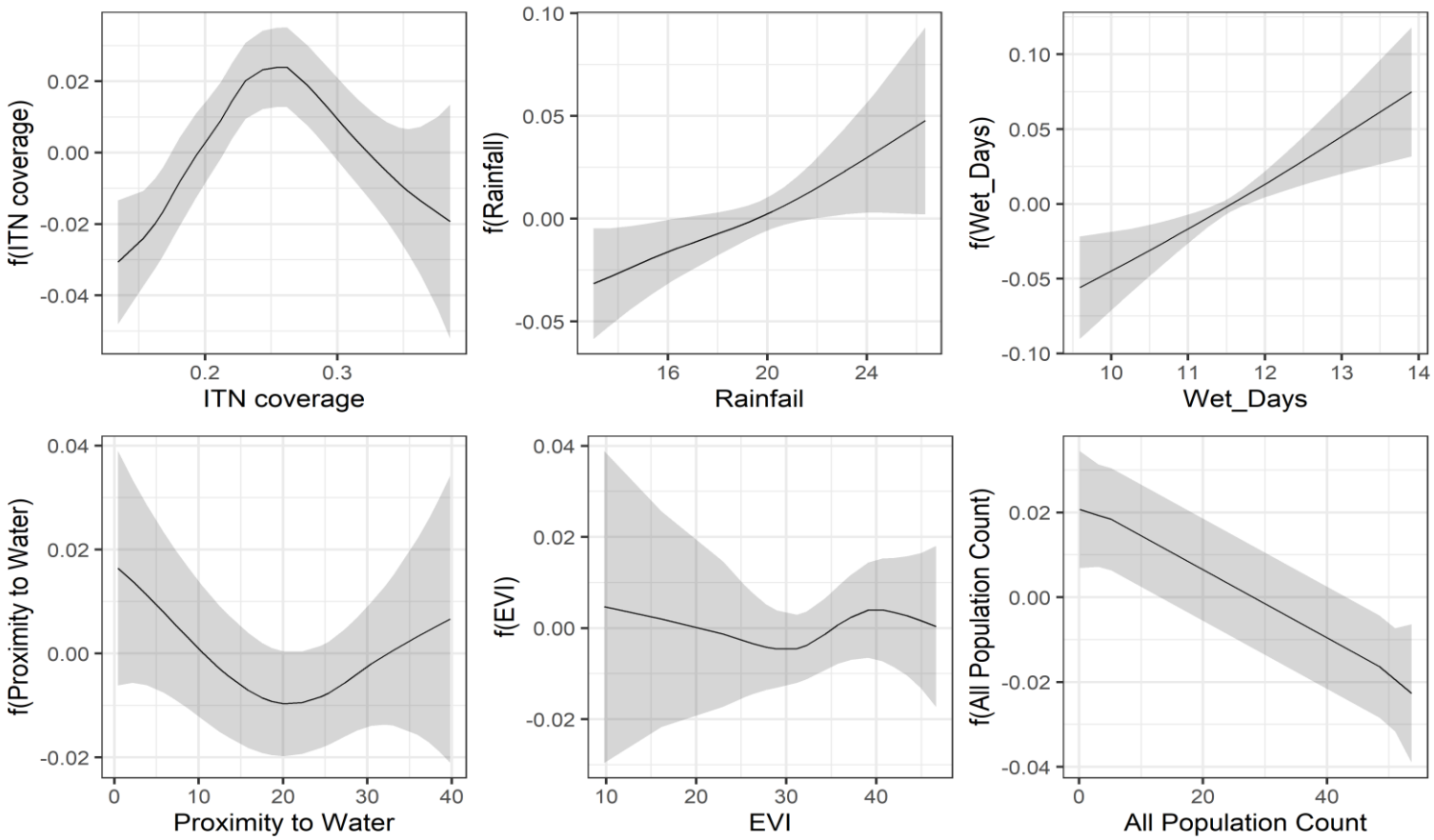


***Figure 24**: GAM on summarized data without year effect*

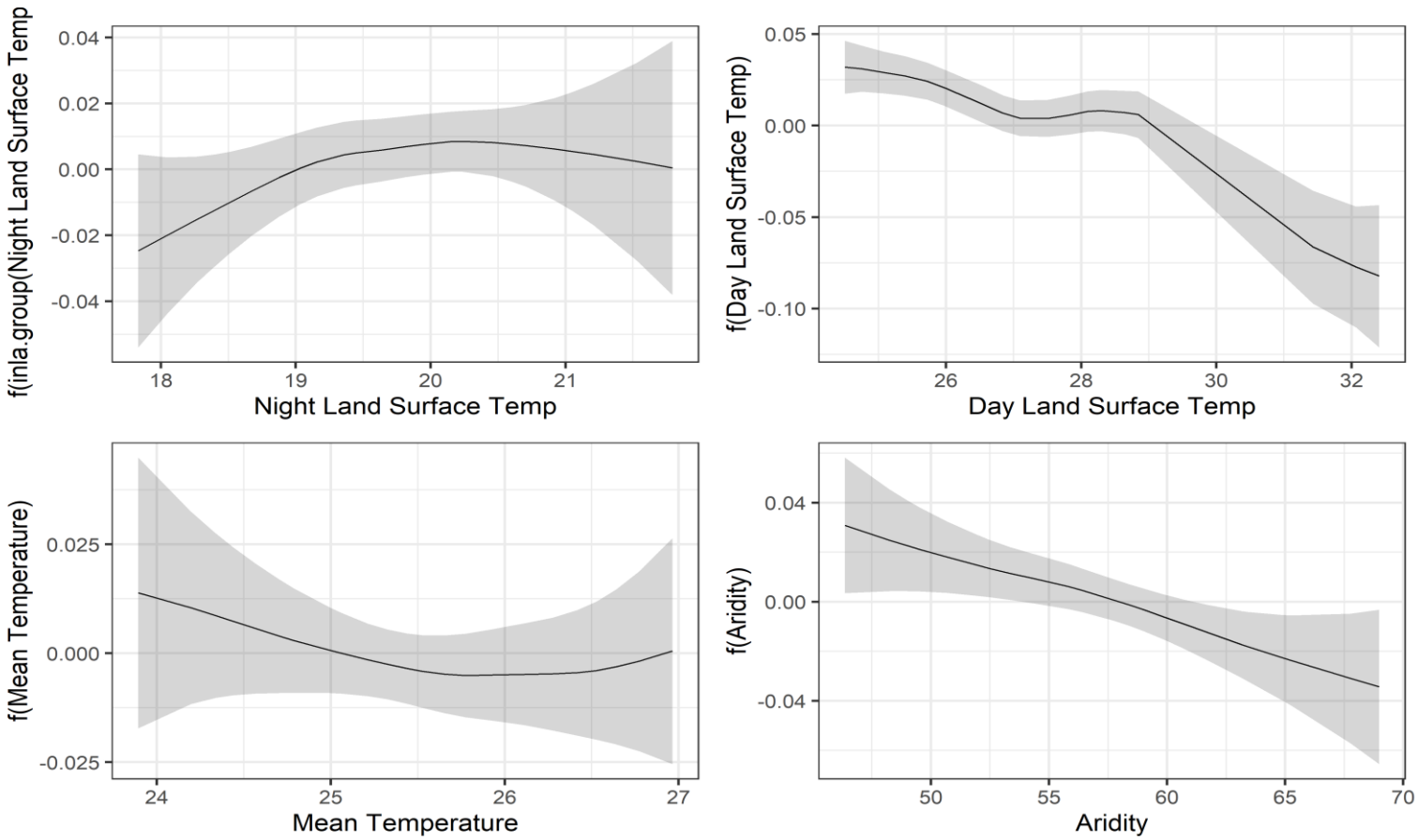


***Figure 25**: GAM on summarized data without year effect*

3.11 Spatial econometric Model

3.11.1 Computing models

As shown in Table 25, the SEM and the SLM suggested that excepted ITNs coverage, all the variables were significantly associated with the prevalence as found with the non-spatial OLS regression, while SDM and SDEM suggested that only wet days, type of residence and day land surface were significantly associated with the prevalence.

***Table 25**: Coefficients and significance from all spatial models*

Covariate	SEM		SLM		SDM		SDEM	
	Coefficient 1	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Mean temperature	-0.0038	0.53	0.00036	0.93	-0.012	0.13	-0.0089	0.21
Wet days	0.03	<0.001	0.022	<0.001	0.033	<0.001	0.032	<0.001
ITN coverage	-0.04	0.47	-0.064	0.05	0.11	0.25	0.14	0.13
Night land surface temp	0.012	0.02	0.0078	0.04	0.0057	0.38	0.0021	0.72
Urban - rural	-0.038	<0.001	-0.035	<0.001	-0.038	<0.001	-0.041	<0.001
Day land surface temp	-0.016	<0.001	-0.0092	<0.001	-0.018	<0.001	-0.015	<0.001
Wealth index	-0.000000041	0.26	-0.000000076	0.06	-0.000000049	0.17	-0.000000059	0.14
Enhanced vegetation Index	0.0012	0.02	0.00061	0.11	0.00083	0.16	0.001	0.06
Rainfall	0.0057	<0.001	0.0023	0.05	0.0017	0.55	0.0012	0.65
Proximity to water	-0.0013	<0.001	-0.00073	0.01	-0.0024	0.11	-0.0024	0.08
Aridity	-0.0029	<0.001	-0.0019	<0.001	-0.0024	0.15	-0.0018	0.19
All population count	-0.00054	0.12	-0.00017	0.55	-0.00069	0.06	-0.00043	0.21

In Table 26, the test of spatial dependence given by the Moran index, or by looking to the Lagrange multiplier (LM) test for the spatial dependence in the error term (not presented here), the test was significant for the SML only. This suggested that this model still had unaccounted spatial dependence despite the spatial lag on the outcome. After using the SDM, by adding the lag on the covariates, the p-value became greater than 0.05. Therefore, no evidence of spatial dependence was observed again. After that, the SDEM, combining the lag on the errors and the covariates was run.

***Table 26**: Checking spatial autocorrelation on the residuals*

Model	Statistic	P-value	Method	Alternative
SEM	-0.07	0.81	Monte-Carlo simulation of Moran I	greater
SLM	0.15	0.03	Monte-Carlo simulation of Moran I	greater
SDM	-0.03	0.68	Monte-Carlo simulation of Moran I	greater
SDEM	-0.03	0.64	Monte-Carlo simulation of Moran I	greater

As shown in Table 27, the DIC of the spatial lag model was smaller than the DIC of the GAM (-3125.58 vs -1744.26) ran previously. Therefore, the spatial lag model was the best. Using the ML (Maximum Likelihood), the best model was the SDEM with the lowest AIC. Comparing the maximum likelihood and the Bayesian fit with INLA, approximately the same values for fixed or rho were obtained. All spatial autocorrelation coefficients were positive and significant. A great difference between INLA and ML was observed using the SLM.

***Table 27**: Comparison of rho from ML vs INLA*

Method	Mean rho ML	P-value[a]	P-value[b]	AIC ML	Mean rho INLA (SD)	95% CI	DIC
SEM	0.51	<0.001	<0.001	-1242	0.50 (0.04)	0.42-0.58	-3125.57
SLM	0.35	<0.001	<0.001	-1210	0.84 (0.00)	0.84-0.85	-3125.46
SDM	0.50	<0.001	<0.001	-1250	0.47 (0.03)	0.39-0.52	-3125.57
SDEM	0.51	<0.001	<0.001	-1252	0.50 (0.04)	0.42-0.58	-3125.58

[a]P-value from Likelihood Ratio test;
[b]P-value from Wald test. Both tests are for evaluating if the spatial correlation coefficient is different from 0

3.11.2 Impact

Lastly, the non-linear model with SDEM was run and the models M1, M2, M3 and M4 were evaluated by looking to the spatial variation or the spatial parameters. As shown in Table 28, as the indirect effect of ITNs coverage was positive and significant, therefore for a particular cluster, increasing the ITNs coverage in its nearby clusters was found to be significantly associated with the decrease of the prevalence in this cluster by 21%. However, no evidence of the direct effect was found. The indirect effect of the mean temperature was positive and significant. Therefore, for a particular cluster, an increase of the mean temperature in its nearby clusters was significantly associated with an increase of the prevalence in that cluster. The direct and the total effect was not significant. The direct effect for wet days was positive and significant, hence, an increase of wet days in a cluster was associated with an increase of the prevalence in the same cluster by 3%. Increasing the day land surface temperature in a cluster was significantly associated with a decrease of the prevalence in the same cluster by 2%. A significant association between the prevalence of malaria and the type of residence was found. There was a significant difference between the type of residence. For aridity, as the total effect was positive and significant, on average, the prevalence was decreased very slightly (0.3%) when aridity was increased at any cluster. No evidence of significant impact of night land surface temperature, EVI, rainfall, proximity to water and population count was found.

***Table 28**: Comparison INLA and Maximum likelihood for SLM and SDM - impacts for SDEM*

	SDEM from ML			SDEM from INLA		
Covariates	**Total effect**	**Direct effect**	**Indirect effect**	**Total impact (95% CI)**	**Direct impact (95% CI)**	**Indirect impact (95% CI)**
Mean temperature	0.01	-0.0089	0.019	0.01 (-0.0063;0.027)	-0.009 (-0.025;0.0069)	0.019 (0.0012;0.037)
Wet days	0.03	0.032	-0.0016	0.03 (0.014;0.047)	0.032 (0.012;0.052)	-0.0018 (-0.023;0.02)
ITN coverage	-0.07	0.14	-0.21	-0.07 (-0.2;0.059)	0.14 (-0.067;0.34)	-0.21 (-0.4; -0.0055)
Night land surface temp	0.0033	0.0021	0.0012	0.0034 (-0.011;0.018)	0.0023 (-0.011;0.015)	0.0011 (-0.012;0.015)
Urban - rural	-0.049	-0.041	-0.0083	-0.049 (-0.072; -0.026)	-0.041 (-0.054; -0.028)	-0.0082 (-0.022;0.0052)
Day land surface temp	-0.0088	-0.015	0.0065	-0.0089 (-0.015; -0.0025)	-0.015 (-0.02; -0.011)	0.0065 (0.0016;0.011)
Wealth index	-0.00000011	-0.000000059	-0.00000051	-0.00000011 (-0.00000027;0.000000046)	-0.00000006 (-0.00000015;0.000000029)	-0.000000053 (-0.00000015;0.000000041)
Enhanced vegetation Index	0.00068	0.001	-0.00033	0.00068 (-0.00079;0.0021)	0.001 (-0.0002;0.0022)	-0.00032 (-0.0017;0.001)
Rainfall	0.0035	0.0012	0.0024	0.0035 (-0.00078;0.0078)	0.0012 (-0.0045;0.007)	0.0023 (-0.0033;0.0078)
Proximity to water	-0.00055	-0.0024	0.0018	-0.00056 (-0.0017;0.00057)	-0.0024 (-0.0054;0.00063)	0.0018 (-0.0013;0.0049)
Aridity	-0.0032	-0.0018	-0.0013	-0.0032 (-0.005; -0.0013)	-0.0019 (-0.005;0.0013)	-0.0013 (-0.0044;0.0019)
All population count	-0.000065	-0.00043	0.00036	-0.000055 (-0.0011;0.001)	-0.00042 (-0.0012;0.00033)	0.00037 (-0.00035;0.0011)

3.11.3 Estimation

As shown in Table 29, from M0, the spatial variation was reduced by 33%, 13%, 8% when considering the full model, the model with ecological variables and the model without ecological variables respectively.

***Table 29**: Spatial autocorrelation coefficient of SDEM linear and non-linear compared for effect of variables*

Covariate	**Parameter**	**Mean rho (SD)**	**95% CI**
Linear covariates	M1	0.58 (0.032)	(0.51; 0.64)
	M2	0.52 (0.04)	(0.44; 0.6)
	M3	0.42 (0.048)	(0.33; 0.51)
	M4	0.39 (0.049)	(0.29; 0.48)
Non-linear covariates	M1	0.58 (0.032)	(0.51; 0.64)
	M2	0.53 (0.037)	(0.46; 0.6)
	M3	0.5 (0.04)	(0.42; 0.58)
	M4	0.5 (0.04)	(0.42; 0.58)

As shown in Figure 26, as the population was increasing, the prevalence of malaria was decreasing. It was the same as aridity, day land surface temperature, mean temperature, ITNs coverage and the proximity to water were increasing. Observing the curve for the ITNs coverage, it appeared that malaria prevalence started to decrease only after 20% of the coverage. It was suggesting also that malaria prevalence was increasing as rainfall, EVI, wet days, and night land surface temperature were increasing.

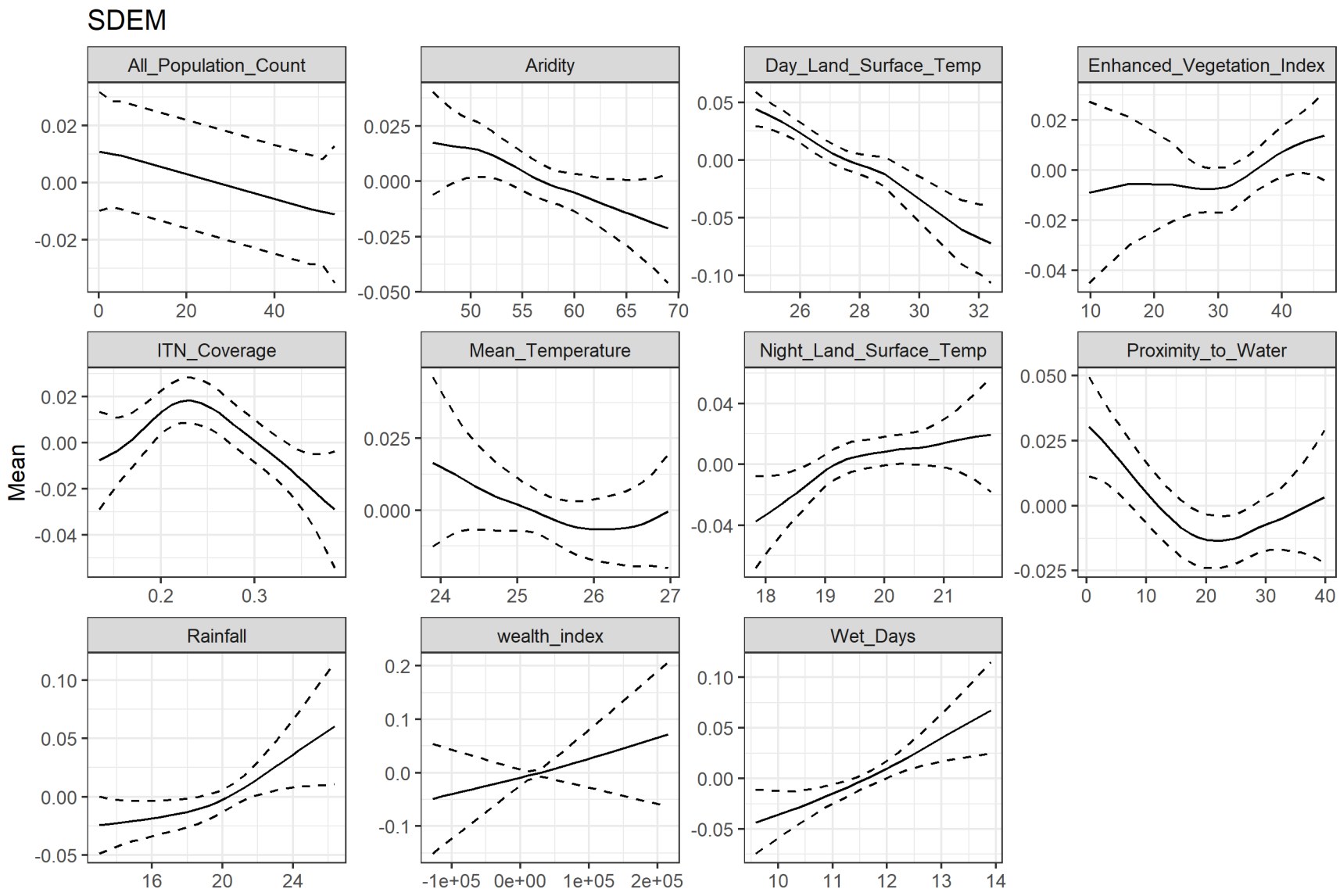


Figure 26: SDEM with non-linear relationship between malaria and covariate

CHAPTER 4 GENERAL DISCUSSION AND CONCLUSION

4.1 Discussion

The aim in this work was to use Bayesian spatio-temporal analysis to identifying the spatial patterns over time by 1) estimating the prevalence of malaria when borrowing strength from neighbors, 2) identifying hot spots and cold spots, and 3) to evaluate the factors that were influencing the prevalence of malaria in Gabon for children between 2-10 years of age, using DHS survey data. This is important in order to have an insight on the spatial and temporal distribution and to inform public health policy in Gabon. As a complex survey was used, the importance of including the design of the survey in the study when doing the estimation was highlighted.

This study considered ecological factors and variables such as ITNs coverage, population count, wealth index and type of residence. To grasp the complexity of malaria prevalence and the surveys indicators, various model strategies were used, as one suggested in the study of Giorgi et al (69). Space-time smoothing model based on the areal data and the unit-level data were used for small area estimation (SAE). The comparison to find the model fitting well the data on these two types of data were made. Spatial geo-statistical analysis was conducted to assess the risk factors using a Bayesian framework. Different types of spatial model were employed. Spatio-temporal models using SPDE approach was used including also the GAM, in order to allow for the flexibility in the relationship of the factors and malaria prevalence. It was also extended to SDEM for spillover effect, and all based on INLA approximation to avoid long-runs of the computer-intensive MCMC method. SPDE and INLA approaches were used in many studies for infectious disease (48,70). In this approach (SPDE), the distance between locations reflects the spatial correlation, assuming then an isotropic Gaussian process.

Before the estimation of the prevalence, the models including the space-time interaction and the one without it were compared. Four sub-models were used to compare the effect of one group of variables before and after adjusting with another group of variables in the prevalence. This was done from the model containing only the intercept and the variable year. All variables included in the regression methods were considered with a VIF value less than 10, from the gaussian random effect model. Some variables were evaluated on the basis of their contribution in the model in the case they were found to have strong correlation with some important variables. The significance was based on the fact that the credible interval did not contain 0 or was not of different sign for Bayesian approach, and p-value less than 5% for frequentist approach. Mainly the prior used was the PC prior with parameters determined as a rule of thumb or the default values in INLA. This is shown in the work of Sampson et al (63). This study investigated for the first time the spatio-temporal distribution of malaria taking into account the whole area in Gabon. The model introducing the interaction between the time and the space was used with the areal data because of the smallest DIC observed.

The main findings while estimating the prevalence over time was that, the prevalence is still high. The highest prevalence of malaria was observed in Estuaire province, in the northwest part. It was decreasing between 2000 and 2010, and increasing slightly after

2010. The lower values of the prevalence were found at south of the country including Ogooue-Lolo, Nyanga and Ngounie province. This decrease between 2000 and 2010 was shown in a study investigating the changing risk of PfPR. It was conducted based on 49 African countries, and 82% of them showed this decrease (65). The pattern of U-shape was also shown in a study conducted in Malawi. They found a decrease of 50%, then followed by a rebound of 25% (71). For almost all the areas, the prevalence seemed to increase again after 2010. According to residence type, the prevalence was found to be higher in rural zones over time. Similar results for type of residence were found in a study conducted in Mozambique (72). They found that children living in urban had lower risk of malaria compared to those living in rural zones. The decrease of malaria prevalence can be explained by the fact that new interventions were introduced around 2000 and 2006 in urban area. In addition, the access is sometimes easy for those who live in urban areas, with equipped houses, but also, by the fact that in some regions some people have free medical care from private health structure like from oil companies (8).

For the high prevalence in rural area, often in some areas there is lack of amenities which implies that some health workers are discouraged to go there when sent by the government. Therefore, there is isolation of some parts and the health care services delivery are very poor (3). For the small trend of increase, there is need to investigate with more data. However, it was also noticed the slowdown of activities from National Malaria Control program (NMCP) to support the small trend of increase during some years, or a phenomenon known as ‘delayed malaria’, when the interest is in a particular age group, and it happened that the intervention is not sustained (73). In fact, in the studies conducted in Zambia and Zanzibar, as the intervention was maintained, the rebound was not shown (77, 78). The difference of the prevalence between the north and the south part was also found in many studies estimating the prevalence of malaria (14). This can possibly be explained by variables like the coverage of control tools, density of the vectors, climatic variables, immunity or other possible raisons.

The contribution of space in the variation of this prevalence was high in 2015, the year during which the new trend of the increase was observed. It suggested that, using the unit-level model, up to 46% of the variation of malaria prevalence could be explained by a spatial component, while 41% with the areal-level model. Although our confidence interval was wide, this result was in the line with a study conducted in Rwanda (49). They found that the spatial component was contributing by 49% at 2014 and 41% at 2018 when estimating malaria relative risk. Our results demonstrated also that it is better to use the survey design of the study and the smoothing model, as it reduced the variance due to the fact of introducing another random effect for borrowing strength from neighbors. This was in line with many studies in the literature (51). The use of the unit-level model reduced the uncertainty in the credible interval of the estimate, while approximately the same estimate was obtained with the areal model. In fact, in the areal model the geographic unit used was province. Hence, the number of units was very small. However, the unit-level model was based on clusters, therefore the number of units in the analysis was increased from 9 to 316.

The finding suggesting that hotspots and cold spots were changing over time and space was also supported by the study conducted in Senegal, Tanzania and Kenya (50). In their study, they analyzed the spatio-temporal distribution of hot spots and the association with some environmental variables. They found that the variation of hot spots was associated with rainfall and EVI. However, for our study, this relationship was not evaluated, and it seemed like hotspots were high in the same period of high prevalence.

From the OLS regression, all the variables were related significantly to malaria prevalence, with high spatial autocorrelation in the residuals. However, using some spatial methods such as GAM or SDEM, only few of those variables were related to malaria prevalence. This difference on the significance, emphasized the fact that, in the presence of spatial autocorrelation, the results of the OLS were biased as the standard error was not correct. Therefore, underestimate the parameters, as pointed by some work (59). Also, as the I Moran's index was positive, close to 1 and significance, clusters who were near tend to have similar values. This was also found in the study conducted in Nigeria using cluster as the unit of the analysis through DHS data (14). In all spatio-temporal model ran, the change of the year did not provide evidence of impacting the prevalence, while it was showing a decrease. This can be due to a small observation of time. Maybe also on the time the data were collected, as malaria is well known to be affected by seasonality.

Among the all models performed to assess the effect of covariates on the prevalence, SDEM was the best as it includes the lag on both the covariates and the error. The findings revealed that, some ecological variables were significantly related with malaria prevalence. They were also impacting considerably the spatial distribution when they were introduced in the model before, and slightly changed after adjusting with the other variables. However, the magnitude on the prevalence were small. More investigation needs to be done with other studies on this magnitude in Gabon. It was found from SPTOLS that, after adjusting the ecological variables with the other variables, the spatial distribution changed by 14% with SPTOLS model, less than 1% with the GAM, and by 7% with the SDEM. All these methods showed that the spatial distribution of the prevalence is importantly affected by the ecological variables in the cluster level, as after introducing the other variables the spatial variance did not change too much. The difference showed could proof the weakness of SPTOLS, as it was considering that all the variables had linear relationship with the outcome.

Using the SDEM model, the findings of this study revealed that, ITNs coverage was the most important variable affecting the variation of malaria prevalence in Gabon with an indirect effect. This result was similar to the finding of a study conducted by Hawley et al (74). In their study, by investigating the effect of ITNs in the nearby households, they found that ITNs have a protective effect on households that did not receive ITNs but were situated in 300 meters around the ones that did. These results from our study highlighted two important points: firstly, the consequence of the spatial effect between two points or locations close. In fact, when the close neighboring was involved in the prevention against the disease, it could protect the close neighbors; secondly, the important role of malaria control policy on the prevalence of malaria even on the cluster level. In the non-linear relationship showed in the graph, when the summarized data on year were used, surprisingly, our study suggested that malaria prevalence started to decrease

once ITNs coverage was above 20%. This is quite different with the estimation when the panel data was used. This surprising result was also highlighted by the GAM. This finding highlighted the fact that under 20% of coverage, ITNs coverage associated with the decrease of the malaria prevalence in Gabon. This need to be ascertained by other studies too.

Mean temperature was found to have a negative indirect effect on the prevalence. This negative effect was consistent with many studies that did not use spatial econometrics model such as the SDEM used here (75–77). While these results were supported by these studies, this result may be unexpected because the increase of the temperature should pose a problem in the cycle of *P. falciparum*. In fact, when the temperature is above 28°C or 30°C the sporogonic cycle is affected, consequently the number of vector which would survive after this cycle would be reduced, then would lead possibly to the decrease of transmission or malaria prevalence. However, the mean temperature in Gabon, according to this study was between 24°C and 27°C which is in the range of the optimal temperature for both vector and parasite, that is, 25°C and 27°C. Thus, this results were suggesting that the country is a suitable area for mosquitoes development. Observing this, it is legitimate to say that the number of mosquitoes under this mean temperature may increase, implying an increase of the probability of being bitten, and then the increase of the prevalence among children. As an indirect effect, this was suggesting that the increase was from the change in the neighborood clusters. This can be due to the fact that when surounded by many clusters in which the mean temperature has changed, even if the temperature did not change inside the considered cluster, the increase of the number of mosquitoes or infected people coming from the nearby clusters will certainly imply that those vectors or hosts of malaria will invade the cluster of households and increase the transmission rate in this cluster, because of interaction of people in the close clusters. Therefore this will likely be reflected by the increae in the PfPR.

Aridity was found to decrease the prevalence of malaria. The increase of aridity in any cluster led to a small and significant decrease (0.3%) of malaria in another cluster. This was also supported by the non-linear relationship. It showed a fast decrease of malaria prevalence with the increase of the aridity. This can be explained by the fact that, with the increase of the aridity, there will be lack of the suitable sites, hence this threaten the lifespan of the adults (78). The same result was found in a study using the SEM and the SLM (14).

On average, day land surface was significantly associated with a slight decrease of malaria prevalence. In fact, when day land surface temperature was increased in some clusters, the prevalence decreased slightly. This finding was supported by the study conducted in Uganda (79). In their study, they found a decline of malaria at higher temperature 35°C, but they did not use the same model to identify the spillover effect.

The finding suggesting that wet days was significantly associated directly with a slight (3%) increase of malaria prevalence, was also found in a study conducted in the Democratic republic of Congo (77). In this study, they focused on the association of climatic factors

in malaria morbidity based on data from 2001 to 2019. They found that one-day increase of the rainy day was associated with 7% of increase in malaria cases. In fact, wet days is also related with rainfall as the number of rainy days, and it was found here to have a significant positive correlation with the rainfall. Rainfall leaves sites for mosquito to grow larvae, therefore contribute in expanding the population. When rainfall is not heavy, the odds of being bitten is possibly increased. In contrast, when there is heavy rainfall and long dry periods, there is possibility to restrain the availability of moisture important for larvae (80).

The population count and the night land surface temperature were significantly associated with a slight decrease of malaria prevalence for a 1% increase, using the spatio-temporal model for OLS. The result of night land surface temperature was expected and was in line with some studies (81). For the population count, the decrease in malaria was surprising and not expected. However, can be supported by the idea that, usually, there is a lot of focus at urban areas, and it is where there is easy access to many things including access to health care. However, in our study, this relationship could be biased as the relationship suggested with malaria prevalence was linear. These associations disappeared when running other spatial models.

A significant relation between rainfall and malaria prevalence was expected. However, no evidence of the relationship was found. This lack of evidence was also found in study conducted in Bangladesh (82). This can be due to the presence of heavy rainfall in some periods in Gabon, which can wash away the breeding sites of mosquitoes, and therefore hide this relationship. In our data the mean rainfall was between 1400 and 2000 mm. Using the spatio-temporal GAM, the spatial correlation between the clusters is decreasing slowly and becomes less than 0.1 after 290 km. This suggested a very slow decrease of the spatial effect compared to SPTOLS. In fact, in Gabon there is an aggregation of households almost everywhere, and the distance between those households are very small, hence allowing the influence of close neighbor to be very strong up to longer distances. No significant association was found for proximity to water, wealth index, and EVI. However, the non-linear relationship suggested that population count and proximity to water decreased the malaria prevalence, and these results were supported by many studies (33,83).

Our results on spatial effect in the estimation of malaria prevalence, while using survey data with the ecological variables, showed the evidence of the strong spatial autocorrelation in malaria prevalence at the cluster level, the presence of hot spots and cold spots to be considered for the future studies or resources allocation, and the effect of environmental variables on the spatial distribution of malaria prevalence all over the country.

The strength

The strong part in this work was the use of the nationally representative survey of good quality which included climate variables in the whole country. The use of advanced spatio-temporal methods to analysis malaria prevalence allowed to account for the design complexity of malaria data.

Limitation

The study was focused on estimation of the prevalence, detection of hotspots and evaluation of ecological variables on cluster of households. The following points were limitation: i)The lack of the GPS data at individual or household level in the DHS. In fact, these data can help to model the effect associated with each individual or household in the population and therefore increase the quality of the estimate by including other individual (age, sex, number of children sleeping under the bed, etc.). (ii) the lack of quality data in all the whole country, to assess the trend of the prevalence after 2015. (iii) This study was cross-sectional at only four points. It would be better to have long time point so that the time trend can be well understood. (iv) The lack of some other variables like health facilities quality and proportion of drug resistance.

Perspective

In the next work the following points will be of interest: (i) to evaluate the impact of those factors in predicting malaria prevalence at unknown location by employing geo-statistical variable selection; (ii) to assess the effect of those variables in spatio-temporal variation of malaria hot spot status (risk of being hot spot) using spatial SEM (structural equation model); (iii) to predict malaria prevalence for the next years by using several methods of prediction like Bayesian additive regression trees (BART), gradient boosted trees (GBM) and compare their performance. (iv) there will be also need to ascertain if the spatial effect observed here may be also observed in the department level or any other level relevant in allocating resources, by using other source of data.

4.2 Policy implication

(1) Strong spatial effect

Despite the fact that spatio-temporal analysis has been widely used in many studies, particularly in Africa, it is unlickely to be widely used in Gabon. In fact, in several studies, an estimate of an outcome may not reflect the real value of the outcome in the population targeted, or may be too far from the acceptable range, depending on many points. Such an estimate will not be helpfull in achieving an important goal such as to target control interventions in areas in needs in order to reduce the mortality in children. To put in the context of this study, in small areas, the number of events may not appear as it is effectively in the areas observed. This may lead to a bias in the estimation, hence, resources will not be distributed according to the actual burden of malaria in different areas, therefore there will be less effect of interventions in the population or even waste of resources. However, the use of spatial component

can be important in obtening a good estimates of an outcome of interest in the small area, when there is a strong spatial effect. Then one can use the effect of neighboors in the estimation to increase the quality of the estimate by reducing the uncertainty, to have an estimate that reflects the true burden of the disease. Therefore, the finding of this study will allow the estimate to be usefull for policy makers or resource allocations in order to rightly target and prioritize areas with high burden.

(2) Spillover effect

The spatial effect observed in the level of cluster of households can be important to implify the effect of an intervention to reach areas not targeted through the local spillover effect. This local spillover effect is the fact that close clusters are influencing each other strongly even until very long distance. It allows an indirect benefit of interventions in cluster not covered by the interventions. ITNs coverage plays a protective effect wich is, in magnitude, much better than the negative spillover effect of mean temperature and wet days. The finding in this study allowed to give strength to the expectation that by taking into account the spillover effect, the interventions may play a significant role in the way of eradicating malaria by 2030 or to reduce the transmission of malaria, and reduce the mortality significantly among children in the low- and middle-income countries, as it can offer protection to uncovered areas.

(3) ITNs coverage and ecological factors

Despite the detrimental effect of some ecological factors, ITNs coverage had a much stronger effect which is to decrease importantly the prevalence. Relating the control intervention and those ecological variables associated with the decrease of the prevalence, significant improvement on the planning and the action of the interventions can be achieved. This in order to impact strongly the clinical incidence or to reduce the risk of mortality in children as a target of sustainable development goals (SDGs) in its section 3 (SDG3).

4.3 Conclusions and recommendations

In summary, the findings of this study helped to show the strong spatial autocorrelation of malaria prevalence which decreased slowly. It provided also information on the local spatial spillover effect in the household clusters level, and gave an insight on the influence of the ecological factors and the intervention, with the ITNs coverage, on the transmission of malaria in Gabon. The results highlighted the high prevalence of malaria in the country until 2015 and the presence of hot spots and cold spots which can change by time. The findings showed that malaria prevalence was found to be associated with ITNs coverage, day land surface temperature, wet days and mean temperature in Gabon which were consistent with previous studies. Although ITNs coverage had shown to have a great effect in reducing the prevalence of malaria even after adjusting for all the other factors, it was decreasing between the two last years of the DHS survey. Almost all models including only ecological variables were better in reducing the observed spatial

variation and only a very slight change was showed after adjusting with the other variables. This observation can lead to suggest that some of those ecological variables with the spillover effect can be important in the determination or the formation of the hotspots at the level of the cluster of households. The reduction of the spatial variation can help possibly in reducing the magnitude or the presence of the hot spots in the country. However, the variation of this distribution according to these variables was not evaluated.

Based on these findings, the results may be considered in two ways. Firstly, research trying to estimate the prevalence of malaria in different areas are encouraged to include the spatial or spatio-temporal component in their estimation for small area estimation in the Bayesian framework, in the case of the small number of units for the analysis. Secondly, despite the fact that some ecological variables were found to increase with the prevalence of malaria, and knowing that likely their monitoring are not easy, because influenced probably by global warming, it will be better to focus on all strategies allowing to improve the ITNs coverage at least at 20%, and to sustain its increase over time in the cluster of households along with the health care access to all rural areas in Gabon. This strategy may help to strengthen the control intervention program of malaria in order to reduce significantly the clinical incidence of malaria, because of the significant impact of ITNs coverage in the PfPR.

Additionnally, the increase of the ITNs coverage associated with the spatial effect on the cluster of households level, spreading over long distance, may enhance the impact of the ITNs coverage on the clinical burden of malaria through the local spillover effect in close areas. If this is not considered, the decrease of ITNs coverage, and other factors like quality of public health facilities, associated with the increase of some ecological factors, will probably lead to a significant upsurge of malaria, and formation of hotspots.

The national direction of meteorology (DGM) and the NMCP need to work together to mitigate or control adverse climatic change effect on malaria if possible by developing a good early warning system (EWS) for malaria, which can help in detecting variations in some ecological variables favorising the upsurge of malaria prevalence. Focusing also research on this EWS for malaria, will help in informing population on different initiatives to undertake when there are changes in environmental or ecological variables, in order to better fight malaria at community or individual level. These recommendations should be applied particularly in targeted cluster of households in the north west, and specifically in Estuaire province and the south part of the country where there are hot spots or cold spots, and high prevalence of malaria. By targeting this on the level of the cluster of households, it is expected that the positive effect of the intervention will spread over long distance to benefit the non-targeted areas because of this spillover effect. These results can be considered as a baseline for the future surveys trying to evaluate the progress of intervention program or studies trying to evaluate the risk factors associated with malaria prevalence considering the ecological variables and the time.

5. Appendices

APPENDIX 1

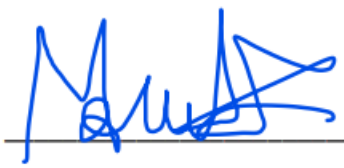
PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I LOTOLA MOUGENI Fabrice (Student number: 2485529) am a student registered for the degree of MSc Biostatistics in the academic year 2021.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature:  Date: 01/02/2023

APPENDIX 2: Malaria prevalence distribution and diagnostics for linear mixed model

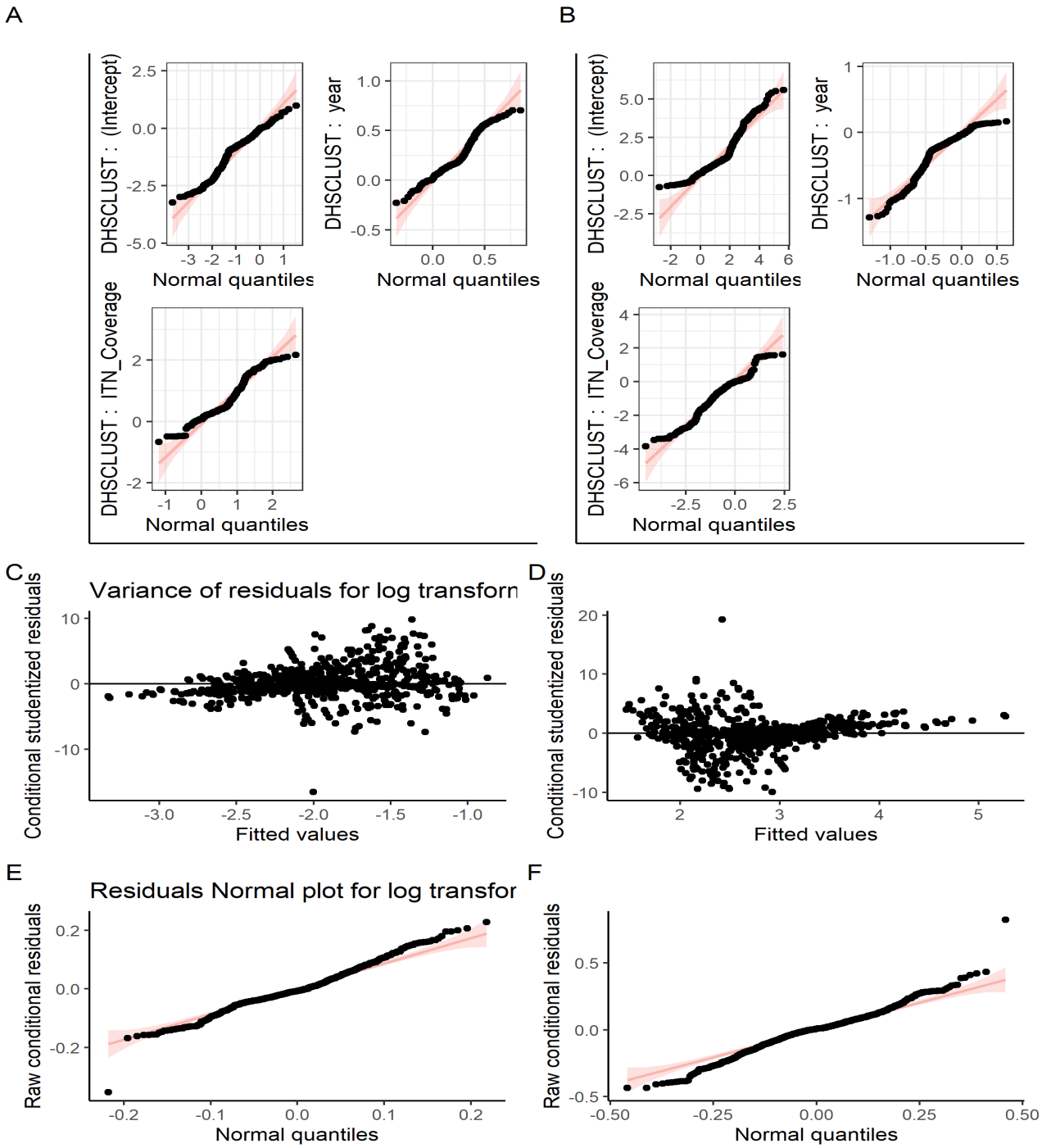


Figure 27: Log transformation vs 1/square root

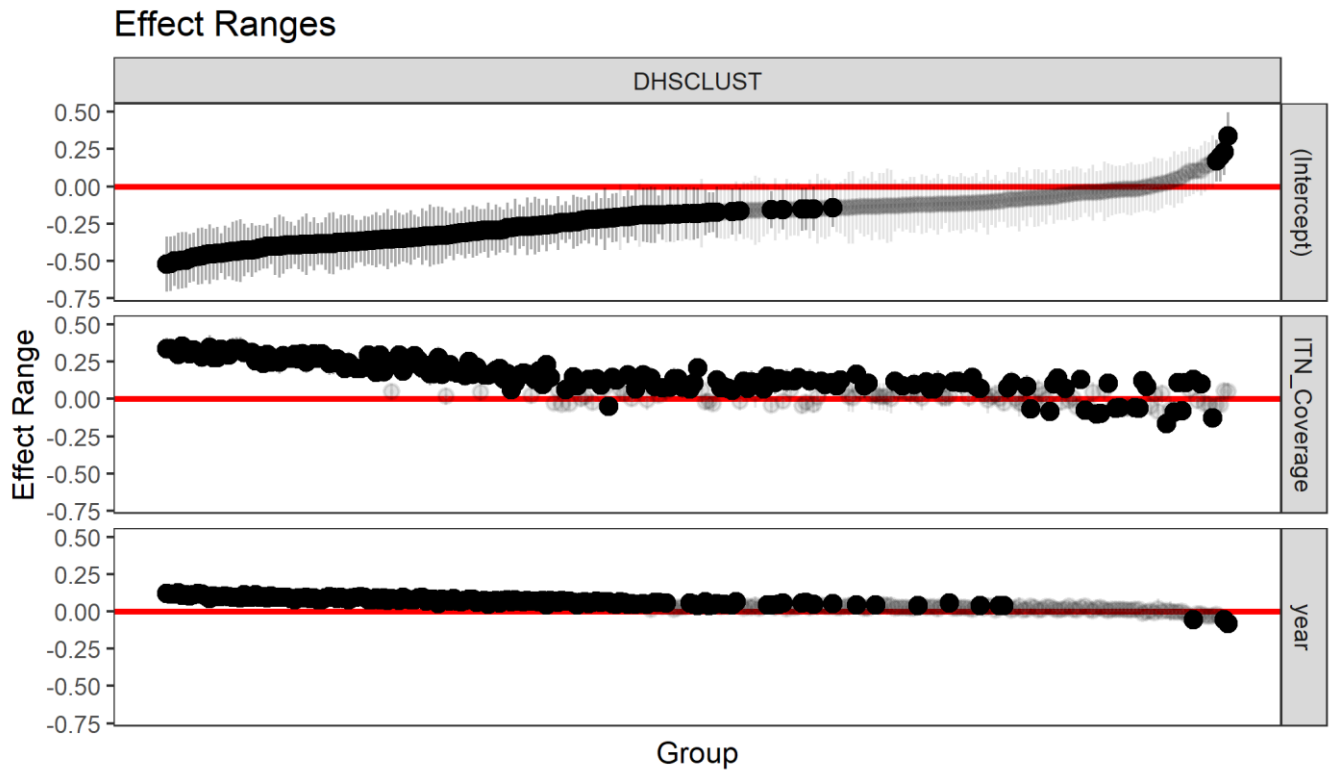


Figure 28: *Effect of ITN and year on the prevalence from each cluster: square root transformation of malaria prevalence*

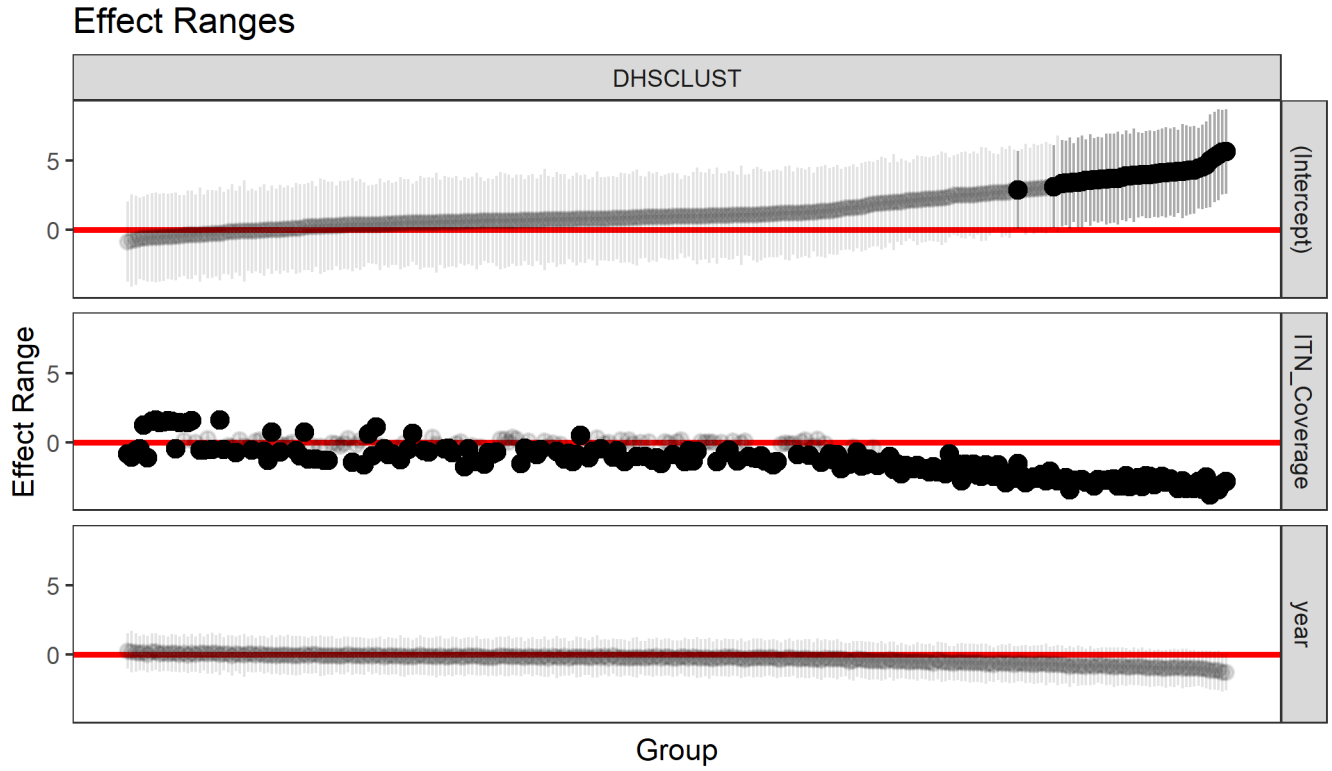


Figure 29*: Effect of ITN and year on the prevalence from each cluster: inverse square root transformation of malaria prevalence*

APPENDIX 3: Certificate Ethics

UNIVERSITY OF THE
WITWATERSRAND
JOHANNESBURG

R49 Mr LM Fabrice

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M220254

NAME: (Principal Investigator)	Mr LM Fabrice
DEPARTMENT:	School of Public Health Division Epidemiology and Biostatistics Medical School University
PROJECT TITLE:	*Bayesian spatio-temporal analysis of malaria prevalence in children aged 2-10 years from 2000-2015 in Gabon*
DATE CONSIDERED:	2022/02/25
DECISION:	Approved unconditionally
CONDITIONS:	
NOTE:	If contact information regarding student study participants is required, please contact the Registrar's office - <Nicoleen.Potgieter@wits.ac.za>
SUPERVISOR:	Professors T Chirwa and B Lell; Mr K Ngianga-Bakwin
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2022/07/06

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **February** and therefore reports and re-certification will be due in the month of **February** each year. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

Signature of Principal Investigator — Date: 08/07/2022

APPENDIX 4: Plagiarism

Indiana University Confirmation Certificate: Advanced Level for Master's and Doctoral Students

Unique Test ID: 13313882204589739438116

An **Advanced Level Certificate** indicates that the person listed below has passed a Certification Test intended for master's and doctoral students. The person below has passed a test by correctly answering at least 9 out of 10 questions selected randomly from a large inventory. Each question provides source material from another author and a sample of student writing. The test taker must determine whether the student version is word-for-word plagiarism, paraphrasing plagiarism, or not plagiarism. Many questions exemplify subtle forms of plagiarism which represent incomplete or incorrect understanding of plagiarism, carelessness, or attempts to disguise actual plagiarism.

Please read the information below carefully. You can submit this confirmation certificate to your academic department or program area, or to your instructor, if required to do so.

I have studied the pages of the "How to Recognize Plagiarism" tutorial site. After doing so, I have taken the test available on the site. I am confident that I know what plagiarism is and how to avoid it. If I had questions after finishing the tutorial, this document confirms that I have sought help from my academic advisor and I do now understand what plagiarism is and how to avoid it.

I understand that there will be no acceptable excuse for plagiarism if it is discovered in my work.

The person(s) at IP # **154.0.186.201** passed the test on: Wednesday, 20th of January 2021, 05:37:12 PM
Unique Test ID: 13313882204589739438116
Time it took to pass the test: 24.9 minutes

Printed Name: **Fabrice Lotola Mougeni**

Signature: 

APPENDIX 5: Turnitin report

Turnitin Originality Report

Processed on: 12-Feb-2023 1:33 PM SAST
ID: 2012021893
Word Count: 24764
Submitted: 1

Research report_turnitin-2.docx By Fabrice Mougeni

Similarity Index	Similarity by Source	
8%	Internet Sources:	7%
	Publications:	6%
	Student Papers:	2%

include quoted | include bibliography | excluding matches < 8 words | mode: quickview (classic) report | print | download

1% match (Internet from 05-Oct-2022)
https://spatialdata.dhsprogram.com/references/DHS%20Covariates%20Extract%20Data%20Description%202.pdf

<1% match (Internet from 05-Oct-2022)
https://spatialdata.dhsprogram.com/references/DHS_Covariates_Extract_Data_Description_3.pdf

<1% match ()
Chigozie Louisa Jane Ugwu, Temesgen Zewotir. "Evaluating the Effects of Climate and Environmental Factors on Under-5 Children Malaria Spatial Distribution Using Generalized Additive Models (GAMs)", Journal of Epidemiology and Global Health

<1% match (Internet from 13-Jun-2013)
http://www.ncbi.nlm.nih.gov

<1% match (Internet from 10-Mar-2013)
http://www.ncbi.nlm.nih.gov

<1% match ()
Vicente Fuenzalida, Gonzalo, Goicoa Mangado, Tomás, Ugarte Martínez, María Dolores. "Bayesian inference in multivariate spatio-temporal areal models using INLA: analysis of gender-based violence in small areas", Stochastic Environmental Research and Risk Assessment, 2020

<1% match ()
Shanka, Tarekegn Solomon. "Bed nets for malaria prevention under field conditions in Ethiopia: : Durability, use, and impact on spatial variation of malaria", The University of Bergen, 2020

<1% match ()
Wambui, KM. "Malnutrition related child morbidity and mortality: A space-time based analysis using Kilifi County Hospital Data 2002 to 2015", 2017

<1% match (student papers from 12-Sep-2022)

Tobias Chirwa _____ _____

6. References

1. Abeku TA. Malaria Epidemics in Africa: Prediction, Detection and Response [Internet]. 2006 [cited 2022 May 20]. Available from: https://repub.eur.nl/pub/7986/

2. World malaria report 2021 [Internet]. [cited 2022 Sep 16]. Available from: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021

3. Mawili-Mboumba DP, Akotet MKB, Kendjo E, Nzamba J, Medang MO, Mbina JRM, et al. Increase in malaria prevalence and age of at risk population in different areas of Gabon. Malar J. 2013 Jan 2;12(1):3.

4. Zoleko Manego R, Koehne E, Kreidenweiss A, Nzigou Mombo B, Adegbite BR, Dimessa Mbadinga LB, et al. Description of Plasmodium falciparum infections in central Gabon demonstrating high parasite densities among symptomatic adolescents and adults. Malar J. 2019 Nov 21;18(1):371.

5. Bouyou-Akotet MK, Mawili-Mboumba DP, Kendjo E, Mabika-Mamfoumbi M, Ngoungou EB, Dzeing-Ella A, et al. Evidence of decline of malaria in the general hospital of Libreville, Gabon from 2000 to 2008. Malar J. 2009 Dec 17;8(1):300.

6. Bejon P, Warimwe G, Mackintosh CL, Mackinnon MJ, Kinyanjui SM, Musyoki JN, et al. Analysis of immunity to febrile malaria in children that distinguishes immunity from lack of exposure. Infect Immun. 2009 May;77(5):1917–23.

7. Assele V, Ndoh GE, Nkoghe D, Fandeur T. No evidence of decline in malaria burden from 2006 to 2013 in a rural Province of Gabon: implications for public health policy. BMC Public Health. 2015 Feb 4;15(1):81.

8. Maghendji-Nzondo S, Nzoughe H, Lemamy GJ, Kouna LC, Pegha-Moukandja I, Lekoulou F, et al. Prevalence of malaria, prevention measures, and main clinical features in febrile children admitted to the Franceville Regional Hospital, Gabon. Parasite. 23:32.

9. Smith DL, Guerra CA, Snow RW, Hay SI. Standardizing estimates of the Plasmodium falciparum parasite rate. Malar J. 2007 Sep 25;6:131.

10. Patil AP, Okiro EA, Gething PW, Guerra CA, Sharma SK, Snow RW, et al. Defining the relationship between Plasmodium falciparum parasite rate and clinical disease: statistical models for disease burden estimation. Malar J. 2009 Aug 5;8(1):186.

11. Moukandja IP, Essone JCBB, Sagara I, Kassa RFK, Ondzaga J, Douki JBL, et al. Marked Rise in the Prevalence of Asymptomatic Plasmodium falciparum Infection in Rural Gabon. PLOS ONE. 2016 May 26;11(5):e0153899.

12. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet Lond Engl. 2006 May 27;367(9524):1747–57.

13. Yeshiwondim AK, Gopal S, Hailemariam AT, Dengela DO, Patel HP. Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia. Int J Health Geogr. 2009 Jan 26;8(1):5.

14. Okunlola OA, Oyeyemi OT. Spatio-temporal analysis of association between incidence of malaria and environmental predictors of malaria transmission in Nigeria. Sci Rep. 2019 Nov 25;9(1):17500.

15. Gabon weather and climate Best time to visit Temperature [Internet]. [cited 2022 May 20]. Available from: https://www.besttimetovisit.co.za/gabon/

16. World Bank Climate Change Knowledge Portal [Internet]. [cited 2022 May 20]. Available from: https://climateknowledgeportal.worldbank.org/

17. Gabon climate: average weather, temperature, precipitation, when to go [Internet]. [cited 2022 May 20]. Available from: https://www.climatestotravel.com/climate/gabon

18. Libreville Climate, Weather By Month, Average Temperature (Gabon) - Weather Spark [Internet]. [cited 2022 May 20]. Available from: https://weatherspark.com/y/61793/Average-Weather-in-Libreville-Gabon-Year-Round

19. Mayala B, Fish TD, Eitelberg D, Dontamsetti T. The DHS Program Geospatial Covariate Datasets Manual. :50.

20. Adeyemi RA. Bayesian spatial modeling of malnutrition and mortality among under-five children in sub-Saharan Africa. [Internet] [Thesis]. 2019 [cited 2022 May 28]. Available from: https://researchspace.ukzn.ac.za/handle/10413/17952

21. Corsi DJ, Neuman M, Finlay JE, Subramanian SV. Demographic and health surveys: a profile. Int J Epidemiol. 2012 Dec;41(6):1602–13.

22. Bray RS, Garnham PC. The life-cycle of primate malaria parasites. Br Med Bull. 1982 May;38(2):117–22.

23. 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. Malar J. 2018 Feb 20;17(1):87.

24. Mordecai EA, Paaijmans KP, Johnson LR, Balzer C, Ben-Horin T, de Moor E, et al. Optimal temperature for malaria transmission is dramatically lower than previously predicted. Ecol Lett. 2013 Jan;16(1):22–30.

25. Ndlovu N. Analysis of the Geographical Patterns of Malaria Transmission in KwaZulu-Natal, South Africa using Bayesian Spatio-Temporal Modelling. :115.

26. Craig MH, Snow RW, le Sueur D. A climate-based distribution model of malaria transmission in sub-Saharan Africa. Parasitol Today Pers Ed. 1999 Mar;15(3):105–11.

27. Alonso PL, Lindsay SW, Armstrong JR, Conteh M, Hill AG, David PH, et al. The effect of insecticide-treated bed nets on mortality of Gambian children. Lancet Lond Engl. 1991 Jun 22;337(8756):1499–502.

28. Lengeler C. Insecticide-treated bed nets and curtains for preventing malaria. Cochrane Database Syst Rev. 2004;(2):CD000363.

29. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. Spat Spatio-Temporal Epidemiol. 2013 Dec;7:39–55.

30. Pamba R, Sevidzem S, Koumba A, Koumba C roland, Ovono A, Mintsa Nguema R, et al. Preventive Measures Against The Vectors Of Malaria In Akanda, Southwest Gabon: Knowledge, Attitudes, Practices, And Beliefs. Eur Sci J ESJ. 2021 Jun 30;17:169.

31. Devi NP, Jauhari RK. Climatic variables and malaria incidence in Dehradun, Uttaranchal, India. J Vector Borne Dis. 2006 Mar;43(1):21–8.

32. Gallup JL, Sachs JD. The economic burden of malaria. Am J Trop Med Hyg. 2001 Feb;64(1-2 Suppl):85–96.

33. Adigun AB, Gajere EN, Oresanya O, Vounatsou P. Malaria risk in Nigeria: Bayesian geostatistical modelling of 2010 malaria indicator survey data. Malar J. 2015 Apr 14;14(1):156.

34. Kandala NB, Ghilagaber G. A Geo-Additive Bayesian Discrete-Time Survival Model and its Application to Spatial Analysis of Childhood Mortality in Malawi. Qual Quant. 2006 Dec 1;40(6):935–57.

35. Nobre AA, Schmidt AM, Lopes HF. Spatio-temporal models for mapping the incidence of malaria in Pará. Environmetrics. 2005;16(3):291–304.

36. Zacarias OP, Andersson M. Spatial and temporal patterns of malaria incidence in Mozambique. Malar J. 2011 Jul 13;10(1):189.

37. McCord GC. Malaria ecology and climate change. Eur Phys J Spec Top. 2016 May 1;225(3):459–70.

38. Elliott P, Wartenberg D. Spatial Epidemiology: Current Approaches and Future Challenges. Environ Health Perspect. 2004 Jun;112(9):998–1006.

39. DHS Covariate Extraction [Internet]. The DHS Program; 2022 [cited 2022 Aug 24]. Available from: https://github.com/DHSProgram/DHS-covariate-extraction

40. Seyoum D, Yewhalaw D, Duchateau L, Brandt P, Rosas-Aguirre A, Speybroeck N. Household level spatio-temporal analysis of Plasmodium falciparum and Plasmodium vivax malaria in Ethiopia. Parasit Vectors. 2017 Apr 20;10(1):196.

41. Elkasabi M, Ren R, Pullum T. Multilevel Modeling Using DHS Surveys: A Framework to Approximate Level-Weights. 2020.

42. The Geospatial Covariate Datasets Manual · PDF file 2020-03-17 · 1 The DHS Program Geospatial Covariate Datasets Manual Second Edition Benjamin Mayala1 Thomas D. Fish1 David Eitelberg2 - [PDF Document] [Internet]. fdocuments.net. [cited 2022 Aug 24]. Available from: https://fdocuments.net/document/the-geospatial-covariate-datasets-manual-2020-03-17-1-the-dhs-program-geospatial.html

43. Smith DL, Guerra CA, Snow RW, Hay SI. Standardizing estimates of the Plasmodium falciparum parasite rate. Malar J. 2007 Sep 25;6(1):131.

44. Rue H, Martino S, Lindgren F, Simpson D, Riebler A, Krainski E. INLA: Functions Which Allow to Perform Full Bayesian Analysis of Latent Gaussian Models Using Integrated Nested Laplace Approximaxion. R package version 0.0-1389624686. 2014.

45. Moraga P, Dean C, Inoue J, Morawiecki P, Noureen SR, Wang F. Bayesian spatial modelling of geostatistical data using INLA and SPDE methods: A case study predicting malaria risk in Mozambique. Spat Spatio-Temporal Epidemiol. 2021 Nov 1;39:100440.

46. Mitchell CL, Ngasala B, Janko MM, Chacky F, Edwards JK, Pence BW, et al. Evaluating malaria prevalence and land cover across varying transmission intensity in Tanzania using a cross-sectional survey of school-aged children. Malar J. 2022 Mar 9;21(1):80.

47. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B Stat Methodol. 2009;71(2):319–92.

48. Musenge E, Chirwa TF, Kahn K, Vounatsou P. Bayesian analysis of zero inflated spatiotemporal HIV/TB child mortality data through the INLA and SPDE approaches: Applied to data observed between 1992 and 2010 in rural North East South Africa. Int J Appl Earth Obs Geoinformation ITC J. 2013 Jun;22(100):86–98.

49. Semakula M, Niragire F, Faes C. Bayesian spatio-temporal modeling of malaria risk in Rwanda. PLOS ONE. 2020 Sep 10;15(9):e0238504.

50. Dieng S, Ba EH, Cissé B, Sallah K, Guindo A, Ouedraogo B, et al. Spatio-temporal variation of malaria hotspots in Central Senegal, 2008–2012. BMC Infect Dis. 2020 Jun 17;20(1):424.

51. Mercer L, Wakefield J, Chen C, Lumley T. A comparison of spatial smoothing methods for small area estimation with sampling weights. Spat Stat. 2014 May 1;8:69–85.

52. Fay RE, Herriot RA. Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. J Am Stat Assoc. 1979 Jun 1;74(366a):269–77.

53. Hierarchical Modeling and Analysis for Spatial Data [Internet]. Routledge & CRC Press. [cited 2022 Apr 25]. Available from: https://www.routledge.com/Hierarchical-Modeling-and-Analysis-for-Spatial-Data/Banerjee-Carlin-Gelfand/p/book/9781439819173

54. Moraga P. Chapter 9 Spatial modeling of geostatistical data. Malaria in The Gambia | Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny [Internet]. [cited 2022 Jun 13]. Available from: https://www.paulamoraga.com/ /sec-geostatisticaldataexamplespatial.html

55. Miller DL, Glennie R, Seaton AE. Understanding the Stochastic Partial Differential Equation Approach to Smoothing. J Agric Biol Environ Stat. 2020 Mar 1;25(1):1–16.

56. Bivand R, Gómez-Rubio V, Rue H. Spatial Data Analysis with R-INLA with Some Extensions. J Stat Softw. 2015 Feb 16;63:1–31.

57. Nkurunziza H, Gebhardt A, Pilz J. Bayesian modelling of the effect of climate on malaria in Burundi. Malar J. 2010 Apr 29;9(1):114.

58. Anselin L. Spatial Econometrics: Methods and Models. Springer Science & Business Media; 1988. 308 p.

59. Duque Y. LeSage & Pace (2009) Introduction to Spatial Econometrics Statistics A Series of Textbooks and Monographs. 2020.

60. Gomez-Rubio V, Bivand RS, Rue H. Estimating Spatial Econometrics Models with Integrated Nested Laplace Approximation [Internet]. arXiv; 2021 May [cited 2022 Jun 2]. Report No.: arXiv:1703.01273. Available from: http://arxiv.org/abs/1703.01273

61. Toma SA, Eneyew BW, Taye GA. Spatial Modelling of Risk Factors for Malaria Prevalence in SNNP Regional State, Ethiopia. Ethiop J Health Sci. 2021 Jul;31(4):731–42.

62. Gollini I, Lu B, Charlton M, Brunsdon C, Harris P. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. J Stat Softw. 2015 Feb 16;63:1–50.

63. Simpson DP, Rue H, Martins TG, Riebler A, Sørbye SH. Penalising model component complexity: A principled, practical approach to constructing priors [Internet]. arXiv; 2015 Aug [cited 2022 May 20]. Report No.: arXiv:1403.4630. Available from: http://arxiv.org/abs/1403.4630

64. Sørbye SH, Rue H. Penalised complexity priors for stationary autoregressive processes [Internet]. arXiv; 2016 Aug [cited 2022 May 20]. Report No.: arXiv:1608.08941. Available from: http://arxiv.org/abs/1608.08941

65. Noor AM, Kinyoki DK, Mundia CW, Kabaria CW, Mutua JW, Alegana VA, et al. The changing risk of Plasmodium falciparum malaria infection in Africa: 2000–10: a spatial and temporal analysis of transmission intensity. The Lancet. 2014 May 17;383(9930):1739–47.

66. Franco-Villoria M, Ventrucci M, Rue H. A unified view on Bayesian varying coefficient models [Internet]. arXiv; 2019 [cited 2022 Jun 3]. Available from: http://arxiv.org/abs/1806.02084

67. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. 2nd edition. New York, NY: Springer; 2016. 767 p.

68. Schielzeth H, Dingemanse NJ, Nakagawa S, Westneat DF, Allegue H, Teplitsky C, et al. Robustness of linear mixed-effects models to violations of distributional assumptions. Methods Ecol Evol. 2020;11(9):1141–52.

69. Giorgi E, Fronterrè C, Macharia PM, Alegana VA, Snow RW, Diggle PJ. Model building and assessment of the impact of covariates for disease prevalence mapping in low-resource settings: to explain and to predict. J R Soc Interface. 18(179):20210104.

70. Cameletti M, Lindgren F, Simpson D, Rue H. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. AStA Adv Stat Anal. 2011 Nov 30;97.

71. Roca-Feltrer A, Kwizombe CJ, Sanjoaquin MA, Sesay SSS, Faragher B, Harrison J, et al. Lack of Decline in Childhood Malaria, Malawi, 2001–2010. Emerg Infect Dis. 2012 Feb;18(2):272–8.

72. Ejigu BA. Geostatistical analysis and mapping of malaria risk in children of Mozambique. PLOS ONE. 2020 Nov 9;15(11):e0241680.

73. Greenwood B, Zongo I, Dicko A, Chandramohan D, Snow RW, Ockenhouse C. Resurgent and delayed malaria. Malar J. 2022 Mar 9;21(1):77.

74. Hawley WA, Phillips-Howard PA, ter Kuile FO, Terlouw DJ, Vulule JM, Ombok M, et al. Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western Kenya. Am J Trop Med Hyg. 2003 Apr;68(4 Suppl):121–7.

75. Mohammadkhani M, Khanjani N, Bakhtiari B, Sheikhzadeh K. The relation between climatic factors and malaria incidence in Kerman, South East of Iran. Parasite Epidemiol Control. 2016 Jun 23;1(3):205–10.

76. Mohammadkhani M, Khanjani N, Bakhtiari B, Tabatabai SM, Sheikhzadeh K. The Relation Between Climatic Factors and Malaria Incidence in Sistan and Baluchestan, Iran. SAGE Open. 2019 Jul 1;9(3):2158244019864205.

77. Panzi EK, Okenge LN, Kabali EH, Tshimungu F, Dilu AK, Mulangu F, et al. Geo-Climatic Factors of Malaria Morbidity in the Democratic Republic of Congo from 2001 to 2019. Int J Environ Res Public Health. 2022 Mar 23;19(7):3811.

78. Kibret S, Glenn Wilson G, Ryder D, Tekie H, Petros B. Environmental and meteorological factors linked to malaria transmission around large dams at three ecological settings in Ethiopia. Malar J. 2019 Feb 26;18(1):54.

79. Ssempiira J, Kissa J, Nambuusi B, Mukooyo E, Opigo J, Makumbi F, et al. Interactions between climatic changes and intervention effects on malaria spatio-temporal dynamics in Uganda. Parasite Epidemiol Control. 2018 Aug 1;3(3):e00070.

80. The Influence of Climate on Malaria Incidence in Malawi [Internet]. [cited 2022 Jun 3]. Available from: https://www.climatelinks.org/resources/influence-climate-malaria-incidence-malawi

81. Adeola AM, Botai JO, Olwoch JM, Rautenbach HCJ de W, Adisa OM, Jager C de, et al. Predicting malaria cases using remotely sensed environmental variables in Nkomazi, South Africa. Geospatial Health [Internet]. 2019 May 14 [cited 2022 Aug 26];14(1). Available from: https://geospatialhealth.net/index.php/gh/article/view/676

82. The Role of Climate Variability in the Spread of Malaria in Bangladeshi Highlands | PLOS ONE [Internet]. [cited 2022 Jun 7]. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0014341

83. Ugwu CLJ, Zewotir T. Spatial distribution and sociodemographic risk factors of malaria in Nigerian children less than 5 years old. Geospatial Health [Internet]. 2020 Dec 29 [cited 2022 May 13];15(2). Available from: https://geospatialhealth.net/index.php/gh/article/view/819