Geostatistical modelling and survey sampling designs for malaria control and surveillance

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Salomon Gottlieb Massoda Tonye

aus Cameroon

Basel, 2022

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Jürg Utzinger, PD Dr. Penelope Vounatsou, und Prof. Dr. Fredrick Edward Makumbi

Basel, den 17.11.2020

Prof. Dr. Martin Spiess

Dekan

Praise to the Lord, the almighty

Infinite gratitude to those who contributed to this work

Summary

In many low-and middle-income countries, malaria is endemic and remains a serious health threat and a significant contributor to mortality. According to the World malaria report 2019, more than ninety percent of malaria cases occurred in Africa. Children remain the most vulnerable group. Cameroon belonged to the list of countries most affected by malaria. The country is subdivided into several ecological zones in which malaria prevalence is spatially heterogeneous. The disease transmission is highly seasonal in the North, perennial in the southern and eastern parts and relatively low in the high mountains of West and Adamawa regions.

Substantial efforts made by international donors and the Cameroon government led to a decline in the malaria burden over the last decade. Many health interventions and actions were implemented to fight against the spread of disease. National surveys such as the Demographic and health surveys (DHS), malaria indicator surveys (MIS), and multiple indicator cluster surveys (MICS) are conducted every three to five years collecting individual and household level data on malaria, disease interventions and socio-economic factors to measure the progress achieved in the control of the disease. The data are georeferenced at the centroid of clusters consisting of groups of around 25 households. For confidentiality reasons the cluster coordinates are reported with jittering, that is they are misplaced within a buffer around the actual location. Bayesian geostatistical models are commonly used to predict the geographical distribution of disease prevalence and quantify the effects of interventions. Predictions are improved by including in the models climatic and environmental proxies available at high spatio-temporal resolution from remote sensing sources. Despite the large number of survey data that have been analysed using geostatistical models, there are few studies of the effects of survey design factors on model-based estimates.

The overall goal of the thesis is to evaluate and further strengthen the methodology of the malaria survey designs used for disease monitoring and evaluation, in particular the aspects related to the timing of the survey, the jittering in the coordinates of the reported cluster locations, and the seasonal monitoring of malaria data. Furthermore, we evaluated the ability of the survey data to estimate the malaria-related deaths and compared estimates of the effects of malaria interventions using data from malaria surveys and the Health Management Information System. More specifically, the thesis pursues the following objectives: i) assess the influence of the survey timing by comparing the DHS and MIS geostatistical model-based malaria risk estimates obtained at different malaria transmission seasons; ii) assess the effects of the DHS jittering algorithm on the prediction of malaria risk and on the estimates of the disease risk factors; iii) evaluate the effects of malaria interventions on the geographical distribution of disease incidence after adjusting for the effects of climatic and environmental factors; iv) improve malaria disease and vector survey sampling designs by optimizing the selection of survey locations and v) evaluate the ability of malaria surveys to estimate malaria-related deaths.

The above objectives were addressed by analysing DHS, MIS and MICS survey data from Cameroon as well as malaria incidence data from the Health Management Information System. The methodology and results for each objective are included in five main chapters.

In Chapter 2, Bayesian stationary geostatistical model are employed to analyse MIS and DHS, national surveys conducted in 2011 during the rainy and dry seasons, respectively. Geostatistical variable selection was applied to identify the most important climatic factors and malaria intervention indicators. The results showed that the timing of the malaria survey influences estimates of the geographical distribution of disease risk, especially in settings with seasonal transmission. In countries with different ecological zones and thus different seasonal patterns, a single survey may not be able to identify all high-risk areas.

Chapter 3 presents the influence of jittering on the assessment of intervention effects and the spatial estimates of disease risk distribution at high spatial resolution. Based on original MIS cluster locations, a set of a hundred (100) shifted cluster locations were generated using the DHS jittering algorithm. Bayesian variable selection applied in the original dataset locations as well as for each one of the hundred jittered datasets to select climatic/environmental predictors, socio-economic factors and malaria intervention indicators. Geostatistical models were applied to the original as well as the simulated data using the selected covariates. The results indicated that the selections of important climatic predictors and of intervention indicators were influenced by the jittering, while estimates of the disease risk at high geographical resolution were slightly affected.

Chapter 4 focused on the selection of relevant cluster locations for an efficient assessment of intervention effects. Based on MIS data, a Bayesian geostatistical model was applied to the most important climatic predictors to estimate the malaria risk, and associated uncertainty over a high resolution gridded surface across Cameroon. An adaptive algorithm was proposed to select survey locations based on a multi-criteria objective approach. The adapted algorithm was able to identify a meaningful subset of cluster locations based on their contribution to the uncertainty and the needs of the national malaria program.

In chapter 5, we evaluated the effects of interventions on the spatiotemporal dynamic of malaria incidence and the capability of HMIS data to capture disease pattern. From 2012 to 2016, confirmed malaria data were extracted and aggregated by month at the district level. During the same period, climatic factors were obtained from satellites and averaged over the district surface. Bayesian variable selection was applied to identify the most important lag time for each continuous climatic factor. A Bayesian spatiotemporal of the relationship between malaria incidence and intervention was fitted and adjusted with the important climatic factor. The percentage of households having one ITN per two persons was identified

as the most important coverage indicators, while the normalized difference vegetation index, rainfall estimates were selected among climatic predictors. Having an ITN for every two persons was negatively associated to malaria cases. The incidence maps drawn at district level were able to capture patterns of disease risk that were not estimated by the DHS and MIS data.

Chapter 6 assessed the relationship between malaria prevalence and all-cause mortality in infants and in children under-5 years old, by considering seasonal influence of malaria transmission as well as socio-economic factors. Bayesian geostatistical Bernoulli and zero-inflated Bernoulli models were fitted on the mortality risk data. A statistically important relation was estimated between infants (excluding neonates), under-five years old mortality and malaria risk. The effects of malaria parasite risks on under-five mortality became more statistically important in the absence of neonates. Mortality in the under-five group was reduced during the dry season.

Acknowledgments

It was a great honor and privilege to have benefited from the support and commitment of several talented people in the achievement of this work.

Our special gratitude goes to the supervisor PD. Dr. Penelope Vounatsou for the scientific lead, patience and great commitment. Her technical skills have been decisive to conclude this work. She had always provided tools to facilitate our work. It was an advantage and a great asset to work with her.

I would also like to express my sincere recognition to Prof. Dr. Jürg Utzinger the faculty representative of this thesis. My Ph.D thesis was the result of the fruitful collaboration of these two people.

Thank you to Prof. Dr. Fredrick Edward Makumbi for his kind agreement to be the external reviewer, and Prof. Dr. Pascal Mäser who accepted to chair my defense.

To the personnel at the Swiss TPH, a sincere thanks for their dedication, diligence and friendship. In addition to creating a suitable work environment, their assistance in the administrative procedures had greatly facilitated our stay in Basel. Our thanks are particularly addressed to Laura Innocentti, Nora Bauer and Dagmar Batra for their professional administrative support throughout the Ph.D programme. Great thanks to Christine Mensch, for her professionalism and warmest attitudes. Many thanks also to my previous and current colleagues at the Swiss TPH for their support and friendship: Abbas adigum, Anton Beloconi, Betty, Nambuusi, Carla Grolimund, Christian Herrmann, Christos Kokaliaris, Eric Diboulo, Elizaveta Semenova, Isidoros Papaioannou, Julius Sempira, Oliver Bärenbold, Ourohire Millogo, Sabelo Dlamini and Sammy Khagayi.

My gratitude to the family members that have always been unswerving support in the difficult times. Particularly, I thank Flore Doumi, my wife, for her encouragement and prayers. To my beloved children Julien Massoda, Esther Massoda and Anael Massoda, thank you for your

love. A special thought to my mothers and fathers Marie Francoise Ngo Ngoubi, Hermine Ngo Nkoll, Marlyse Engamba, Marie-joseph Ossual, Tonye Tonye Charles, and Nguea Victor. My recognition to my sisters and brothers, Hélène, Flambeau, Philippe, Elisabeth, Jenny, Fabien, Jeunesse and the rest of the family. To friends, who like brothers have always been there, Bertrand Okon Mine, Benedith Minka, Calvin Messina Ndjoh, Casimir Ketchoum, Dieudonné Kinkielele, Etienne Mboui Billong, Eithel Mekongo, Emmail Wanjo, Francky Fouedjio, Gaston Manga, Guy Raoul Mbangue, Francine Nzokem, Magloire Kuetche, Maturin Nkamgnia, Marcel Opoumba, Merline Ngnietcha, Modeste Nana, Njoya Ndam Arouna and Peter Tendoh.

To Claude Hariton, Normand Tremblay, Bernard Guimont, Yves Lachance and the team members of diaMentis project, I would like to express my sincere gratitude.

To the wonderful personnel of Cameroon national malaria control program and the national institute of statistics my respect to their commitment and contributions to fighting against malaria and for giving us a free access to databases.

To the other contributors that were not clearly mentioned here, I also express our sincere gratitude to those who have contributed in any way to this work.

Finally, the greatest thank goes to God and our Lord Jesus Christ, for his mercy, healthy and protection, before, during, and after this thesis.

Contents

List of tables

List of figures

Abbreviations

ACT: Artemisinin-based Combination Therapy

AIDS: Acquired Immune Deficiency Syndrome

BCI: Bayesian Credible Interval

CAR: Conditional Auto Regressive

DHS: Demographic and Health Survey

GLM: Generalized Linear Models

GMRF: Gaussian Markov Random Field

GRUMP: Global Rural Urban Mapping Project

HIV: Human Immunodeficiency Virus

INLA: Integrated Nested Laplace Approach

INS : « Institut National de la Statistique », National institute of Statistics

IPT: Intermittent Preventive Treatment

IRS: Indoor Residual Spraying

ITN: Insecticide-Treated Net

LLIN: Long Lasting Insecticide Net

LST: Land Surface Temperature

MARA: Mapping Malaria Risk in Africa

MCMC: Markov Chain Monte Carlo

MDG: Millennium Development Goal

MIS: Malaria Indicator Survey

Minsante : Ministry of Public Health in French

NDVI: Normalized Difference Vegetation Index

NMCP: National Malaria Control Programme

OR: Odds Ratio

PCR: Polymerase Chain Reaction

PNLP : « Programme National de Lutte contre le Paludisme »

RBM: Roll Back Malaria

RDT: Rapid Diagnostic Test

SES: Socio Economic Status

SPDE: Stochastic Partial Differential Equation

Swiss TPH: Swiss Tropical and Public Health (institute)

UNICEF: United Nation Children Emergency Fund

USAID: United State Agency for International Development

WHO: World Health Organization

Chapter 1 Introduction

1.1 Rationale

The fight against communicable and infectious diseases remains a global concern for human and animal life. Countries have developed strategic action plans to structure the control and response to disease communicable and infectious diseases, with the aim of reducing their burden. Malaria is clearly among the dangerous infectious diseases in the world with a highly negative impact on the poorest populations. It contributes to the cause of premature mortality among children in the low-and middle-income countries (Liu et al., 2015).

In sub-Saharan countries, malaria national control programs were created by Governments to ensure the consistent implementation of evidence-based health interventions already endorsed at the global level (WHO, 2016). These interventions were continuously implemented with the support of international and local non-governmental organizations. Different indicators were defined to assess the effects and impacts of health interventions on disease trends. Standardized household surveys were also designed to collect and estimate intervention indicators (Roll Back Malaria, USAID, CDC, UNICEF, WHO, 2013). However, these surveys were not conducted regularly as desired and it was therefore always difficult to obtain reliable and up-to-date data. In addition, there was a need for spatial analyses applied to existing data collected at national and regional levels through household surveys and health information systems for targeted, cost-effective disease control. The lacks of georeferenced data and human resources possessing the skills have long been the main reason for the scarcity of geostatistical studies. In the field of malaria, scientific collaboration such as the MARA (Mapping Malaria Risk in Africa) project was initiated to address this weakness, (Swiss TPH et al, 1996). Data from entomological and household surveys were inventoried and stored in a central repository to facilitate spatial analyses of disease risk at a continental

level. Since then, several West and East African countries like Nigeria, Uganda, and Burkina Faso have done interesting work using the Bayesian geostatistical models to assess the intervention effects and distribution of malaria risk at high geographical resolution (Adigun et al., 2015; Diboulo et al., 2015; Nambuusi et al., 2019; Ssempiira et al., 2017). Despite all these, many important research questions remain unresolved. For instance, what is the optimal way to select the locations and timing for a survey to best capture the geographical distribution of the disease risk? For confidentiality reasons the coordinates of the survey locations are reported within a buffer around their actual position. How does this spatial displacement influence the geostatistical model-based predictions and estimates of the effects of risk factors?

We addressed the aforementioned questions using Cameroon as a case country and analysing data from household surveys and from the Health Management Information System. Cameroon is made up of heterogeneous geographical and ecological areas with high disparities of malaria risk among areas. The length and intensity of malaria transmission season vary depending on the ecological zone from four to twelve months. Since 2006, the country has benefited from funding from The Global Fund to Fight Aids, Tuberculosis and Malaria (The Global Fund, 2006). In 2011, Malaria control interventions and actions were scaled-up at the country level. In particular, two mass campaigns of eight and twelve million insecticide-treated bed nets were organized in 2011 and 2015, respectively. In addition, an intervention like seasonal chemoprevention of malaria in children under five is being implemented in the North and Sahelian parts of the country since 2016 (Bowen, 2013; Antonio-Nkondjio et al., 2019).

1.2 Malaria

1.2.1   Global overview

Despite massive investments in combatting malaria and progress made in reducing disease's burden, malaria remains a global threat. Malaria is a vector-borne infectious disease, and its transmission strongly dependent on ecological conditions and human activities. The disease is persistently present on three continents: Africa, South-America and Asia. In 2019 more than 80 countries and territories were affected by this disease. More than 3.3 billion of people were exposed. The disease's risk was rated as highest for one billion of people that lived mainly in sub-Saharan African countries, where about 93% of cases were declared by health authorities, as well as 94% of malaria-related deaths. Children were the most affected group representing 67% of malaria deaths (WHO, 2019). Also, around 228 million cases were reported yearly with 80% of occurring in fifteen African countries and in India. The estimated number of global malaria-related deaths was about 405,000 in 2019.
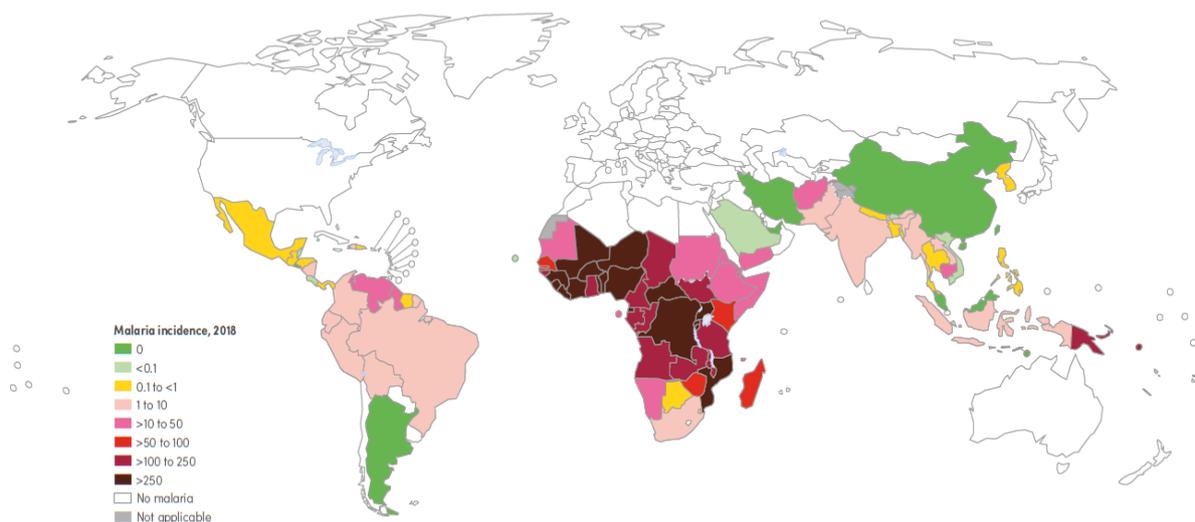
Figure 1.1 : Map of malaria cases incidence per 1000 population at risk, WHO, 2019

The past decade was characterized by the contribution of research in the production of tools and drug combinations for the fight against the spread of malaria. The universal coverage and effective management of cases for the entire population at risk were promoted. The following key preventive actions were implemented and scaled-up in many affected countries: purchase and free distribution of insecticide-treated bed nets (ITNs), intermittent preventive treatment (IPT) during pregnancy, and seasonal chemoprevention. Case management involved systematic confirmation of all suspected fevers. This was possible thanks to the introduction of malaria rapid diagnostic tests, supplemented with the use of microscopy, which remains the "diagnostic gold standard"(Fançony et al., 2013; Berzosa et al., 2018). The treatment of simple malaria was improved through the use of artemisinin-based combination therapies (ACTs). Between 2012 and 2017, more than one billion of ITNs were distributed during free mass campaigns. In 2017 more than 276 million malaria rapid diagnostic tests (RDTs) were sold globally (WHO, 2018a).

1.2.2   Disease transmission and symptoms

Malaria infection is due to the presence of a *Plasmodium* parasite in human blood. More than two hundred (200) species of the genus *Plasmodium* have been identified and they can also be found in reptiles, birds, and mammals(Rich and Ayala, 2006). Human malaria is caused by five of these *Plasmodium* species: *P. falciparum, P. vivax, P. malariae, P. ovale* and *P. knowlesi* (WHO, 2015a). *P. falciparum is* largely predominant in Sub-Saharan countries, while *P.vivax* is widespread all over the world. The *P.falciparum* is more aggressive and responsible for most of the malaria infection and mortality in Africa. More than four hundred (400) species of the Anopheles mosquito have been identified, with seventy (70) having malaria vector capabilities, but only females were involved (Sharma, 2006).

Transmission of malaria parasites to the human host occurs through the bite of the female Anopheles mosquitoes during a blood meal with inoculation of the parasite in the sporozoïte stage to the human blood system (Menkin-Smith and Winders, 2019; WHO, 2015b). Therefore, the lifecycle of malaria parasites depends on both hosts; mosquitoes and humans. Biological specificities of humans, interventions, characteristics of the female Anopheles mosquitoes, and the environmental conditions can contribute to differences in the transmission duration and intensities observed in different countries(Fowkes et al., 2016; Lelisa et al., 2017; Pothin et al., 2016).
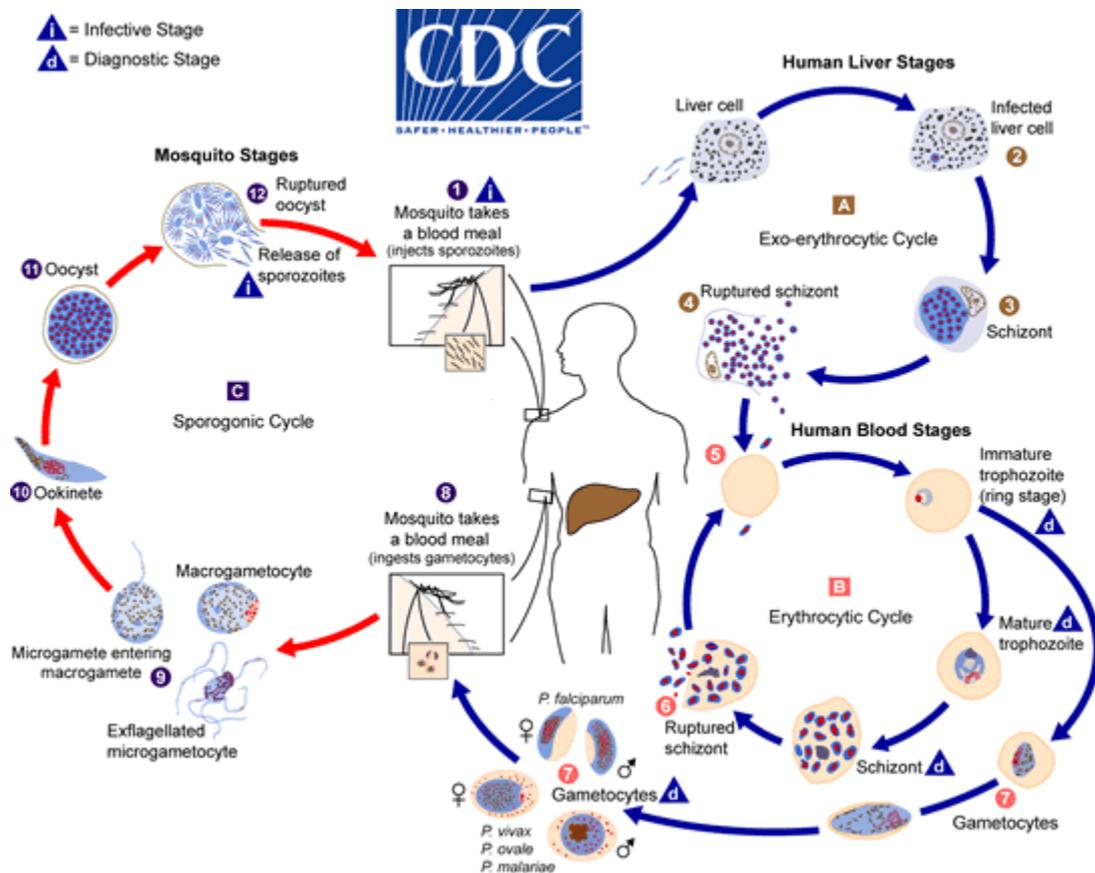
Figure 1.2: Malaria parasite life cycle in humans and in mosquitoes

Interventions that either provide a physical barrier between humans and mosquitoes, or interrupt the parasite's development cycle, or reduce mosquito population are encouraged and combined to maximize their impact on disease transmission (Korenromp et al., 2016). Since human and mosquito populations share environmental conditions, therefore a change in

climatic and environmental factors may also influence malaria transmission. It has been showed that rainfall and temperatures play an important role in the development cycle of *Plasmodium* parasites (Arab et al., 2014; Christiansen-Jucht et al., 2014; Okuneye et al., 2019).

Being infected can lead to a simple or a severe stage of the disease. Simple malaria symptoms are generally fever, headache and chills. Some of the following symptoms can be seen in severe malaria: respiratory distress, severe anemia, and multi-organ failure. The severe stage of the disease could be reached for many reasons, such as the absence of antimalarial treatment within a day (24 hours) of the onset of symptoms. After years of constant exposure to malaria parasites, some people develop partial immunity which induces the presence of asymptomatic groups in the population (Doolan et al., 2009). Simple and severe malaria cases are preventable and curable.

1.3 Health interventions

Core actions and their related interventions are organized into the management and prevention of malaria cases. They are associated with supportive interventions such as communication, administration, or monitoring and evaluation. The communication actions are useful to promote the prompt treatment of malaria cases and sensitize on the effective use of ITNs by households.

1.3.1   Malaria diagnosis and treatment

The World Health Organization recommends using microscopic laboratory analysis and rapid diagnostic tests for malaria diagnosis. Recent technological advances brought a new group of accurate diagnostic tools based on molecular detection called PCR (Polymerase chain reaction) test. However, those materials are relatively still expensive and required some technical skills (Berzosa et al., 2018; Siwal et al., 2018). Treatment of simple malaria cases is

currently ensured with the use of artemisinin-based combination therapy (ACT), a set of drugs that can affect the lifecycle of malaria parasites. Case management of malaria allowed purchasing and distributing recommended medicines and diagnostic tools to the entire health facilities.

1.3.2   Malaria prevention

Malaria prevention interventions may be categorized into:

- Physical protection: which include insecticide-treated bed nets.

- Chemo preventive: that can preserve humans from the disease status. Precisely, the Sulfadoxine-Pyrimethamine (SP) drug was administrated as intermittent preventive treatment (IPT) during pregnancy and also during the seasonal chemoprevention of the population at risk. In Cameroon, because of malaria drug resistance risk, the SP was combined with Amodiaquin (SP-AQ) in areas with strong seasonal malaria transmission (WHO, 2013a, 2013b).

- Insecticide spraying: that target reduction of mosquito population by using indoor residual spraying (IRS) in households and breeding sites.

1.4 Data sources

1.4.1   Household survey data

Significant global funding to fight against diseases like HIV/AIDS, Tuberculosis, and Malaria has heightened the need to provide reliable data for proper monitoring and evaluation of health interventions. Standardized sampling methods and tools were designed to harmonize data collection and enable cross-country comparison of indicators (Roll Back Malaria, USAID, CDC, UNICEF, WHO, 2013). Three types of national household surveys were mainly used during the last decade: Demographic and Health Survey (DHS), malaria

indicators survey (MIS) and multiple indicator cluster survey (MICS). DHS was the principal instrument for monitoring and evaluation of health and socioeconomic progress within a defined period that varies with the funding availability. The malaria indicators survey (MIS) was focused on estimating malaria prevalence during peak malaria transmission season and assessing the effects of interventions. The MICS was used to evaluate the situation of women and children in low- and middle-income countries (Roll Back Malaria, USAID, CDC, UNICEF, WHO, 2013; UNICEF, 1995; USAID, 1997)

These surveys were supported by the Government and international partners as the United States Agency for International Development (USAID), the United Nations International Children's Emergency Fund (UNICEF), the Global Fund against Aids, Tuberculosis and Malaria (GFATM) and others. From 2000 to 2020, four DHS, five MICS and only one MIS surveys were conducted in Cameroon. The data used in the thesis are mainly based on the 2011 DHS, the 2011 MIS and the 2014 MICS.

1.4.2 Health information data

In Cameroon, the malaria cases are self-reported by the patients to health centers that were public, private or confessional. Fever is often the triggering symptom for medical visits. Data are collected routinely on a daily basis and recorded on medical forms. Each health facility are located into a unique health district area and under the administrative coordination of that district. Malaria data collected at health facility level were compiled from medical forms and transcribed into the monthly malaria report at the district level. The chief of health district was responsible for ensuring that collected data were entered into the national malaria database. The questionnaire used to transfer data at the health district was differentiating clinical diagnosed cases findings from suspected cases of malaria.

1.4.3   Pilot site setting

Cameroon shares borders with Gabon, Congo and Equatorial Guinea in the south, Nigeria in the West, Chad in the North and Central African Republic in the East. Several ecological zones characterized by specific climatic and environmental factors cover the country. The country's administration was subdivided into 10 regions, 58 subdivisions and 189 health districts. In 2018, the population was estimated at 24 million and children under-five represented 16 percent. Urban dwellers accounted for about 49 per cent of the population and there were significant economic disparities between rural and urban areas (Minsante, INS and UNFPA, 2016). The country had six main ecological zones, that correspond to specific malaria transmission facets : the dry Sahelian in the Far North region and the Sudano-Guinean in the North region where the period of malaria transmission was between 4 and 6 months; the highlands of Adamawa and West regions where the length of transmission was from 7 to 12 months; the equatorial forests which includes Centre, East and part of South regions where the transmission was stable; and the Atlantic coastal covering the Littoral and a part of South and South-West regions where the malaria was perennial with seasonal variations. The country belongs to the fifteen countries that contribute to 80% of Malaria cases worldwide. The *Plasmodium falciparum* was the predominant species and responsible for more than 95% of confirmed infection cases in Cameroon. Five Anopheles species were the major responsible of the malaria transmission in Cameroon: An. *funestus*, An. *gambiae*, An. *moucheti*, An. *nili*, An. *Arabiensis* (Antonio-Nkondjio et al., 2006).

1.4.4   Environmental factors

Malaria transmission is driven by suitable climatic and environmental factors, therefore proxies of climatic conditions are used to predict the geographical distribution of the disease. Such data are generally stored as image files that can be extracted using appropriate software

tools. They were used as proxies of conditions prevailing on the ground and to assess climate changes. In the field of epidemiology, commonly used factors are: Land Surface Temperature during the Day and Night (LSTD, LSTN), Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), land cover surface, permanent water bodies, Rainfall Estimates (RFE) and altitude. Those climatic and environmental data could be retrieved from multi-satellite sources, and MODIS (or Moderate Resolution Imaging Spectroradiometer) that captures land and aquatic conditions was a relevant data source used in many scientific studies (Dlamini et al., 2019; U.S. Geological Survey, 2017)

1.5 Bayesian and spatial analysis

Malaria data collected usually by HMIS or household survey have space-time characteristics and depending of their sources, those data are associated to georeferenced data or space surface. Therefore, it becomes crucial to consider the presence of correlation in space and often in time. Models using Bayesian approaches known to be intensive are flexible, comprehensive and highly recommended in small sampling size cases. Furthermore, Bayesian statistics provide an appropriate frame for spatial prediction, estimation of parameters and their uncertainties.

Bayesian modeling has been increasingly applied in epidemiological analyzes and made relevant contribution to the monitoring of environmental diseases. In particular, to estimate the geographical distribution of disease risk and to assess intervention effects (Cressie, 1993; Green et al., 2015; MacLehose and Hamra, 2014; Diboulo et al., 2015). As part of the fight against malaria infection, several disease risk maps were produced at the national and continental levels, based on the relationship between individual parasitological results, climatic factors, and intervention indicators. Bayesian spatial modeling was useful to identify the heterogeneity of areas resulting from intervention effects, human behavioral and vector

resistance to insecticides (Bernardinelli et al., 1995; MacLehose and Hamra, 2014; Diboulo et al., 2015; Adigun et al., 2015; Khagayi et al., 2019).

The dependence between climatic factors has been proven and documented. As an example, a relationship was observed between rainfall and land surface temperature, and there was also an association between rainfall and vegetation coverage (Pandey et al., 2018; Schmidt et al., 2014). Thus, a Bayesian modeling of malaria risk that involves such climatic/environmental predictors should perform variable selection that take into account spatial correlation to identify the most important climatic and environmental factors, as well as the intervention indicators associated with the disease risk.

1.5.1 Variable selection

Different methods of variable selection have been proposed in the Bayesian analysis(O'Hara and Sillanpää, 2009). The stochastic search variable selection approach was mostly used with a slight adaptation in our thesis (Ishwaran and Rao, 2005). Bayesian regularized ridge regression was also used to shrink predictors towards zero and identified the most important ones (Hoerl and Kennard, 1970; Hsiang, 1975; Polson and Scott, 2011). Only the selected climatic, environmental and intervention indicators were included in the final models (Gemperli et al., 2004; Diboulo et al., 2016; Ssempiira et al., 2018; Giardina et al., 2014).

1.5.2 Bayesian spatial models

In the framework of Bayesian modeling, the generalized linear geostatistical models were built by including a spatial random effect in a Generalized Linear Models (GLM). Spatial random effects assumed to arise from a multivariate Gaussian distribution and a Matèrn correlation function that involved the distance between spatial positions was used to determine the covariance matrix (Matern, 1986; Diggle et al., 1998). Bayesian conditional autoregressive models were also fitted in this work (Gelfand and Vounatsou, 2003). Bayesian

inference was mainly implemented using Markov chain Monte Carlo (MCMC) approach. The Integrated Nested Laplace Approximations (INLA) was also used (Besag and Green, 1993; Rue et al., 2009).

1.6 Thesis objectives

The overall goal of the thesis is to evaluate and further strengthen the methodology of the malaria survey designs used for disease monitoring and evaluation, in particular the aspects related to the timing of the survey, the jittering in the coordinates of the reported cluster locations, and the seasonal monitoring of malaria data. Furthermore, we evaluated the ability of the survey data to estimate the malaria-related deaths and compared estimates of the effects of malaria interventions using data from malaria surveys and the Health Management Information System.

The thesis pursues the following specific objectives:

1. Assess the influence of the survey timing by comparing the DHS and MIS geostatistical model-based malaria risk estimates obtained at different malaria transmission seasons.

2. Assess the effects of the DHS jittering algorithm on the prediction of malaria risk and on the estimates of the disease risk factors.

3. Evaluate the effects of malaria interventions on the geographical distribution of disease incidence after adjusting for the effects of climatic and environmental factors.

4. Improve malaria disease and vector survey sampling designs by optimizing the selection of survey locations.

5. Evaluate the ability of malaria surveys to estimate malaria-related deaths.

Details on the methodology and results for each objective are given in the following five chapters, each written in the form of a publication or manuscript for peer-reviewed scientific journal.

Chapter 2 Challenges of DHS and MIS to capture the entire pattern of malaria parasite risk and intervention effects in countries with different ecological zones: the case of Cameroon

Salomon G Massoda Tonye[1,2,3], Celestin Kouambeng[3], Romain Wounang[4], Penelope Vounatsou[1,2,*]

[1]Swiss Tropical and Public Health Institute, Basel, Switzerland

[2]University of Basel, Switzerland

[3]National Malaria Control Programme, Yaoundé, Cameroon

[4]National Institute of Statistics, Yaoundé, Cameroon

Corresponding author

Email: Penelope.vounatsou@swisstph.ch

Abstract

Background

In 2011, the demographic and health survey (DHS) in Cameroon was combined with the multiple indicator cluster survey (MICS). Malaria parasitological data were collected, but the survey period did not overlap with the high malaria transmission season. A malaria indicator survey (MIS) was also conducted during the same year, within the malaria peak transmission season. This study compares estimates of the geographical distribution of malaria parasite risk and of the effects of interventions obtained from the DHS and MIS survey data.

Methods

Bayesian geostatistical models were applied on DHS and MIS data to obtain georeferenced estimates of the malaria parasite prevalence and to assess the effects of interventions. Climatic predictors were retrieved from satellite sources. Geostatistical variable selection was used to identify the most important climatic predictors and indicators of malaria interventions.

Results

The overall observed malaria parasite risk among children was 33% and 30% in the DHS and MIS data, respectively. Both datasets identified the normalized difference vegetation index and the altitude as important predictors of the geographical distribution of the disease. However, MIS selected additional climatic factors as important disease predictors. The magnitude of the estimated malaria parasite risk at national level was similar in both surveys. Nevertheless, DHS estimates lower risk in the North and Coastal areas. MIS did not find any important intervention effects, although DHS revealed that the proportion of population with an insecticide-treated nets (ITN) access in their household was statistically important. An important negative relationship between malaria parasitaemia and socioeconomic factors,

such as the level of mother's education, place of residence and the household welfare were captured by both surveys.

Conclusion

Timing of the malaria survey influences estimates of the geographical distribution of disease risk, especially in settings with seasonal transmission. In countries with different ecological zones and thus different seasonal patterns, a single survey may not be able to identify all high-risk areas. A continuous MIS or a combination of MIS, health information system data and data from sentinel sites may be able to capture the disease risk distribution in space across different seasons.

2.1 Introduction

Malaria is an endemic disease and a public health issue in Cameroon. It is a major cause of morbidity and mortality among children less than five years. In 2014, the morbidity of malaria was 30% in children and 18% in adults (MoH, 2001; PNLP, 2014a). Conscious of this situation, the government has considered the fight against malaria to be a national priority and part of the health strategic plan (INS and International ICF, 2012). Since 2002, the National Malaria Control Programme (NMCP) was created under the coordination of the ministry of public health. The aim was to improve the quality of strategic actions and to raise resources. During the last ten years, huge investments have been deployed by donors, the international community and the government, to develop strategies and tools for reducing the burden of malaria in the country. According to the national malaria strategic plan of 2014 - 2018 (PNLP, 2014b), the NMCP is implementing interventions to sustain and scale up malaria control. Those interventions include distribution of insecticide-treated nets (ITN) to populations at risk and of Sulfadoxine-Pyrimethamine (SP) to pregnant woman, parasitological confirmation of suspected malaria cases (microscopy or rapid diagnostic test), and treatment of uncomplicated malaria cases by Artemisinin-based Combination Therapy (ACT). Until 2011, the NMCP has distributed ITNs only to vulnerable groups. In 2012, the distribution policy has changed and more than eight million of long-lasting insecticide nets (LLIN) was given to populations at risk (RBM et al, 2008; WHO, 2015c). Before the LLIN mass campaign distribution, two representative surveys were carried out by the National Institute of Statistics: a Demographic and Health Survey (DHS) combined with Multiple Indicator Cluster Survey (MICS) and a Malaria Indicator Survey (MIS).

The DHS was the first national malaria survey to collect prevalence data across the country, however for logistic reasons data were collected outside the malaria high transmission season. The NMCP and partners have decided to conduct the MIS during the second and most

important rainy season (September-October), when the highest peak of malaria transmission occurs in order to assess the ability of DHS to estimate the malaria burden in the country (Tchinda et al., 2012). Hence, the objective of this study is to assess the influence of the survey period on the detection of risk pattern by comparing estimates of the malaria parasite risk and the effects of interventions obtained from both surveys. The analysis was carried out using Bayesian geostatistical logistic regression models similar to the ones that have been used for spatial analyses of other DHS and MIS data such as Angola, Senegal, Nigeria, Burkina Faso, Uganda and Sudan (Adigun et al., 2015; Diboulo et al., 2016; Giardina et al., 2012; Gosoniu et al., 2010; Noor et al., 2012; Ssempiira et al., 2017).

2.2 Methods

2.2.1 Country profile

Cameroon is a central Africa country, bordered with Nigeria to the West, Chad to the North, Central African Republic to the East, Congo, Gabon and Equatorial Guinea to the South. The country is decentralized and organized around 10 regions, 58 divisions and 360 communal areas. English and French are the official languages. Yaoundé is the political capital and Douala is the economic town. The global surface of the country is 475 650 km², population is around 22 million inhabitants (2010, 2005) and index of human development is 0.512 in 2015 (UNDP, 2015). The percentage of the population living in urban areas is 49%. Children under five years old represent 17% of the population (INS, 2013; INS and International ICF, 2012). Despite of the presence of natural resources as oil, gas, iron, gold, and favourable climatic situations for agriculture, the national income per inhabitant is still low (<2000$ per year) with important disparity between urban and rural areas (The World Bank, 2016). The country has different geographic and ecological zones, that generate six epidemiologic facets of malaria transmission (Antonio-Nkondjio et al., 2006; Ayala et al., 2009; PNLP, 2007), corresponding to different ecological systems: the dry Sahelian in the Far North region and

the Sudano-Guinean in the North region where malaria transmission period is between 4-6 months; the highlands of Adamawa and West regions with length of malaria transmission between 7-12 months; the equatorial forests which includes Centre, East and part of South regions where the transmission is stable; and the Atlantic coastal covering the Littoral and a part of South and South-West regions where the malaria is perennial with seasonal variations. The malaria transmission in the North part of Cameroon is characterized by seasonal pattern linked to rainy season which cover the period from August to October. Like many Africa countries, the *Plasmodium falciparum* is also the predominant species and responsible of more than 95% of confirmed infection cases in this study (Gething et al., 2011; Hay et al., 2009).

2.2.2 Malaria parasitological data

DHS-MICS 2011 Survey

DHS are nationally representative household based surveys commonly carried out by the National Institute of Statistics and ICF International in Africa or elsewhere collecting socioeconomic, demographic, disease and intervention related data. MICS is another standardized household survey carried out by UNICEF, compiling health related data on children and women. Both DHS and MICS were carried jointly in Cameroon during January to August 2011. A sample of 15 050 households living in 580 clusters was selected using a two-stage sampling approach, 291 clusters were in urban zone (Figure 2.1). Blood samples were taken in 50% of households inside the cluster surveyed and 5 515 children that are under five year were tested by a Rapid Diagnostic Test (SD BIOLINE Malaria Antigen Pf/Pan) (INS and International ICF, 2012).

MIS 2011 Survey

The MIS was carried out between September and November 2011, during the malaria high transmission season, one month after the increase of rains in the country. The MIS was conducted on 6 040 households within 257 clusters randomly selected out of the 580 clusters of the DHS 2011 (Figure 2.1). The sample size was determined using the same calculations as DHS; however it was based on the proportion of children aged 0 to 59 months using ITN in comparison to DHS that considered the proportions of a range of indicators. Malaria screening was performed in 4 939 children under five years old living in selected households and with the approval of adult in charge using a Rapid Diagnostic Test (First Malaria Response Antigen) (Minsante, INS, 2012).

2.2.3 Environmental and Climate factors

Environmental and climate predictors were extracted from satellite sources (Table 2.1). In particular, data were compiled on Land Surface Temperature during the day and night (LSTD, LSTN), Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Land cover surface and permanent water bodies obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) Terra satellite. Rainfall estimates (RFE) and altitude were retrieved from FEWS (or Famine Early Warning Systems Network) and SRTM (or Shuttle Radar Topographic Mission) web sites, respectively (John Weier and Herring David, 2000; U.S. Geological Survey, 2018). Climatic proxies with weekly and bi-weekly temporal resolution were averages over the one year period prior to the survey.

Table 2.1: Sources, spatial and temporal resolution of predictors

Data	Spatial resolution (km²)	Period	Temporal resolution
Land surface temperature (LST) for day and night [1]	1 × 1	2010-2011	8 days
Normalized difference vegetation index (NDVI) [1]	0.25 × 0.25	2010-2011	8 days
Enhanced vegetation index (EVI)[1]	1 × 1	2010-2011	16 days
Rainfall estimates(RFE) [2]	8 × 8	2010-2011	10 days
Land cover [1]	1 × 1	2010-2011	year
Digital elevation (Altitude) [3]	1 × 1	2011	NA
Permanent water bodies (rivers, lakes, wetlands) [1]	1 × 1	2011	NA

[1] https://reverb.echo.nasa.gov/reverb
[2] https://earlywarning.usgs.gov/fews/
[3] http://glcfapp.glcf.umd.edu/data/srtm/

2.2.4 Socio economic factors

Socio-economic data were included in both, the DHS and the MIS surveys. Two socio-economic proxies were used, education of women in reproductive age and household asset index. The education level was treated as a categorical variable with three levels (primary, secondary and university). The household asset index was included in the database and used in categorical form, grouped into quintiles corresponding to the poorest, poor, middle, rich and richest segments of the population. Rural and urban area information was available in the database for the observed survey locations and it was extracted from the GRUMP (or Global Rural and Urban Mapping Project) database at the locations of predictions (1997).
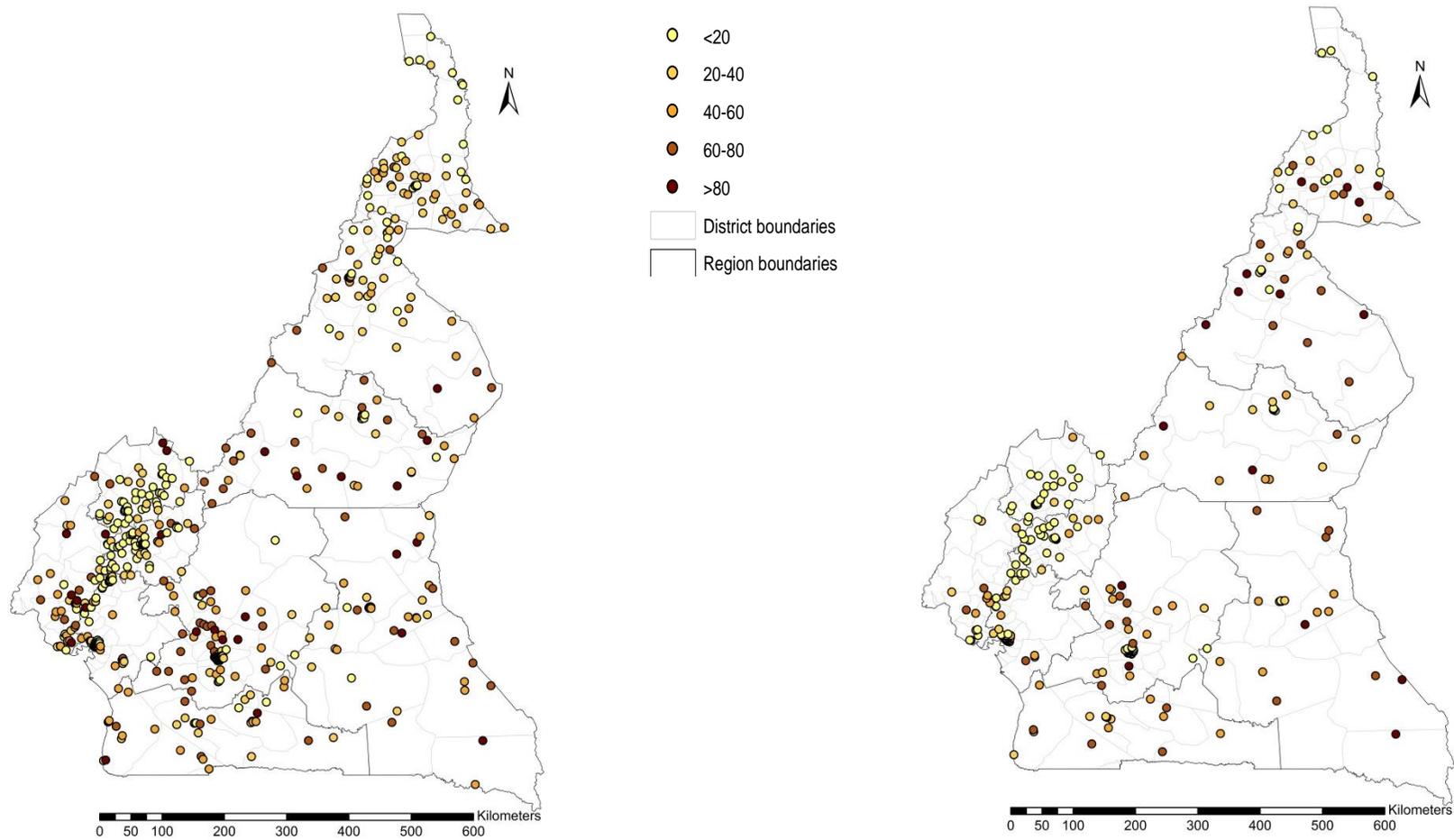
Figure 2.1: Observed malaria parasite risk in children under five years at 580 DHS locations (left) and at 257 MIS locations (right).

2.2.5 Intervention indicators

To capture the effects of interventions at national level, output indicators were generated using data available in the DHS and MIS, according to the household survey indicators tool for malaria control developed by Roll Back Malaria and partners. In particular, the following coverage indicators of use and access to ITN interventions was calculated: (a) proportion of children under five years old who slept under an ITN the previous night; (b) proportion of households in the cluster with at least one ITN; (c) proportion of households in the cluster with at least one ITN for every two people; (d) proportion of population with access to an ITN within their household. Furthermore, a health system performance indicator was calculated to measure the proportion of children under five years old with fever in the last two weeks who seek treatment at hospital, tested and treated with recommended ACT (Roll Back Malaria, USAID, CDC, UNICEF, WHO, 2013).

2.3 Bayesian geostatistical modelling

Bayesian geostatistical binomial models fitted on cluster level malaria aggregated data were used to estimate parasitaemia risk at high spatial resolution based on climatic predictors (Model 1). Climatic variables were categorized in groups with cut-offs defined from quintiles and exploratory analysis. Geostatistical variable selection was carried out to identify the most important climatic and environmental predictors, including their best fitting functional form (Giardina et al., 2014; O'Hara and Sillanpää, 2009). For each predictor a categorical indicator was introduced with values 0, 1 and 2 corresponding to exclusion of the predictor from the model or inclusion in linear or categorical form, respectively. The indicator was assumed to arise from a multinomial distribution with probabilities defining the variable-specific exclusion/inclusion probabilities (in linear/categorical forms) in the model. A threshold of 50% was considered for the probability of inclusion (i.e. posterior inclusion probability) into the predictive geostatistical model. In the final model, the effect of a predictor was considered

to be statistically important if the 95% Bayesian Credible Interval (BCI) of the coefficient did not include the one on the odds ratio scale. Validation of Model 1 was performed to assess the model's predictive performance. In particular, we divided the sample into a training set which included 80% of the data which was used for model fit and a test set consisting of the remaining data. Model validation compared the mean error between the observed parasitaemia at the locations of the test set with the model-based predicted risk. The model predictive performance was also evaluated by calculating the proportion of test locations correctly predicted within the 95% of BCI. Bayesian kriging was applied using Model 1 to predict the parasitaemia risk over a gridded surface of 117 192 cells and obtain pixel-level risk estimates at 2x2 km² resolution (Diggle et al., 1998; Handcock and Stein, 1993).

Geostatistical variable selection was also applied to select the most important coverage indicators of malaria interventions. A Bayesian geostatistical Bernoulli model was fitted on the parasitaemia status of each child to estimate the effect of selected malaria interventions (Model 2) after adjusting for potential confounding effects of the climatic factors used in Model 1 and of the socioeconomic factors. The same methodology was employed separately on the DHS and the MIS data. Model fit and prediction were conducted in R (R Core Team, 2016) and OpenBUGS version 3.2.3 (Imperial College and Medical Research Council, London, UK) (Gelfand and Smith, 1990; Sturtz et al., 2005). Convergence of parameters was assessed by the Geweke statistic and by visually inspecting the traceplots (Geweke, 1991). Computations were performed in the parallel scientific computing (sciCORE) platform of Basel University. Different maps were produced by ESRI's ArcGIS version 10.2.1 for Desktop (http://www.esri.com/).

2.4 Results

Table 2.2: Descriptive information of the DHS and MIS data

Survey information		DHS	MIS
Rainy period		-Low rainy season: April, May, June. -High rainy season : August, September, October	
Survey period		January to August	September to November
Numbers of locations		580	257
Numbers of households		15 050	6 040
Numbers of children aged 0-59 months surveyed		5 515	4 939
Parasitaemia prevalence		30 (12 - 53)	33 (6 - 57)
Socio economic			
Education level of mothers (%)	No education	20	23.2
	Primary	33.8	30.2
	Secondary	40.7	39.9
	University	5.5	6.7
Wealth index (%)	Most poor	22	25
	Very poor	21.8	22.2
	Poor	20.2	21.2
	Less poor	19.7	17.5
	Least poor	16.3	14
ITN ownership			
Percentage of households with at least one ITN		36.4 (26.5 - 52.3)	46.3 (30 – 61.9)
Percentage of households with at least one ITN for every two person		14 (8.6 - 27.6)	15 (6.8 - 26.9)
ITN use, ACT and Indoor Residual Spray coverage			
Percentage of children aged 0-59 months who slept under an ITN the night before the survey		21 (5.4 - 38.6)	35 (9.7 - 56.2)
Percentage of population with access to an ITN in their household		5 (3 – 9.7)	9 (2.7 - 13. 8)
Indoor residual spray		2.3 (0 - 8.9)	1.9 (0.2 - 5.2)
Percentage of children with fever in the last two weeks who seeked treatment and received ACT		6.1 (0.7 - 17.8)	12.6 (1.1 - 29.8)

DHS results

The observed parasitaemia risk in children under five years old was 30% at national level, 37% in the rural and 20% in the urban areas. In urbanized cities, such as Yaoundé and Douala the malaria parasite risk was among the lowest in the country, i.e. 12% and 13% respectively. Five per cent of the population had access to an ITN within their household and 21% of children slept under an ITN during the night preceding the survey. The percentage of children under five years old with fever in the last two weeks that treated with ACT was 6%. The

proportion of children from the poorest and poor quintiles was 63% and the proportion of mothers with at least primary education was 80% (Table 2.2).

The geostatistical variable selection identified NDVI and altitude (in categorical form) as the most important predictors of parasitaemia risk, using the cluster level model (Model 1). The proportion of the population with access to an ITN in their household and the proportion of children under five years old with fever in the last two weeks who seeked treatment at hospital, tested and treated with ACT were selected from Model 2 (Table 2.3).

Table 2.3: Posterior inclusion probabilities (%) of the climatic predictors and intervention coverage indicators obtained by the geostatistical variable selection applied to DHS and MIS data.

Model	Variable	DHS			MIS		
		Excluded	Continuous form	Categorical form	Excluded	Continuous form	Categorical form
Model 1,2 : Cluster level	Rainfall	58	42	0	36	18	46
	NDVI *	14	**86**	0	12	**83**	5
	LSTD	80	20	0	55	23	22
	EVI*	56	32	12	16	17	**67**
	Distance to water body*	41	33	26	23	25	**52**
	Altitude*	0	0	**100**	1	**96**	3
	LSTN	43	10	47	54	29	17
	Forest*	69	-	31	34	-	**66**
	Savana	58	-	42	69	-	31
	Cropland	82	-	18	72	-	28
Model 2 : Individual level	% population with access to an ITN in their household*	17	**83**	-	42	**58**	-
	% households with at least one ITN	62	38	-	64	36	-
	% children slept under ITN previous night	61	39	-	36	**64**	-
	% of households with one ITN per two persons*	52	48	-	33	**67**	-
	% of children with fever who received recommended antimalaria drugs (ACT) *	46	**54**	-	70	30	-

* Variable with posterior inclusion probability (continuous or categorical) above 50%.

Posterior estimates of the model's parameters are shown in Tables 2.4 and 2.5. The climatic cluster level model confirmed known relations between malaria parasite risk and climatic predictors, i.e. a positive association with NDVI and a negative relation with altitude. A malaria parasite risk map was generated using the climatic predictors identified from the DHS data (Figure 2.3).

The individual level model (Model 2 in Table 2.4) shows that children in urban areas or those living in households with higher socioeconomic level were less affected by malaria. Children

to mothers with high educational level or aged below twelve months had low malaria parasite risk. The proportion of population with access to an ITN in their household was able to capture a statistically important effect on parasitaemia risk.

MIS results

The national observed malaria parasite risk was 33% with substantial disparities between rural (43%) and urban (19%) areas. The North-West region and the towns of Yaoundé and Douala had registered low prevalence of 10%, 6% and 16%, respectively. According to the survey data, 9% of the population had access to an ITN within their household and 15% of households possessed one ITN per two persons. The percentage of children under five years old with fever in the last two weeks that received ACT was 12%. The percentage of children from the poorest and poor quintiles was 68 % and the proportion of mothers with at least primary education was 76% (Table 2.2).

The geostatistical variable selection applied to the cluster level model (Model 1) had identified NDVI, the categorical forms of EVI and of distance to water, the presence of forest and altitude as the most important predictors of parasitaemia risk. The individual level model estimated high posterior inclusion probabilities for the following ITN coverage proxies i.e. the proportion of population with access to an ITN in their household, the proportion of children who slept under an ITN in the previous night and the proportion of households with one ITN per two persons. However, the paired correlations between the above ITN indicators were ranging from 0.6 to 0.8; therefore the indicator included in the final model (Model 2) was the last ITN coverage measure which had the highest inclusion probability (Table 2.3).

Parameter estimates for Model 1 and 2 are shown in Table 2.5. The cluster level predictive model indicated that malaria parasite risk was positively related with NDVI, EVI and presence of forest, and it was negatively associated with altitude. A malaria parasite risk map was drawn with climatic predictors selected by the MIS (Figure 2.4).

The individual level model (Model 2 in Table 2.5) showed that children in rural areas as well as those living in households with lower socioeconomic status are more vulnerable to parasitaemia risk. Children aged below twelve months have low risk. The educational level of mother was not statistically associated with malaria parasite risk. ITN coverage was statistically important and had a negative effect on malaria parasite risk.

The proportion of test locations falling into the BCIs of the predictive posterior distributions with probability coverage varying from 50-95% was comparable for both surveys (Model 1), but the accuracy of estimates was higher for the DHS data as it is shown by the smallest BCI width (Figure 2.2) which indicates the difference between the upper and the lower values of the interval. Validation of the cluster level Model 1 on the DHS and MIS data showed a mean absolute error of 0.050 and 0.038, respectively.
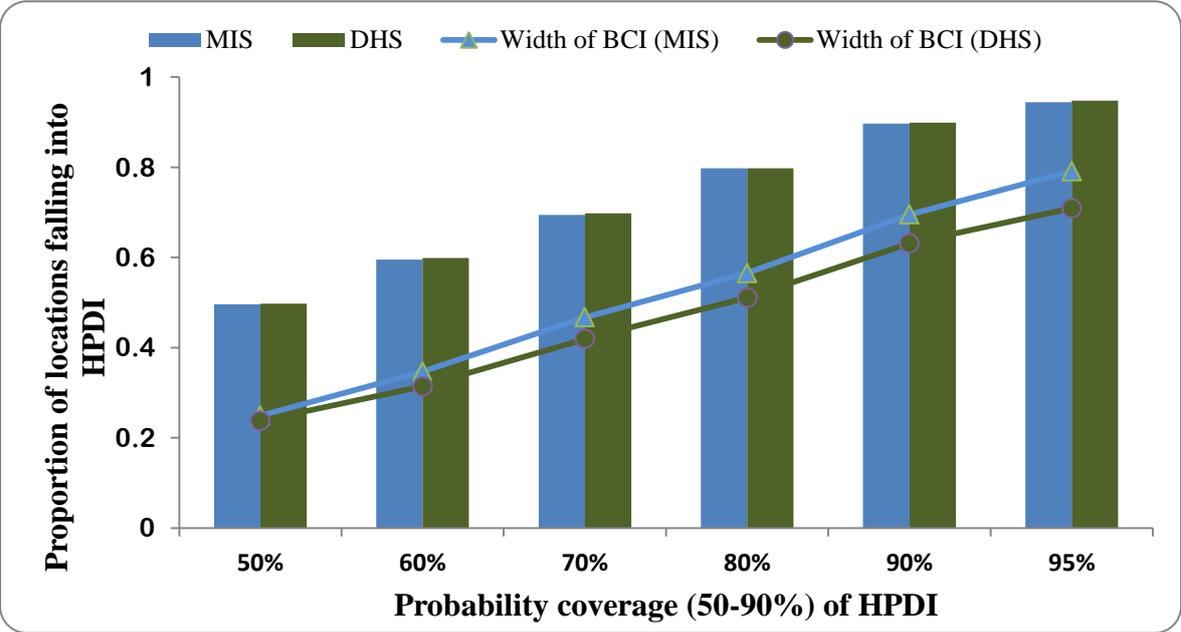
Figure 2.2: Proportion of test locations falling within highest posterior density intervals (HPDIs) of varying probability coverage

Table 2.4 : Estimates (posterior median and 95% BCI) of the geostatistical model parameters based on the cluster level climatic (Model 1) and the individual level model (Model 2), DHS 2011.

Factor		Model 1 OR (95% BCI)	Model 2 OR (95% BCI)
DHS			
NDVI		1.41 (1.19 ; 1,68)	1.22(0.99;1.5)
Altitude	<1000m	1	1
	1000-1500m	0.43 (0.26 ; 0.67)	0.37(0.22; 0.61)
	>1500m	0.267 (0.11 ; 0.60)	0.14(0.05; 0.35)
Sex	Female		1
	Male		1.03(0.90; 1.18)
Area type	Rural		1
	Urban		0.72(0.53; 0.96)
Wealth Index	Most poor		1
	Very poor		0.84(0.66; 1.08)
	Poor		0.91(0.67; 1.23)
	Less poor		0.78(0.54; 1.12)
	Least poor		0.32(0.21; 0.49)
Education level of mothers	No education		1
	Primary		0.83(0.69; 0.99)
	Secondary		0.68(0.55; 0.85)
	University		0.47(0.24; 0.89)
Age	0-1+		1
	1-2		1.83(1.43; 2.37)
	2-3		2.29(1.78; 2.94)
	3-4		2.92(2.28; 3.77)
	>4		3.10(2.41; 3.99)
% population access to an ITN in their household			0.23(0.07; 0.74)
% of children with fever in the last two weeks who received ACT			1.33(0.99 ;1.76)
Spatial parameters		Posterior median (95% BCI)	Posterior median (95% BCI)
σ^2		1.42 (1.09; 1.55)	1.43 (1.08; 1.97)
Range (km)		65.09 (44.90; 74.29)	81.63 (54.20; 134.54)

+: Children less than 6 months were not surveyed

Table 2.5: Estimates (posterior median and 95% BCI) of the geostatistical model parameters based on the cluster level climatic (Model 1) and the individual level model (Model 2), MIS 2011.

Factor		Model 1 OR (95% BCI)	Model 2 OR (95% BCI)
MIS			
NDVI		1.55 (1.12; 2.12)	1.33(0.97 ; 1.82)
EVI	<0.21	1	1
	0.21-0.38	1.90 (1.03; 3.51)	1.38(0.82 ; 2.33)
	>0.38	1.25 (0.51 ; 3.02)	0.92(0.41 ; 2.1)
Distance to water body	<70m	1	1
	>=70m	1.82 (1.005 ; 3.45)	1.60(0.90 ; 2.86)
Altitude		0.39 (0.26 ; 0.57)	0.37(0.25 ; 0.53)
Forest	No	1	1
	Yes	1.55 (1.002 ; 2.39)	1.17(0.77 ; 1.79)
Sex	Female		1
	Male		0.99(0.86 ; 1.15)
Area type	Rural		1
	Urban		0.55(0.38 ; 0.80)
Wealth Index	Most poor		1
	Very poor		0.6(0.46 ; 0.76)
	Poor		0.66(0.49 ; 0.89)
	Less poor		0.46(0.32 ; 0.66)
	Least poor		0.39(0.25 ; 0.61)
Education level of mothers	No education		1
	Primary		1.15(0.92 ; 1.43)
	Secondary		0.92(0.70 ; 1.22)
	University		1.03(0.57 ; 1.84)
Age	0-1+		1
	1-2		1.31(0.96 ; 1.77)
	2-3		2.29(1.70 ; 3.10)
	3-4		2.57(1.90 ; 3.48)
	>4		3.49(2.62 ; 4.65)
% households with 1 ITN per 2 persons			0.16(0.05 ; 0.47)
Spatial parameters		Posterior median (95% BCI)	Posterior median (95% BCI)
σ^2		1.81 (1.24 ; 2.92)	1.62 (1.10 ; 2.76)
Range (km)		154.8 (89.50; 292.96)	188.09(100.35 ; 353.63)
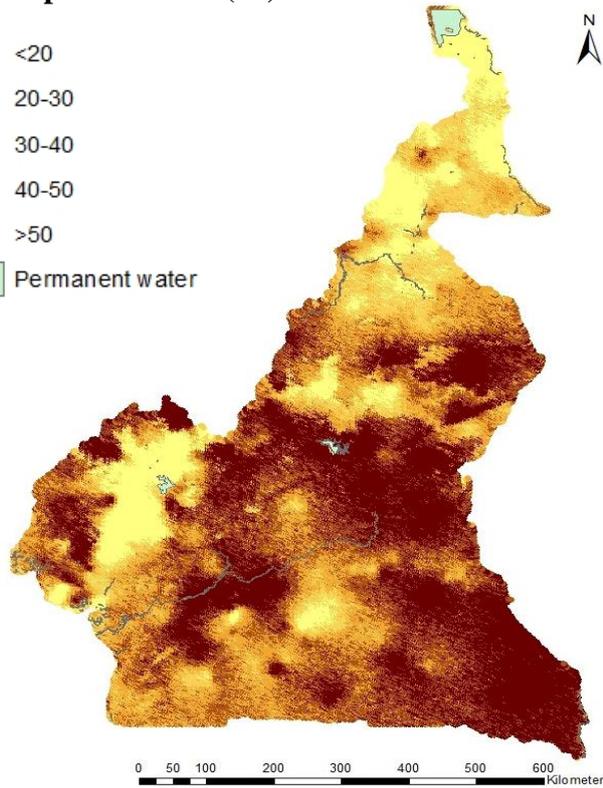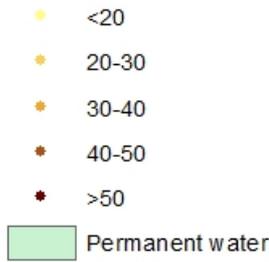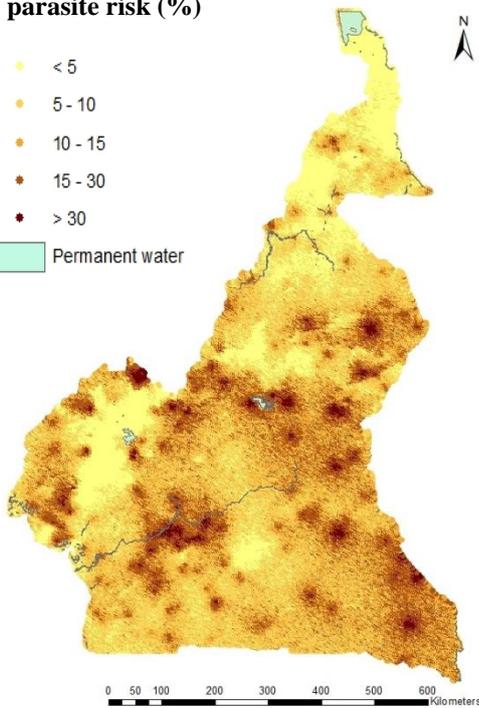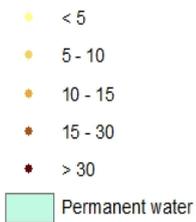
Malaria parasite risk (%)

Figure 2.3: Malaria parasite risk estimates among children less than 5 years, obtained from Model 1 using the DHS 2011; median (top), 2.5th percentile (bottom left) and 97.5th percentile posterior predictive distribution (bottom right).

Malaria parasite risk (%)

- <20
- 20-30
- 30-40
- 40-50
- >50
- Permanent water

2.5% percentile of the posterior distribution of predictive malaria parasite risk (%)

- < 5
- 5 - 10
- 10 - 15
- 15 - 30
- > 30
- Permanent water

97.5% percentile of the posterior distribution of predictive malaria parasite risk (%)

- <50
- 50-70
- 70-80
- 80-90
- >90
- Permanent water
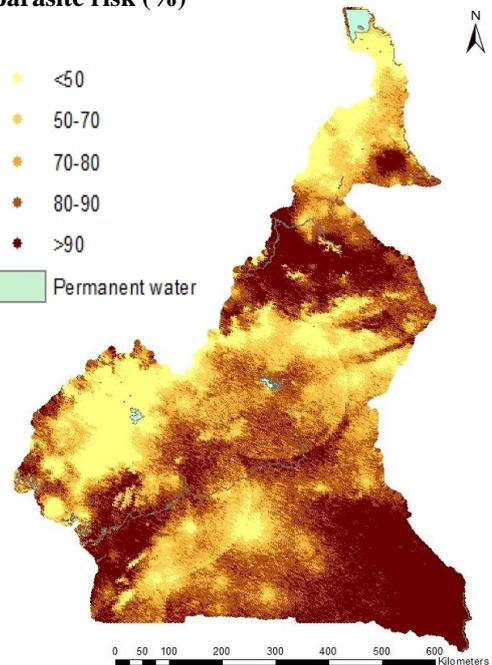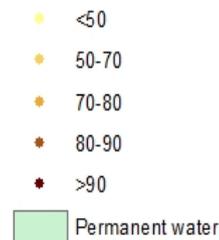
Figure 2.4: Malaria parasite risk estimates among children less than 5 years, obtained from Model 1 using the MIS 2011; median (top), 2.5th percentile (bottom left) and 97.5th percentile posterior predictive distribution (bottom right).

2.4 Discussion

This study is the first to assess the influence of survey season on the estimates of the geographical distribution of malaria parasite risk and of the effects of interventions, using data collected by DHS and MIS carried out at the same locations and year, but at different malaria transmission seasons. The analysis employed Bayesian geostatistical models because this study was interested in comparing the estimates of the risk pattern across the country rather than at the observed locations.

The DHS collects a large number of indicators on diverse sectors and huge logistics are involved to guarantee the coverage of all clusters, in particular those in rural areas with difficult access. Moreover, the planning of DHS usually avoids the rainy season in Africa because of road's degradation which challenges the survey implementation. The constraints described above have often an impact on the schedule and duration of DHS. The DHS and MIS surveys in Cameroon provide a unique opportunity to assess the effect of season on malaria survey-based estimates.

Both surveys showed low level of parasitaemia risk (under five percent) in West and Adamawa highlands. These areas are suitable for elimination interventions. Also, both data indicated that, the parasitaemia risk in East region was the highest in the country and above 50%. This high risk level is explained by the important coverage of forest, the predominance of rural areas and the low educational level of the population.

DHS data did not identify a cluster of high malaria parasite risk in the North and Far-North regions as estimated by the MIS. However, evidence from the upsurge of malaria cases that over strain the capacity of the health system during the rainy season and the high malaria mortality risk among children in the northern part of the country does not support the DHS finding (2013; White et al., 2014). The non-concomitance between DHS and the malaria

seasonal transmission in the north regions may explain the underestimation of malaria parasite risk in that area.

Furthermore, DHS could not capture a malaria cluster in the coastal part which is the estuary of the biggest rivers in the country that pour into the Atlantic Ocean. During the long rainy season that begins in August, some areas are flooded and large ponds of stagnant water are created (Bigoga et al., 2007, 2014; Nahum et al., 2010; Snow et al., 2015). The high transmission occurs just within the rainy season which is characterized by the increase of mosquito population. The water availability is among the key criteria for mosquito breeding, especially in the North part of Cameroon which is covered by the Sahel and in coastal towns, such as Douala because of poor condition of the pluvial drainage system. The model has identified additional climatic factors in MIS compared to the DHS.

In the Adamawa, North-West, West and Centre regions of the country, MIS estimated lower risk compared to DHS. In the capital, Yaoundé, parasitaemia risk was 6% based on MIS that is half the one obtained by the DHS data. The coverage of household by an ITN among the population of Yaoundé and Douala was 31% and 37%, respectively. But among households with at least one ITN, the percentage of those who use ITN in Yaoundé and Douala was among the highest in the country i.e. 43% and 52%, respectively. Human behaviour at the beginning of the rainy season changes and people are likely to increase the use of preventive tools, such as ITN, mosquito spraying devices or repellents (Frey et al., 2006; Atieli et al., 2011; Apinjoh et al., 2015).

The altitude and NDVI were identified as important predictors in the cluster level models of both surveys. The presence of forest, EVI and distance to water body were found to be important in modelling the MIS data. As known, the altitude has a negative effect on malaria parasite risk. The effect of distance to water was not linear and households located more than 70 meters away from water bodies are at higher risk of malaria compared to those households

49

close to them for a number of reasons including the wind direction and the availability of human hosts (Thomas et al., 2013). Rainy season has an influence on vegetation and on human activities, such as farming which exposes people to mosquito bites, and that could be the reason of the positive association between the EVI, NDVI, the presence of forest and the parasitaemia risk (Kar et al., 2014; Kimbi et al., 2013).

The analysis of the MIS data showed that the proportion of households with one ITN per two persons was statistically important with a negative effect indicating that the household coverage had an influence on malaria parasite risk among children(Larsen et al., 2014). According to the DHS, the ITN coverage indicator with a statistically important and protective effect was the population with access to an ITN. The use of ACTs among children under five years old with fever in the last two weeks before the survey was positively associated to the malaria parasite risk but not statistically important. Similar results regarding ACTs have been obtained from the MIS in Uganda and in Burkina-Faso (Diboulo et al., 2016; Ssempiira et al., 2017).

The disease risk resembles the pattern of socioeconomic inequalities in the country. In both surveys, the place of residence had an important effect and was negatively associated to malaria parasite risk. The DHS data showed that the effect of only the least poor category of the wealth index was statistically important compared to the most poor baseline category, however the MIS data estimated statistically important effects in all socio-economic categories. The educational level of mothers had a protective effect which was however statistically important only for the DHS. These results suggest that during the high malaria transmission season, the quality of the household environment is more important than the mother's education. Obviously, children from wealthy households can benefit from additional vectors control tools, such as appropriate malaria treatment, ITNs, sprays products and the sanitized neighbourhood. Wanzirah et al and Tusting et al (2017) have also shown that high

house quality reduces the entry of mosquito vectors and therefore lessens the risk of infection (Wanzirah et al., 2015; Tusting et al., 2017).

A gradient of malaria parasite risk was associated to the age and as expected the gender effect was not statistically important. Younger children were at lower risk than older ones, which may be a consequence of the passive immunity given by mothers (Doolan et al., 2009).

The high residual spatial correlation estimated by the models, especially those that used the MIS data indicates the presence of unmeasured spatially structured factors that influence the geographic distribution of the parasitaemia risk. It is likely that the climatic proxies considered in the model such as day and night LST or NDVI and EVI were not able to capture the entire ground climatic conditions. Similar analyses of other MIS data estimated relative high residual spatial correlation, particularly in recent surveys that climatic factors are confounded from malaria interventions (Adigun et al., 2015; Ssempiira et al., 2017; Giardina et al., 2014). The BCI width of the estimated parameters obtained with DHS were tighter than those of MIS, most likely due to the smaller number of survey clusters in the later (2002; Heckmann et al., 2014).

Both, DHS and MIS were used a RDT. RDTs could remain positive for few weeks after a malaria treatment. Therefore our estimates of parasitaemia risk may be slightly overestimated than those based on diagnosis by microscopy (Fançony et al., 2013; Golassa et al., 2013; WHO, 2015b).

DHS and MIS are based on a two-stage cluster sampling design. In the first stage, the number of clusters that are selected at regional level is proportional to the population. This design oversamples clusters in places with high population density and can selects fewer clusters over larger regions with small populations (i.e. East region) where the disease may vary more compared to the urban areas and big cities such as Yaoundé and Douala. Therefore, the DHS/MIS survey design may provide lower precision of the estimates in rural areas.

51

Since 2011, Cameroon has implemented two mass campaigns of LLINs, introduced preventive treatment of children against malaria in the North region and built two large dams in the East and South regions. There is currently a DHS on-going in Cameroon and the results of our study will serve as a baseline to assess the changes in malaria risk as a result of disease interventions, climatic effects and environmental modifications (Edlund et al., 2012; Kibret et al., 2015).

2.5 Conclusion

Timing of the malaria survey influences estimates of the geographical distribution of the disease risk, especially in settings with seasonal transmission. The DHS and MIS in Cameroon provide information about the geographical distribution of malaria parasite risk and of the effects of interventions in a country that different ecosystems cohabitate. In countries where malaria transmission is affected by seasonality, a single survey may not be able to identify all high risk areas. A continuous MIS similar to the one running for example in Senegal or a combination of MIS, health information system data and data from sentinel sites may be able to capture the disease distribution in space across different seasons. However, in countries with no variation in the malaria transmission season, a single survey may be sufficient.

2.6 Appendix

2.6.1 Geostatistical models formulation

Model 1: Estimating malaria risk at cluster level

Let y_i be the number of parasitaemia positive children among the n_i screened at cluster $s_i, (i = 1, \ldots, N)$. We adopt the Bayesian framework of inference and assume that Y_i has arisen from binomial distribution, i.e. $Y_i|n_i, p_i \sim Bin(n_i, p_i)$, where the malaria parasite risk at cluster i p_i, on the logit scale is a function of climatic factors $X_i = (1, X_{i1}, X_{i2}, \ldots X_{ip})$, and spatial random effects U_i, that is $\text{logit}(p_i) = \sum_{k=0}^{p} \beta_k X_{ik} + U_i$. We assume that U_i arise from a spatial process i.e. i.e. $U \sim MVN(\mathbf{0}, \Omega)$, where $U = (U_1, U_2, \ldots, U_n)^T$ and Ω is the variance-covariance matrix *NxN*, so that $\Omega_{ij} = \sigma^2 \exp(-\rho d_{ij})$. d_{ij} is the euclidian distance between cluster s_i and s_j, σ is the spatial variance also called the partial sill and ρ the smoothing parameter to control correlation. We consider non-informative prior for β_j defined such as $\beta_j \sim N(0,100)$.

Model 2: Estimating malaria risk at individual level

Let Y_{ij} be the status of malaria parasitaemia of child j who lives in cluster $s_i, (i = 1..N)$, We assume a Bernoulli distribution, i.e. $Y_{ij}|p_{ij} \sim Be(p_{ij})$, and model individual level function covariates on the logit scale of child specific risk, p_{ij}, i.e. $\text{logit}(p_{ij}) = \sum_{k=0}^{p} \beta_k X_{ijk} + U_j$ where, $X_{ij.} = (1, X_{ij1}, X_{ij2}, \ldots X_{ijp})$, the list of covariates including climatic factors and intervention coverage indicators, U_i is the dependence spatial effect of cluster s_i. We use the similar prior distributions for spatial and regression coefficients parameters as above in Model 1.

Prior distributions of the spatial process parameters

We assigned to σ^2 an Inverse Gamma prior distribution, $\sigma^2 \sim IG(2.01, 1.01)$ and adopt a Uniform prior distribution for ρ, i.e. $\rho \sim U(\frac{-\log(0.05)}{d_{max}}, \frac{-\log(0.05)}{d_{min}})$. The later prior assumes that the spatial correlation bellow 0.05 is negligible, d_{max} and d_{min} were associated to the maximum and minimum (non-zero) Euclidean distance between the survey locations.

2.6.2 Bayesian variable selection

We applied stochastic search variable selection assuming a normal mixture prior distribution for the regression coefficient of each predictor. A categorical indicator I_k is used to exclude the associated predictor from the model when $I_k = 0$, to select its linear form when $I_k = 1$ or to select the categorical form when $I_k = 2$. We assume a multinomial distribution prior for I_k with probability function $\prod_{l=0}^{2} \pi_l^{\delta_l(I_k)}$, where π_l represent the inclusion probability of each form ($l = 0, 1, 2$) and $\delta_k(.)$ is the Dirac function : $\delta_k(I_i) = \begin{cases} 1 & if \ I_i = k \\ 0 & if \ I_i \neq k \end{cases}$. We adopt a spike and slab prior for the regression coefficients $\beta_{k,l}$, which is a mixture of normal prior distributions i.e. $\beta_{k,l} \sim \delta_l(I_k)N(0, \tau_{k,l}^2) + (1 - \delta_l(I_k))N(0, c\tau_{k,l}^2)$ with a mixing proportion equal to the inclusion probability of the corresponding coefficient. The spike component shrinks the regression coefficient to zero when the variable is excluded and the slab assumes a non-informative, normal prior distribution when the inclusion probability of the specific functional form is high (i.e. larger than 50%). A non-informative prior Dirichlet distribution with hyper-parameter $\alpha = (1,1,1)^T$ is used, then $\pi = (\pi_1, \pi_2, \pi_3)^T \sim Dirichlet(3, \alpha)$, the constant c is fixed at 10^{-5} and inverse Gamma prior distribution are considered for $\tau_{k,l}^2 \sim IG(0.01, 0.01)$.

2.6.3 Bayesian kriging

Let $S_0 = \{S_{01}, S_{02} \ldots, S_{0k}\}$ be the centroids of the gridded surface. We predict the parasitaemia risk at the grid from the predictive posterior distribution (Sudipto Banerjee, 2015):

$$P(\widetilde{Y}_0|\widetilde{Y}) \ \alpha \int P(\widetilde{Y}_0|\widetilde{Y}, \tilde{\beta}, \widetilde{U_0}, \widetilde{U})P(\widetilde{U_0}|\widetilde{U}, \sigma^2, \rho)P(\widetilde{U}, \sigma^2, \rho|Y)P(\tilde{\beta}|\widetilde{U}, \rho, Y)P(\tilde{\rho}, \widetilde{U}|Y)d\beta d\rho dUd\sigma^2$$

$\widetilde{Y}_0 = Y(S_0)$, where $Y(s_o) \sim Be(N, p_{s_o})$ is the predicted number of children tested positive in the new location S_o and the associated covariates X_0. Conditional to the spatial process and model parameters, with the link relation between the risk and spatial covariate defined as $\text{logit}(p_{s_o}) = \sum_{k=0}^{p} \beta_k X_{s_o k} + U_0$.

Chapter 3 The influence of jittering DHS cluster locations on geostatistical model-based estimates of malaria risk in Cameroon

Salomon G Massoda Tonye[1,2,3], Romain Wounang[4], Celestin Kouambeng[3] and Penelope Vounatsou[1,2]

1. Swiss Tropical and Public Health Institute, Basel, Switzerland

2. University of Basel, Basel, Switzerland

3. National Malaria Control Program, Yaoundé, Cameroon

4. National Institute of Statistics, Yaoundé, Cameroon

Corresponding author

E-mail: Penelope.vounatsou@swisstph.ch

Abstract

Background

In low-and-middle income countries, national representative household surveys such as the Demographic Health Surveys (DHS) and the Malaria Indicator Surveys (MIS) are routinely carried out to assess the malaria risk and the coverage of related interventions. A two-stage sampling design is used to identify clusters and households within each cluster. To ensure confidentiality, DHS makes the data available after jittering the geographical coordinates of the clusters, shifting their original locations within a radius of 10 km. Our study assesses the influence of jittering on the estimates of the geographical distribution of malaria risk and on the effects of malaria control interventions using data from the latest MIS in Cameroon.

Methods

We generated hundred datasets by jittering the original MIS data. For each dataset, climatic factors were extracted at the jittered locations and Bayesian geostatistical variable selection was applied to identify the most important climatic predictors and malaria intervention coverage indicators. The models were adjusted for potential confounding effects of socio-economic factors. Bayesian kriging based on the selected models was used to estimate the geographical distribution of malaria risk. The influence of jittering was analysed using results of the variable selection and the Bayesian credible intervals of the regression coefficients.

Results

Geostatistical variable selection was sensitive to jittering. Among the important predictors identified in the true data, distance to water bodies and presence of forest were mostly influenced by the jittering. Altitude and vegetation index were the least affected predictors. The various sets of selected environmental factors were able to capture the main spatial patterns of the disease risk, but the jittering increased the prediction error. The parameter

56

estimates of the effects of socio-economic factors and intervention indicators were relatively stable in the simulated data.

Conclusion

In Cameroon, the malaria risk estimates obtained from the jittered data were comparable to the ones generated using the true locations, however jittering modified our interpretation upon the relationship between environmental predictors and malaria transmission.

Keywords: Bayesian inference, demographic and health survey (DHS), malaria indicator survey (MIS), jittering, geostatistics, malaria interventions.

3.1 Introduction

During the last decade, Bayesian geostatistical models have increasingly been used to determine spatio-temporal patterns of malaria risk, capture the effects of control interventions, and identify environmental and socioeconomic factors that are related to changes in the distribution of malaria risk (Gosoniu et al., 2012; Adigun et al., 2015; Diboulo et al., 2016; Ssempiira et al., 2017). In most low- and middle-income countries, the data used to fit geostatistical models are mainly collected by national household surveys such as the Demographic and Health Surveys (DHS) and the Malaria Indicators Survey (MIS) (Giardina et al., 2014). A two-stage sampling design is used to select survey clusters and households within clusters. The clusters include typically around 25 households and are georeferenced according to their centroid. However, to ensure confidentiality of the health status of the enrolled individuals, the longitude and latitude of the cluster centroids are randomly jittered from their original positions within a radius of 0 to 10 km according to the type of location (rural / urban) (Burgert CR, Colston J, Roy T, Zachary B., 2013).

Some studies have assessed the influence of imprecise geographical locations on model fit (Kyriakidis and Dungan, 2001; Cressie and Kornak, 2003; Gryparis et al., 2009; D. Li et al., 2012). In particular, studies on jittering DHS data have investigated the impact of spatial displacement on the estimates of the effects of distance-based covariates such as proximity to health services or areal covariates such as poverty measures defined in areas around a cluster location. These studies have been conducted in the field of HIV infection (Warren et al., 2016a, 2016b) and using simulated and real data assessed the potential effects of locations shift on model parameter estimates. However, within the Bayesian geostatistical modelling framework, studies assessing the effects of the cluster displacement on the pixel-level

predictions of disease risk such as malaria and on the estimates of the covariates, for example climatic factors or control intervention effects are rather lacking.

The fourth DHS in Cameroon was combined with the Multiple Indicators Cluster Survey (MICS) in 2011 and carried out between January and August, a period which unfortunately did not overlap with the high malaria transmission season. In the same year, the National Malaria Control Program (NMCP), the National Institute of Statistics (NIS) and other partners have conducted a MIS from September to November within the high malaria transmission season on a subset of clusters previously surveyed by DHS. The geographical coordinates of the DHS clusters involved in the MIS were registered without any alteration (Minsante, INS, 2012; Massoda Tonye et al., 2018; INS and International ICF, 2012).

Our study aims to assess the influence of jittering of cluster locations on geostatistical model-based malaria risk estimates at high spatial resolution and on the estimates of the control interventions effects. A large simulation study was carried out based on the MIS cluster locations and the random displacement procedure of DHS. Bayesian geostatistical models were applied on the simulated data and the results were compared with the non-jittered data.

3.2 Methods

3.2.1 Country settings

Cameroon, a country in Central Africa has a population of around 24 million inhabitants with an annual population growth of 2.5% within the territory surface of 475 650 km^2 (BUCREP, 2011). Fifty one percent of the population lives in urban areas(INS, Minsante et UNICEF, 2015). In 2017, the gross domestic product rate was 3.1% and the last estimates of human development index done in 2014 was 0.518(UNDP, 2015). The country is spanned by different ecological environments with various length of malaria transmission, namely: the dry Sahelian in the Far North region and Sudano-Guinean in the North region (4-6 months

length), the highlands of Adamawa region and West (7-12 months length); the equatorial forests in Centre, East and South regions; the Atlantic coastal in Littoral, South-West and part of South regions where malaria was perennial (Gemperli et al., 2006; PNLP, 2014b, 2007).

3.2.2 Data

3.2.2.1 Malaria indicator survey data

The Cameroon Malaria Indicator Survey (MIS) of 2011 was nationally representative and funded by the Global fund against AIDS, Tuberculosis and Malaria with the aim to collect malaria indicators additional to those in DHS and to compare the overall malaria parasite prevalence obtained by the MIS and DHS data (Ministry of Public Health, 2010). MIS was conducted in 257 clusters randomly selected out of the 580 clusters of the Cameroon DHS 2011 and involved 6.040 households and 4.939 children aged between 6 to 59 months (Figure 3.1). A Rapid Diagnostic Test (First Malaria Response Antigen) was used for malaria screening of children with the approval of adult in charge (Minsante, INS, 2012). Apart of the malaria parasite data, the survey collected information on malaria interventions and socio-economic proxies.

Data on malaria interventions were processed to create the following intervention coverage indicators as proposed by the Global Malaria Action Plan and Roll Back Malaria monitoring and evaluation group: (a) proportion of children in the households who slept under an insecticide treated-net (ITN) the night before the survey, (b) proportion of households in the cluster with at least one ITN, (c) proportion of households in the cluster with one ITN per two persons, (d) proportion of population with access to an ITN in their household. Adherence to the health system was calculated by the proportion of children with fever who sought treatment at hospital, tested and treated with the recommended anti-malaria drugs (ACT) during the last two weeks (Roll Back Malaria, USAID, CDC, UNICEF, WHO, 2013).

The analysis included the education level of women in reproductive age and the household welfare index as socio-economic proxies. The education level was categorized into three levels (primary, secondary and university). The household asset index was available in the database and classified households into the poorest, poor, middle, rich and richest categories. The area type (urban or rural) was extracted from the MIS data.

3.2.2.2 Simulated data

Hundred datasets were generated from the original MIS data, each with randomly jittered cluster locations from the MIS coordinates according to the jittering algorithm used by the DHS program. In particular, clusters in urban areas were randomly displaced within a radius of 2 kilometers; whilst 99% of those in rural areas were shifted within a radius of 5 kilometers from their original locations. The remaining 1% of rural clusters were displaced up to a radius of 10 kilometers (Burgert CR, Colston J, Roy T, Zachary B., 2013). The simulated data differ from the MIS data in the cluster coordinates. The prevalence, intervention and socio-economic information maintained the same as at the original locations.
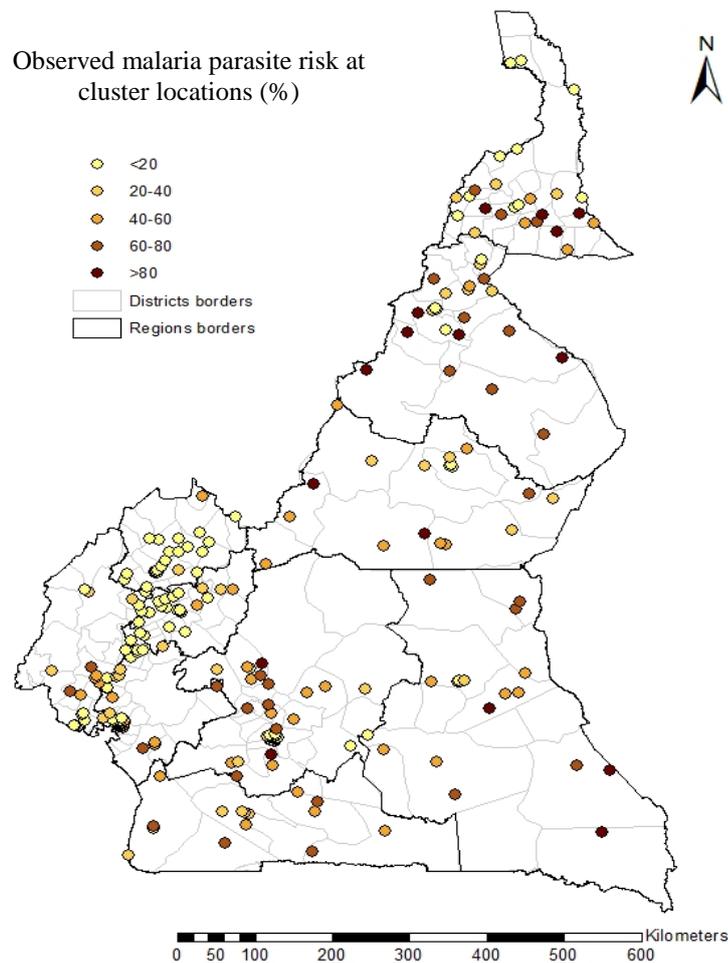
Figure 3.1: Observed malaria parasite risk in children under five years at 257 MIS locations.

3.2.2.3 Environmental and climate data

Environmental and climate proxies were obtained from satellite sources (Table 3.4 in Appendix). Day and night Land Surface Temperature (LSTD, LSTN), Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and Rainfall estimates (RFE) were averaged over the year prior to the survey. The covariates forest, savannah, cropland and distance to permanent water bodies (DWB) were retrieved from Land cover. The data were extracted at the MIS cluster locations and at the locations of simulated datasets.

3.3 Bayesian geostatistical modelling

Bayesian geostatistical binomial regression models were fitted on the malaria parasite data (MIS and simulated ones) aggregated at cluster locations. The models incorporated geostatistical variable selection to identify the most important climate and environmental covariates including their functional forms (i.e. continuous or categorical). In particular, a categorical indicator was created from each climatic predictor, taking the values 0, 1 and 2 which corresponds to the exclusion of the predictor from the model or its inclusion in linear or categorical form, respectively. It was assumed that the indicator arose from a multinomial distribution with probabilities defining the variable-specific exclusion/inclusion probabilities (in linear/categorical forms) in the model. A threshold of 50% was considered for the probability of inclusion (i.e. posterior inclusion probability) into the predictive geostatistical model (Barbieri and Berger, 2004). The predictive performance of the models obtained from each simulated dataset was evaluated using the log predictive score comparing model-based predictions at the MIS locations with the observed MIS survey data (Krnjajić and Draper, 2014).

Bayesian kriging (Diggle P. J. et al., 2002) was applied on the observed data as well as on the simulated data with the best and least predictive performance (i.e. maximum and minimum log predictive score, respectively) employing the corresponding models identified by the data, namely Model 1a (MIS data), Model 1b (simulated data with the best predictive performance) and Model 1c (simulated data with the worst predictive performance). For each one of on the above models, a gridded surface of malaria parasite risk was estimated over 117 192 cells/pixels of 2x2 km² spatial resolution covering the country.

To assess the effect of jittering on individual level covariates such as the ITN coverage indicators, geostatistical Bernoulli models were fitted on individual level data obtained from the observed MIS and the simulated data. As described above, three models were fitted i.e

Model 2a (applied to MIS data), Model 2b and 2c (applied to simulated data) with the best and worse predictive ability, respectively. We implemented geostatistical variable selection to identify the most important ITN indicators (Giardina et al., 2014). The individual level models were adjusted for the confounding effects of the climatic predictors selected by the corresponding cluster-level model and the socio-economic proxies. Due to high correlation among the ITN indicators, only one indicator allowed into the model.

Covariates were statistically important when the corresponding Bayesian Credible Interval (BCI) did not include the one in the odds ratio scale. Computation was performed on a dual processor workstation (2x2.6 Ghz, 128GB RAM). OpenBUGS version 3.2.3 (Imperial College and Medical Research Council, London, UK) was used for Bayesian model fit and prediction (Lunn et al., 2000). Data management and analysis were carried out in R statistical software (R Core Team, 2016; Sturtz et al., 2005). Convergence was assessed by the Geweke statistic, visual inspection of the traceplots and achieved in less than 200 000 iterations (Geweke, 1991). Maps were drawn in ArcGIS version 10.2.1 (http://www.esri.com/)(ESRI, 2013).

3.4 Results

Descriptive analysis

The overall malaria prevalence estimated by the MIS data was 33%. In the rural areas, 43% of children were tested positive, meanwhile this proportion was 19% in urban areas. The most affected areas were located in the North, East and South regions of Cameroon with malaria risk 57.2%, 56.5% and 50.9%, respectively. The proportion of mothers that had attended university was 6.7% and those without any education were 23.3%. The proportion of households with at least one ITN was 46%. Only nine per cent of the population had access to an ITN in their household and 12% of children with fever who sought treatment at hospital

received a recommended ACT during the last two weeks. Sixty-eight percent (68 %) of households were most poor or poor.

.
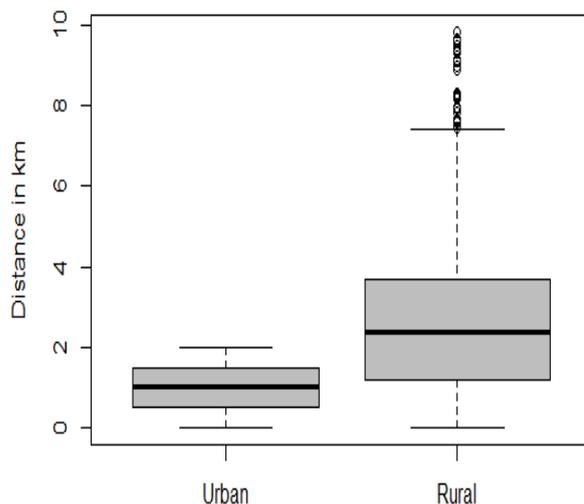
Figure 3.2 : Distribution of the distances (km) between the original and shifted locations across the 100 simulated datasets according to urban and rural cluster

Figure 3.2 displays the distribution of the distances between the original and shifted locations across the 100 simulated datasets. As expected by the DHS jittering algorithm, the median distance between the true and their jittered locations was larger in rural than urban clusters.

Geostatistical analysis of MIS data

The geostatistical variable selection performed at the cluster level model (Model 1a) fitted on the original MIS data identified NDVI and altitude (in continuous form), EVI and distance to water (in categorical form) and the presence of forest (binary) as the most important predictors of parasitaemia risk (Table 3.1). Estimates of the final geostatistical model (Table 3.2) indicated that the malaria parasite risk was positively associated with NDVI, EVI, and presence of forest, and negatively associated with altitude. The individual level model (Model 2a) fitted to the original MIS data selected the proportion of households with 1 ITN per 2

persons as the most important predictor (Table 3.1). The association of this predictor with the parasitaemia risk was negative as shown in the final geostatistical model (Table 3.2).

Table 3.1: Posterior inclusion probabilities (%) of the climatic predictors and intervention coverage indicators based on geostatistical variable selection applied on the three datasets i) MIS (Model 1a, Model 2a) ii) simulated data with best predictive ability (Model 1b, Model 2b) and iii) simulated data with worst predictive ability (Model 1c, Model 2c)

Model	Variable	MIS data			Simulated data with best predictive ability			Simulated data with worst predictive ability		
		Excluded	Continuous	Categorical	Excluded	Continuous	Categorical	Excluded	Continuous	Categorical
Model 1a, 1b, 1c: Cluster level	RFE	36	18	46	56	18	26	26	10	**64**
	NDVI *	12	**83**	5	2	**98**	0	20	**58**	22
	LSTD	55	23	22	59	19	22	60	17	23
	EVI*	16	17	**67**	0	0	**100**	12	42	46
	DWB*	23	25	**52**	33	26	41	10	14	**76**
	Altitude*	1	**96**	3	3	**79**	18	2	**98**	0
	Forest*	34	-	**66**	41	0	**59**	36	0	**64**
	Savannah	69	-	31	68	0	32	74	0	26
	Cropland	72	-	28	81	0	19	75	0	25
	LSTN	54	29	17	46	**54**	0	24	10	**66**
Model 2a,2b,2c : Individual level	% population access to an ITN in their household	84	16	-	87	13	-	82	18	-
	% households with at least one ITN	100	0	-	84	16	-	98	2	-
	% of households with one ITN per two persons*	47	**53**	-	42	**58**	-	77	23	-
	% children slept under ITN previous night*	69	31	-	87	13	-	43	**57**	-
	% of children with fever who received recommended anti-malaria drugs (ACT)	73	27	-	76	24	-	70	30	-

Geostatistical analysis of simulated data

Geostatistical variable selection applied to each one of the simulated data identified 26 sets of climatic and environmental predictors that were included in the selected model (Table 3.5 in

Appendix). Two models had among the highest posterior inclusion probabilities equal to 19% and 18%; that is they were selected in 19% and 18% of the simulated data, respectively. Both models included NDVI and altitude (continuous), EVI (categorical) and forest presence. Furthermore, the DWB was included in the second most frequent model (with inclusion probability of 18%).

In accordance with the original (un-jittered MIS data), all simulated data identified the altitude (in continuous form) as an important predictor, while the LSTD, savannah and cropland land use types were excluded from all the selected models (Table 3.3). LSTN was rarely included in the set of important predictors.

Table 3.2: Estimates (posterior median and 95% BCI) of the geostatistical model parameters based on the cluster level (Models 1a, 1b, 1c) and the individual level models (Models 2a, 2b, 2c)

Factor		MIS data		Simulated data with the best predictive ability		Simulated data with the worst predictive ability	
		Model 1a OR (95% BCI)	Model 2a OR (95% BCI)	Model 1b OR (95% BCI)	Model 2b OR (95% BCI)	Model 1c OR (95% BCI)	Model 2c OR (95% BCI)
RFE	0-30 mm					1	1
	30-60 mm					0.83(0.24 ; 2.49)	2.54(0.86 ; 7.97)
	>60 mm					0.36(0.08 ; 1.65)	1.07(0.30 ; 3.98)
NDVI		1.55 (1.12; 2.12)	1.33(0.97 ; 1.82)	1.63(1.18 ; 2.29)	1.31(0.96 ; 1.82)	1.6 (1.23 ; 2.09)	1.32(1.03 ; 1.70)
EVI	<0.21	1	1	1	1		
	0.21-0.38	1.90 (1.03; 3.51)	1.38(0.82 ; 2.33)	1.95(1.03 ; 3.67)	1.33(0.79 ; 2.22)		
	>0.38	1.25 (0.51 ; 3.02)	0.92(0.41 ; 2.1)	1.24(0.49 ; 3.08)	0.9(0.40 ; 1.98)		
Distance to water body	<70m	1	1			1	1
	>=70m	1.82 (1.005 ; 3.45)	1.60(0.90 ; 2.86)			1.98(1.09 ; 3.95)	1.74(0.98 ; 3.17)
Altitude		0.39 (0.26 ; 0.57)	0.37(0.25 ; 0.53)	0.53(0.3 ; 0.91)	0.42(0.25 ; 0.72)	0.38(0.24 ; 0.6)	0.35(0.23 ; 0.54)
Forest	No	1	1	1	1	1	1
	Yes	1.55 (1.002 ; 2.39)	1.17(0.77 ; 1.79)	1.49(0.95 ; 2.34)	1.18(0.77 ; 1.81)	1.46(0.93 ; 2.28)	1.06(0.68 ; 1.62)
LSTN_continuous				1.37(0.83 ; 2.29)	1.19(0.76 ; 1.89)		
LSTN_categrical	0-14					1	1
	14-18					1.82(0.29 ; 18.01)	2.06(0.63 ; 14.04)
	>18					1.45(0.21 ; 16.22)	1.87(0.53 ; 13.42)
Gender	Female		1		1		1
	Male		0.99(0.86 ; 1.15)		0.99(0.86 ; 1.15)		1(0.87 ; 1.15)
Area type	Rural		1		1		1
	Urban		0.55(0.38 ; 0.80)		0.54(0.38 ; 0.78)		0.56(0.39 ; 0.81)
Wealth Index	Most poor		1		1		1
	Very poor		0.60(0.46 ; 0.76)		0.6(0.47 ; 0.76)		0.6(0.47 ; 0.76)

Factor		MIS data		Simulated data with the best predictive ability		Simulated data with the worst predictive ability	
		Model 1a *OR (95% BCI)*	Model 2a *OR (95% BCI)*	Model 1b *OR (95% BCI)*	Model 2b *OR (95% BCI)*	Model 1c *OR (95% BCI)*	Model 2c *OR (95% BCI)*
	Poor		0.66(0.49 ; 0.89)		0.66(0.48 ; 0.88)		0.65(0.48 ; 0.87)
	Less poor		0.46(0.32 ; 0.66)		0.45(0.31 ; 0.65)		0.46(0.32 ; 0.66)
	Least poor		0.39(0.25 ; 0.61)		0.39(0.25 ; 0.61)		0.38(0.25 ; 0.60
Education level of mothers	No education		1		1		1
	Primary		1.15(0.92 ; 1.43)		1.14(0.91 ; 1.42)		1.15(0.92 ; 1.44)
	Secondary		0.92(0.70 ; 1.22)		0.92(0.7 ; 1.21)		0.93(0.7 ; 1.22)
	University		1.03(0.57 ; 1.84)		1.03(0.56 ; 1.83)		0.98(0.53 ; 1.73)
Age	0-1+		1		1		1
	1-2		1.31(0.96 ; 1.77)		1.32(0.97 ; 1.79)		1.34(0.99 ; 1.81)
	2-3		2.29(1.70 ; 3.10)		2.30(1.71 ; 3.10)		2.32(1.73 ; 3.12)
	3-4		2.57(1.90 ; 3.48)		2.59(1.92 ; 3.49)		2.60(1.93 ; 3.51)
	>4		3.49(2.62 ; 4.65)		3.38(2.51 ; 4.54)		3.40(2.54 ; 4.57)
% households with 1 ITN per 2 persons			0.16(0.05 ; 0.47)		0.14(0.05 ; 0.44)		
% of children with fever in the last two weeks who received ACT							0.35(0.18 ; 0.66)
Spatial parameters		Posterior median (95% BCI)	Posterior median (95% BCI)	Posterior median (95% BCI)	Posterior median (95% BCI)	Posterior median (95% BCI)	Posterior median (95% BCI)
Spatial variance		1.81 (1.24 ; 2.92)	1.62(1.10 ; 2.76)	1.88(1.22 ; 3.6)	1.64(1.17 ; 2.5)	1.87(1.29 ; 3.13)	1.64(1.1 ; 2.8)
Range (km) [1]		154.8 (89.50; 292.96)	188.09(100.35 ; 353.63)	111.43(62.26 ; 217.13)	214.98(120.37 ; 487)	143.29(75.72 ; 301.36)	158.87(94.18 ; 321.1)

1: Smallest distance that spatial correlation is <5%

Table 3.3: Relative frequencies of the climatic predictors and their functional forms identified by the geostatistical variable selection across the 100 simulated data. The predictors selected by the original data are in bold.

	Climatic predictors									
Functional form	RFE	NDVI	LSTD	EVI	DWB	Altitude	Forest	Savannah	Cropland	LSTN
Continuous	0%	**95%**	0%	24%	0%	**100%**	0%	0%	0%	1%
Categrorical	25%	1%	0%	**51%**	**64%**	0%	**90%**	0%	0%	3%
Excluded	75%	4%	**100%**	25%	36%	0%	10%	**100%**	**100%**	**96%**

The estimates of the effects of climatic predictors based on the selected models were overlapping between the simulated datasets (Figures 3.3-3.4). Altitude (in continuous form) was always statistically important and negatively associated with the parasitaemia risk. NDVI was positively associated and statistically important for the malaria parasite risk in most simulated data. Malaria parasite risk had a positive and most often statistically important relationship with the presence of forest, EVI and DWB. On the other hand, the importance of RFE, LSTN or LSTD on parasitaemia risk varied with the data.

RFE(30-60mm vs <30mm)   RFE (>60mm vs RFE <30mm)
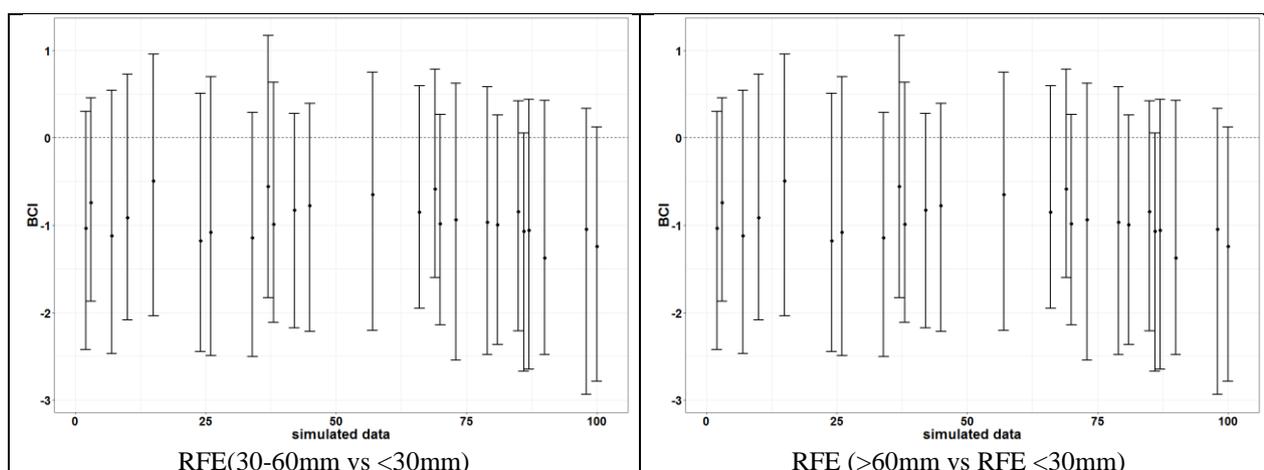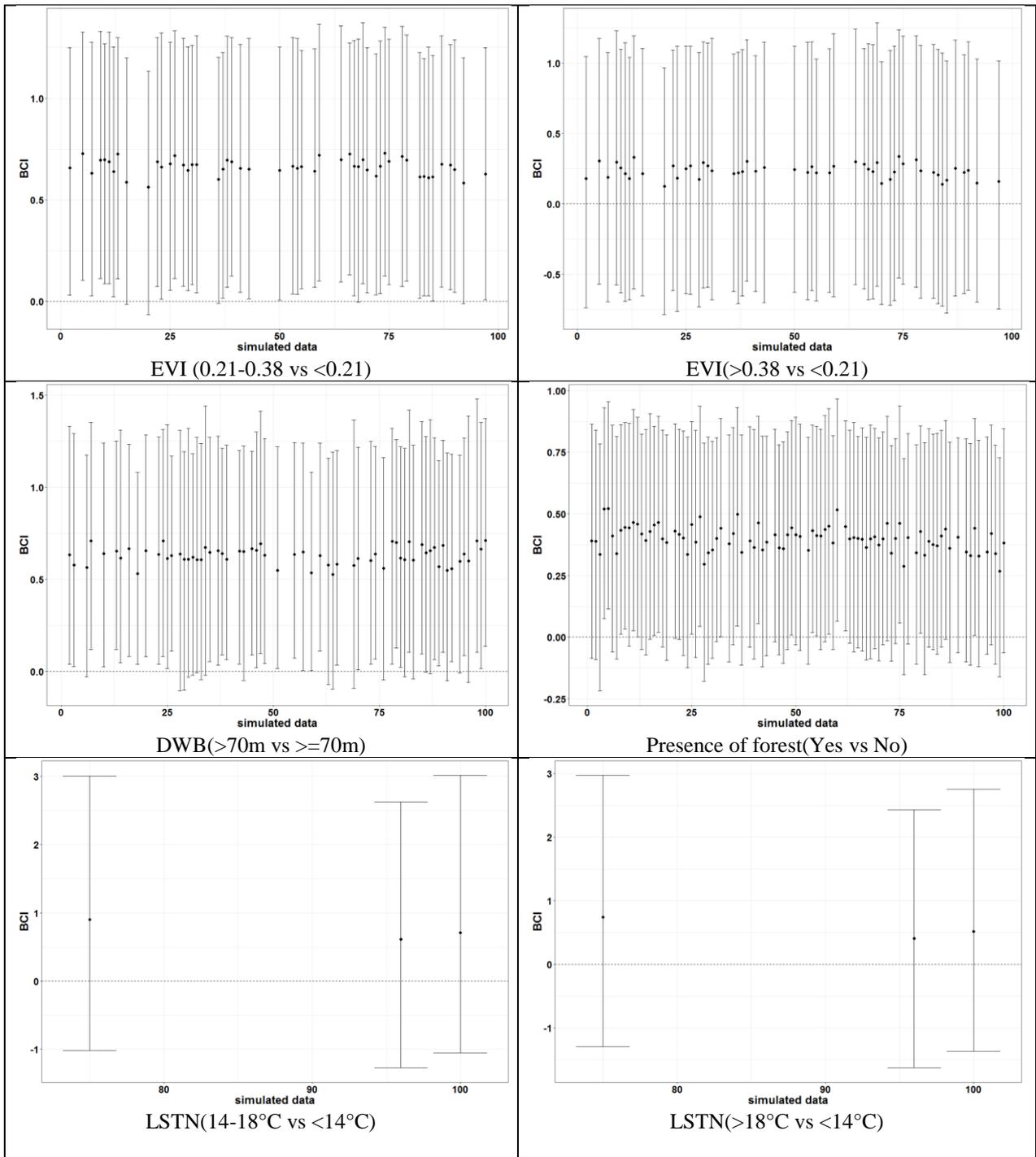
Figure 3.3: Effects (posterior median, 95% BCI) of categorical covariates estimated by the selected geostatistical model across to the simulated data ordered according to the logarithm predictive score values.
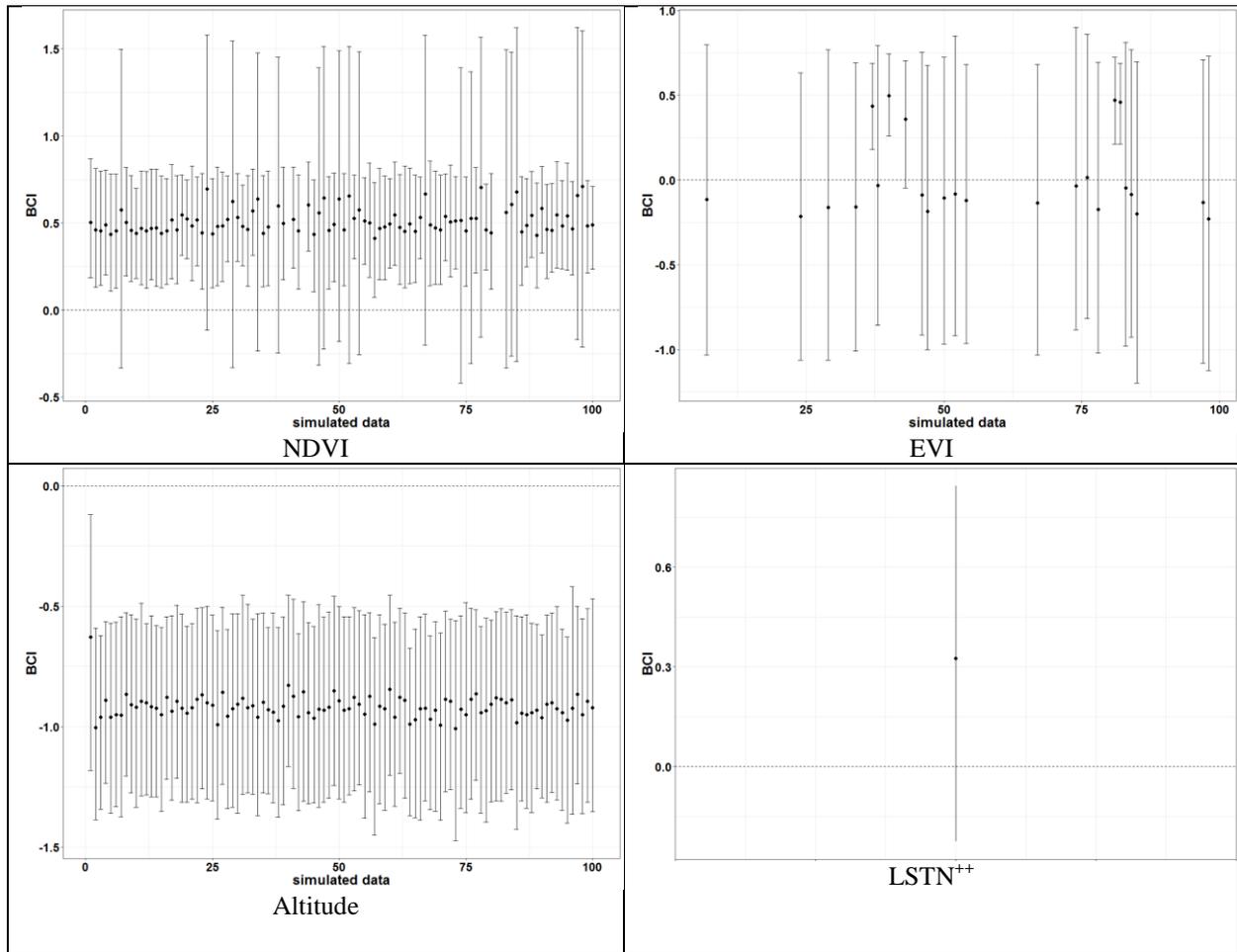
++ LSTN was selected in continuous form in only one simulated data

Figure 3.4: Effects (posterior median, 95% BCI) of continuous covariates estimated by the selected geostatistical model across to the simulated data ordered according to the logarithm predictive score values.

Table 3.1 presents posterior inclusion probabilities of the selected models based on the simulated data with the best (Model 1b) and worst predictive performance (Model 1c) and on the observed MIS data (Model 1a). The difference between Model 1b and Model 1a is that the former includes LSTN and excludes DWB. Model 1c includes RFE, LSTN which are not in Model 1a and excludes EVI. Regarding the selection of intervention indicators from the

individual-level model, the Model 2b with the best predictive performance among the simulated data gave similar results to the true model (Model 2a). The model with the worse performance among simulated data (Model 2c) was not able to capture the statistically important effect of the malaria intervention indicator (i.e. proportion of households with 1 ITN per 2 persons and the proportion of children slept under an ITN the previous night). The directions of the effects were estimated by the Bayesian geostatistical models (Table 3.2).

The global spatial patterns of disease risk in the East, North and Coastal parts of the country were well captured by the three cluster-level models. Maps drawn on the same scale clearly indicate similar geographic patterns predicted by the three models (Model 1a, Model 1b, and Model 1c), therefore the models with best and worst predictive performance were able to capture the disease risk distribution of the MIS dataset. However, the prediction uncertainties of Model 1b and Model 1c over the gridded surface were greater than the ones obtained from Model 1a (Figure 3.5-3.6).

The spatial variance estimate and uncertainty obtained from the simulated data with the worst and best predictive abilities were close to the ones produced by the observed MIS data. The residual spatial correlation estimated by the different cluster and individual models remained high, indicating the presence of unmeasured spatially structured factors related to the geographic distribution of the parasitaemia risk.
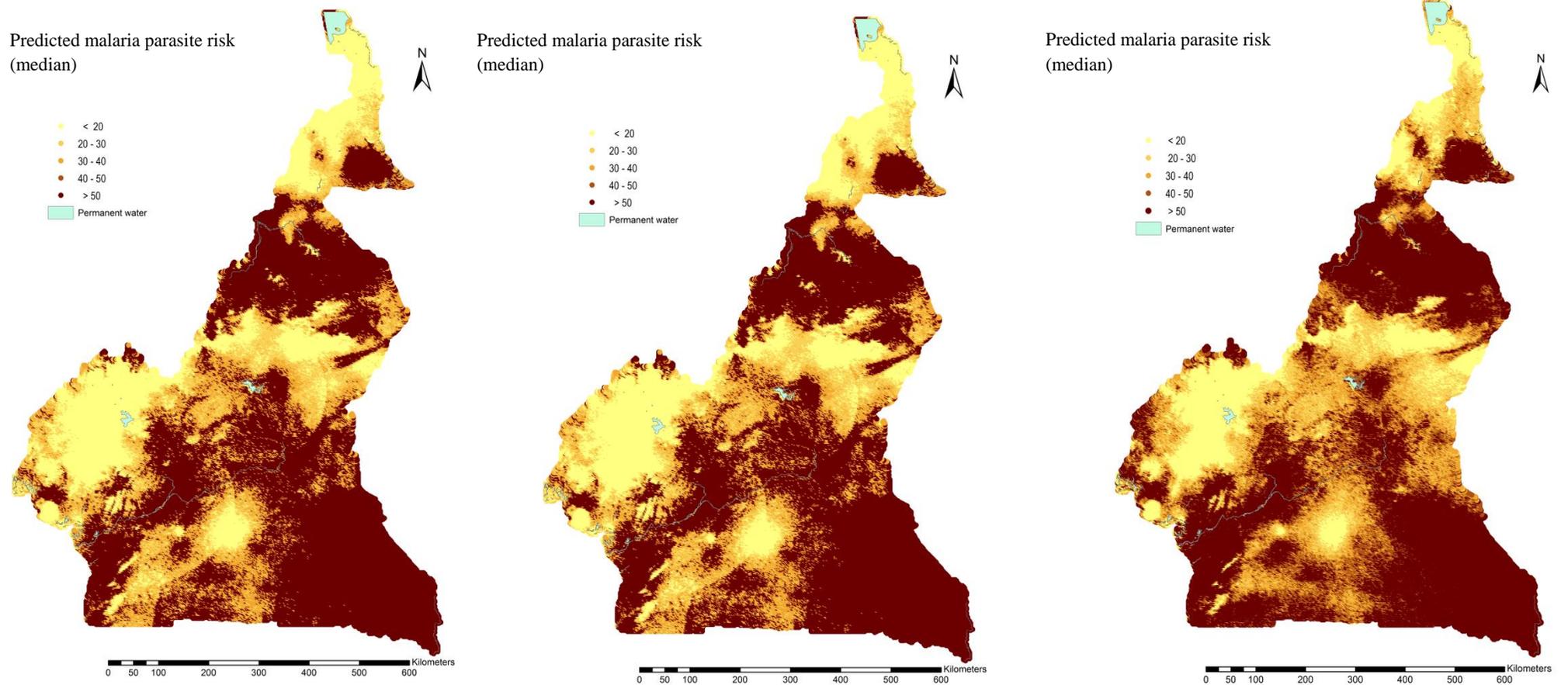
Figure 3.5: Malaria parasite risk estimates among children less than 5 years, obtained from i) Model 1a (left), ii) Model 1b (center) and iii) Model 1c (bottom right).
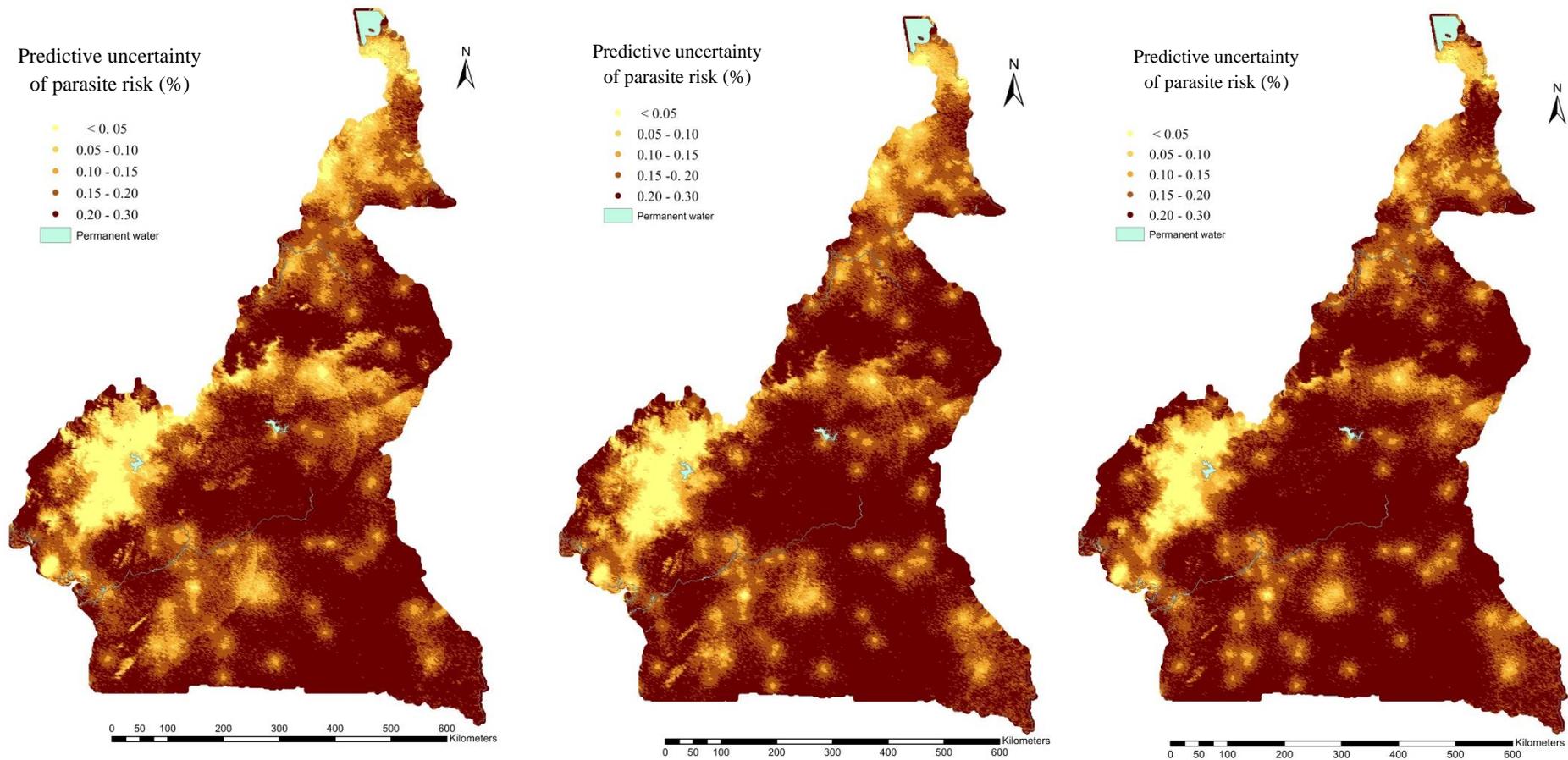
Figure 3.6 : Predictive uncertainty (standard deviation of predictive posterior distribution) of estimated parasite risk among children less than 5 years, obtained from i) Model 1a (left), ii) Model 1b (center) and iii) Model 1c (bottom right).

3.5 Discussion

This study is the first to assess the effects of jittering of DHS/MIS cluster locations on the estimates of the geographical distribution of malaria risk and of the intervention effects obtained by Bayesian geostatistical modelling (Adigun et al., 2015; Diboulo et al., 2016; Gosoniu et al., 2012; Giardina et al., 2014). A large number of jittered datasets were simulated from real data and geostatistical variable selection was applied to determine the impact of jittering on the model formulation.

Different subsets of climatic factors in the simulated data were identified as important predictors of malaria risk. However, in 18% of the datasets, the models included the same predictors with the true model obtained by the observed MIS data, while the model with the highest posterior inclusion probability (19%) could not capture the statistically importance of the DWB predictor. NDVI and altitude were selected in more than 95% of the simulated data and DWB was identified in 64%. Furthermore, the jittering of cluster locations had an influence on the selected functional form of the climate predictors (continuous/categorical). These results showed that spatial displacement can influence the risk factor analysis and the estimation of the effects of malaria interventions on the disease risk. Similar findings have been reported for distance-based covariates in the study presented by Warren et al(Warren et al., 2016b).

The direction of the relation between parasite risk, NDVI and altitude was remained the same in all simulated data. In particular, the continuous form of NDVI was statistically important in most of simulations and as expected, the altitude was always negatively related and statistically important to the malaria parasite risk. Those associations were confirmed with the estimates obtained from the true dataset and findings from others studies(Adigun et al., 2015; Samadoulougou et al., 2014). However, the shifting of cluster locations had an effect on the uncertainty estimates of the covariate effects and therefore on their statistical importance.

Warren et al have also concluded that displacement of clusters led to an increase in the estimated uncertainty of the regression coefficient (Warren et al., 2016a). In addition, Cressie et al proved that in the presence of spatial location error the prediction estimates and regression coefficients were influenced(Cressie and Kornak, 2003).

In most simulations, the BCIs of the altitude and NDVI (continuous forms), EVI and DBW (categorical forms) were overlapping. The majority of datasets were able to capture the statistical importance of those covariates. Altitude and vegetation index change little in space within the radius corresponding to the random displacement of cluster's coordinates, most likely due to small environmental gradient within each ecological zone of the country. Thus, locations inside the displacement buffer shared in most cases the same environmental conditions and therefore their parameter estimates were not affected by jittering (Kwan, 2012).

The simulated data with the highest predictive performance was the one with location configuration among the closest to the true data. The parameter estimates of this model were also similar to the true model.

According to the simulation, the proportion of households with 1 ITN per 2 persons or the proportion of children slept under an ITN the precious night before the survey was identified as important predictors of the individual-level malaria risk model. Variable selection applied on the intervention coverage indicators revealed that jittering influenced their posterior inclusion probabilities into the model and therefore the inference about the effects of malaria interventions. This result could be due to the confounding effects of climatic predictors.

All the wealth index categories were statistically important and negatively associated to the malaria parasite risk in the true model. The three individual level models showed that, posterior parameter estimates of socioeconomic factors were relatively stable, irrespective of the model. The socioeconomic factors are related to the individual risk rather than the malaria

prevalence at the location level, and thus estimates of the socio-economic effects are not much influenced by the displacement of the clusters. On the order hand, demographic factors also related to the individual were not affected by the jittering. The gender was not statistically important and a gradient of risk was noted in the age groups as already supported by other studies (Adigun et al., 2015; Diboulo et al., 2016; Gosoniu et al., 2012; Ssempiira et al., 2017).

The effect of the selected intervention indicator was statistically important and negatively associated to the parasitaemia risk regardless of the simulated dataset. Similar to the socioeconomic status and demographic factors, intervention effects are more likely to be higher at the individual and household than the community; therefore the changes of cluster locations did not influence the direction of the relationship between ITN coverage indicators and malaria parasite risk after adjusting for socioeconomic factors.

The BCIs of the spatial correlation parameters of the true, best and worst models were overlapping and their spatial variances were not dramatically changed. Spatial range parameters depend on the cluster locations and are sensitive to the distance between locations. The individual level models overestimated spatial correlation especially for the model having the worst predictive ability. The change of cluster locations could lead to a misspecification of the spatial dependence structure of the disease risk (Perez-Haydrich et al., 2013; Warren et al., 2016b).

The geographical patterns obtained from the simulated data with the highest and lowest predictive performance were similar to the ones obtained from the true data. The relationship between malaria risk and the climatic factors is rather stable within the same ecological zone and therefore is not strongly influenced by the jittering of the locations. This result was expected since several studies showed a local interdependency between climatic factors within small buffer zones. (Berg et al., 2013; Haerter et al., 2018; Hou et al., 2018).

These results are based on the assumptions of a stationary and isotropic spatial process of the malaria risk. Violation of these assumptions may influence the results of geostatistical variable selection and therefore the impact of jittering on model specification.

3.6 Conclusion

Moderate spatial modifications in the geographical positions of the clusters surveyed might not dramatically changes the estimation of the spatial patterns of disease risk in Cameroon, especially when the climate and environmental conditions are similar within the radius of the random displacement. Nevertheless, the alteration of cluster locations have an influence on the climate predictors selected to draw the disease risk maps at high geographical resolution, and could affect the interpretation of relationship between malaria infection and environmental and climate factors that support the disease transmission. A wise compromise should be done to ensure confidentiality and relative robustness of disease risk estimates. Simulations can be realized to identify the suitable subset of jittered clusters that preserve environmental predictors and their estimates.

3.7 Appendix

Additional file 1: Spatial configuration of four initial points (squares in red) and their random jittering coordinates (triangles in black), where the farthest is in green.

(a)

(b)

(c)

(d)
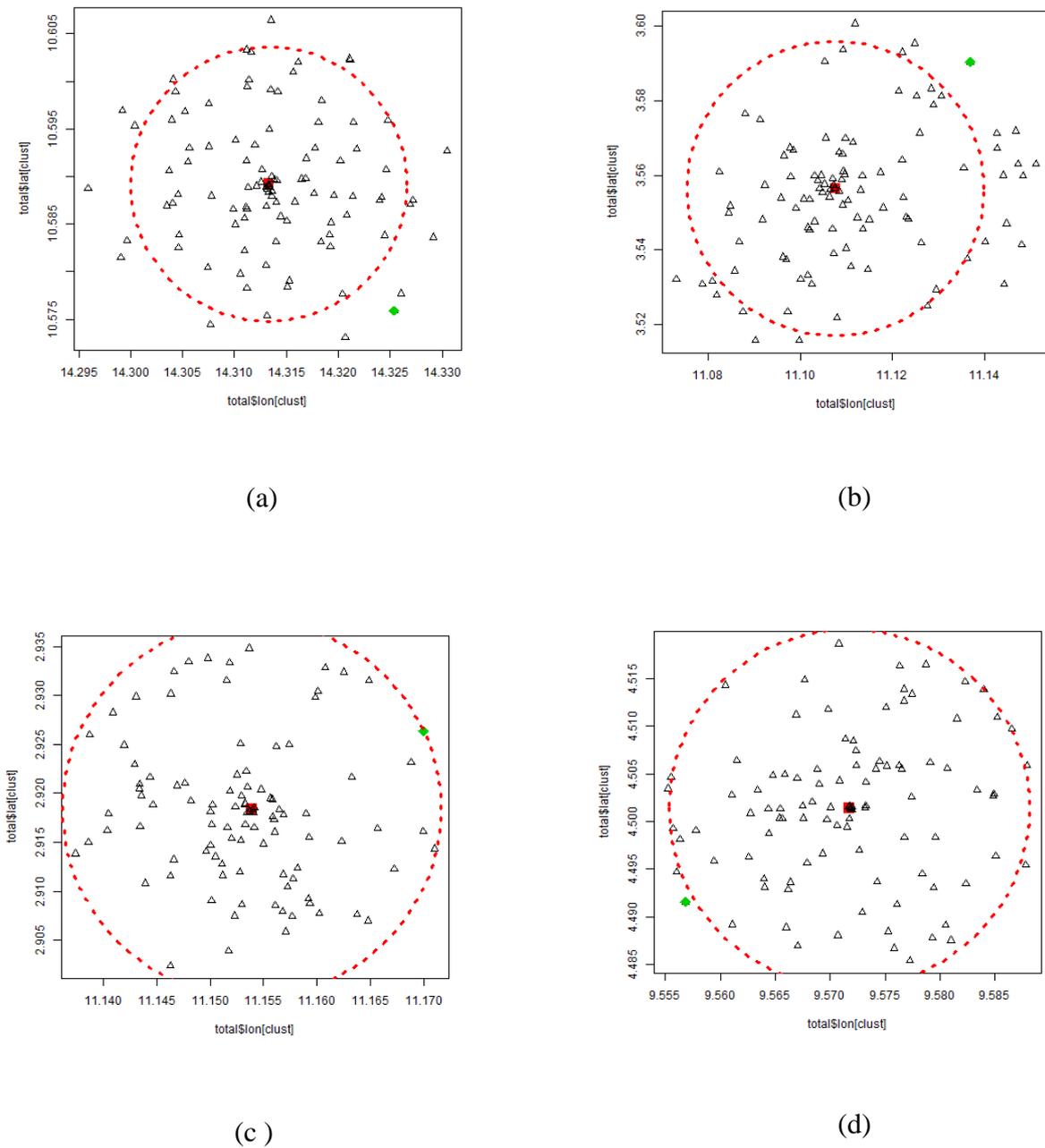
Figure 3.7 : Cluster locations of code identification 12 (a), 49 (b), 36(c) and 126 (d) including their jittered locations
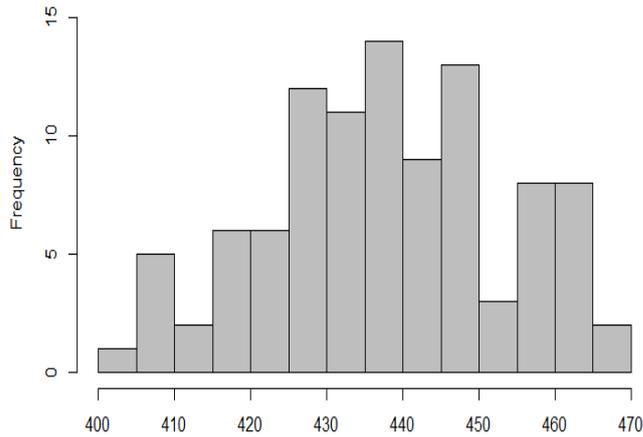
Figure 3.8: Distribution of the sum of distances between original and jittered positions in each simulation

Table 3.4: Spatial and temporal resolution of climate and environmental factors

Data	Source	Spatial resolution (km²)	Temporal frequency
Land surface temperature (LST) for day and night[1]	Moderate resolution Imaging Spectroradiometer (MODIS) Terra	1 × 1	8 days
Normalized difference vegetation index (NDVI) [1]	MODIS	0.25 × 0.25	16 days
Enhanced vegetation index (EVI)[1]	MODIS	1 x 1	16 days
Rainfall estimates (RFE) [2]	U.S. Geological Survey-Earth Resources Observation Systems (USGSS)	8 × 8	10 days
Land cover [1] : Cropland, Forest, Savannah	MODIS	1 x 1	year
Digital elevation (Altitude) [3]	Shuttle Radar Topographic Mission(SRTM)	1 × 1	NA
Permanent water bodies (rivers, lakes, wetlands) [1]	MODIS	1 × 1	NA

1. http://reverb.echo.nasa.gov/reverb

2. http://earlywarning.usgs.gov/fews/

3. http://glcfapp.glcf.umd.edu/data/srtm/

Table 3.5: Frequency distribution of the best models identified by the geostatistical variable among the 100 simulated data. Model M1a includes the combination of predictors included in the best model obtained from the observed MIS data.

						Climate and environmental factors *					
Model	RFE	NDVI	LSTD	EVI	DWB	Altitude	Forest	Savannah	Cropland	LSTN	Frequency
1	**1**	**2**	**1**	**3**	**1**	**2**	**3**	**1**	**1**	**1**	**19%**
2	**1**	**2**	**1**	**3**	**3**	**2**	**3**	**1**	**1**	**1**	**18%**
3	3	2	1	3	3	2	3	1	1	1	9%
4	1	2	1	1	3	2	3	1	1	1	8%
5	1	2	1	2	3	2	3	1	1	1	6%
6	3	2	1	2	3	2	3	1	1	1	5%
7	1	2	1	2	1	2	3	1	1	1	5%
8	1	2	1	1	1	2	3	1	1	1	5%
9	3	2	1	1	3	2	3	1	1	1	4%
10	1	2	1	1	3	2	1	1	1	1	3%
11	3	1	1	2	3	2	3	1	1	1	2%
12	1	2	1	3	3	2	1	1	1	1	2%
13	3	2	1	3	1	2	3	1	1	1	1%
14	3	2	1	2	3	2	1	1	1	1	1%
15	3	2	1	1	3	2	3	1	1	3	1%
16	3	2	1	1	3	2	1	1	1	1	1%
17	3	2	1	1	1	2	3	1	1	1	1%
18	1	3	1	2	3	2	3	1	1	1	1%
19	1	2	1	3	1	2	3	1	1	3	1%
20	1	2	1	3	1	2	3	1	1	2	1%
21	1	2	1	2	3	2	1	1	1	1	1%
22	1	2	1	2	1	2	1	1	1	1	1%
23	1	2	1	1	3	2	3	1	1	3	1%
24	1	2	1	1	1	2	1	1	1	1	1%
25	1	1	1	2	3	2	3	1	1	1	1%
26	1	1	1	2	1	2	3	1	1	1	1%
MIS	**1**	**2**	**1**	**3**	**3**	**2**	**3**	**1**	**1**	**1**	

*Legend: 1=Excluded; 2=Continuous; 3=Categorical form

3.7.1 Geostatistical models formulation

Different assumptions were made about the probability density functions used at cluster and at individual levels. Models describe below were applied for the true and simulated data, where the index $k \epsilon \{1 \dots 101\}$ represents the original data for the first value and the next others the simulated data.

Model 1: Estimating malaria risk at cluster level

Considered Y_i^k the variable of number of parasitaemia positive children among the n_{ik} screened at the cluster i of the simulation k, s_i^k, $(i = 1, \dots, N)$. In the Bayesian inference context, we had assumed that Y_{ik} arise from a binomial distribution, i.e. $Y_i^k \sim Bin(n_i^k, p_i^k)$, where the malaria parasite risk at cluster i of the simulation k, p_{ik}, on the logit scale is a function of climatic factors $X_i^k = \left(1, X_{i1}^k, X_{i2}^k, \dots X_{ip}^k\right)$ and spatial random effects U_{ik}, such as $\text{logit}(p_{ik}) = \sum_{j=0}^{p} \beta_j^k X_{ij}^k + U_i^k$.

We assume that U_i^k arise from a stationary isotropic spatial process i.e. $U^k \sim MVN(\mathbf{0}, \Omega^k)$, where $U^k = (U_1^k, U_2^k, \dots, U_N^k)^T$ and each Ω^k is the variance-covariance matrix NxN, so that $\Omega_{ij}^k = \sigma_k^2 \exp(-\rho_k d_{ij}^k)$. d_{ij}^k is the euclidian distance between cluster s_i^k and s_j^k, σ_k^2 is the spatial variance also called the partial sill and ρ_k the smoothing parameter to control correlation. We consider non-informative prior for β_j^k defined such as $\beta_j^k \sim N(0,100)$.

Model 2: Estimating malaria risk at individual level

Let Y_{ij}^k be the status of malaria parasitaemia of child j who lives in cluster i of the simulation k, s_i^k, $(i = 1, \dots, N)$. We assume a Bernoulli distribution, i.e. $Y_{ij}^k \sim Be(p_{ij}^k)$, and model individual level function covariates on the logit scale of child specific risk, p_{ij}^k i.e. $\text{logit}(p_{ij}^k) = \sum_{l=0}^{p} \beta_j^k X_{ijl}^k + U_i^k$ where, $X_{ij}^k = \left(1, X_{ij1}^k, X_{ij2}^k, \dots X_{ijp}^k\right)$, the list of covariates including climatic factors and intervention coverage indicators, U_i^k is the dependence spatial

effect of cluster i of the simulation k, s_i^k. Similar prior distributions in Model 1 were also used for spatial and regression coefficients parameters.

Prior distributions of the spatial process parameters

We assigned to σ_k^2 an Inverse Gamma prior distribution, $\sigma_k^2 \sim IG(2.01, 1.01)$ and adopt a Uniform prior distribution for ρ_k, i.e. $\rho_k \sim U(\frac{-\log(0.05)}{d^k{}_{max}}, \frac{-\log(0.05)}{d^k{}_{min}})$. The later prior assumes that the spatial correlation bellow 0.05 is negligible, $d^k{}_{max}$ and $d^k{}_{min}$ were corresponding to the maximum and minimum (non-zero) Euclidean distance between the survey locations in the same simulations or in the true data.

3.7.2 Bayesian variable selection

For each simulated and true data, we have applied stochastic search variable selection assuming a normal mixture prior distribution for the regression coefficient of each climate and intervention factors. A categorical indicator I_p^k was used to exclude the associated factor p from the model of data set k, when $I_p^k = 0$, to select its linear form when $I_p^k = 1$ or to select the categorical form when $I_p^k = 2$. We assume a multinomial distribution prior for I_p^k with a probability function $\prod_{l=0}^2 \pi_l^{k \, \delta_l^k(I_p^k)}$, where π_l^k represent the inclusion probability of each form ($l = 0, 1, 2$) and $\delta_l^k(.)$ is the Dirac function: $\delta_l^k(I_i^k) = \begin{cases} 1 \ if \ I_i^k = l \\ 0 \ if \ I_i^k \neq l \end{cases}$. We adapted a spike and slab prior for the regression coefficients $\beta_{p,l}^k$, which is a mixture of normal prior distributions i.e. $\beta_{p,l}^k \sim \delta_l^k(I_p^k)N\left(0, \tau_{p,l}^{k \, 2}\right) + (1 - \delta_l^k(I_p^k))N\left(0, c \times \tau_{p,l}^{k \, 2}\right)$ with a mixing proportion equal to the inclusion probability of the corresponding coefficient. The spike component shrinks the regression coefficient to zero when the variable is excluded and the slab assumes a non-informative, normal prior distribution. A climate or intervention factor was identified as important and thus selected in the best model when the inclusion probability

of a specific functional form (linear or categorical) is larger than 50% (Ishwaran & Rao, 2005; O'Hara & Sillanpää, 2009).

For each simulated and true data, a non-informative prior Dirichlet distribution with hyper-parameter $\alpha = (1,1,1)^T$ is used, then $\pi^k = (\pi_1^k, \pi_2^k, \pi_3^k)^T \sim Dirichlet(3, \alpha)$, the constant c is fixed at 10^{-5} and inverse Gamma prior distribution was adopted for $\tau_{p,l}^{k}{}^2 \sim IG(0.01, 0.01)$.

3.7.3 Bayesian kriging

Let $S_0 = \{S_{01}, S_{02} \dots, S_{0m}\}$ be the centroids of the gridded surface. Children parasitaemia risks were predicted over the gridded surface from the predictive posterior distribution (Sudipto Banerjee, 2015):

$$P\left(\tilde{Y}_0^k | \tilde{Y}^k\right) \alpha \int P\left(\tilde{Y}_0^k | \tilde{Y}^k, \tilde{\beta}^k, \tilde{U}_0^k, \tilde{U}^k\right) \times P\left(\tilde{U}_0^k | \tilde{U}^k, \sigma_k^2, \rho_k\right) \times P\left(\tilde{U}^k, \sigma_k^2, \rho_k | \tilde{Y}^k\right)$$

$$\times P(\tilde{\beta}^k | \tilde{U}^k, \tilde{Y}^k) d\tilde{\beta}^k d\tilde{U}^k d\sigma_k^2 d\rho_k$$

For the dataset k, $\tilde{Y}_0^k = Y^k(S_0)$, where $Y(s_o) \sim Bin(N, p_{s_o}^k)$ is the predicted number of children tested positive in the new location S_o and the associated covariates X_0. Conditional to the spatial process and model parameters, with the link relation between the risk and spatial covariate defined as $\text{logit}(p_{s_o}^k) = \sum_{j=0}^{p} \beta_j^k X_{s_o j} + \tilde{U}_0^k$.

Chapter 4 Spatio-temporal patterns of malaria incidence in Cameroon.

Salomon G Massoda Tonye[1,2,4], Maturin Nkamgnia[1], Magloire Kuetche[1], Celestin Kouambeng[1], Hugues abomabo[1], Jean Christian Kouontchou[5], Achu Dorothy[1], Etienne Fondjo[1], Prosper Ndong A Bessong[1], Mounia Hocine[3], Penelope Vounatsou[2,4]

1. National Malaria Control Program.

2. Swiss Tropical and Public Health institute

3. Conservatoire Arts et Métiers de Paris

4. University of Basel, Basel, Switzerland

5. World Health Organization

Corresponding author:

E-mail: Penelope.vounatsou@swisstph.ch

Abstract

Background

Cameroon malaria incidence data are collected at health facilities level and centralised in a national database. The completeness of data has substantially improved since 2012. Around that year Cameroon has scaled up malaria interventions. Our study aims to evaluate the effects of the intensified malaria control on the spatio-temporal distribution of the disease incidence adjusting for climatic and environmental factors.

Method

Monthly, confirmed malaria cases aggregated by district were extracted from the Health Management Information System (HMIS) in Cameroon during 2012-2016. Climatic and environmental predictors were retrieved from remote sensing sources. They were averaged over lagged periods of 1 to 3 months between climatic suitability and malaria incidence. A Bayesian, spatio-temporal, conditional autoregressive model was applied to assess spatio-temporal patterns of the disease. Bayesian variable selection was used to identify the most appropriate lag time for the climatic factors. The analysis was adjusted for the coverage of malaria interventions collected by household surveys.

Results

Variable selection of climatic factors identified Normalized Difference Vegetation Index (NDVI), Rainfall estimates (RFE) and Land surface temperature day (LSTD) as the most important predictors. Confirmed malaria cases were negatively associated to the proportion of households owning at least one ITN for every two household members. Statistically important and positive relations were captured between malaria cases and the two-months lagged values of RFE and NDVI, respectively. A negative association was also observed between malaria cases reported and previous month values of LSTD. Among the children under five, the

spatial distribution of malaria cases showed similar areas to those identified by the cross-sectional household surveys such as Malaria Indicator Survey (MIS) or Demographic Health Survey (DHS). In addition, the geographical spatial patterns were identified in periods not covered by MIS or DHS data.

Conclusion

Bayesian conditional autoregressive models applied to health data have identified meaningful spatial patterns not captured by the DHS/MIS. Therefore, although MIS/DHS data remain a standard to monitor progress by assessing of health intervention effects and providing a high-resolution disease risk distribution, health data should be more used to meet gaps in information between national and representative household surveys and to ensure continuous monitoring of disease trends and identification of abnormal patterns.

Keywords: Bayesian inference, demographic and health survey (DHS), malaria indicator survey (MIS), conditional autoregressive, confirmed malaria cases, Health facilities.

4.1 Introduction

In 2019, more than 80 countries were reported in endemic stage and more than three billion of people were at risk. The Sub-Saharan African countries were very affected with 93% of cases and 94% of deaths (WHO, 2019). In 2019, More than 4 million of consultations were related to malaria and representing 40% of the patients received at health facilities level. The percentage of suspected cases tested by RDT or microscopy was estimated around 640%. Children under-five years were representing 33% of confirmed malaria cases while they were 17% of the entire population (PNLP, Minsante, 2020).

Since 2011, Cameroon y has scaled-up malaria control interventions to the entire population at risk with the support of various donors such as the Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM), Bill and Melinda Gates foundation, United Nations specialized structures, other non-governmental organizations. In particular, rapid diagnostic tests (RDTs) and artemisinin-based combination therapies (ACTs) were introduced in the health system. Drugs and diagnostics tools were subsidized for patients aged above the age of five, while the treatment of uncomplicated and severe malaria cases in children under five, received at hospital, were free of charge since January 2011 and July 2014, respectively(Bowen, 2013).   In addition, two national mass distribution campaigns of approximately 8 million and 12 million of long-lasting insecticides bed-nets (LLINs) were conducted in January 2012 and June 2016, respectively. From August to October 2016, the first Seasonal Malaria Chemoprevention (SMC) intervention was achieved in favour of children living in Sahelian and Tropical Dry facets (Bowen, 2013; GFATM, 2018; PNLP, 2012, 2011).

During the 2012-2016 period, the decline of the malaria cases was observed in Cameroon. Health facilities data showed a decrease of malaria mortality from 17.6% to 12.4% but also a reduction of malaria morbidity from 31% to 27%. However, children under five years were

still more vulnerable than other age groups, they accounted for about 35% of malaria cases when they represented only 17% of population (PNLP, 2012; BUCREP, 2011; PNLP, 2016; Minsante, INS and UNFPA, 2016). However, despite the reduction of malaria burden, the country was still in the control stage and aims to accelerate health interventions towards elimination. The transition from control to elimination and ultimately eradication requires an efficient health surveillance system that is useful for planning and detection of abnormal changes in the malaria trends (World Health Organization, 2016, 2017). In 2011, the country setup a computerized malaria data entry system covering all health districts, called 'BD CamMalaria 1.0'. Currently, more than four thousand health facilities (HF) regularly send data reports that are stored into the malaria database located at the district and regional levels. The modernization of malaria data management has improved the availability and quality of data, even though health system performance issues such as population attendances rate and completeness of medical reports continue to affect data (Cibulskis et al., 2011).

Despite the progress made in malaria data quality, their use for decision-making remains low. Therefore, to change this situation and strengthen the malaria surveillance system, it appears interesting to perform suitable statistical models to analyse the effect of interventions on the spatiotemporal dynamic of malaria incidence. Bayesian conditional autoregressive models seems appropriate (Gelfand and Vounatsou, 2003; Moraga and Lawson, 2012).

This study sought to assess the spatiotemporal relationship between confirmed malaria cases, and malaria intervention adjusted by the climatic and environmental factors, (Arab et al., 2014; Caminade et al., 2014; Tanser et al., 2003). Bayesian conditional autoregressive models assuming that malaria cases arise from a negative binomial distribution and including Fourier terms, climatic factors and intervention coverage indicator will be performed. Such models were already used to fit malaria incidence in many African countries such as Ethiopia,

Namibia, Mozambique and Uganda (Zacarias and Andersson, 2011; Alegana et al., 2013; Ssempiira et al., 2018; Laguna et al., 2017).

4.2 Methods

4.2.1 Country setting

Cameroon is localized in Central Africa, the country had 10 regions, 58 divisions, 189 health districts (HD) and the population was around 23.5 million with annual growth rate of 2.5% in 2016 (Ministere de la sante publique, 2011; Ministry of Economy, Planning and regional development, 2009). There geographical distribution of health facilities and medical personal was unequal over the country (Tandi et al., 2015). There are three types of health facilities (public, private and confessional) but the majority are public. Malaria is endemic over the country with variations due to the different epidemiological zones: the dry Sahelian in the Far North and the Tropical in the North, the highlands of Adamawa and West, the central plateau that covering the Centre, East and part of South regions, the Atlantic coastal including the Littoral, South-west and coastal part of South regions (Ayala et al., 2009; Gemperli et al., 2006). Four malaria parasite species are present in the country : *plasmodium malariae, plasmodium vivax, plasmodium ovale* and *plasmodium falciparum* which is predominant (PNLP, 2007). An annual average of 1.9 million of confirmed malaria cases were declared within the period of our study (BUCREP, 2011; PNLP, 2013, 2012, 2011).

4.2.2. Data

4.2.2.1 Malaria incidence data

Confirmed malaria cases collected in Health facilities (HF) between January 2012 and December 2016 were extracted from the malaria database. From 2012 to 2016, the number of HF involved in the malaria data transmission was 3 121, 3 261, 3 452, 3 587 and 4 057 respectively. The completeness of health data reports varied from 81 to 85% (PNLP, 2013, 2012, 2011). The total number of confirmed malaria cases were 9 825 827. Malaria

indicators were divided according to age (less than five and above five years) (PNLP, Minsante, 2013). Microscopy and rapid diagnostics tests were used to confirm malaria cases at health facilities(WHO, 2015b).
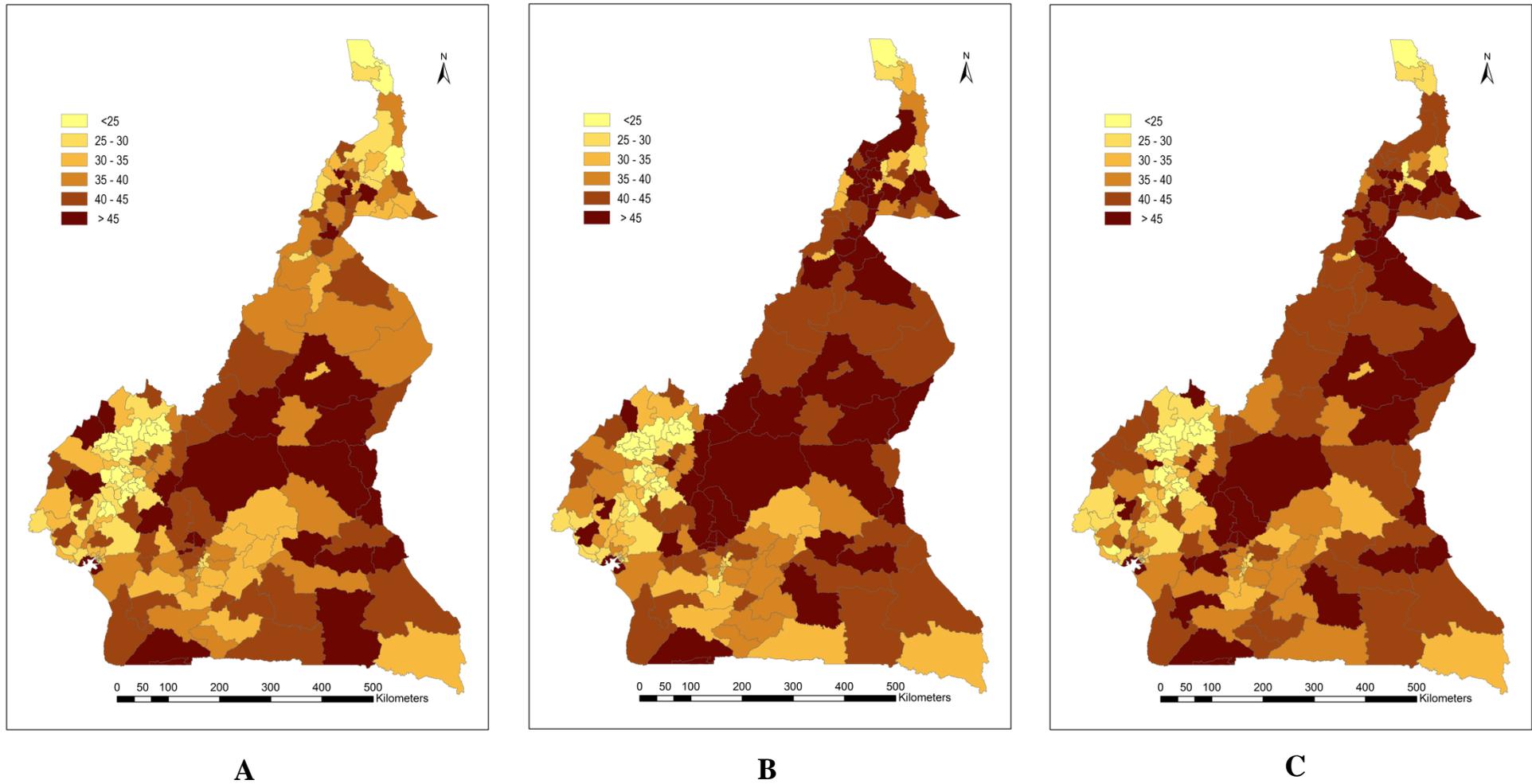
Figure 4.1 : Observed ratio of children under five malaria cases on above five malaria cases per 100 and by health districts in 2016 (A), 2015 (B) and 2014 (C).
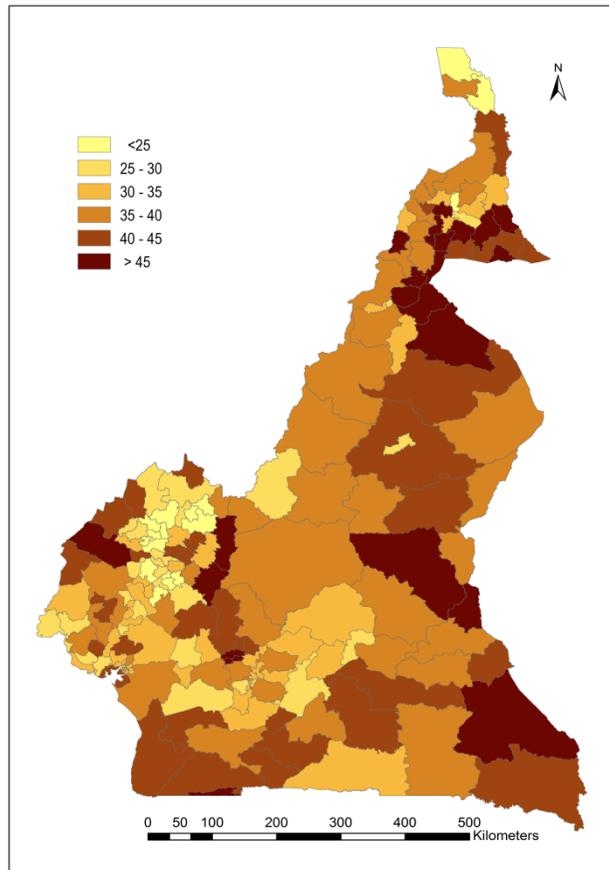
D
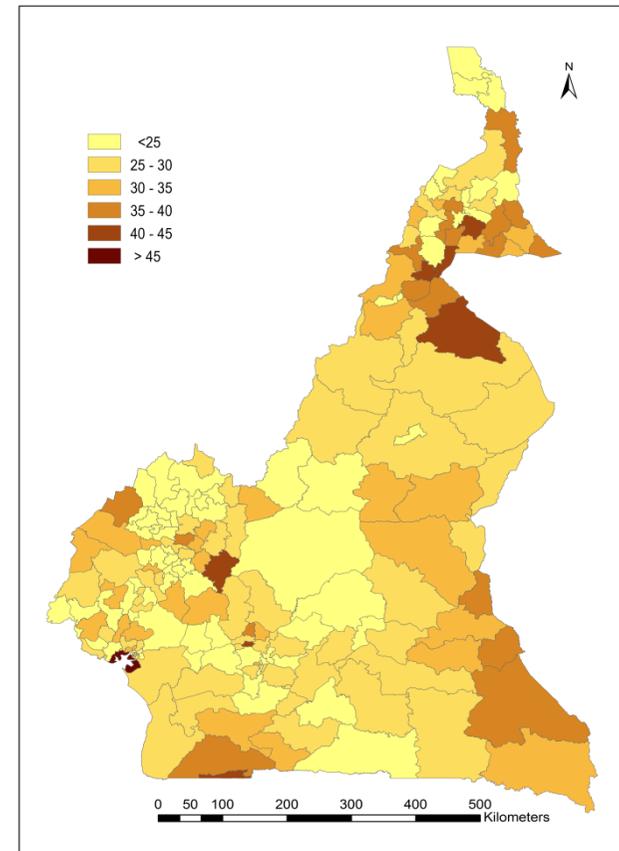
E

Figure 4.2 : Observed ratio of children under five malaria cases on above five malaria cases per 100 and by health districts in 2013 (D) and 2012 (E)

4.2.2.2 Climatic and environmental data

Remote sensing data obtained from satellite images were used as proxies for climatic and environmental conditions prevailing on the ground. The MODIS (or Moderate Resolution Imaging Spectroradiometer) web site was used for extraction of the Land Surface Temperature (LST) day and night (LSTD, LSTN), Normalized difference vegetation index (NDVI) and the Enhanced vegetation index (EVI) (U.S. Geological Survey, 2017). The Rainfall estimates were obtained through the FEWS (or Famine Early Warning Systems Network) web site (U.S. Geological Survey, 1985). Their spatial and temporal resolutions were given in Table 4.1. Many studies showed a relation between climate covariates of the previous months and malaria incidence of the current month, therefore lagged factors were created for each covariates up to the two previous months (Ikeda et al., 2017; Guo et al., 2015; Kipruto et al., 2017; Zhao et al., 2014). The covariates values were averaged over a grid surface covering the corresponding health district.

Table 4.1: Sources, spatial and temporal resolution of predictors

Data	Spatial resolution (km²)	Period	Temporal
Land surface temperature (LST) for day and night	1×1	2011-2016	8 days
Enhanced vegetation index (EVI)	0.25×0.25	2011-2016	16 days
Normalized difference vegetation index (NDVI)	0.25×0.25	2011-2016	8 days
Rainfall Estimates (RFE)	8×8	2011-2016	10 days
People per grid square	0.1×0.1	2010	5 years

4.2.2.3 Intervention indicators

Four national and representative surveys were carried out by the National Institute of Statistics during the period from 2011 to 2016: the demographic and health survey and malaria indicators survey held in 2011, the post-campaign survey on the use of LLINs in 2013, the multiple indicators and clusters survey in 2014 and the post-campaign survey on

LLINs use in 2016. Data collected include indicators on possession and use of LLINs among the population. Their aggregated values were representative up to the regional level. The proportion of households owning at least one ITN for every two household members was extracted. (INS, Minsante et UNICEF, 2015; National Institute of Statistics, 2017, 2013, 2011)

4.2.2.4 Demographic data

The entire population was at risk of malaria infection. The spatial distribution of demographic data at the district level retrieved through satellite sources were adjusted by census data and population growth rate (BUCREP, 2011; Minsante, INS and UNFPA, 2016; Tatem et al., 2013; University of Southampton, 2013). The population attendance rate to health system was applied to estimate the expected population at risk captured by the health system.

4.3 Bayesian spatiotemporal modelling

Bayesian conditional autoregressive negative binomial models including the most important climate covariates, seasonality terms, temporal correlation, spatial structured random effect (clustering) and unstructured random effect (heterogeneity) were fitted on the confirmed malaria cases outcome at health district level. Models were applied on the total number of malaria cases and also in age-specific group (i.e. children aged less than five years old and the others). The corresponding population size in the district was considered as an offset term.

The most important climate factors were identified using variable selection applied on the confirmed malaria incidence data regardless of age-groups. The climate factors and their associated lagged variables were involved in the variable selection. Due to the existing correlation between current and lagged values of a given climate factor, a categorical indicator was associated to each climatic factor with values 0, 1, 2 and 3 indicating the exclusion from the model or inclusion of current month values or inclusion of the previous

month current values, or inclusion of last two previous month values, respectively (Ishwaran and Rao, 2005). Intervention coverage was also considered in the model by including the proportion of households owning at least one ITN for every two people.

Based on the selected climate predictors, models were fitted on national population and on age-specific groups. Let consider Y_{it} the confirmed malaria cases at district $i \in \{1 \ldots 189\}$ at time $t \in \{1 \ldots 60\}$. The variable time combines year and month data. It was assumed that confirmed malaria case arises from negative binomial distribution $Y_{it} \sim NegBin(p_{it}, r)$, where $p_{it} = \dfrac{r}{r+\mu_{it}}$ and μ_{it}, r are respectively the mean and dispersion parameter of the negative binomial distribution. The relationship between the expected mean of confirmed malaria cases, population at risk and the climatic and intervention covariates was defined *as* $log(\mu_{it}) = log(p_{it}) + \beta_0 + X_{it}^T\beta + g(i,t)$, where p_{it} was the expected health population, β was regression coefficients vector and g the specific function that includes into separate parts, the spatial and temporal components (Briët et al., 2013; Lindén Andreas and Mäntyniemi Samu, 2011).

The function g was defined *as* $g(i,t) = f_k(t) + \beta_1 t + \varepsilon_t + \varphi_i + \gamma_i + h(r,t)$, where f, ε, φ captured the seasonal, temporal and spatial variations, respectively. While γ was the unstructured random effect and h the space-time effect with $r \in \{1 \ldots 10\}$ was the epidemiological region strata index in which a district belongs (Bernardinelli et al., 1995).

Seasonal structure of malaria cases was modelled using the Fourier functions, $f_k(t) = Acos(w(t)-\vartheta) = aCos\left(\dfrac{2\pi}{T}t\right) + bSin(\dfrac{2\pi}{T}t)$, where, T is the period and it was set to T=12, A is the amplitude estimating the peak i.e. $A = \sqrt{a^2 + b^2}$, $w(t) = \left(\dfrac{2\pi}{T}t\right)$ and $\vartheta = \arctan(b/a)$. A random effect ε_t was assumed a correlation such as $\varepsilon_t \sim N\left(\rho\varepsilon_{t-1}, \dfrac{\tau^2}{1-\rho^2}\right)$ and $\rho \in [0,1[$ where $\varepsilon_{t-1} = \varepsilon_1 \sim N\left(0, \dfrac{\tau^2}{1-\rho^2}\right)$.

Spatial correlation of malaria cases at health district was introduced using a conditional autoregressive Gaussian distribution (CAR). Each φ_i conditional to the neighbour $\varphi_{j \in V_i}$ follow a normal distribution, $\varphi_{i|j \in V_i} \sim N\left(\sum_{j \in V_i} \rho_{ij} \varphi_j, \frac{\sigma_i^2}{n_i}\right)$, with V_i the set of district neighbouring to i, and ρ_{ij} the contribution of district j on the spatial variation and n_i number of district neighbours. The regional space-time random effect was defined as $h(r,t) = \delta_{ry}$, $r = 1..10, y = round((t-1)/12)$, assuming that $\delta_{ry}, \sim N\left(0, \tau_{ry}^2\right)$ and the district unstructured random effect was designed as $\gamma_i \sim N(0, \gamma^2)$.

During the identification process of the best models, the terms used to capture the seasonal, temporal and spatial components of the model were removed sequentially and their contribution on the goodness of fit was evaluated using the deviance information criterion (DIC). The best model (i.e. lowest DIC) was used to generate the maps of the malaria incidence at beginning of each semester from 2012 to 2016. A selected covariate was considered as statistically important when the Bayesian Credible Interval (BCI) of its related parameter estimate did not include zero.

Variable selection was carried out by OpenBUGS version 3.2.3 (Imperial College and Medical Research Council, London, UK) and data management process realized in R(Lunn et al., 2000; R Core Team, 2016). Bayesian inference was performed by the Integrated Nested Laplace Approximation (Lindgren and Rue, 2015; Rue et al., 2009; Schrödle and Held, 2011). The maps generated by ESRI's ArcGIS version 10.2.1 for Desktop and R were based on the administrative health district structure of 2016 (ESRI, 2013; R Core Team, 2016).

4.4 Results

The overall malaria incidence rate was estimated at 67, 83, 98, 103 and 100 per 1000 respectively from 2012 to 2016. Despite of the annual increase of incidence ratio between children and others noted from 2012 and 2015, the health facilities in West highlands had a

stable annual ratio of incidence between children under five years old over patients above five (Figure 4.1, Figure 4.2). The evolution of confirmed malaria (all ages) over time has indicated the month of October as the peak of transmission, whilst the lower number of cases was recorded between December and February (Figure 4.3).
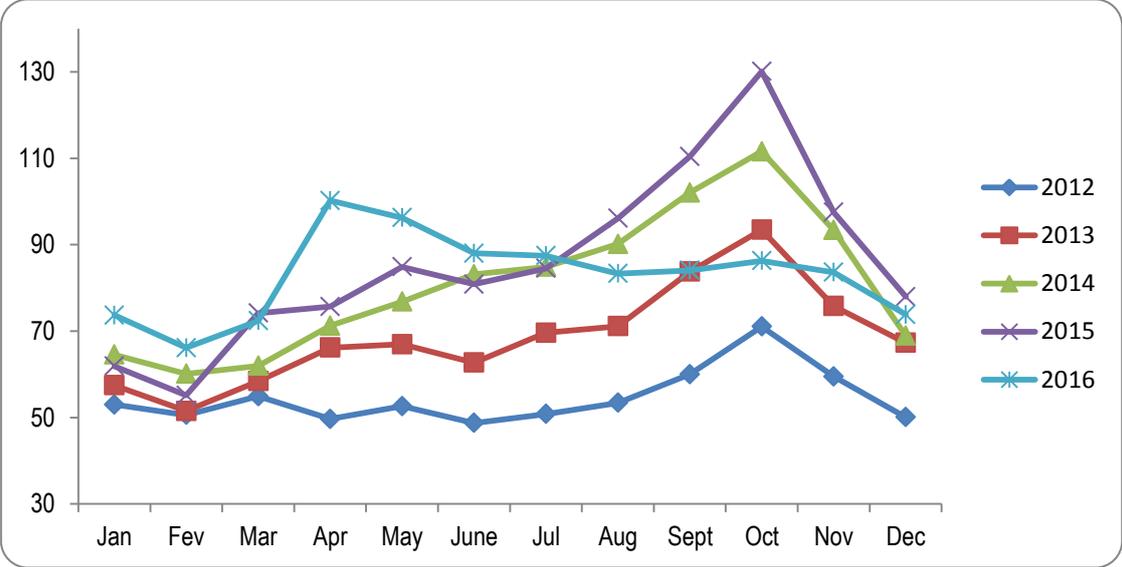
Figure 4.3 : Confirmed malaria cases (all ages) per 10 000 inhabitants and per months from 2012 to 2016.

A seasonal cycle appears within the period of September to November, where the highest number of malaria cases are notified in the country. In 2016, this periodical cycle was broken by the combined actions of free distribution of more than 12 million of LLINs to households and the implementation of SMC amongst children in the Sahelian and Tropical Dry facets before and within malaria high transmission season(WHO, 2017a).
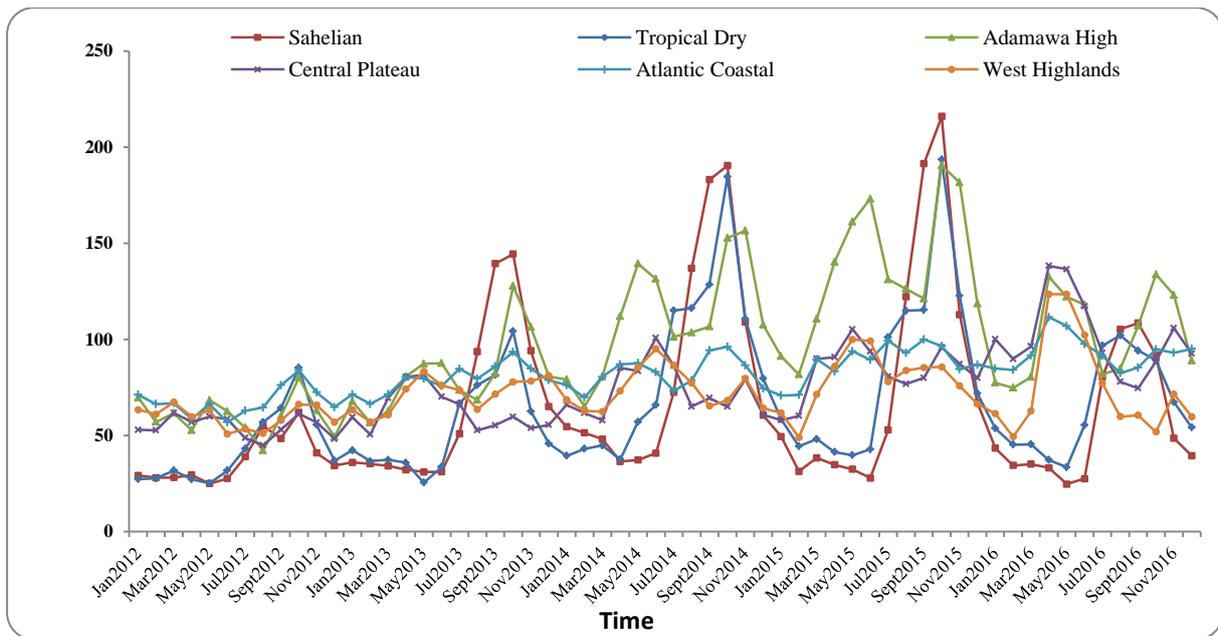
Figure 4.4 : Confirmed malaria cases (all ages) per 1000 inhabitants by epidemiological facets

According to the epidemiological strata, the temporal variation of malaria cases can be classified in different groups. The first group was characterized with strong seasonality of transmission with the peaks within the months of September and October. The health districts in the Coastal, Sahelian and Tropical Dry facets belong to this group. The second group in Central plateau and West had the peak in malaria cases between the months of April to June, The malaria transmission in the Adamawa highlands was not stable over time, the periods from April to July and September to November were marked by an increase of disease cases (Figure 4.4).

Based on RFE, two important rainy seasons were observed per year, the first occurred between April and May and the second in the period of August to October. Malaria peaks appearing in October (exception of year 2016) were preceded by the second rainy. The LST day and night as well as NDVI were varying with the malaria cases (Figures 4.10-4.12 in appendix).

The pair correlation analysis of covariates indicates a strong linear relation between NDVI and EVI for the current and lagged values. The EVI was more correlated to the LSTD than NDVI and therefore EVI was removed from the predictors list (See Figures 4.13 - 4.13 in Appendix A).

Table 4.2: Lag times of predictors selected by the variable selection

Climatic predictors	Inclusion probabilities (%)			
	Excluded	Lag 0	Lag 1	Lag 2
RFE	0	0	0	**100**
LST,day	3.08	1.89	**94.98**	0.045
LST,night	**77.07**	1.45	19.12	2.35
NDVI	0	0	0	**100**

Lag 0: Average over the current month.
Lag 1: Average over the previous month.
Lag 2: Average over the second previous month.

Bayesian variable selection indicated that RFE and NDVI averaged over the last two months previous to the current as the most important predictors of malaria incidence (Table 4.2). The LSTD of the previous month was also identified as an important climate factor; meanwhile the LSTN was excluded.

Four models were fitted. Parameter estimates of models including the selected climate covariates, spatial dependency, unstructured random effect at district level and the Fourier terms are presented in Table 4.3. The Fourier terms were always statistically important. The climate factors NDVI and RFE of the two previous months were statistically important and positively associated to the malaria cases of the current month. The LSD, day of the previous month was negatively correlated to the confirmed malaria cases (see Table 4.3: Models 1 to 4). The adjusted autoregressive effect appeared statistically important and having an influence on the goodness to fit. In the final models (Table 4.4), the proportion of households owning at least one ITN for every two people was negatively associated to confirmed malaria cases.

The goodness of fit of the final models was done graphically by plotting observed and fitted confirmed malaria cases (Appendix A, Figures 4.16-4.17). The trend of estimated malaria incidence over the five years was close to the observed incidence. In 2016, the peak transmission of October was not identified and the most important discrepancy between fitted and observed was noticed in 2016 (Figure 4.5).
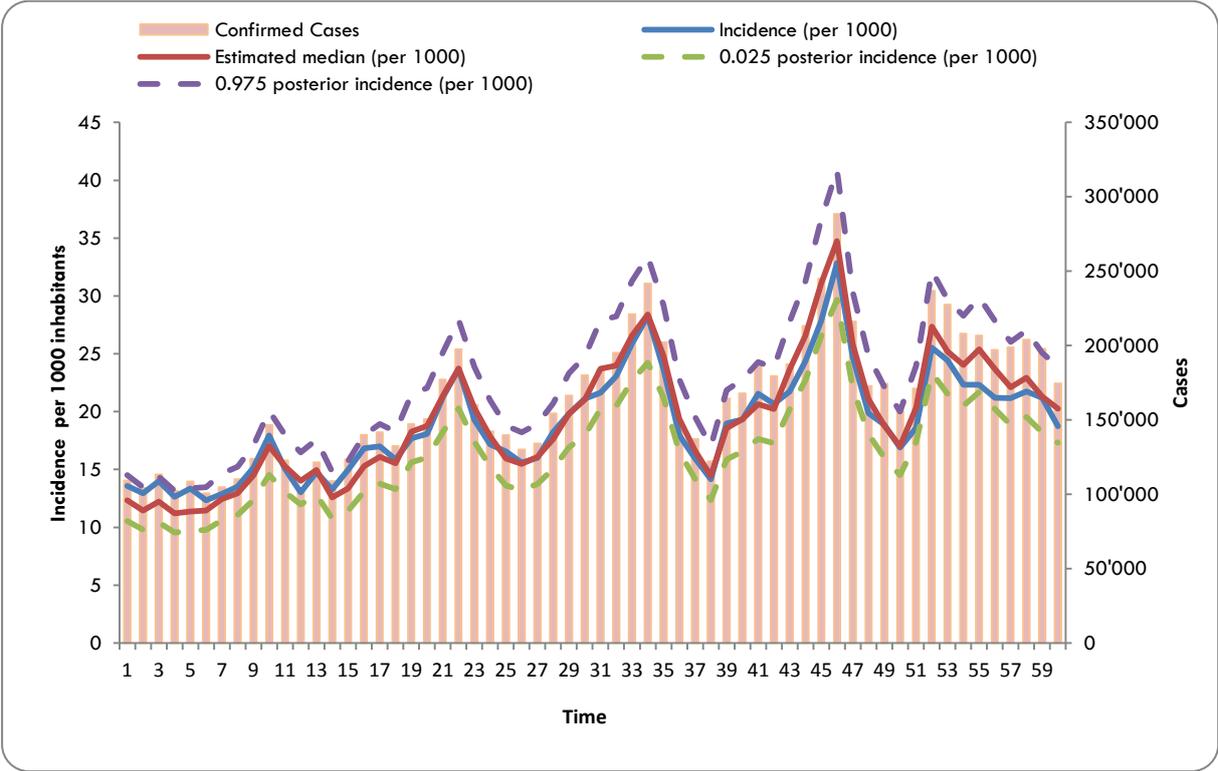
Figure 4.5 : Observed and estimated malaria incidence including the 2.5%, 50% and 97.5% posterior estimates malaria incidence obtained by Model 4.

Table 4.3: Bayesian Negative Binomial CAR models with the most important predictors of malaria risk in all groups.

Parameters	Model 1 Posterior median (95% BCI)	Model 2 Posterior median (95% BCI)	Model 3 Posterior median (95% BCI)	Model 4 Posterior median (95% BCI)
Intercept	-4.345 (-4.541 ; -4.149)	-4.048 (-4.357 ; -3.743)	-4.355 (-4.648 ; -4.052)	-4.265 (-4.587 ; -3.925)
NDVI (Lag 2)*	0.336 (0.239 ; 0.432)	0.585 (0.476 ; 0.693)	0.585 (0.476 ; 0.693)	0.568 (0.461 ; 0.675)
Rain (Lag 2)*	0.07 (0.059 ; 0.082)	0.115 (0.098 ; 0.132)	0.115 (0.098 ; 0.131)	0.115 (0.098 ; 0.131)
LSTD (Lag 1)*	-0.054 (-0.059 ; -0.05)	-0.042 (-0.047 ; -0.037)	-0.042 (-0.047 ; -0.037)	-0.045 (-0.05 ; -0.04)
Cosinus	-0.044 (-0.059 ; -0.028)	-0.041 (-0.145 ; 0.062)	-0.047 (-0.139 ; 0.046)	-0.053 (-0.145 ; 0.04)
Sinus*	0.116 (0.09 ; 0.142)	0.117 (0.008 ; 0.226)	0.132 (0.033 ; 0.23)	0.151 (0.051 ; 0.25)
Twoitn*	-0.002 (-0.004 ; -0.0003)	-0.001 (-0.003 ; 0.001)	-0.001 (-0.003 ; 0.001)	-0.005 (-0.008 ; -0.003)
Time*	0.013 (0.012 ; 0.014)	-	0.011 (0.004 ; 0.017)	0.013 (0.005 ; 0.02)
ρ(autoregressive)	-	0.842 (0.717 ; 0.938)	0.691 (0.493 ; 0.857)	0.711 (0.507 ; 0.874)
DIC	161 095.65	160 154.64	160 155.01	159 740.24

Twoitn corresponds to the proportion of households owning at least one ITN for every two household members was extracted.

*: the symbol indicates parameters that were statistically important.

 Model 1: Without autoregressive and time-region interaction;

Model 2: Without time trend and time-region interaction;

Model 3: Without time-region interaction;

Model 4: with all covariates.

Table 4.4: Bayesian Negative Binomial of the best model (Model 4) applied to malaria risk in children and more than five years groups.

Parameters	Under five years Model 1 posterior median (95% BCI)	Above five years Model 2 posterior median (95% BCI)
Intercept	-3.658 (-4.902 ; -2.378)	-4.48 (-4.774 ; -4.171)
NDVI (Lag 2)*	0.589 (0.476 ; 0.702)	0.54 (0.434 ; 0.645)
Rain (Lag 2)*	0.125 (0.107 ; 0.143)	0.106 (0.089 ; 0.122)
LSTD (Lag 1)*	-0.047 (-0.053 ; -0.042)	-0.042 (-0.047 ; -0.037)
Cosinus	-0.045 (-0.168 ; 0.076)	-0.048 (-0.132 ; 0.036)
Sinus*	0.169 (0.042 ; 0.296)	0.133 (0.042 ; 0.223)
Twoitn*	-0.006 (-0.009 ; -0.003)	-0.005 (-0.007 ; -0.002)
Time*	0.01 (-0.016 ; 0.035)	0.013 (0.006 ; 0.018)
ρ(autoregressive)	0.944 (0.838 ; 0.992)	0.667 (0.448 ; 0.857)
time-region interaction*	31.663 (16.623 ; 53.913)	66.123 (33.193 ; 122.638)

Twoitn corresponds to the proportion of households owning at least one ITN for every two household members was extracted.

*: the symbol indicates parameters that were statistically important.

Model 1 and Model 2 fitted with all covariates.

- **January**
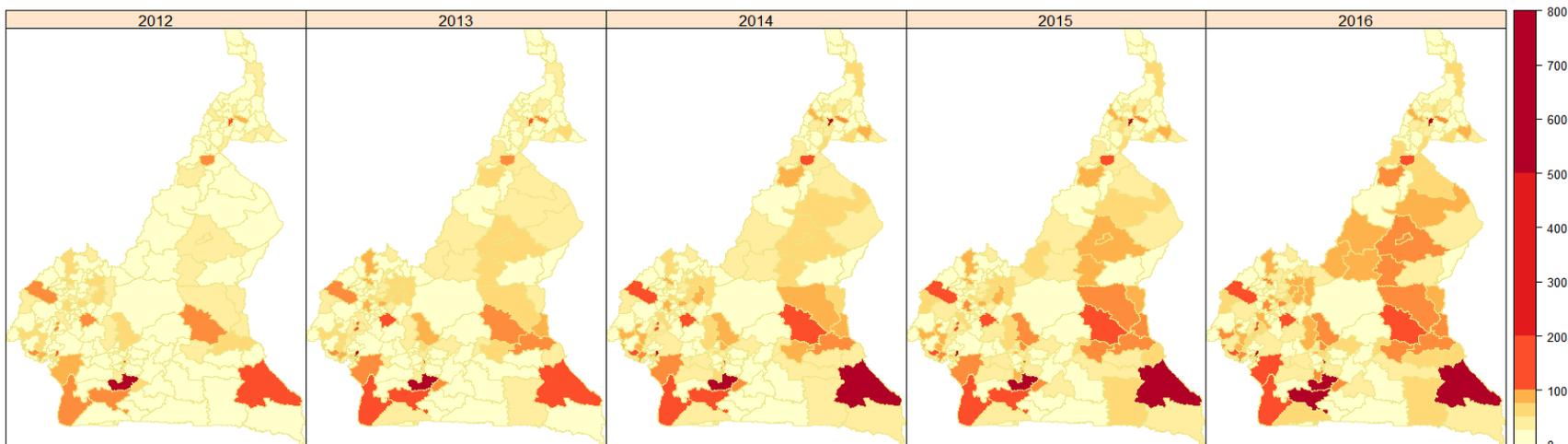
- **April**
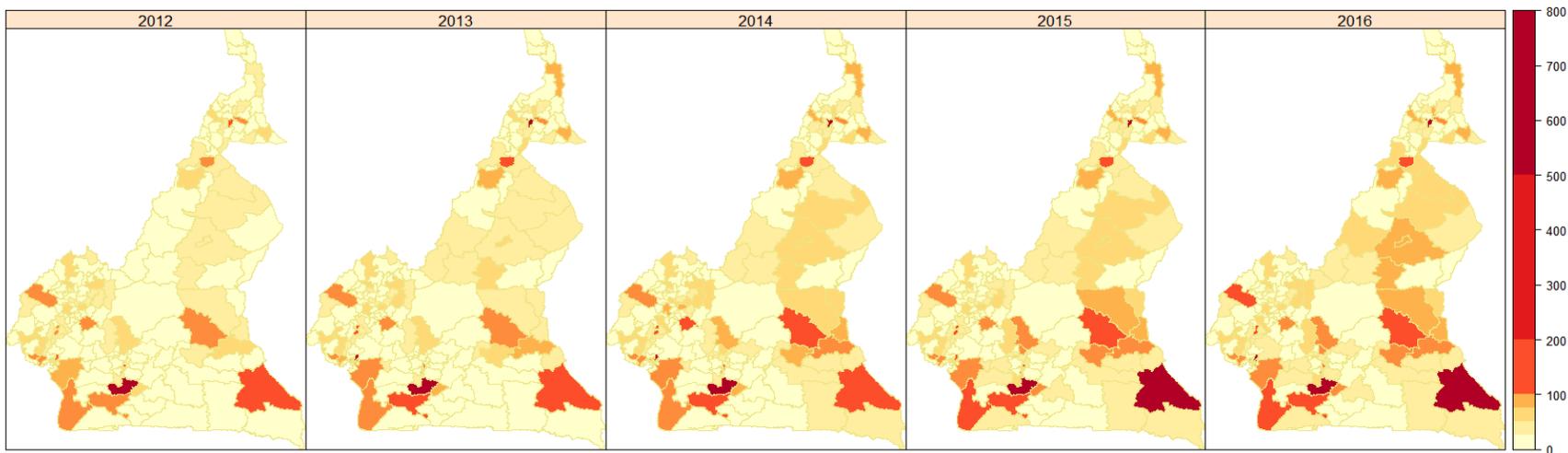
Figure 4.6 : Estimated children under-five malaria incidence per 1 000 inhabitants in Cameroon from 2012 to 2016 (January and April).

105
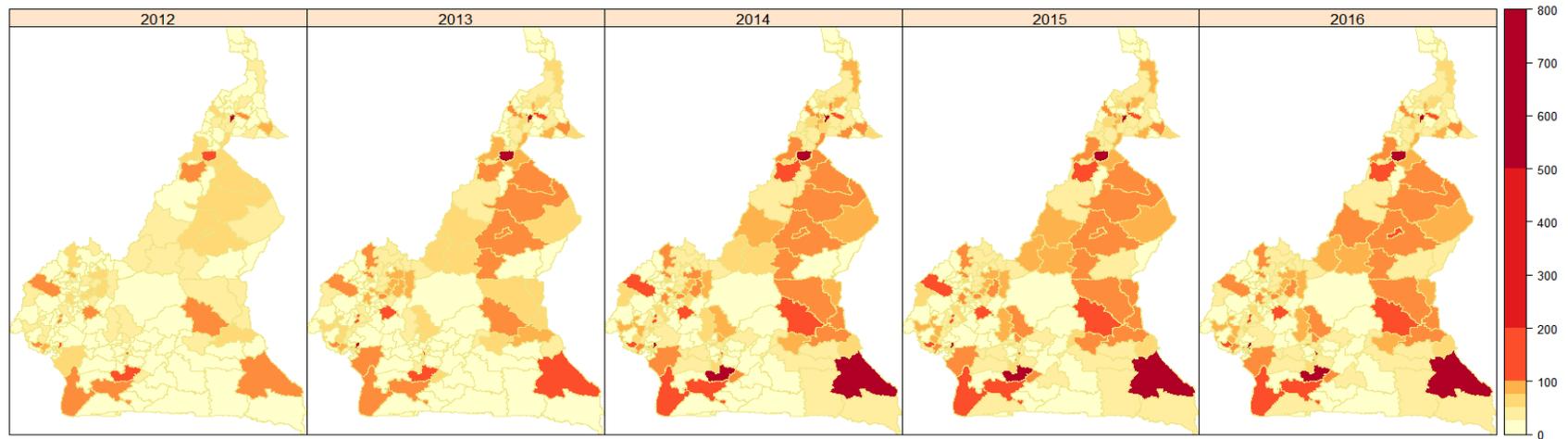
- **July**

- **October**

Figure 4.7: Estimated children under-five malaria incidence per 1 000 inhabitants in Cameroon from 2012 to 2016 (July and October).

- **January**
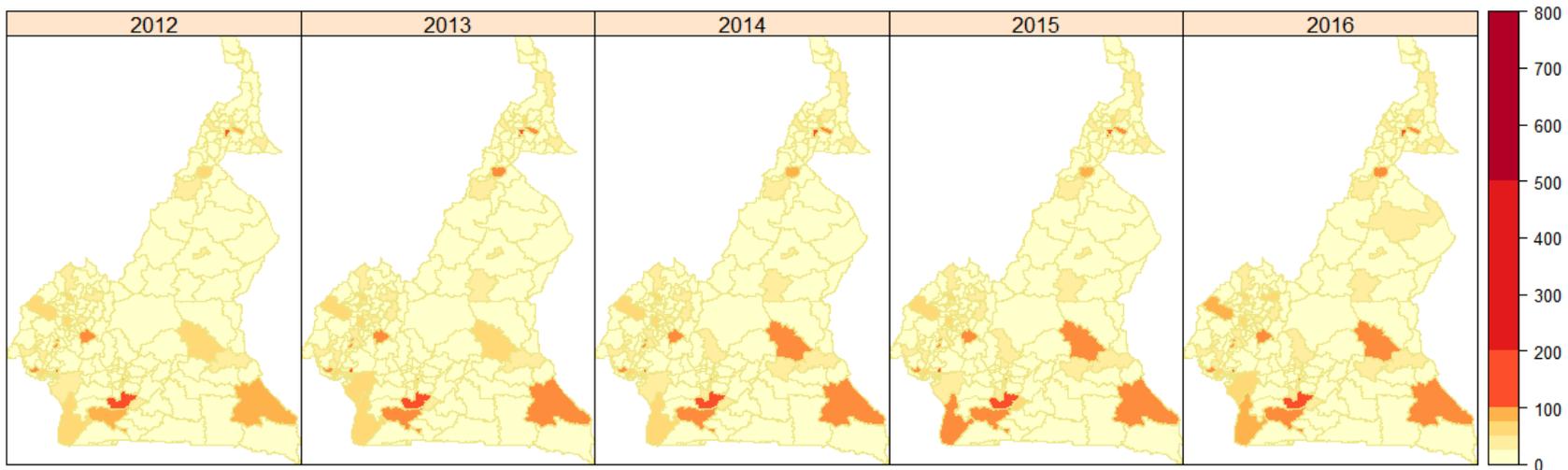
- **April**
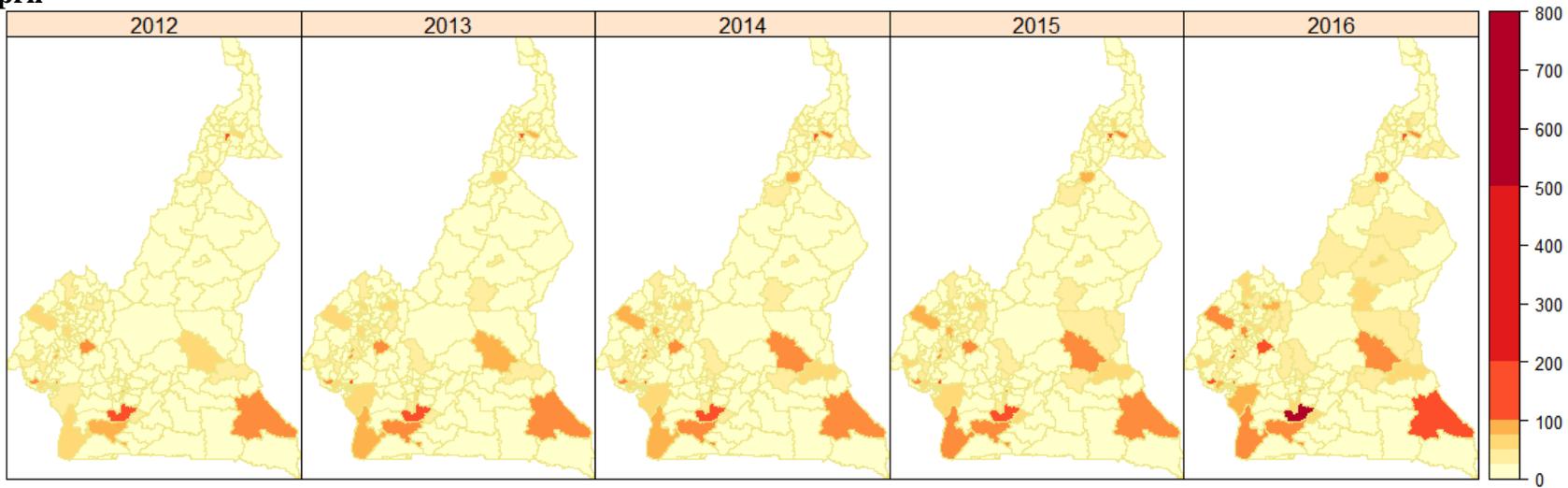
Figure 4.8: Estimated above five years malaria incidence per 1 000 inhabitants in Cameroon from 2012 to 2016 (January and April).

- **July**
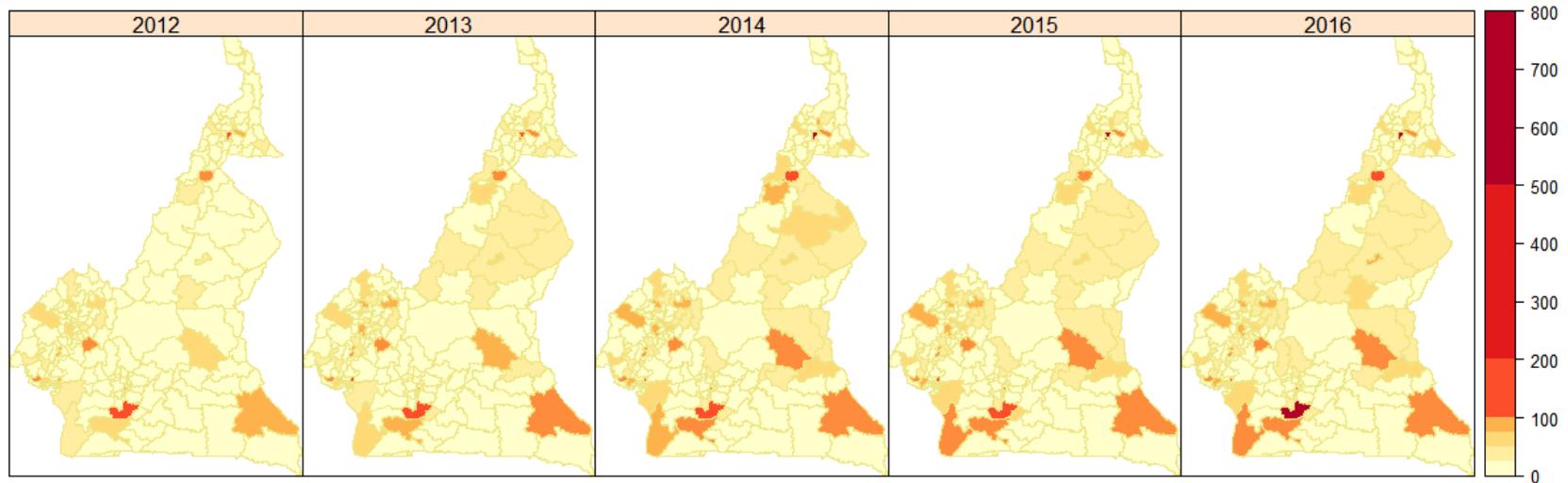
- **October**
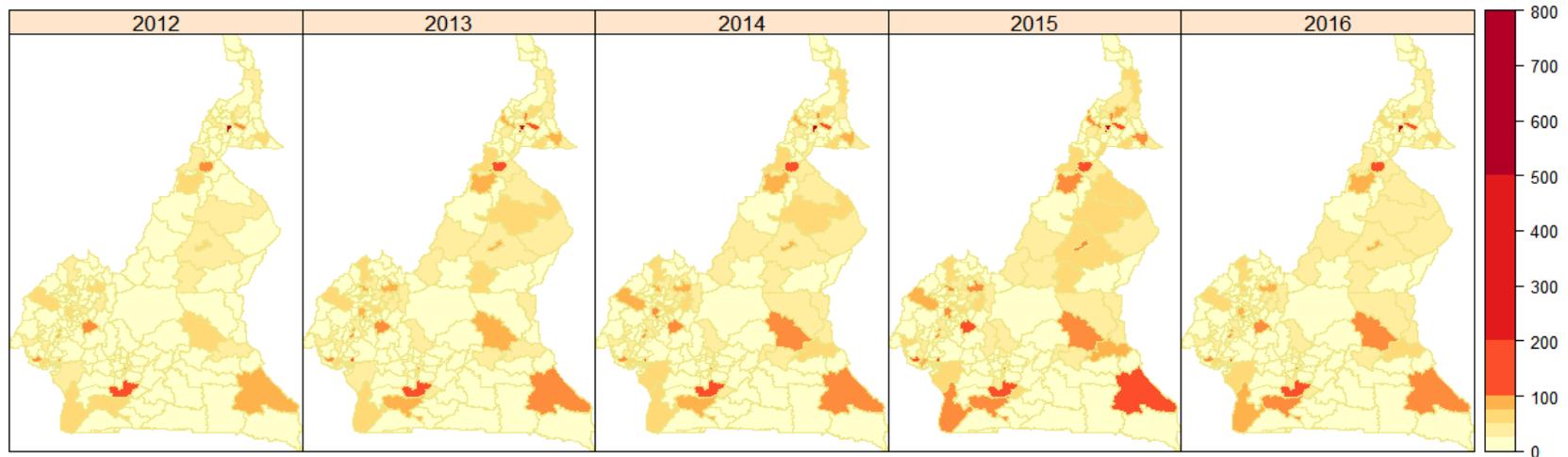
Figure 4.9: Estimated above five years malaria incidence per 1 000 inhabitants in Cameroon from 2012 to 2016 (July and October).

4.5 Discussion

This study was the first to highlight spatial patterns of malaria incidence among children below five years old and those above five using the HMIS data in Cameroon. Climate and environmental factors retrieved from satellites sources were used as covariates into a Bayesian conditional autoregressive model to produce maps of malaria incidence estimates at district level.

Compared to cross-sectional survey data such as MIS or DHS, health facility data are collected on a daily basis. Malaria data are integrated in the routine health activities and do not require additional budgeting. Moreover, population of all age groups, with symptoms, that is seeking malaria treatment at health facilities is included instead of the MIS and DHS malaria parasitological data that usually include only children aged between 6 months and five years. The use of confirmed malaria cases avoids mixing up between clinical malaria and asymptomatic cases. This choice contributes to a reduction in bias due to asymptomatic cases among above five years, which could lead to overestimate disease risk as in cross-sectional surveys (Bonaventure Tientche, Damian Nota Anong, Smith Asaah, Jerome Fru-Cho,Theresa K. Nkuo-Akenji and Damian Nota Anong, 2016; Dalrymple et al., 2018; Fru-Cho et al., 2014).

During the study period, the completeness of reports submitted by the health facilities was above 80% and covered all regions (10), divisions (52) and districts (189) of the country. Reporting was higher in urban than rural areas due to political management and the difference of socio-economic levels, which resulted in varying degrees of population attractions. However, the geographical coverage of health facilities at district level was higher than the one obtained by the DHS sampling method in which many health districts are generally not represented by cluster locations.

The direction of the effects between the malaria cases and the selected climate factors, i.e. NDVI, RFE, and LSTD confirmed previous analyses. However, only the NDVI was captured in the analyses of the DHS and MIS data collected in 2011 and used into Geostatistical models to predict the disease risk distribution in Cameroon (Stresman, 2010; Diboulo et al., 2015; Massoda Tonye et al., 2018). Altitude was not included in our climatic dataset due to the ecological fallacy that can influence measures at district level, particularly when high irregularity over district surface are recorded (Chapman Hall CRC, 2010).

From 2011 to 2016, the percentage of households with at least one LLIN had increased from 33 to 77 %. Two national mass-campaign organized in the country could explained that result. However, the percentage of households owning one LLIN per two people just increased from 32 to 47 % during the same period and this indicator was negatively associated to the malaria incidence. Further efforts need to be made for an equitable distribution of LLINs and to achieve universal coverage in the country(National Institute of Statistics, 2017, 2013).

A gradual upsurge of overall malaria incidence was observed between 2012 and 2015, which could be explained by the progressive decrease of effectiveness of long-lasting insecticides nets distributed in 2011 with a shelf-life of three years. The month of July was characterized by increased of malaria cases among children in the north where the peak occurred in October, with an exception observed in 2016 where there was an overall drop of incidence in the country and specifically in the north. This result can be explained by the combination of the second nationwide LLIN mass-campaign and the first implementation of the SMC in children under 5 in the North and Far North regions (WHO, 2017a; Scheme, 2013; Boussougou-Sambe et al., 2017).

Despite of the scaling-up of malaria interventions related to case management and prevention, some permanent pattern of risk were identified in the East region (Yokadouma, Garoua boulai, Kette, Betare oya, Bertoua and Doume districts) bordering to Central African Republic, in coastal districts (Kribi and Sanaga-Maritime districts), in North region (Pitoa, Maroua 3 and Gazawa districts), South -West region (Mamfe district) and Centre (Mbalmayo district) and South region (Ebolowa district). Most of the previous cited health districts have specific socio-economic and infrastructure challenges (Figures 4.6-4.9).   In the north, the Pitoa district has a water retention dike that could positively influenced the mosquitoes population and explain the persistence of high malaria incidence among children. Other persistent high malaria incidence were located in Health districts of the East region may be due to overlapping of noxious factors such as insufficient coverage of interventions, the continued influx of refugees from Central African Republic due to a civil war and the abundance of forests which is a favourable ecosystem for mosquitoes(Kar et al., 2014). The geographical distribution of malaria incidence resembles that of the disease risk estimates, obtained with DHS and MIS surveys that were carried out at different malaria transmission periods and were based on parasitological test among children under five in 2011 (Massoda Tonye et al., 2018). These consistent results suggest that in Cameroon, health facility data are a reliable source of information for malaria monitoring possibly due to the introduction of computerized systems dedicated to capture malaria data from health facilities.

Spatial and temporal analysis of confirmed malaria incidence showed a pattern of high transmission period starting in July and ending in October where the peak was observed. In the north region, we identified lower risk of malaria transmission from January to June, while higher risk was observed between July and October (Figures 4.5-4.9). Those findings were aligned to the DHS and MIS data. This study confirms the importance of timing in surveys that aim at estimating malaria transmission risk in Cameroon(Tabue et al., 2014, 2017).

Lack of universal health care at national level and the low-income of population combined with free treatment of uncomplicated and severe malaria cases among children affected the number of children seeking malaria treatment. Due to this favourable case management policy, malaria cases and the disease trend within the children population increased, especially in health districts areas with low socioeconomic assets. (INS, 2014; INS and International ICF, 2012).

Our study assumed that the displacement of patients between health districts was negligible, in fact there is no evidence of the contrary, but this issue should be raised as a potential bias that can influence estimates of the model. Indeed, a patient living in a given health district with specific climatic factors moves to get malaria diagnosis and treatment in another health district were medical infrastructures are more appropriate. Residential information about patient should be integrated in the malaria software to address this issue.

Health facilities were not geographically referenced by coordinates (latitude and longitude) into our study period. An on-going WHO project was launched in 2015 to collect geographical coordinates of health facilities. This database will be available in the future and can enable the use of models as Log-Gaussian Cox Processes to improve the quality of malaria incidence maps (Y. Li et al., 2012; Diggle et al., 2013).

Cameroon ministry of public health and its partners supported periodic health facility surveys to assess data quality and malaria supply chain management of drugs and diagnosis tools. Instead of random sampling of health districts and facilities, it may be efficient to select them based on such models that could identify areas with abnormally high or low malaria incidence.

4.6 Conclusion

Our study showed the contribution of the scaling up of control activities on the reduction of malaria cases recorded at health facility level. Despite necessary adjustments of the Cameroon health system that can contribute to ensure a continuous improvement of malaria data, this study had been able capture the influence of climatic factors on the malaria cases changes over space (health district) and time (month). These results indicate that malaria data obtained from health facilities are also able to be helpful in framework of a continuous follow-up of specific interventions at local and national levels, and to target areas with high spot of malaria cases.

4.7 Appendix A

Pearson correlation structures of climatic/environmental factors from 2012 to 2016.
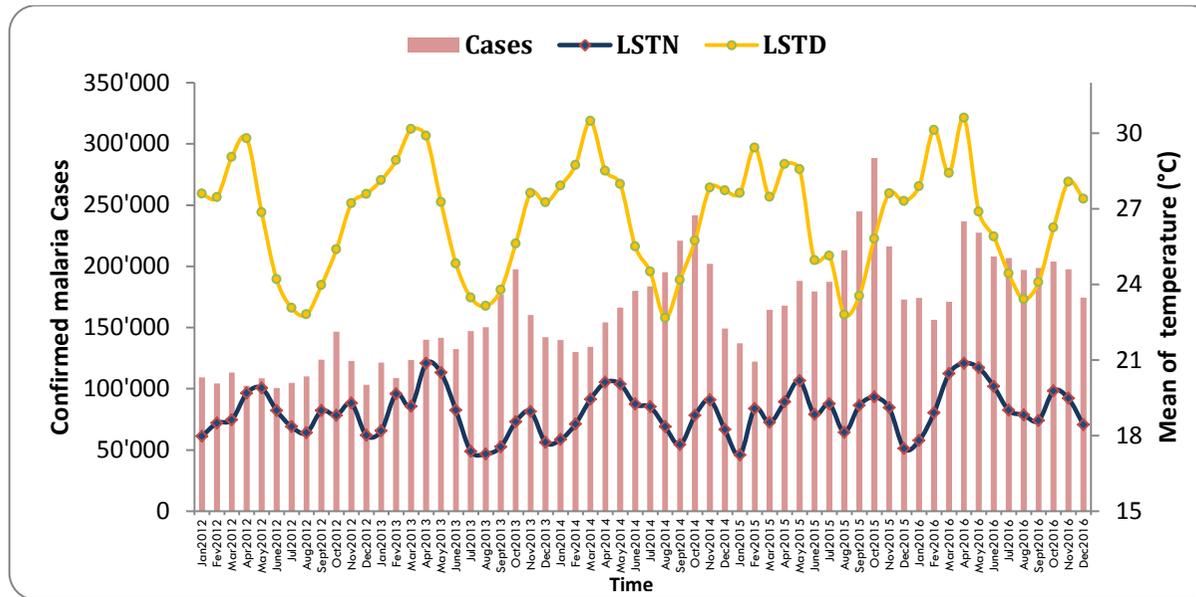
Figure 4.10 **:** LST, night and day trends with malaria cases registered from 2012-2016
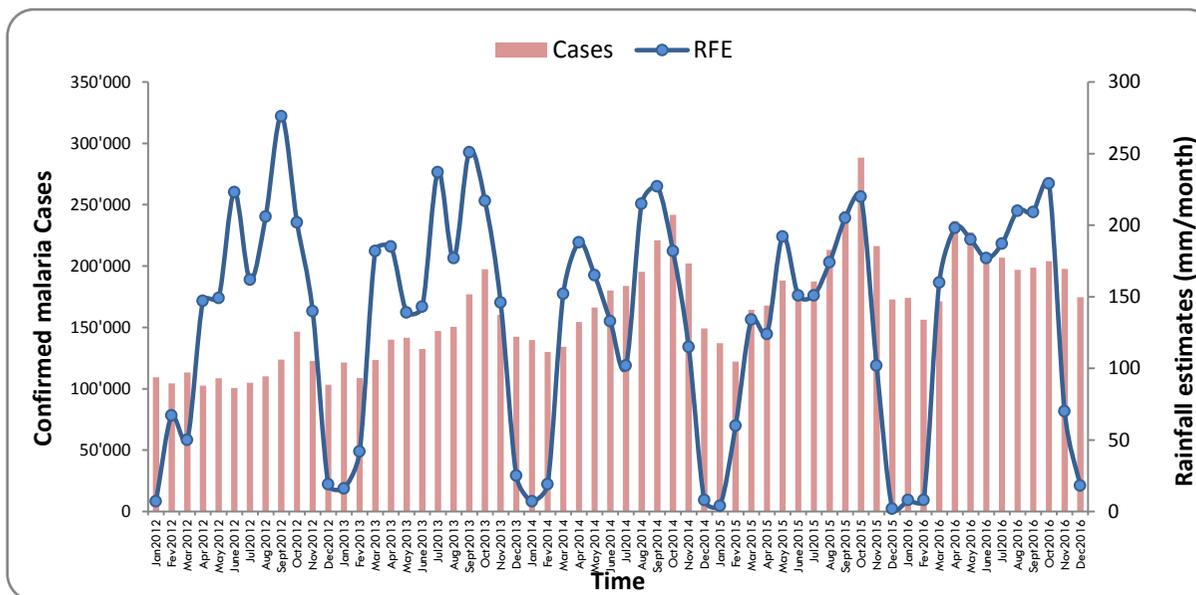
Figure 4.11: Cross-trends of malaria cases and rainfall estimates.
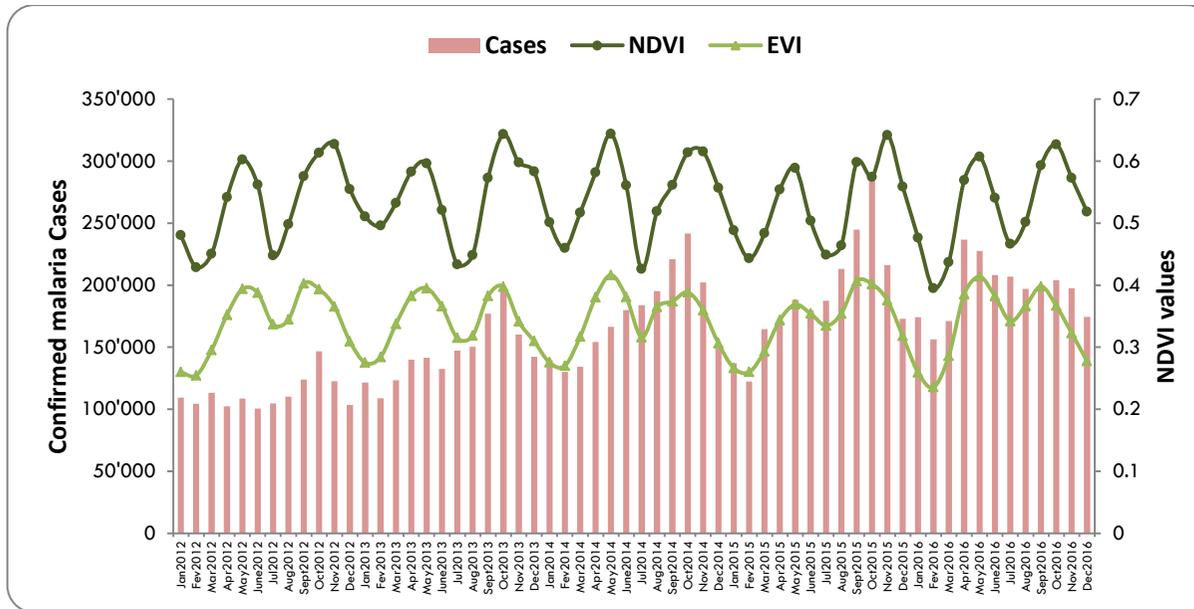
Figure 4.12: Cross-trends of malaria cases and NDVI, EVI estimates.
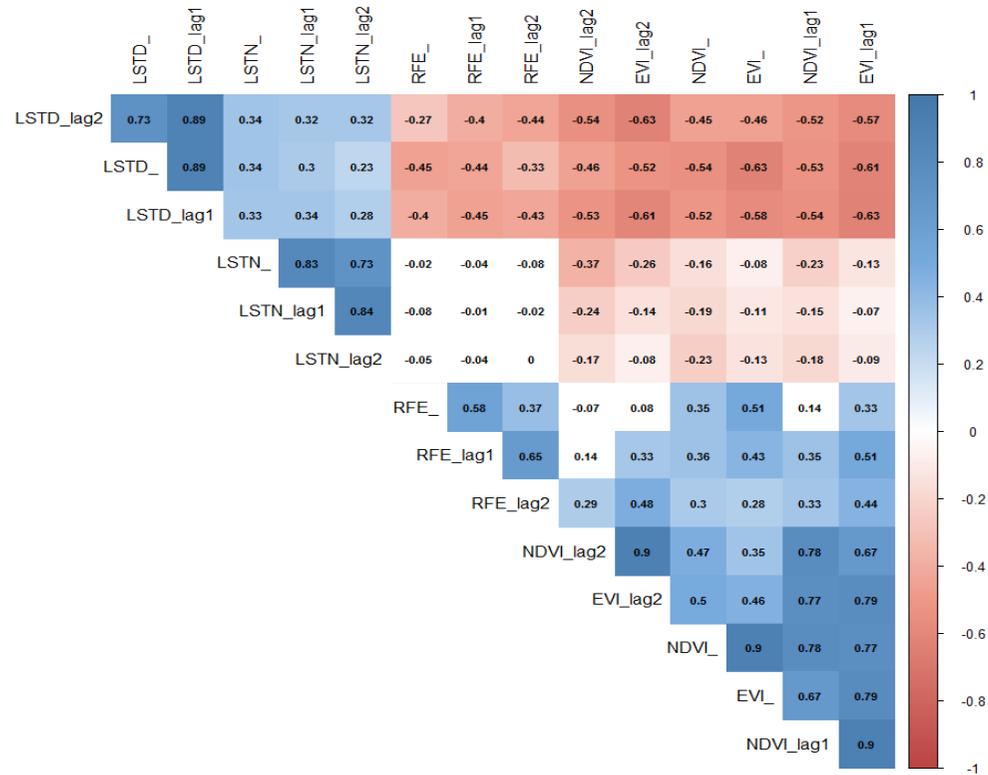
Figure 4.13: Pearson correlation structure in the 2012 between climatic covariates and lagged values.
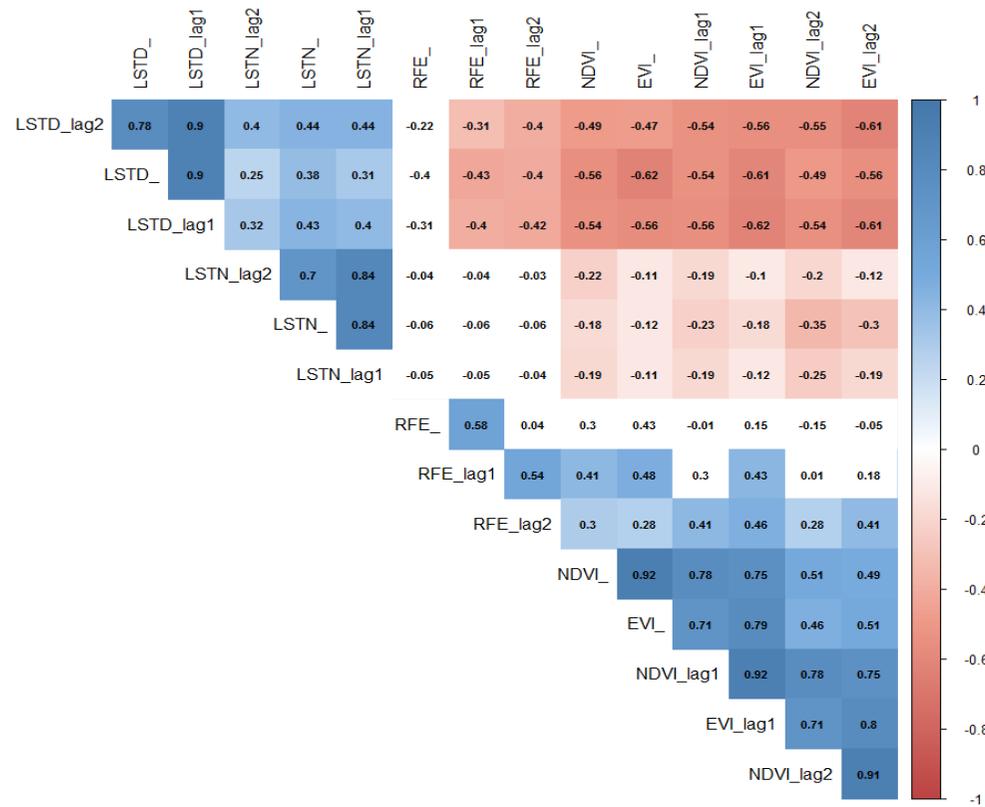
Figure 4.14: Pearson correlation structure in 2014 between climatic covariates and lagged values.
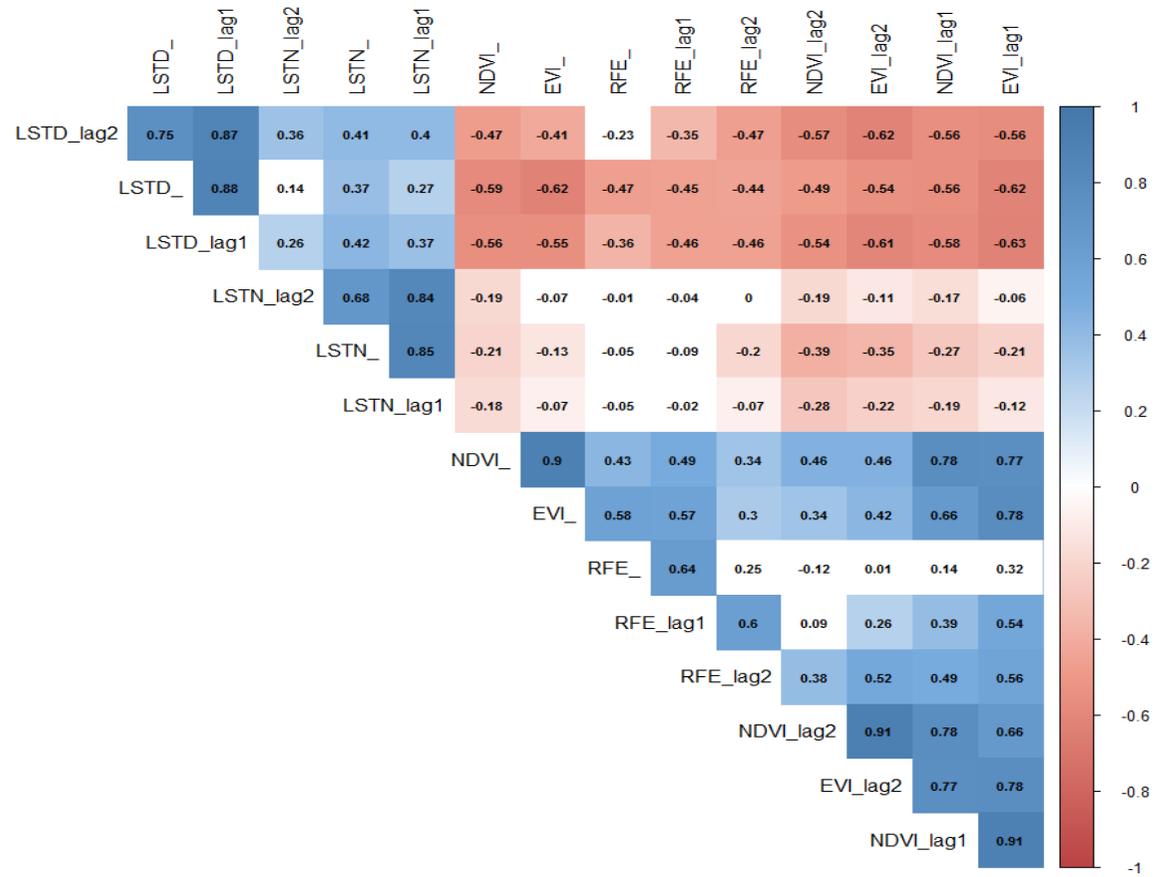
Figure 4.15: Pearson correlation structure in 2016 between climatic covariates and lagged values
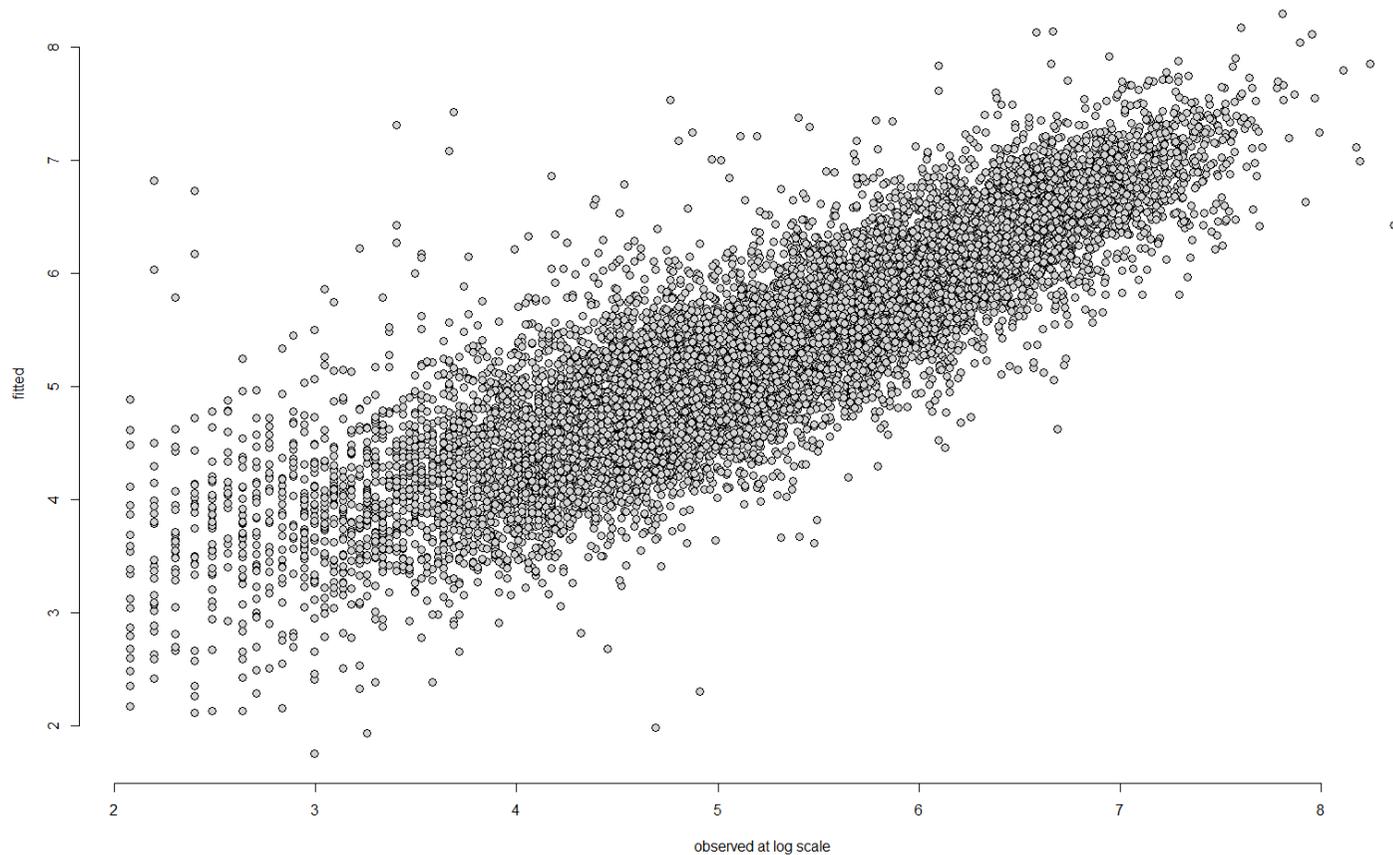
Figure 4.16: Observed values versus posterior means from the model fitted of confirmed malaria cases among children below five years group at log scale level.
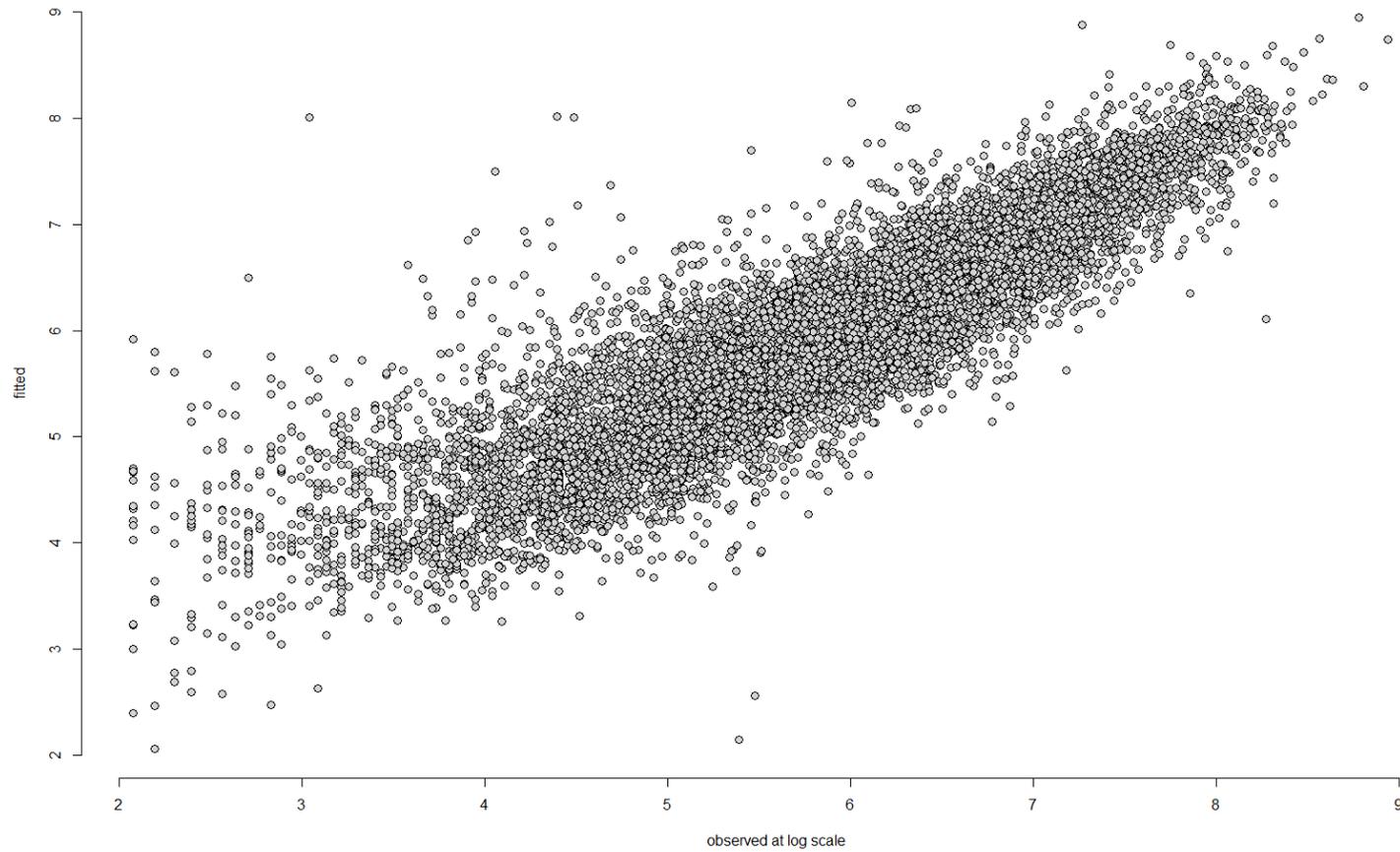
Figure 4.17: Observed values versus posterior means from the model fitted of confirmed malaria cases among people above five years at log scale level.

4.8 Appendix B

Bayesian variable selection

Additional file 2

We assumed that confirmed malaria case at location district i, at time t arises from negative binomial distribution $Y_{it} \sim NegBin$ (p_{it}, r), where $p_{it} = \dfrac{r}{r+\mu_{it}}$ and μ_{it} , r are the mean and dispersion parameter of the negative binomial distribution, respectively.

$$log(\mu_{it}) = log(pop_{it}) + \beta_0 + \beta_1 t + \rho\varepsilon_{t-1} + \varphi_i + \gamma_i + \delta_r t$$

$$+ \sum_{j=1}^{k} \sum_{m=0}^{p} \beta_m^j X_{i(t-m)}^j + a_1 \cos\left(\frac{2\pi t}{T}\right) + b_1 \sin\left(\frac{2\pi t}{T}\right)$$

District code $i = 1 \dots 180$, time t $= 1 \dots 60$ months, frequency $= 1/T=1/12$

$$\varphi_{i|j \in V_i} \sim N\left(\sum_{j \in V_i} \rho_{ij}\varphi_j \, , \frac{\sigma_i^2}{n_i}\right), \quad \varepsilon_t = \rho\varepsilon_{t-1}, 0 < \rho < 1, \delta_r \sim N(0, \tau_r^2)$$

$$\gamma_i \sim N(0, \gamma^2). \, \varepsilon_{t-1} = \varepsilon_1 \sim N\left(0, \frac{\tau^2}{1-\rho^2}\right), \beta_l \sim N(0, \tau_l^2), l = 0..1.$$

Let consider $X_{i(t-m)}^j$, the climatic/environmental factor j collected at district code i during the

time t-m (m=0,1,2) that correspond either to the current month t, or to the previous (t-1) or

two month lagged before (t-2). A Bayesian stochastic search variable selection approach was

applied such as for each climate factor j and the related lagged time values, a single predictor

$X_{i(t-m)}^j$ was captured as the most important predictor. Because EVI and NDVI had a high

linear correlation, a first variable selection level was used to ensure that only one of them will

be included in the final model. For the second level of variable selection, an inclusion

indicator I_j was assigned to each climatic or environmental covariate j, which defines

exclusion of the variable (including lagged time values) from the model ($I_j = 1$). The

inclusions of the current month (t) when $I_j = 2$, of the previous month (t-1) when ($I_j = 3$), of

the second lagged month when $I_j = 4$. Each I_j was related to a probability mass

function $\prod_{k=1}^{4} \pi_k^{\delta_k(I_j)}$, where π_k denotes the inclusion probabilities that are constraints by

$\sum_{j=1}^{4} \pi_j = 1$ and $\delta_k(.)$ is the Dirac function, $\delta_k(I_j) = \begin{cases} 1, \text{if } I_j = k \\ 0, \text{if } I_j \neq k \end{cases}$. We have adapted from the

spike and slab approach to select a covariate by using corresponding regression coefficient $\beta^{j}_{t-m,m\epsilon\{0,1,2\}}$ such as, i.e.

$$\beta^{j}_{t} \sim \left(\delta_1(I_j) + \delta_3(I_j) + \delta_4(I_j)\right) N\left(0, u_0\tau_j^2\right) + \delta_2(I_j)N\left(0, \tau_j^2\right) \text{ (1)}$$

$$\beta^{j}_{t-1} \sim \left(\delta_1(I_j) + \delta_2(I_j) + \delta_4(I_j)\right) N\left(0, u_0\tau_j^2\right) + \delta_3(I_j)N\left(0, \tau_j^2\right) \text{(2)}$$

$$\beta^{j}_{t-2} \sim \left(\delta_1(I_j) + \delta_2(I_j) + \delta_3(I_j)\right) N\left(0, u_0\tau_j^2\right) + \delta_4(I_j)N\left(0, \tau_j^2\right) \text{(3)}$$

A informative normal prior with variance close to zero (i.e. $u_0 = 10^{-5}$) that shrink β^{j}_{t-m} to zero when the factor j (including lagged values) was excluded or once any $X^{j}_{(t-k),k\neq m}$ was included. A non-informative Dirichlet distribution was also adopted with hyper parameters $\alpha = (1,1,1,1)^T$, so that $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)^T \sim Dirichlet(4, \alpha)$. A categorical distribution prior with an equal exclusion or inclusion probabilities was assumed for the inclusion indicator i.e. $I_j \sim Cat(\boldsymbol{\pi})$, and an inverse Gamma priors for each τ_j^2 ~Inv-Gamma (2.01, 1.01).

Chapter 5 Contribution to the identification of relevant locations in malaria multi-setting countries using Bayesian geostatistical sampling design

Salomon G. Massoda Tonye[1,2,3] , Wouang Romain[2], Kouambeng Celestin[3], Penelope Vounatsou[1,2]

1. Swiss Tropical and Public Health Institute, Basel, Switzerland

2. University of Basel, Basel, Switzerland

3. National Institute of Statistics, Yaoundé, Cameroon

4. National Malaria Control Programme, Yaoundé, Cameroon

Corresponding author

Email: Penelope.vounatsou@swisstph.ch

Abstract

Background

The increasing demand on cost-effective health interventions and on informative data had influenced development of sampling design approaches. The identification of appropriate cluster locations for the monitoring of vector control activities as well as intervention effects was among the key issues during the last decade. Our study proposed an algorithm able to identify relevant cluster locations to improve our knowledge on malaria transmission. Households data used in our study were extracted from the 2011 Cameroon Malaria Indicator Survey (MIS) and climatic data were retrieved from satellite sources.

Method

Geostatistical variable selection involving the spike and slap approach was used to identify the most important climatic predictors. Bayesian kriging based on the selected predictors was applied to estimates disease risk and the associated uncertainty at unobserved locations. A random selection of new relevant cluster locations was achieved using their spatial uncertainty and a set of defined target objectives of malaria control program.

Results

At national level, malaria parasite risk among children below five years was 33%. Children in urban areas were less exposed than those in rural areas. The North, East, and Coastal regions were the most affected parts of the country and their contribution to the overall spatial uncertainty was important. Our algorithm had mainly identified cluster locations in rural areas and particularly from East, North and Center regions as important to improve the quality of spatial estimates. Some areas in the West where the disease risk was low also identified.

Conclusion

Our study was able to identify additional cluster locations based on a comprehensive approach that combined need of reducing the spatial uncertainty and of incorporating the national disease control objectives. This algorithm based on model uncertainty can be applied on other disease related to environmental conditions. It is a useful tool to the identification of areas/sites able to generate valuable data to assess the disease intervention effects.

Keywords: Malaria, Malaria Indicator Survey, Demographic and Health Survey, parasitaemia, Statistically important, Sampling design.

5.1 Introduction

Mobilizing financial resources for health interventions that can contribute to the effective reduction of mortality and morbidity, and remains one of the major challenges to global health (Van de Pas et al., 2017; Warren et al., 2013). Malaria control interventions have been implemented on a large scale in affected countries, mostly in the low- and middle-income group. In these countries, social and economic sectors also had investment needs. Therefore, the design of cost-effective interventions has been encouraged over the last decade, and malaria funding has been primarily allocated to interventions with the most favorable financial cost and expected outcomes ratio on the disease burden (Russell et al., 1996; Korenromp et al., 2013; Zelman et al., 2016; Head et al., 2017).

Supportive health interventions such as monitoring and evaluation (M&E) have recently been more recognized as an essential component of the global strategic plan for malaria (WHO, 2016). In fact, M&E activities have produced reliable information to assess progress and impact of interventions defined by the global and national action plans. The most popular M&E activities were probably periodic household surveys, health surveillance reporting system, follow-up of environmental factors, and vector control. However, M&E activities were budget demanding and several countries faced the challenge of ensuring efficient monitoring of health interventions, especially when national or global resources were reduced. Due to budgetary constraints, the national and representative household surveys such as Demographic and Household Survey (DHS), Multiple Indicator Clusters Survey (MICS) or Malaria Indicator Survey (MIS) were often deferred and conducted at different frequencies varying from two to five years according to the country and commitment of donors. Among the reasons for these expensive budgets were standardized sampling design approach based on intervention coverage indicators and cluster population densities to ensure population representativeness at national as well as regional levels (USAID, 1997).

Nevertheless, the spatial predictive information associated to disease risks and estimated by previous surveys were not included in the sampling design of the next survey (INS and International ICF, 2012; INS, Minsante et UNICEF, 2015).

The DHS/MICS/MIS standard sampling method usually takes advantage clusters in populated urban areas, which can lead to an irregular distribution of cluster locations and introduce a redundancy of spatial data. Thus, inducing a selection of non-informative data that can be perceived as a budget inefficiency (Diggle and Lophaven, 2006). In another way, the selection of suitable sites for vector control and malaria surveillance activities was generally based on an entomological criterion that involved environmental and climatic conditions. Bayesian geostatistical sampling design approaches can contribute to improve the cost-effectiveness of household surveys and the identification of sites for vector control by providing a relevant set of locations useful to enhance our knowledge on disease risk distribution, by using spatial predictive data obtained from previous surveys (Consultative Group on Health Systems and Operational, 2011; Travis et al., 2004).

Interesting contributions have been achieved in the field of environmental monitoring. These approaches could be grouped into a non-adaptive class (i.e. retrospective) which aims to diminish clusters locations from an existing network, and the adaptive class (i.e. prospective) that investigate among unobserved locations to improve spatial prediction (Diggle and Lophaven, 2006; Groenigen et al., 1998; Lark, 2002; Mcbratney et al., 1981). Studies on the field of non-adaptive eostatistical design focus on defining a cost function to optimize. *McBratney et al* (1981) had suggested to identify cluster locations that maximize the spatial predictive variance over the surface and *Russo* (1984) preferred those that minimize the semi-variogram information (Mcbratney et al., 1981; Russo, 1984). *Van Groenigen et al* (1998) proposed to maximize the spatial coverage, when *Lark* (2002) chose to maximize the likelihood estimation derived from geostatistical model (Groenigen et al., 1998; Lark, 2002).

A geometrical approach was developed by *Royle et Nychka* (1998), in which they recommend positions that provide a high spatial coverage using Euclidian distance as a cost function (Royle and Nychka, 1998).

Although these approaches have been relatively successful in minimizing the loss of spatial forecasting but maintaining the robustness of geostatistical model parameter estimates remains a problem. It was also observed that nearby locations were more likely to generate a moderate change in estimates of model-based geostatistics, while regular spatial distribution was more interesting for spatial predictive performance(Diggle and Lophaven, 2006).

In the Bayesian framework, *Diggle et al* (2006) suggested to optimize metrics in retrospective and prospective sampling designs. The overall variance of prediction over the surface was used to include or exclude cluster location sequentially from a prior sampling. With a view of reducing sampling, they were able to maintain an acceptable spatial forecast and to minimize degradation of model parameter estimates. By combining a regular distribution of locations with close pair clusters selection algorithms they were able to provide both an accurate spatial prediction and stable model parameter estimates (Diggle and Lophaven, 2006).

The Geostatistical design approaches mentioned above have their optimality based on a specific cost function, and do not involve any strategical or operational disease objectives. They were often computationally expensive especially in Bayesian framework. Our study aims to contribute to adaptive geostatistical design in multi-objectives context of malaria. A Bayesian Geostatistical model was applied on the 2011 Cameroon MIS data to fit the relationship between malaria parasite risk, environmental and climatic factors. Then, based on estimates of spatial uncertainty and on the malaria program objectives, a weight procedure was used prior to selection of cluster locations on the predictive surface of the country. Different sets of cluster locations were identified for monitoring and assessment of malaria

intervention and vector control. The selection of clusters was performed using proportional sampling based on adjusted weights (Skinner, 2016; William G. Cochran, 1977).

5.2. Methods

5.2.1 Country setting

Cameroon is located in Central Africa, neighboring Tchad, Nigeria and the Central African Republic. The country is member of international organizations, bilateral organizations and multilateral organizations such as the World Health Organization (WHO), United Nations International Children's Emergency Fund (UNICEF) and the coordination agency for the fight against epidemics in Central Africa ("OCEAC" in French) whose headquarters is located in the capital city Yaoundé (Djaman et al., 2017; Motta et al., 2017; Munier et al., 2017). Administratively Cameroon is made up of ten (10) regions. These are further split up into 52 divisions, all having a total of 189 health districts. In 2017, the country's population was estimated at 24 million inhabitants, unevenly spread over a surface area of 475,650 km². Six main malaria transmission epidemiological zones exist, associated with specific ecological factors. National health surveillance of vector-borne diseases such as malaria, onchocerciasis, buruli ulcer, or lymphatic filariasis were mostly done passively during medical consultations at health facilities, and the malaria surveillance at sentinel site was not yet operational (WHO, 2018b, 2017b). Different sources of malaria parasitological data collected at individual level were used for surveillance and planning. They were mainly provided through health facilities, household surveys, and entomological studies (Kleinschmidt et al., 2018; Mbakop et al., 2019).
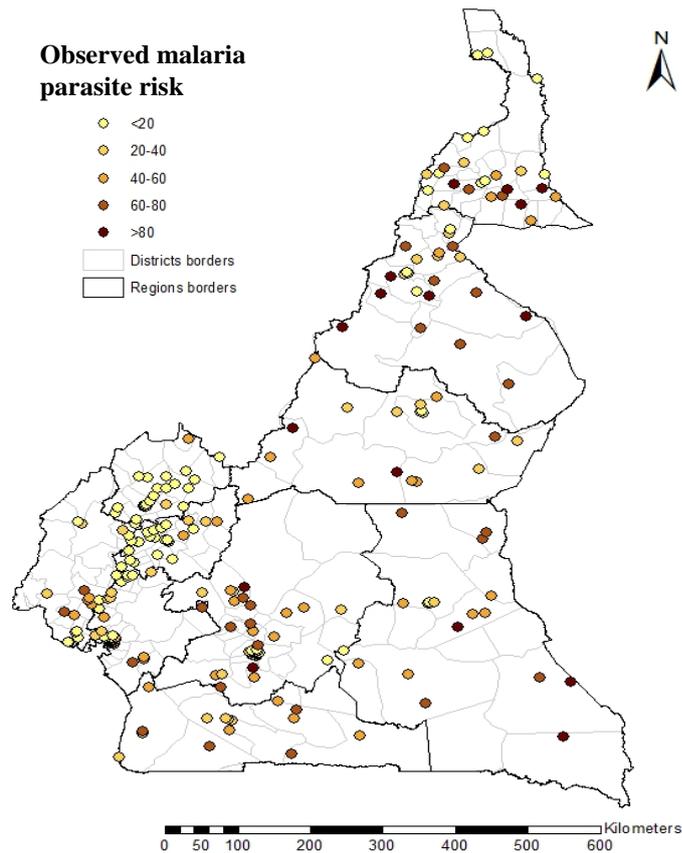
Figure 5.1: Observed malaria parasite risk in children below five years old at 257 MIS locations.

5.2.2. Data

5.2.2.1. Parasitological data

The fourth Cameroon DHS 2011 was combined with MICS and conducted from January to June, out of the malaria high transmission season. 580 clusters were selected using the two-stage random selection strategy and stratified according to the regional administrative unit (National Institute of Statistics, 2011). The cluster locations surveyed in Cameroon MIS 2011 were a subset of those visited in the fourth DHS. Cameroon MIS 2011 was carried out within the high malaria transmission season, between September and November 2011. A random sample of 257 clusters was selected out of the 580 clusters surveyed previously in the DHS, including 128 in rural areas (Figure 5.1). A total of 6 040 households were successfully surveyed during which 4,939 children less than five years old were tested. Following the agreement issued by the national ethical health committee, adult in charge's approval was

clearly asked, obtained and recorded before any interview or screening of children with rapid diagnostic test (INS and International ICF, 2012).

5.2.2.2. Environmental and climate data

Climate covariates were extracted from satellite remote sensing data. They were used as proxies of environmental conditions. Land Surface Temperature during the day and night (LSTD, LSTN), Normalized Difference Vegetation Index (NDVI), and Enhanced Vegetation Index (EVI) were used indicators of the temperature and vegetation coverage on the ground. Categories of the land cover surface were used to capture the forest presence, the savannah, cropland areas and distance to permanent water bodies (John Weier and Herring David, 2000; U.S. Geological Survey, 2017). Rainfall estimates (RFE) and altitude were sourced from FEWS (or Famine Early Warning Systems Network) and SRTM (or Shuttle Radar Topographic Mission) web sites, respectively (U.S. Geological Survey, 1985). The continuous climate proxies were averaged over one year period before the survey. Climatic data was extracted from locations surveyed and also on the gridded country surface represented by 117 192 cluster locations. The distance between grid points was fixed at 2 km.

5.2.3. Algorithm parts

5.2.3.1 Geostatistical modelling

Model-based Geostatistics was defined by *Cressie(1993), Diggle et al* (1998), the model was globally described with three components $\{Y, U, \theta\}$, the spatial Gaussian process $U = \{u(s_i), s_i \epsilon A\}$, the mutually independent observed outcome data $Y = \{y(s_i), s_i \epsilon A\}$ at specific locations $A = \{s_i, i \in 1 \dots n\}$ and the set of unknown parameters $\theta = \{\beta, \delta, \sigma, \tau\}$. In the Bayesian framework, basically the joint distribution was formulated by a product of likelihood, conditional and prior distributions as following $P(Y, U, \theta) = \mathrm{P}(Y|U, \theta)P(U|\theta)P(\theta)$ (Cressie, 1993; Diggle et al., 1998). A set of parameters representing the regression coefficients, length-scale, spatial variance and noise, respectively. An extension of these models commonly known as the generalized linear geostatistical model

that combining non-Gaussian data and unobserved spatial Gaussian random effect was formulated by Diggle et Ribeiro (2006) as follows:

$$Y(s_i) \mid U(s_i) \sim f(\mu(s_i), v),$$

$$g\big(\mu(s_i)\big) = \beta^T X(s_i) + U(s_i),$$

$$Cov\big(U(s_i), U(s_j)\big) = \sigma^2 \rho((s_i - s_j)/\delta, \tau).$$

Where *f* could arise from a binomial or Bernouilli distribution, and *g* is a link function (.i.e. log or logit) (Ribeiro, 2017).

5.2.3.2. Target definitions

Sample selection of cluster locations depends on their contribution to spatial uncertainty and on the objectives of the study. The objectives could be expressed as a function of the spatial Gaussian process estimates (Diggle and Lophaven, 2006). Within the malaria context, *Kabaghe et al* (2017) suggested an adaptive sampling design that focused on the high risk points of the predictive gridded surface, the targeted cluster locations was those that having a disease risk above a predefined threshold, $T = \{s_i \epsilon A, \ p(s_i) > c\}$. The threshold may depend on the definition of a hotspot, or the projection of prevalence given a set of interventions. Threshold can be defined at national, regional or district level and it may depend on strata i.e. ecological zone, urbanity level.

Other rules er low malaria increased mortality was related to the expectation of health intervention effects (Kabaghe et al., 2017). In countries having multiple and various ecological zones in which coexisting extremely low and high disease risk, the target list may be extended to incorporate locations with the higher or lower disease risk estimates obtained by geostatistical modeling. Additional targets could be defined as following : $T_j = \{s_i \epsilon A, c_{j-1} < p(s_i) < c_j\}$, where $c_{j-1}, c_j \in]0,1[$

5.2.3.3. Sample size and stratification

The household survey sampling size is commonly a trade-off between budget constraints and technical needs for survey accuracy. Our study will not address computation of an optimal sample size. DHS or MIS survey are usually stratified using the second administrative level (regional or provincial). Allocation of clusters at regional level was proportional to the population size and to their contribution to the overall spatial uncertainty. The selection of cluster locations is achieved within each region.

5.2.3.4. Selection of cluster locations

Cluster locations identification can be performed using random or deterministic methods. Our choice was oriented towards random selection, an inclusion indicator was designed, and it involves target objectives and the spatial uncertainty information. The uncertainty can be calculated with the variance or the coefficient of variation. A simple mathematical formula for cluster inclusion weight was: $W(s_i) = \sum_{j=1}^{p} a(s_i) I_{T_j}(s_i)$, where p is the total number of targets, a Dirac function $I_{T_j}(s_i) = \begin{cases} 1 & if\ s_i \epsilon T_j \\ 0 & else \end{cases}$, and $a(s_i)$ represents cluster relative contribution to the regional uncertainty, obviously constraints by $\sum_{i=1}^{n} a(s_i) = 1$.

5.2.3.5. Inhibition of clusters

Persons living within the same areas may share similar infection risk. To consider spatial redundancy, a minimum distance called radius *h* (in kilometer) between two clusters was fixed. Entomological studies or analysis of the spatial correlation parameter obtained from geostatistical models can be used to estimate the radius *h*. *Diggle et al* (2006) and *Kabaghe et al* (2017) presented some advantages of cluster inhibitions, however they were not able to indicate the optimal distance *h*, probably because it depends on the study field. In malaria disease transmission, we had to considerate the vector abilities and human behaviours (Diggle and Lophaven, 2006; Kabaghe et al., 2017). In our study the distance *h* was fixed at 4

km. In case of violation of distance between two clusters, the one with the lowest spatial uncertainty was excluded.

5.2.4 Algorithm implementation

Detailed steps of our contribution are described as follows:

Step 1: Bayesian geostatistical modelling

1- Choose an appropriate geostatistical model formulation to explain the disease distribution based on the existing set of cluster locations identified with a standard survey sampling (i.e. DHS or MIS sampling methodology).

We considered a non-complex formulation. Let us consider sites surveyed S=$\{s_i$, i $\in$ 1..N$\}$, where $Y(s_i)$ correspond to observed cases tested positive (microspcopic/rapid diagnostic test) in the cluster s_i. In this study, it was assumed that $Y(s_i)$ has arisen from the binomial distribution, i.e. $Y(s_i)|X_{s_i}^k, p_{s_i} \sim Bin(n_{s_i}, p_{s_i})$. At the logit scale, the linear relationship between prevalence (p_{s_i}), climatic and environmental factors $X_{s_i}^k$, and a spatial random effect $U(s_i)$ was defined as logit$(p_{s_i}) = \sum_{k=0}^{p} \beta_k X_{s_i}^k + U(s_i)$. It was assumed that $U(s_i)$ arises from an isotropic and stationary spatial Gaussian process i.e. U = $(U(s_1), ..., U(s_N))^T \sim$MVN$(0_N, \sigma_u^2 \Omega)$, and $\sigma_u^2 \Omega$ is the variance-covariance matrix $N \times N$, such as $Cov\left(U(s_i), U(s_j)\right) = \sigma_u^2 \Omega_{ij} = \sigma_u^2 \exp(-\delta d(s_i, s_j))$, with $d(s_i, s_j)$: Euclidian distance between clusters s_i and s_j, σ_u^2 is the spatial variance, δ the smoothing parameter that controls the correlation.

2- Divide the dataset into two parts that correspond to the training (70-80%) and test datasets (20-30%).

3- Apply a geostatistical variable selection on continuous and categorical forms of climatic and environmental factors to identify the most important climate covariates. To identify the adequate predictors, each predictor was associated with a categorical

indicator I_k that is used to exclude it or identify the proper functional form. Accordingly, $I_k = 0$ when the factor is excluded, $I_k = 1$ when the linear form is selected or $I_k = 2$ to select the categorical. A multinomial distribution prior is assumed for I_k with probability function $\prod_{l=0}^{2} \pi_l{}^{\delta_l(I_k)}$, where π_l is the inclusion probability of each form (l = 0, 1, 2). $\delta_k(.)$ is the Dirac function : $\delta_k(I_i) = \begin{cases} 1 \; if \; I_i = k \\ 0 \; if \; I_i \neq k \end{cases}$. Based on the spike and slab approach, regression coefficients were modified such as $\beta_{k,l}$ that is a mixture of normal prior distribution i.e. $\beta_{k,l} \sim \delta_l(I_k)N(0, \tau_{k,l}^2) + (1 - \delta_l(I_k))N(0, c\tau_{k,l}^2)$. The inclusion probabilities were drawn from a Dirichlet distribution with hyper-parameter $\alpha = (1,1,1)^T$ is used, then $\pi = (\pi_1, \pi_2, \pi_3)^T \sim Dirichlet(3, \alpha)$ (Ishwaran and Rao, 2005).

4- Fit the final model using the selected predictors and the training dataset and run cross-validation on the test dataset to control the goodness of fit. Perform Bayesian kriging for disease risk prediction at unobserved positions of the gridded surface.

Step 2: Specification of targets and sample size

5- Define relevant rules as described in 5.2.3.2 based on expected results of disease control interventions. For our study the entire country surface was included and clusters with predicted malaria risk under 5% or above 50% were selected.

6- Integrate accuracy needs and budget information related to monitoring and evaluation component to determine the number of cluster locations to be monitored through the project/grant within malaria actions plan. The total number of clusters is estimated by sampling calculation and adjusted by the budget constraint.

7- Stratify the country according to health intervention zones which in Cameroon are correspond to the 10 regions. For each region assigned a cluster number proportional to the contribution on the overall spatial uncertainty and the population size. The Spatial uncertainty was calculated by the spatial variance.

Step 3: Selection of locations

8- Determine the minimum distance (h) based on the spatial range between two locations and compute the inclusion weight indicator for each point of the gridded surface. Within each region, exclude all cluster locations previously observed from the sample. For convenience reason, the distance h is fixed at 4 km.

9- Randomly select a predetermined number of cluster locations by applying probability proportional based on the inclusion weight as described in section 5.2.3.4. The inclusion indicator was calculated as the product of the spatial contribution to uncertainty. Save the selected conform clusters into a set, $B = B \cup \{\text{selected clusters}\}$ (initialiazed as $B = \emptyset$).

10- For each region, if two cluster locations within the set B were less than h km away, remove the one with the lowest spatial uncertainty. Until the expected number of clusters was obtained, the steps from 9 to 10 will be repeated.

5.3. Results

National malaria prevalence among children was estimated at 33% in the 2011 MIS data. Children from rural cluster locations were more exposed to the parasite risk, and the observed disease risks were 43% in rural areas and 19% in urban areas. A disparity was observed on coverage on different regions and ecological areas by cluster locations. Some regions were more represented, because they contain the most urbanized and populated areas of the country as the regions of Centre, Littoral and West.

The country cluster locations are categorised into rural (932) and urban (116 260) areas, it corresponds to 99% in urban areas. Among the 257 selected clusters by our algorithm, only one was in urban area. The regions of Littoral, Centre, East and North were representing 70% of the 257 locations selected.

Table 5. 1: Estimates odds ratio (posterior median and 95% BCI) of the Geostatistical model parameters based on the cluster level model, MIS 2011.

Factor		OR (95% BCI)
NDVI		1.55 (1.12; 2.12)
EVI	<0.21	1
	0.21-0.38	1.90 (1.03; 3.51)
	>0.38	1.25 (0.51 ; 3.02)
DWB	<70m	1
	>=70m	1.82 (1.005 ; 3.45)
Altitude		0.39 (0.26 ; 0.57)
Forest	No	1
	Yes	1.55 (1.002 ; 2.39)
Spatial parameters		Posterior mean (95% BCI)
σ²(variance)		1.81 (1.24 ; 2.92)
φ (range)		1.39 (0.80; 2.63)

Bayesian geostatistical variable selection applied to the cluster level model selected altitude, EVI, DWB, NDVI and presence of forest as the most important predictors of malaria risk. Among the selected predictors that were statistically important, the vegetation index as EVI and NDVI, Forest presence and DWB were positively associated with disease risk, while the altitude was negatively associated (Table 5. **1**). The geostatistical model also identified spatial patterns of the disease risk in the Centre, East, North and Coastal parts, but also an increase of uncertainty in those areas (Figure 5.2).
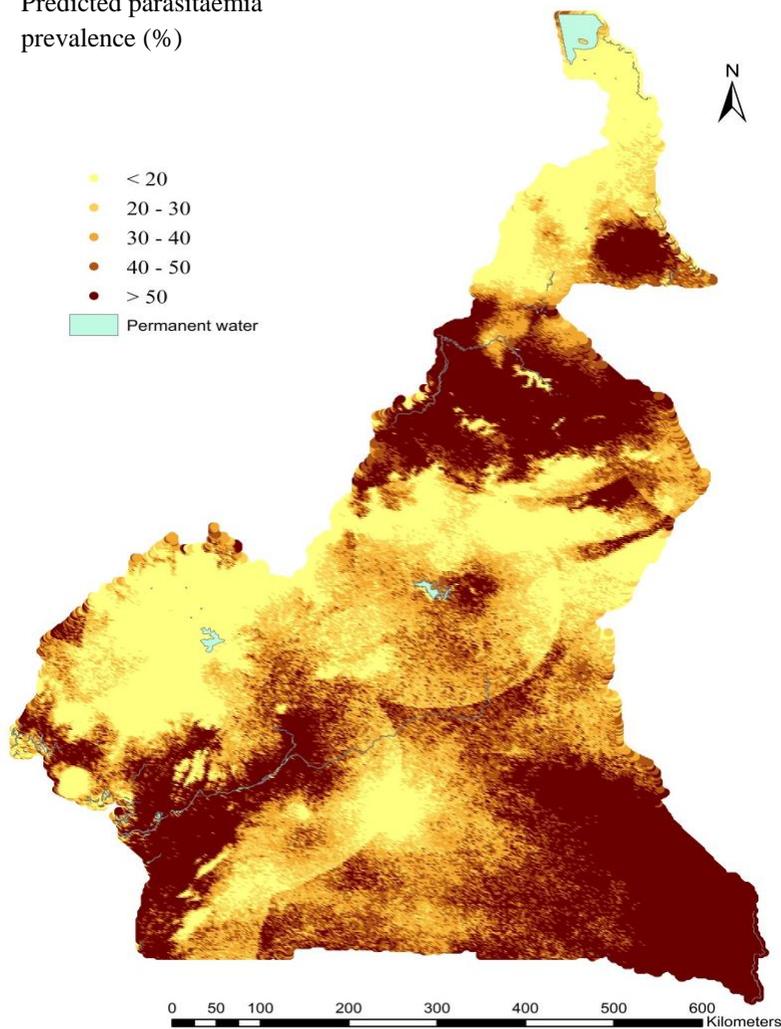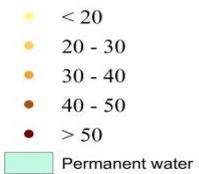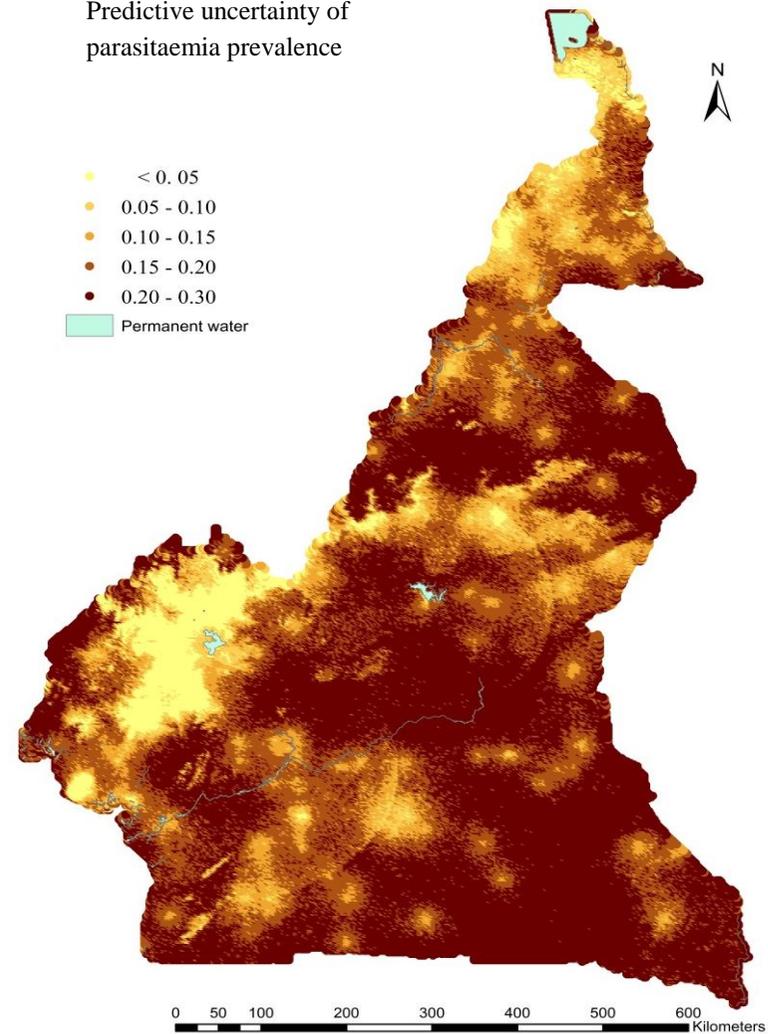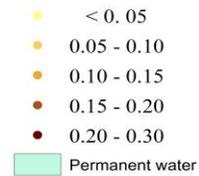
Figure 5.2: Malaria parasite risk estimates among children below 5 years, obtained from MIS 2011 database(left), ii) standard deviation of parasites risk estimates (right).
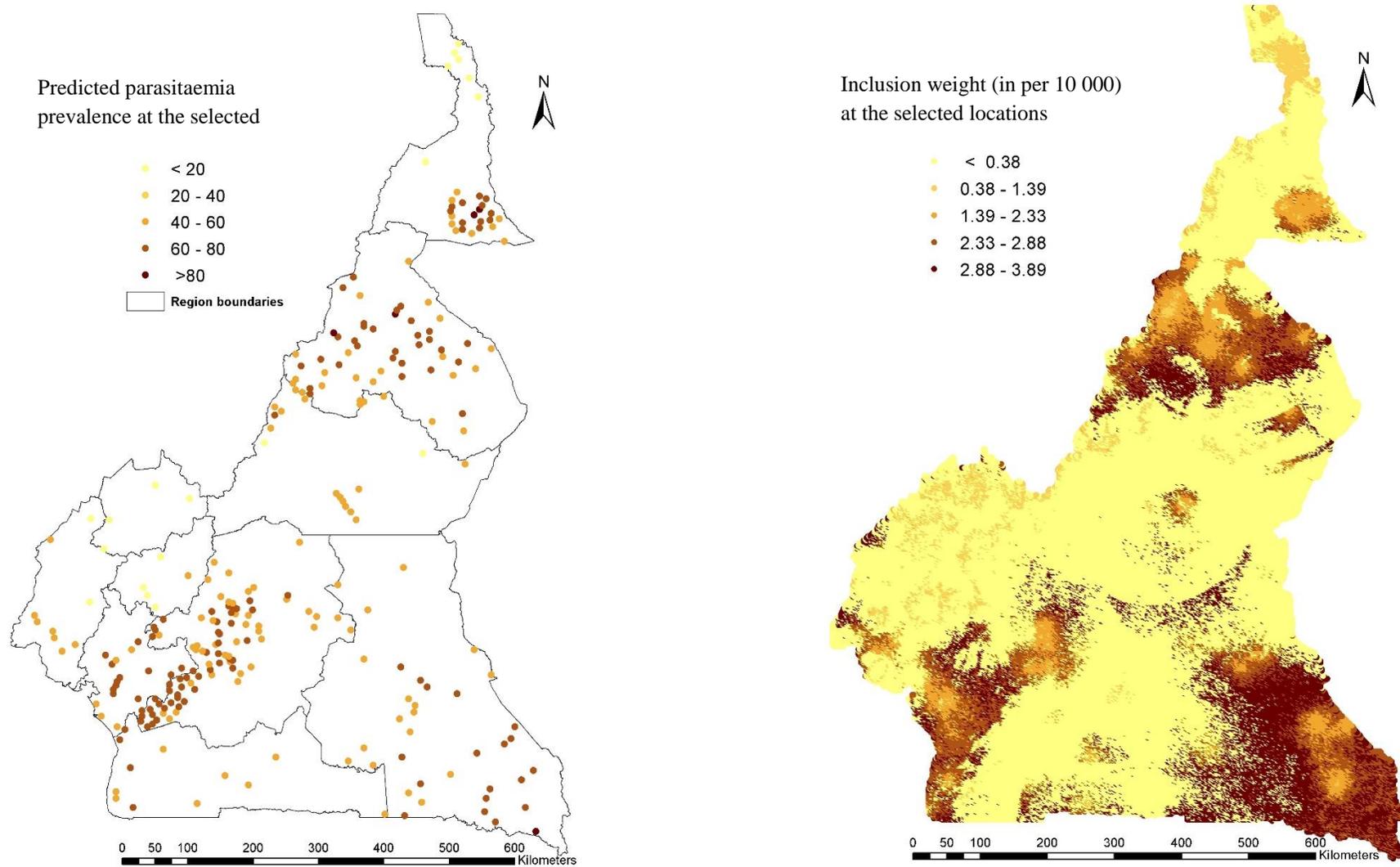
Figure 5.3: Cluster locations selected sample sizes of 257 with the related inhibition distance was fixed at h=4 km (Left). At the right distribution of inclusion weight. Threshold for the targets were fixed at c1=5% and c2=50%.

5.4. Discussion

Compare to the DHS and MIS selection of locations approach, our study has identified additional spatial structures that are suitable for surveyed. The algorithm can contribute to the selection of appropriate clusters for monitoring the disease and strengthening the national health surveillance system and evaluates the effects of intervention through effective selection of cluster locations by applying an adaptive geostatistical sampling design approach (Bousema et al., 2013). Some existing adaptive geostatistical sampling strategies sought to optimize the identification of cluster locations using only spatial uncertainty but did not consider disease control objectives as the identification of clusters in pre-elimination stage, the stratification or population size adjustment as relevant parameters. Our work is therefore an additional tool that may be used in the global fight against malaria and other vector-borne diseases. It is also a contribution to extend the works realized by Diggle et al (2006) and then by Kabaghe et al (2017) (Diggle and Lophaven, 2006; Kabaghe et al., 2017).

Previous works in a Bayesian geostatistical sampling had raised the question of the need for intensive computing when trying to identify efficient locations in a large surface. In the low-resource context such implementation could be a challenge for the disease surveillance team. In addition, they did not address the specific issue of malaria transmission, such as availability of human hosts and disease surveillance objectives. Kabaghe et al (2017) provide an algorithm that focuses on the identification of high-risk disease areas and does not take into account the complex situation of countries in which different heterogeneous ecosystem spaces coexist (Kabaghe et al., 2017) . In countries where the geographical distribution of disease risk is disparate, surveillance of areas for which the risk of disease is considered low is equally relevant in the framework of malaria elimination.

The cluster locations identified by the algorithm can be useful to improve our knowledge on the additional factors that influence the distribution of disease risk in Cameroon. The Littoral, Centre, East and North regions were characterized by areas with larger spatial variations. In these regions an important number of cluster locations was identified (Figure 5.3). This result was expected. However, when increasing the total number of clusters to be monitored at the national level, the selected locations were regularly distributed over the surface in the North region and presented a similar spatial distribution with those of the 2011 Cameroon MIS. The locations identified in the Centre region were more numerous and located in the rural areas and in the borders with Littoral region. Those areas were less represented in the MIS data due to the presence of bigger city of Yaoundé. Eastern borders with Central African Republic seem also to be preferred. Increasing of sample size revealed the need of border surveillance with Nigeria, particularly in South West and in North regions where patterns were identified.

Our algorithm has selected very few cluster samples in the West region. Sites observed from MIS were spatially close and had a relatively low parasite risk estimates compared to other regions. West region was known to record lower temperatures and highland areas predominated. These environmental configurations are unfavorable for malaria transmission. The distribution of spatial variance obtained using Bayesian kriging appeared high in the boundaries of the West region where few clusters were surveyed. Our results suggested that the algorithm is also slightly influenced by the areas with low parasite risk and high uncertainty as well as by those that have not been surveyed yet.

Different samples were identified based on our Bayesian geostatistical model and the assumptions of isotropic and stationary spatial process made, and the functional form of covariance was based on an exponential squared function. These choices could influence the variable selection of climatic predictors, and the Bayesian kriging prediction at unobserved location as well as the distribution of the spatial variance. Stratification structure was based

142

on regional boundaries, there was no obligation about this choice. Other administrative levels may be more optimal depending on disease control needs.

It has been found that the sample size and inhibition distance were hyperparameters that could modify the algorithm results. A sensitivity analysis based on those variables should be achieved by evaluating the expected spatial variance. However, results will clearly depend on the spatial field involved and the computation can become intensive. The availability of funding was identified as a reasonable way to constraint the sample size. The appropriate number of households to visit within a given cluster location were not analysed in this study. DHS recommendations on the optimal household number per cluster and place of residence could be followed.

In contrast to the standardized sampling design commonly applied in household surveys such as DHS or MIS, in which the cluster locations were stratified according to administrative and urbanization structures. Our method used regional stratification and focused on the malaria objectives and on the contribution to the overall spatial variance. Therefore, when target rules do not guarantee the representativeness of the country space, such sampling design can lead to "preferential sampling" over-sampling of some areas of the country. According to the defined objectives, the sampling locations generated can represent the population at risk of disease. The geostatistical model estimates based on a preferential sampling could be biased. Innovative statistical models are proposed to address this issue (Diggle et al., 2010; Zidek et al., 2014; Giorgi et al., 2018).

5.5 Conclusion

Our algorithm did not compete with the DHS standardized sampling design approach already used. It provides a comprehensive and useful tool that can identify additional areas of relevance for the monitoring of disease risk. It was relatively easy to implement and then may be involved in the identification of sentinel sites and in vector control surveillance. It can be

applied in other diseases that are highly dependent on the environment and could be useful for bridging information gaps in countries. In the context of repeated surveys as DHS and MIS, this algorithm applied iteratively can highlight the cluster areas with the most interesting contribution to the spatial disease risk distribution in the country.

Chapter 6 Malaria seasonality and all-cause mortality risk among infant and under-five in Cameroon.

Salomon G. Massoda Tonye[1,2,3] , Wounang Romain[4], Paul Libite[4], Kouambeng Celestin[3], Magloire Kuetche[3], Penelope Vounatsou[1,2]

1. Swiss Tropical and Public Health Institute, Basel, Switzerland

2. University of Basel, Basel, Switzerland

3. National Malaria Control Program, Yaoundé, Cameroon

4. National Institute of Statistics, Yaoundé, Cameroon

Corresponding author

Email: Penelope.vounatsou@swisstph.ch

Abstract

Introduction

In Cameroon, few studies have assessed the relationship between malaria prevalence and mortality risk in under-five years old children. The findings are inconclusive as some studies were not able to capture a statistically important association. Our study aims to improve our understanding of the malaria-mortality association by taking into account malaria seasonality.

Methods

We analysed mortality data from the multiple indicator cluster survey (MICS) of 2014 in Cameroon. Malaria risk during the dry and rainy season was obtained from the Demographic health (DHS) and the malaria indicators (MIS) surveys of 2011, respectively. Bayesian geostatistical logistic, zero-inflated logistic models including spatially varying covariates were fitted to estimate mortality risk in infants and in children under-five years old. The models considered neonates, socioeconomic confounding, and climatic covariates. The geographical distribution of spatial random effects was employed to identify hotspot areas.

Results

We estimated a statistically important relation between infants (excluding neonates), under-five years old mortality and malaria risk. The effects of malaria parasite risks on under-five mortality became more statistically important in the absence of neonates. In particular, the odds ratio of infant mortality due to malaria changed from 1.76 (95 % BCI: 0.97- 3.16) to 2.86 (95 % BCI: 1.38-5.83) when neonates were removed. Similarly, the odds ratio of mortality children under-five years was modified from 1.57 (95 % BCI 1.03-2.38) to 1.98 (95 % BCI: 1.21-3.21). Mortality in the under-five group was reduced during the dry season. However, this effect was not observed in the group of infants. Child's age, maternal age at birth, birth intervals, and multiple births were the most important determinants. Older children

or children whose mothers aged were from 24 to 35 years or mothers having an interval between births of more than 4 years were less affected. The positive values of spatial random effects were identified in the eastern part of the North and Adamawa regions, as well as in the rural areas of the Center, South-West, and Far-North regions. While negative values of spatial random effects which indicated a lower risk of mortality were estimated around urban cities like Yaoundé and Bafoussam.

Conclusion

In a malaria-endemic country like Cameroon, where the risk of contracting malaria is seasonal, it is important to conduct household surveys in different malaria transmission seasons. The role of neonates was also noted in the relationship between malaria prevalence and under-five mortality. As malaria-endemic countries move towards the elimination step, disease control programmes should contribute to implement a continuous household surveillance during different malaria transmission seasons to allow targeting of zones with high variations of under-five mortality risk.

Keywords: Malaria, Malaria Indicator Survey, Demographic and Health Survey, parasitaemia, Statistically important, Under-five mortality.

6.1 Introduction

Under-five mortality is a global health concern and has been included in the 2030 vision for Sustainable Development Goals (SDG). Intervention indicators were included in the SDG scoreboard to measure countries' progresses in different areas such as health prevention, case management and other socioeconomic sectors (United Nations, 2015). Over the last decade, many African countries have been supported and have succeeded in increasing coverage of interventions, increased their health infrastructure, and facilitated access to treated water and formal education. However, compared to developed countries, the child mortality rate has remained high in many low- and middle-income countries, particularly those with endemic malaria transmission(World Health Organization, 2009; Liu et al., 2015; Gething et al., 2016; Golding et al., 2017; Papaioannou et al., 2019).

In Cameroon, the government, civil society organizations, and international organizations are committed to increase life expectancy and reduce the burden of child mortality, therefore they have conducted important health prevention interventions. As part of the fight against malaria, a first national mass vector control campaign provided eight million long-lasting insecticide-treated mosquito nets in 2012. Malaria case management interventions were also strengthened through the introduction of malaria rapid diagnostic tests (RDTs) and free treatment of children under five years with uncomplicated malaria (Bowen, 2013; The Global Fund, 2006). As a result the overall under-five mortality has been decreased from 157 deaths per 1,000 live births in 1990 to 103 in 2014. The percentage of malaria deaths reported by health facilities also decreased from 62% to 42% over the same period (National Institute of Statistics, 2017). These encouraging downward trends could be attributed to existing health and socioeconomic interventions. Despite these results, the country has remained far below the 2030 SDG minimum under-five mortality target of 25 deaths per 1,000 live births in children under-five years old (United Nations, 2015).

There is a dearth of information on the determinants and geographical distribution of mortality risk as well as on the association between mortality and malaria prevalence in Cameroon. Dwyer-Lindgren et al (2018) identified spatial disparities in under-five mortality at the second administrative level (department), while Golden et al (2017) produced similar work at a high geographical resolution in African countries, including Cameroon (Dwyer-Lindgren et al., 2018; Golding et al., 2017). These studies combined mortality data collected by Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS) at different periods to assess the temporal and spatial distribution of under-five mortality risks in Africa. Malaria risk, socioeconomic and climatic factors were included in geostatistical model to estimate child mortality risks. Papaioannou et al (2019) assessed the relationship between under-five mortality and malaria risk in many African-countries by considering the severity of disease that resulted to improve the understanding and significance of the relation. However, in Cameroon and in other African countries, DHS and MICS were mainly conducted out of the peak malaria transmission season. Therefore, malaria parasite risk was probably underestimated in some areas. The use of data from a single malaria transmission season can contribute to a reduction in the statistical importance of malaria risk in under-five mortality as is the case in some African countries, including Cameroon (Kazembe et al., 2007; Papaioannou et al., 2019). Other studies had assessed risk factors that focused on data from health facilities and a specific group as neonatal or infant populations (Tebeu et al., 2015; Ndombo et al., 2017). However, they were not nationally representative.

In 2011, DHS and malaria indicator survey (MIS) were conducted during the dry and rainy seasons, respectively. In addition, due to the international assessment of the Millennium Development Goals (MDGs), the 2014 Cameroon MICS was carried out to assess progress achieved by the country in different sectors, including among others education, child and female abuse, and disability. Reliable mortality data on children below five years was

collected. This study assessed the effects of malaria transmission season on mortality risk among infant and under-five groups from 2007 to 2011. Bayesian zero-inflated Bernoulli and Bernoulli models including a spatial random effect are applied to the 2014 Cameroon MICS data. Such models have already been used to assess mortality risks in African countries (Gemperli et al., 2004; Musenge et al., 2013).

6.2 Methods

6.2.1 Country setting

Cameroon is a tropical country that belongs to the Central African Economic and Monetary Community, it shares borders with Gabon in the south, Nigeria in the west, Chad in the north and the Central African Republic in the east. Cameroon is also called "Africa in miniature", because the country harbors a wide diversity of the African ecology. Economically, Cameroon is ranked in the lower bound of low- and middle-income countries, its gross national income per capita was estimated at $1,330US in 2015 and there were large disparities between populations in rural and urban areas (The World Bank, 2016). The government is the largest employer, followed by some multinational companies. Less than 20% of the population works in the formal sector which implies claiming a regular monthly salary and a potential health insurance. Access to health care was inequitable and characterized by an unequal geographical distribution of health workers and infrastructures (Tandi et al., 2015). The country's political system is based on democratic concepts, recently introduced in 1990, and much remains to improve in terms of transparency and accountability of the leaders. The country is subdivided into 10 regions, 58 subdivisions and 189 health districts. In 2018, the country population was estimated at 24 million inhabitants with children under five years represented 16% of the population. 51.6% of the population lives in urban areas. The largest urban cities are Douala and Yaoundé, accounting for 20% of the population. Malaria was highly seasonal in the northern part of the country and perennial in the southern part. The

highlands of West region and part of Adamawa regions have low malaria transmission areas where disease risk among children was estimated lower than 5% (Massoda Tonye et al., 2018).

6.2.2 Study data

Malaria parasitological data

DHS-MICS 2011 Survey

A single survey was conducted and combined the DHS and MICS data collection tools. The survey was carried out in Cameroon from January to August 2011. A sample of 580 clusters and 15,050 households were selected using a two-stage sampling approach, 291 clusters were in the urban zone (Figure 6.2). Blood samples were collected from 5,515 children in half of the households inside each of the cluster surveyed. Each child was tested by a Rapid Diagnostic Test (SD BIOLINE Malaria Antigen Pf/Pan). National Institute of Statistics (NIS), Ministry of Health (MoH), ICF International, UNICEF, USAID and other partners were involved in the technical and financial components of these surveys (INS and International ICF, 2012).

MIS 2011 Survey

The Malaria Indicator Survey (MIS) was carried out between September and November 2011, during the malaria high transmission season, one month after the increase of rains in the country. The MIS was conducted on 6.040 households within 257 clusters randomly selected out of the 580 clusters of the DHS 2011 (Figure 6.2). Malaria parasite screening was performed on 4.939 children less than five years old using a Rapid Diagnostic Test (First Malaria Response Antigen). The NIS and MoH were in charge of technical aspects and the funding was provided by the Global Funds against AIDS, Tuberculosis and Malaria. Consent approval was always obtained from adults in charge (Minsante, INS, 2012).

Mortality and socioeconomic and geo-referenced data

Cameroon MICS 2014 Survey

The survey was carried out by the National Institute of Statistics in partnership with the Ministry of Public Health and UNICEF. Data collection was conducted from 03 June to 20 August 2014. The country was stratified into twenty-two (22) strata, corresponding to rural and urban areas of the ten (10) regions, as well as the two largest urban cities of Yaoundé and Douala (Figure 6.1). A total of 477 clusters were randomly selected among the 580 clusters previously identified and surveyed in the 2011 Demographic and Health Survey (DHS). Twenty-four and twenty-one households were interviewed into each urban and rural area, respectively.
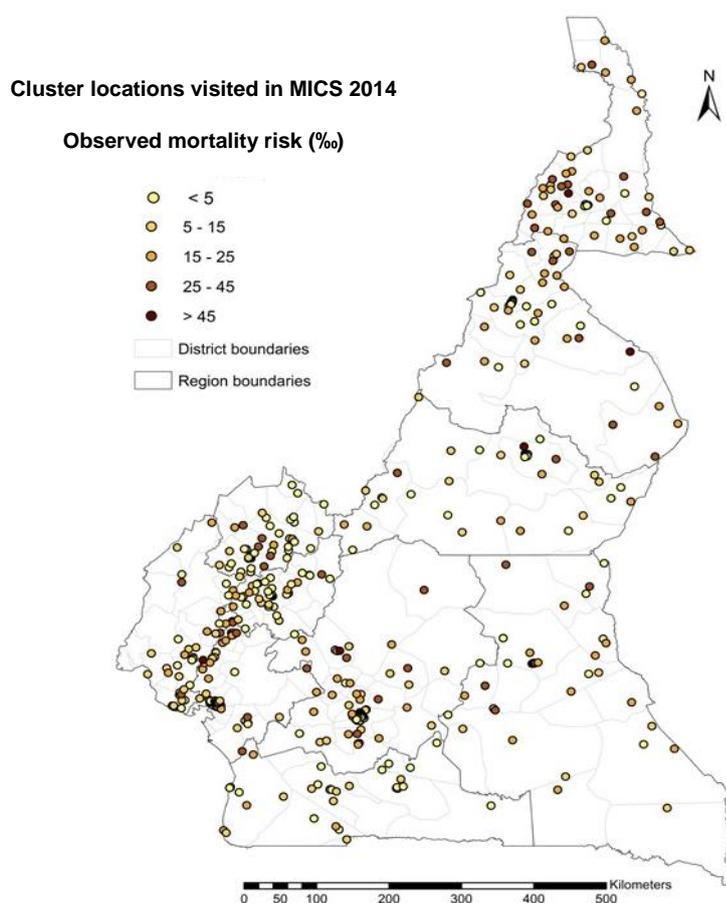
Figure 6.1: Under-five mortality estimates at the 2014 Cameroon (2011-2017) MICS locations.

At the national level, a total of 10259 households were interviewed with a good response rate of 99.6%. Data collected come from 9861 selected women and their 7081 children. Prior to the interview, consent forms were signed by interviewees as required by the national ethics committee. The final database was freely accessible online according to the statistical procedures defined by sponsorship of the global MICS program (https://www.unicef.org/statistics/index_24302.html). Data retrieved from the Cameroon MICS 2014 was hierarchically structured in clusters and regions. For each mother, a retrospective review of births and deaths was performed at the household level based on hospital and administrative data presented during the survey. For each birth, variables about vital status were reported in the birth historical questionnaire in a way that ensures to keep the month and year of death when it occurred during the last decade prior to the survey.

Inclusion criteria

Be a child of less than five years old and be alive or dead between January 2007 and December 2011.

Socioeconomic and confounding indicators

The following socioeconomic variables were considered: mother's level of education, place of residence that corresponds to rural and urban status, administrative region and household wealth index. At the individual level, the confounding indicators retained were mother's age at birth that was split into three, the gender of the child and the mother's level of education categorized into three levels (none, primary, secondary and university) and intervals between births. The household asset index was included as a proxy of socioeconomic status and grouped into three categories representing the poor, middle and rich. Indicators that can approximate the children's conditions of live such as the availability of improved water and sanitation were also used.

Climatic and environmental indicators

Data used to approximate the ground conditions was extracted from satellite remote sensing data. The following indicators were included in our dataset: Land Surface Temperature during the day and night (LSTD, LSTN), Normalized Difference Vegetation Index (NDVI), and Enhanced Vegetation Index (EVI) were used as proxies of the temperature and vegetation coverage on the ground (John Weier and Herring David, 2000; U.S. Geological Survey, 2017). Rainfall estimates (RFE) and altitudes were retrieved from FEWS (or Famine Early Warning Systems Network) and SRTM (or Shuttle Radar Topographic Mission) web sites, respectively. Continuous indicators were extracted for each month of the study period. Those used to predict malaria risks were averaged over one year prior to the MIS and DHS surveys.
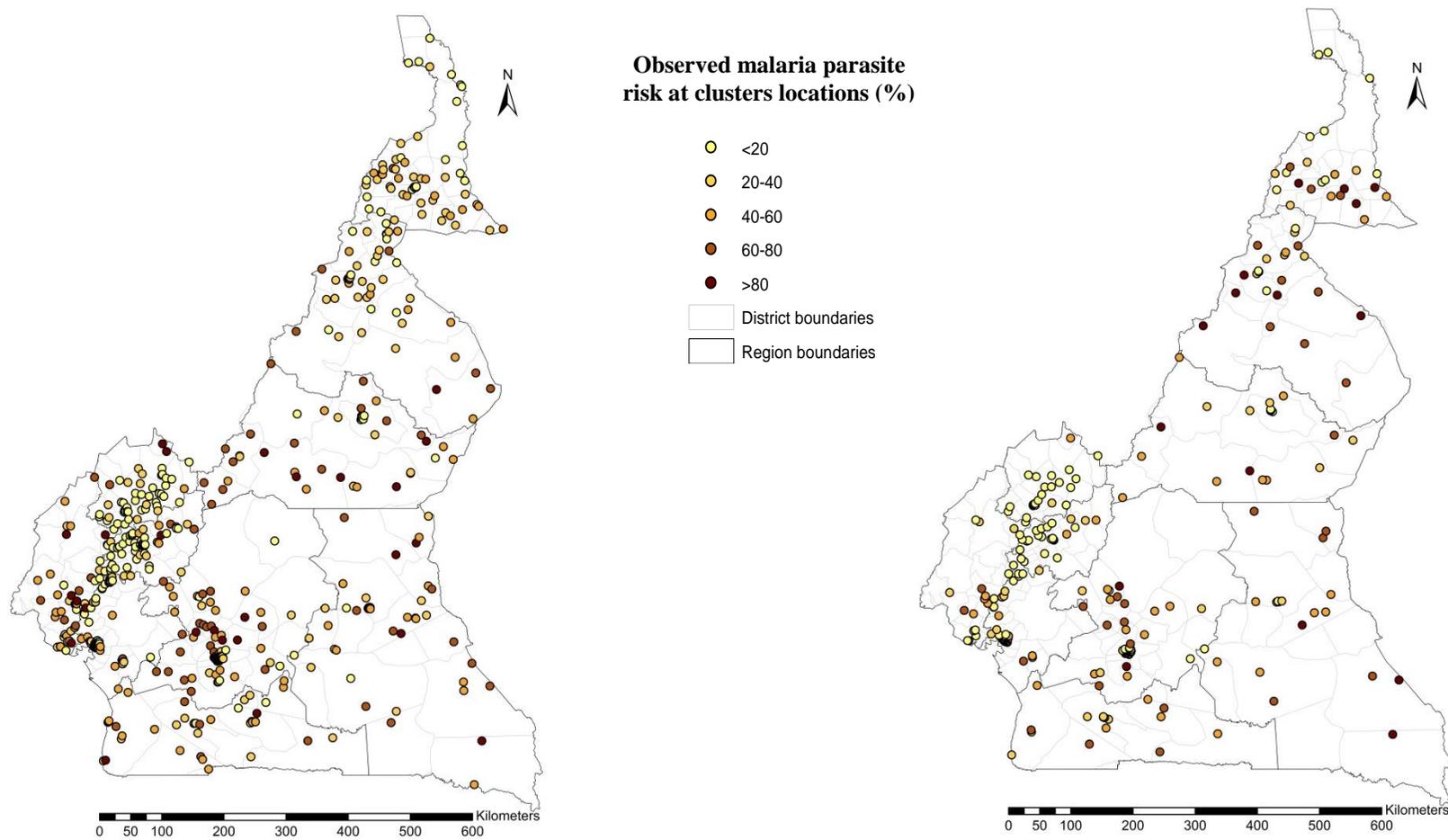
Figure 6.2: Observed malaria parasite risk in children under 5 years at 580 DHS locations (left) and at 257 MIS locations (right).

6.2.3 Statistical analysis

Malaria parasitology results were obtained from individual MIS and DHS data and aggregated at the cluster level. A Bayesian logistic model was applied on MIS data to estimate the relationship between important climatic factors and malaria parasite risk at the cluster level. Bayesian variable selection using a spike and slap approach was applied prior to identifying the most important climatic factors to include into the final model. The Bayesian prediction was performed to predict parasite risk during the rainy season (MIS) at unvisited MICS cluster locations.

We split MICS data into a part of consecutive segment to capture the influence of seasonality and to assign the corresponding malaria risk estimates obtained from DHS and MIS data. It resulted in a longitudinal dataset that contained multiple lines of data on each child until five years old or death occurred. Bayesian logistic models that assumed that vital child status arises from a Zero-inflated Binomial (Bernoulli) were used to fit its relationship with malaria risk. Socio-economic factors commonly used as confounders in literature and insecticide treated-nets ownership were considered in the models. The most important climatic factor was selected using regularized ridge regression and included in the models. The groups of children less than one year as well as those less than five years were analysed by taking into consideration neonatal deaths. Moreover, models that used a Binomial (Bernoulli) distribution were also fitted and the parameter estimates of each group were compared. A random spatial random effect was always included as a frailty effect in our models.

Spatially varying coefficient models were fitted to identify hotspot areas where season influenced mortality risks among infant and under-five. A factor was defined as statistically important when its corresponding Bayesian Credible Interval (BCI) at log scale did not contain the value one. Bayesian inference was performed through the integrated nested Laplace approximation (INLA) and stochastic partial differential equations (SPDE)

approaches (Elias T. Krainski et al, 2018). A sensitive analysis was realized and we used the package R-INLA (Lindgren and Rue, 2015). Maps representing the different distribution of spatial random effects were drawn using the statistical software R (R Core Team, 2016).

6.3 Results

Under-five mortality was estimated at 103 deaths per 1000 live births in the Cameroon by the MICS 2014. Males were more affected, and their corresponding death risk was estimated at 70 deaths per 1000 live births, while females recorded 61 deaths per 1000 live births. The risk of death occurring in the first year was calculated at 60 deaths per 1000 live births and this reduced in children aged between one and five years to 46 deaths per 1000 live births. Infant and under-five mortalities were higher in rural areas compared to urban. The under-five mortality rate was estimated at 103 deaths per 1000 live births in rural areas while it was 74 deaths per 1000 in urban areas. On the other hand, the educational level of the mother and socio-economic status of households contributed to mortality risk. For persons who attended secondary school, under-five mortality rate was estimated at 26 deaths per 1000 live births. This got to 39 deaths per 1000 live births in those without any formal education. Among households with the highest quintile of wealth index, the under-five mortality rate in less than five years was around 29 deaths per 1000 live births while in those with the lowest quintile of wealth index was 39 deaths per 1000 live births. Mortality data collected by MICS showed that most of the death cases among under-five occurred during the dry season, however a decline was observed over time (National Institute of Statistics, 2017).
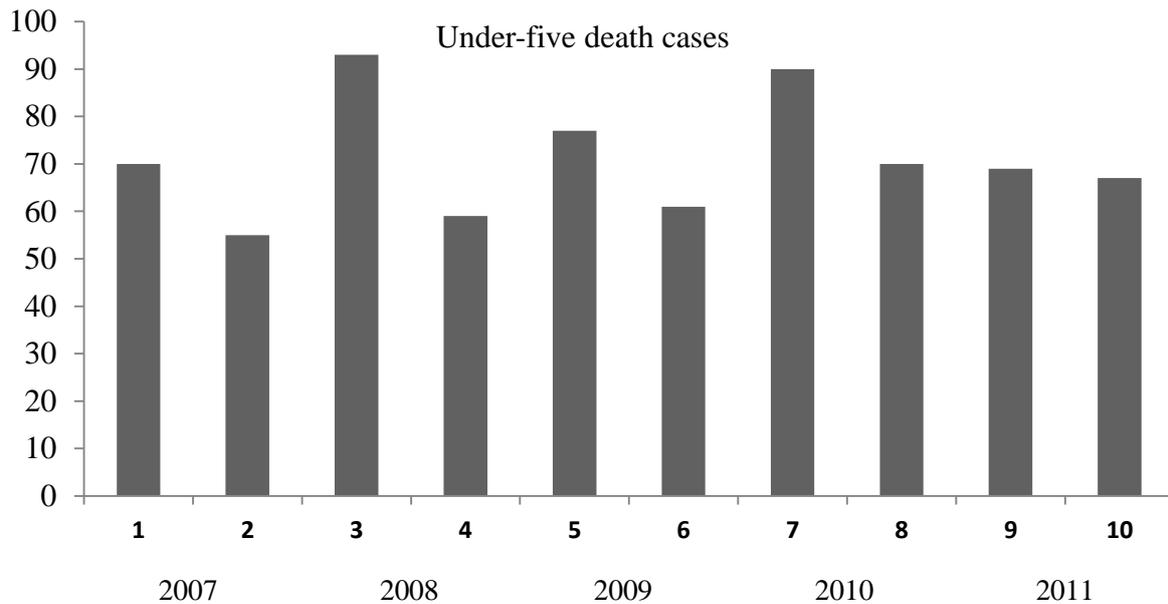
Figure 6.3: All death on children under-five by malaria transmission seasons from 2007 to 2011(odd correspond to dry and even to wet).

The geostatistical variable selection identified NDVI and altitude as the most important predictors of parasitaemia risk, using the cluster level model during the dry season (DHS) while NDVI, the categorical forms of EVI and of distance to water, the presence of forest and altitude were selected as the most important predictors of parasitaemia risk during the rainy season (DHS). Regularised ridge regression selected the Land surface temperature day as the most important climatic to include in the association between under-five mortality and malaria prevalence.

Models showed that the dry season was statistically important and increased the under-five mortality, however this effect was not observed in infants. Malaria risk was positively associated to the mortality risk among children under five years old. However, for the infants (including neonates), the effect of malaria risk was not statistically associated to the mortality (Table 6.1, Table 6.2).

Socioeconomic and demographic factors were statistically important in their relationship with mortality risk for the two groups: the interval between births, multiple births, age, gender, and

mother's age at birth. An odds ratio gradient was observed in the child's age categories, where mortality was lower in older children. Twin children had high risks of death and there was a gender effect, the males are slightly disadvantaged compare to females.

Maternal age at birth, birth intervals, and multiple births were the most important determinants of mortality. Older children or children whose mothers aged were from 24 to 35 years or mothers having an interval between births of more than 4 years were less affected.

The spatial random effect hotspots were identified in the eastern part of the North and Adamawa regions, as well as in the rural areas of the Center, South-West, and Far-North regions. While negative values of spatial random effects which indicated a lower risk of mortality were estimated around urban cities like Yaoundé and Bafoussam (Figure 6.4, Figure 6.5, Figure 6.6, Figure 6.7).

Estimates obtained using the Zero-inflated Bernoulli model were close to Bernoulli regressions and their DIC were similar, their spatial variances were comparable while the spatial range estimates from Bernoulli models were smaller in the group of under-five years. (In appendix, Table 6.4 and Table 6.5). In all ages, the autoregressive effects were not statistically important (Table 6.3). Spatial random effect patterns identified by the Zero-Inflated Bernoulli regressions in the infant and less than five years were similar to the ones captured by the Bernoulli regressions. The spatial random effect maps of Zero-inflated Bernoulli models are in appendix (Figures 6.10-6.13).

Table 6.1: Bayesian geostatistical model-based estimates (posterior median and 95% BCI) obtained from infant mortality risk analysis with and without the neonates.

| | Infants (including neonates) | | Infants (without neonates) | |
| | Bernoulli | | Bernoulli | |
Parameters	Median	OR (95%BCI)	Median	OR(95%BCI)
Malaria risk	1.76	(0.97, 3.16)	2.86*	(1.38, 5.83)
Age groups				
< 6 months	1		1	
others	0.51*	(0.41, 0.63)	1.67*	(1.26, 2.22)
Gender				
Male	1		1	
Female	0.78*	(0.64, 0.96)	0.8	(0.6, 1.05)
Mother age at birth				
<24	1		1	
24-35	0.53*	(0.37, 0.76)	0.59*	(0.36, 0.95)
>35	0.77	(0.47, 1.25)	0.89	(0.46, 1.73)
Place of residence				
Rural	1		1	
Urban	0.91	(0.64, 1.27)	0.94	(0.59, 1.47)
Preceding birth interval				
Firstborn	1		1	
< 2 years	1.76*	(1.28, 2.44)	1.78*	(1.13, 2.82)
2 years	0.8	(0.57, 1.14)	1.08	(0.67, 1.74)
3 years	0.55*	(0.34, 0.86)	0.68	(0.35, 1.26)
=> 4 years	0.59*	(0.38, 0.9)	0.68	(0.36, 1.24)
Socio ecomomic status				
Poor	1		1	
Middle	0.87	(0.64, 1.18)	0.72	(0.47, 1.08)
Rich	0.8	(0.54, 1.18)	0.74	(0.43, 1.26)
Mother education				
No formal education	1		1	
Primary	0.81	(0.63, 1.04)	0.76	(0.55, 1.05)
Above secondary	0.81	(0.59, 1.13)	0.62	(0.39, 0.96)
Improved sanitation				
No	1		1	
Yes	1.01	(0.62, 1.58)	1.03	(0.5, 1.95)
Improved water source				
No	1		1	
Yes	0.89	(0.66, 1.19)	1.03	(0.68, 1.54)
Multiple births				
Twin	1		1	
Single	0.2*	(0.14, 0.27)	0.55	(0.3, 1.12)
Time				

Parameters	Infants (including neonates) Bernoulli		Infants (without neonates) Bernoulli	
	Median	OR (95%BCI)	Median	OR(95%BCI)
2007	1		1	
2008	1.13	(0.82, 1.55)	1.12	(0.72, 1.75)
2009	0.93	(0.67, 1.3)	0.89	(0.56, 1.42)
2010	1.07	(0.78, 1.48)	0.99	(0.63, 1.56)
2011	1.01	(0.72, 1.41)	1.12	(0.72, 1.76)
Season				
Rainy	1		1	
Dry	1.01	(0.8, 1.28)	0.93	(0.67, 1.28)
Mother's year of birth				
<1980	1		1	
1980-1985	0.83	(0.62, 1.12)	0.98	(0.65, 1.48)
>1985	0.6*	(0.42, 0.85)	0.76	(0.46, 1.22)
Latitude	1.13*	(1.03, 1.26)	1.16*	(1.02, 1.31)
Longitude	0.98	(0.86, 1.1)	0.91	(0.79, 1.03)
Land surface temperature day	0.99	(0.96, 1.02)	0.99	(0.95, 1.04)
Proportion households with at least one ITN	0.61	(0.3, 1.28)	0.42	(0.17, 1.07)
Spatial parameters				
Varince	0.18	(0.04, 1.87)	0.62	(0.04, 31.12)
Range	0.73	(0.17, 6.89)	1.32	(0.11, 31.42)

*: Statistically important
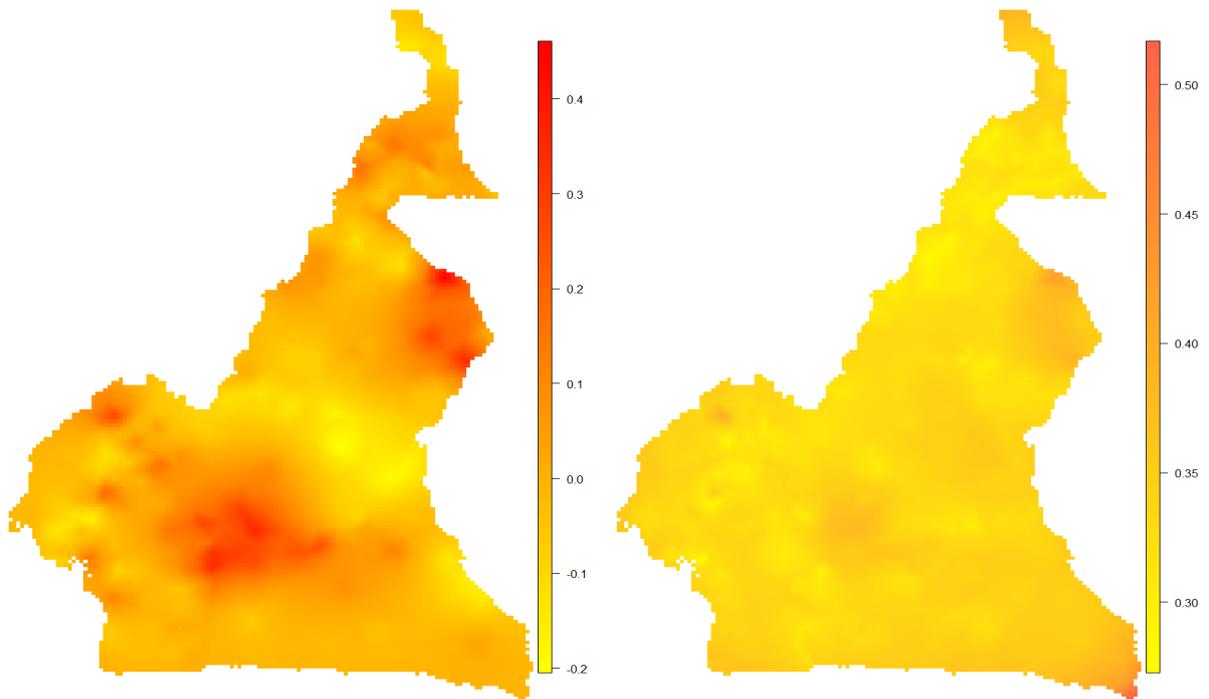
Figure 6.4: Posterior point estimates of spatial random effect (left) and its standard deviation (right) of Bernoulli model on children less than one-year-old group (including neonates).
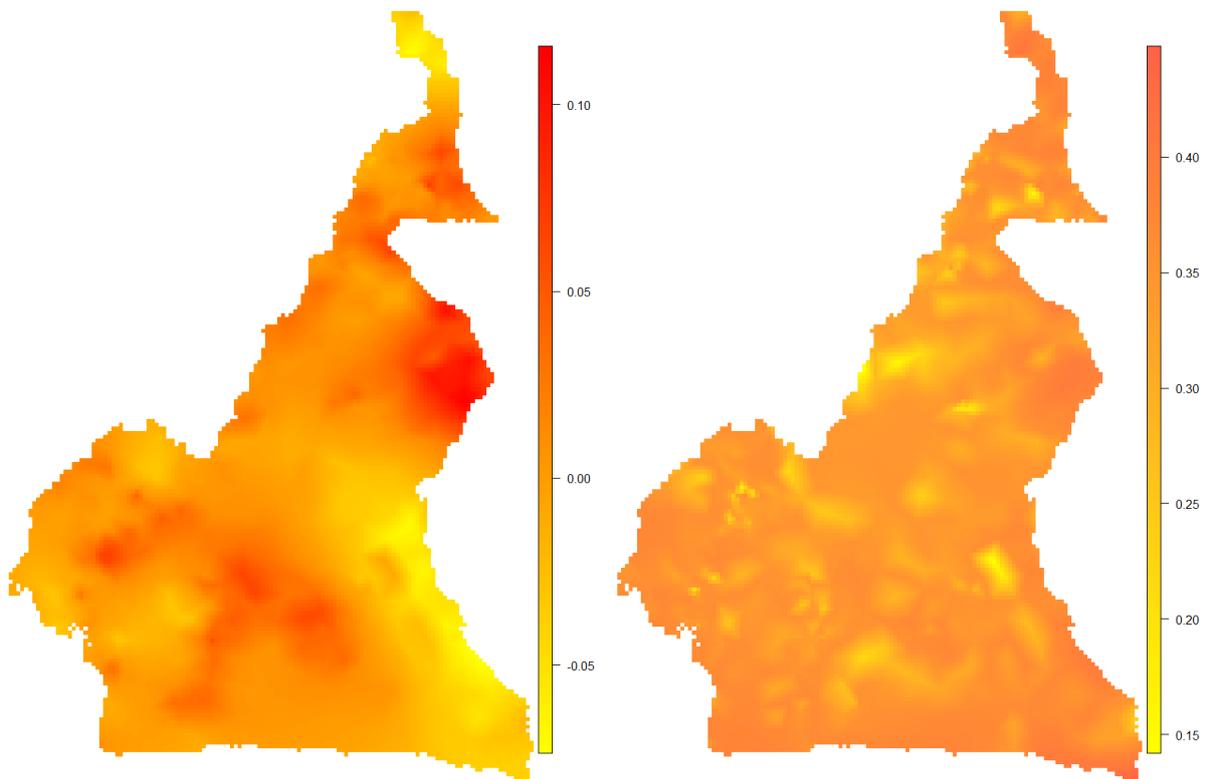
Figure 6.5: Posterior point estimates of spatial random effect (left) and its standard deviation (right) of Bernoulli model on children less than one-year group (without neonates).

Table 6.2 : Bayesian geostatistical model-based estimates (posterior median and 95% BCI) obtained from under-five mortality risk analysis with and without the neonates.

Parameters	Infants (including neonates) Bernoulli		Infants without neonates) Bernoulli	
	Median	OR(95%BCI)	Median	OR(95%BCI)
Malaria risk	1.57*	(1.03, 2.38)	1.98*	(1.21, 3.21)
Age groups				
< 6 months	1		1	
6 months-1 year	0.53*	(0.42, 0.65)	1.77*	(1.34, 2.35)
1- 2 years	0.31*	(0.25, 0.38)	1.05	(0.8, 1.39)
2-3 years	0.21*	(0.16, 0.26)	0.71*	(0.52, 0.96)
3-4 years	0.12*	(0.09, 0.16)	0.42*	(0.29, 0.6)
4-5 years	0.06*	(0.04, 0.09)	0.23*	(0.15, 0.34)
Gender				
Male	1		1	
Female	0.86	(0.74, 1)	0.9	(0.76, 1.07)
Mother age at birth				
<24	1		1	
24-35	0.72*	(0.55, 0.94)	0.87	(0.64, 1.19)
>35	1.15	(0.8, 1.66)	1.49	(0.97, 2.28)
Place of residence				
Rural	1		1	
Urban	0.8	(0.62, 1.03)	0.77	(0.57, 1.04)
Preceding birth interval				
Firstborn	1		1	
< 2 years	1.66*	(1.3, 2.12)	1.63*	(1.22, 2.17)
2 years	1	(0.77, 1.29)	1.2	(0.89, 1.61)
3 years	0.71*	(0.51, 0.98)	0.83	(0.57, 1.21)
=> 4 years	0.59*	(0.42, 0.82)	0.62*	(0.41, 0.92)
Socio ecomomic status				
Poor	1		1	
Middle	0.85	(0.68, 1.07)	0.79	(0.61, 1.03)
Rich	0.81	(0.6, 1.08)	0.8	(0.56, 1.12)
Mother education				
No formal education	1		1	
Primary	0.84	(0.7, 1.02)	0.85	(0.69, 1.06)
Above secondary	0.78	(0.61, 1.004)	0.7	(0.52, 0.941)
Improved sanitation				
No	1		1	
Yes	0.72*	(0.47, 1.06)	0.58*	(0.32, 0.97)
Improved water source				
No	1		1	
Yes	0.85	(0.68, 1.07)	0.89	(0.68, 1.16)

Parameters	Infants (including neonates) Bernoulli		Infants without neonates) Bernoulli	
	Median	OR(95%BCI)	Median	OR(95%BCI)
Multiple births				
Twin	1		1	
Single	0.28*	(0.22, 0.37)	0.54*	(0.37, 0.84)
Time				
2007	1		1	
2008	1.14	(0.9, 1.46)	1.13	(0.85, 1.5)
2009	1	(0.78, 1.28)	0.99	(0.74, 1.32)
2010	1.12	(0.88, 1.43)	1.08	(0.81, 1.43)
2011	0.94	(0.73, 1.21)	0.94	(0.7, 1.25)
Season				
Rainy	1		1	
Dry	1.23*	(1.03, 1.46)	1.27*	(1.04, 1.56)
Mother's year of birth				
<1980	1		1	
1980-1985	0.95	(0.77, 1.18)	1.07	(0.83, 1.37)
>1985	0.88	(0.67, 1.14)	1.14	(0.83, 1.55)
Latitude	1.07*	(1.001, 1.14)	1.07	(0.98, 1.15)
Longitude	1.01	(0.94, 1.09)	0.98	(0.91, 1.07)
Land surface temperature day	0.99	(0.97, 1.02)	0.99	(0.97, 1.03)
Proportion households with at least one ITN	0.8	(0.47, 1.36)	0.79	(0.42, 1.48)
Zero inflated parameter				
Spatial parameters				
Variance	0.25	(0.04, 2.85)	0.14	(0.09, 0.28)
Range	0.56	(0.12, 3.92)	0.37	(0.23, 0.53)

*: Statistically important
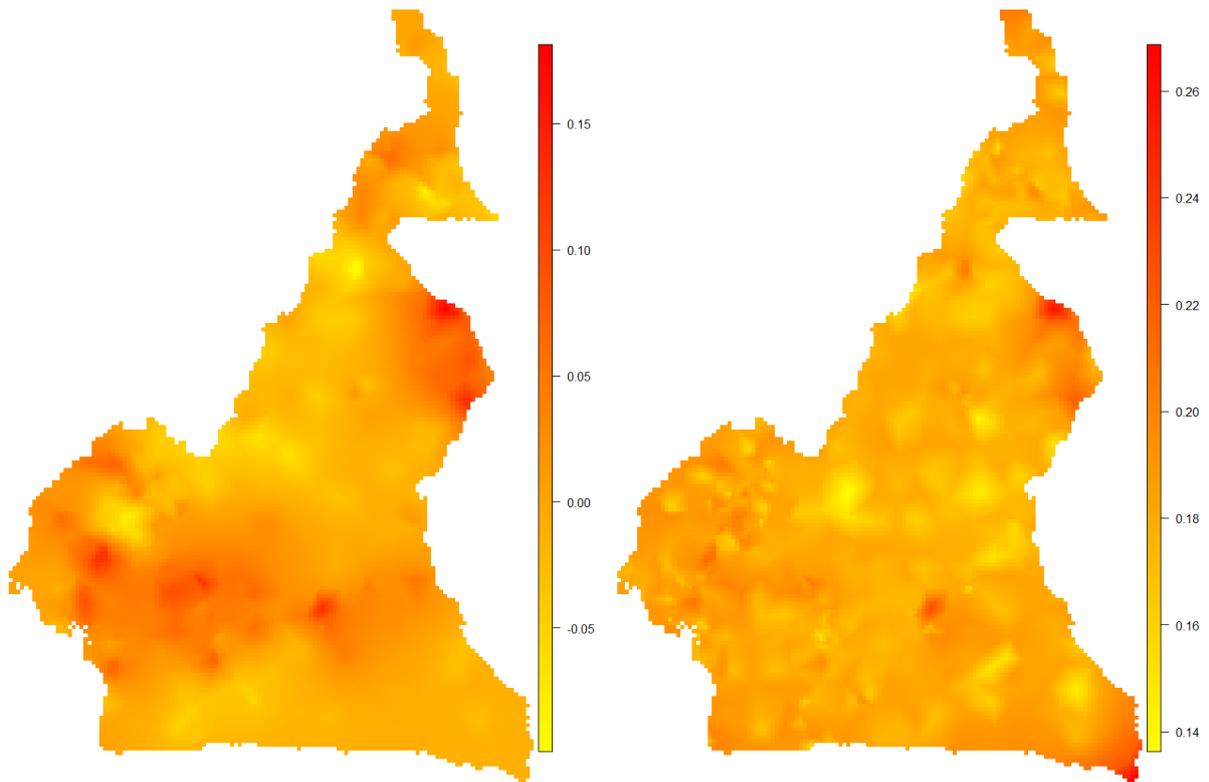
Figure 6.6 : Posterior point estimates of spatial random effect (left) and its standard deviation (right) of Bernoulli model on children less than five years old group (including neonates).
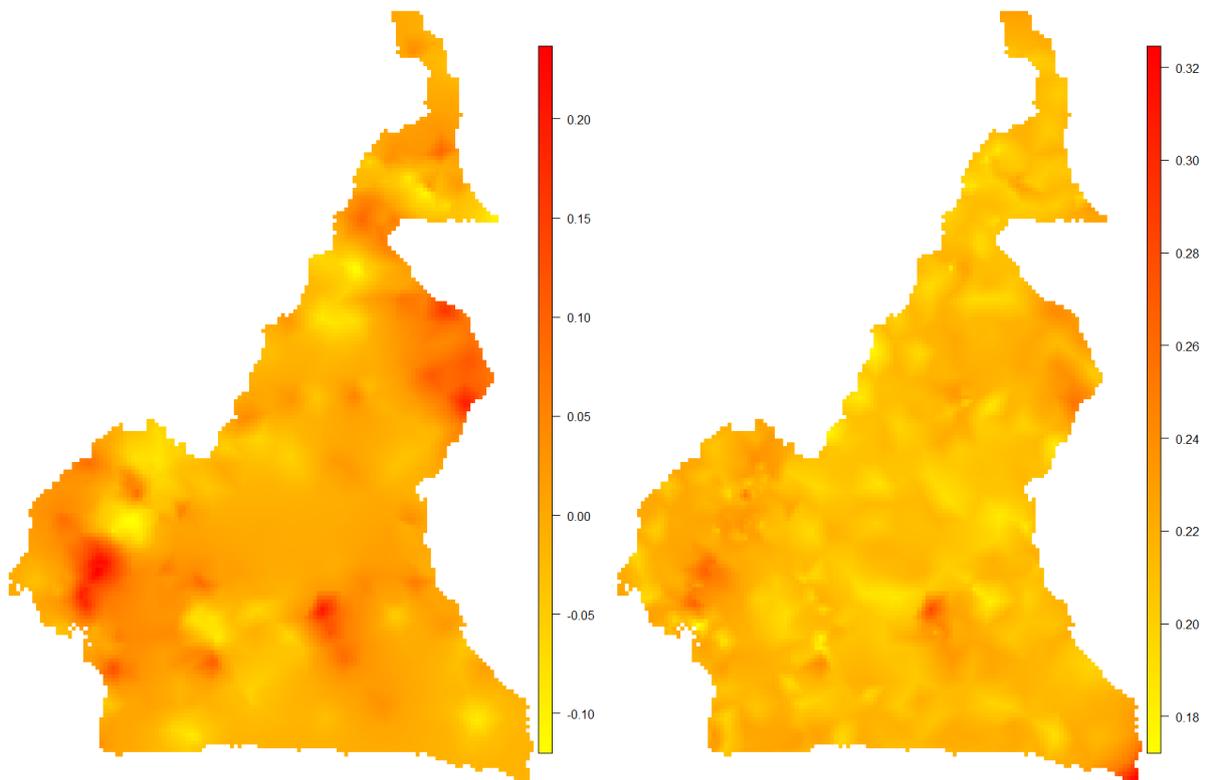
Figure 6.7: Posterior point estimates of spatial random effect (left) and its standard deviation (right) of the Bernoulli model on children less than five years old group (without neonates).

Table 6.3: Bayesian geostatistical model-based estimates (posterior median and 95% BCI) obtained from infants and under-five mortality risk analysis without the neonates.

Parameters	Infants (without neonates) Bernoulli		Under five years (without neonates) Bernoulli	
	Median	OR (95 % BCI)	Median	OR (95 % BCI)
Malaria risk	2.92*	(1.39, 6.13)	1.96*	(1.16, 3.29)
Age groups				
< 6 months	1		1	
6 months-1 year	1.67*	(1.26, 2.23)	1.76*	(1.33, 2.34)
1- 2 years			1.05	(0.8, 1.39)
2-3 years			0.71*	(0.53, 0.97)
3-4 years			0.42*	(0.29, 0.61)
4-5 years			0.18*	(0.1, 0.29)
Gender				
Male	1		1	
Female	0.8	(0.6, 1.05)	0.91	(0.76, 1.09)
Mother age at birth				
<24	1		1	
24-35	0.59	(0.36, 0.95)	0.86	(0.62, 1.17)
>35	0.9	(0.46, 1.75)	1.48	(0.95, 2.28)
Place of residence				
Rural	1		1	
Urban	0.94	(0.59, 1.48)	0.71*	(0.51, 0.97)
Preeceding birth interval				
Firstborn	1		1	
< 2 years	1.78*	(1.13, 2.83)	1.65*	(1.23, 2.22)
2 years	1.08	(0.67, 1.74)	1.26	(0.94, 1.71)
3 years	0.68	(0.35, 1.26)	0.89	(0.6, 1.29)
=> 4 years	0.68	(0.36, 1.24)	0.64*	(0.42, 0.95)
Socio ecomomic status				
Poor	1		1	
Middle	0.72	(0.47, 1.09)	0.79	(0.6, 1.04)
Rich	0.75	(0.43, 1.27)	0.88	(0.62, 1.25)
Mother education				
No formal education	1		1	
Primary	0.76	(0.55, 1.06)	0.86	(0.69, 1.07)
Above secondary	0.62*	(0.39, 0.97)	0.7*	(0.52, 0.95)
Improved sanitation				
No	1		1	
Yes	1.03	(0.5, 1.95)	0.6	(0.33, 1.01)
Improved water source				
No	1		1	

| Parameters	Infants (without neonates) Bernoulli		Under five years (without neonates) Bernoulli	
	Median	OR (95 % BCI)	Median	OR (95 % BCI)
Yes	1.02	(0.67, 1.54)	0.89	(0.67, 1.17)
Multiple births				
Twin	1		1	
Single	0.55	(0.3, 1.12)	0.54*	(0.36, 0.84)
Time				
2007	1		1	
2008	1.12	(0.72, 1.75)	1.14	(0.86, 1.53)
2009	0.89	(0.56, 1.42)	1	(0.74, 1.34)
2010	0.99	(0.63, 1.56)	1.06	(0.79, 1.42)
2011	1.12	(0.72, 1.76)	0.93	(0.69, 1.26)
Season				
Rainy	1		1	
Dry	0.93	(0.67, 1.28)	1.25*	(1.02, 1.54)
Mother's year of birth				
<1980	1		1	
1980-1985	0.98	(0.65, 1.48)	1.03	(0.79, 1.33)
>1985	0.76	(0.47, 1.22)	1.14	(0.83, 1.56)
latitude	1.16*	(1.02, 1.33)	1.07	(0.99, 1.15)
longitude	0.9	(0.77, 1.04)	0.99	(0.91, 1.08)
Land surface temperature day	0.99	(0.95, 1.04)	0.99	(0.96, 1.02)
Proportion households with at least one ITN	0.42	(0.17, 1.1)	0.68	(0.35, 1.31)
Spatial parameters				
Season varying effect				
Variance	0.92	(0.07, 28.67)	0.15	(0.05, 0.51)
Range	1.33	(0.11, 15.65)	0.45	(0.21, 1.17)
Autoregressive effet				
rho	0.02	(-0.99, 0.99)	-0.03	(-0.98, 0.99)

*: Statistically important
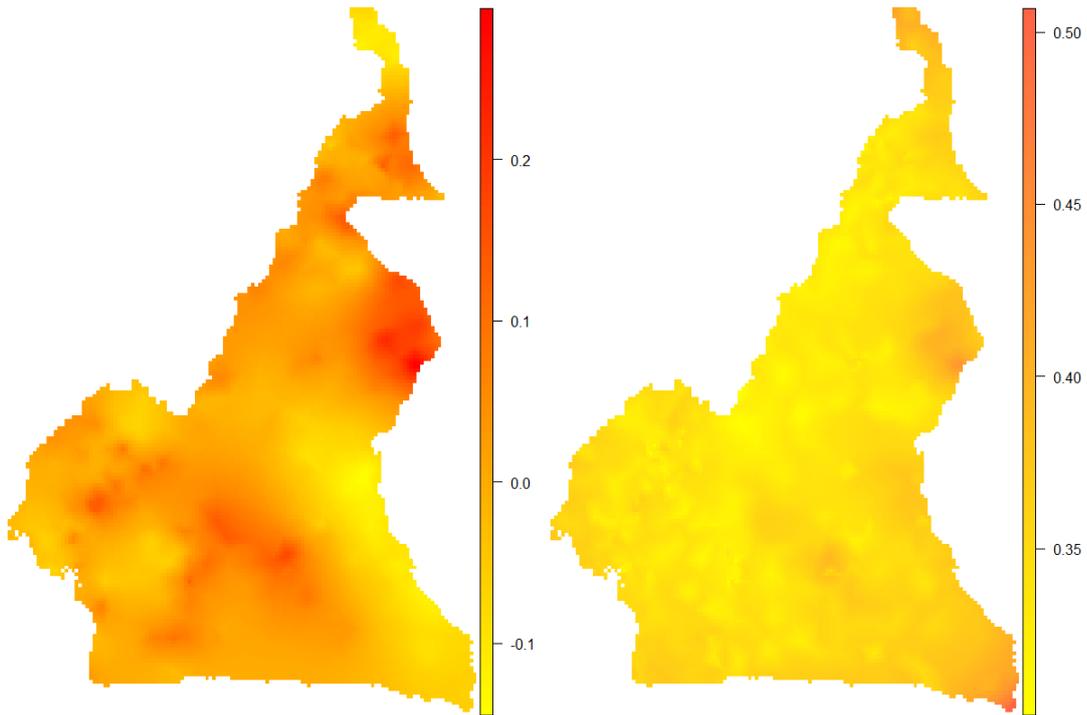
Figure 6.8 : Posterior point estimates of spatial random effects (left) and its standard deviation (right) of the Bernoulli model on children less than one-year old group (without neonates).
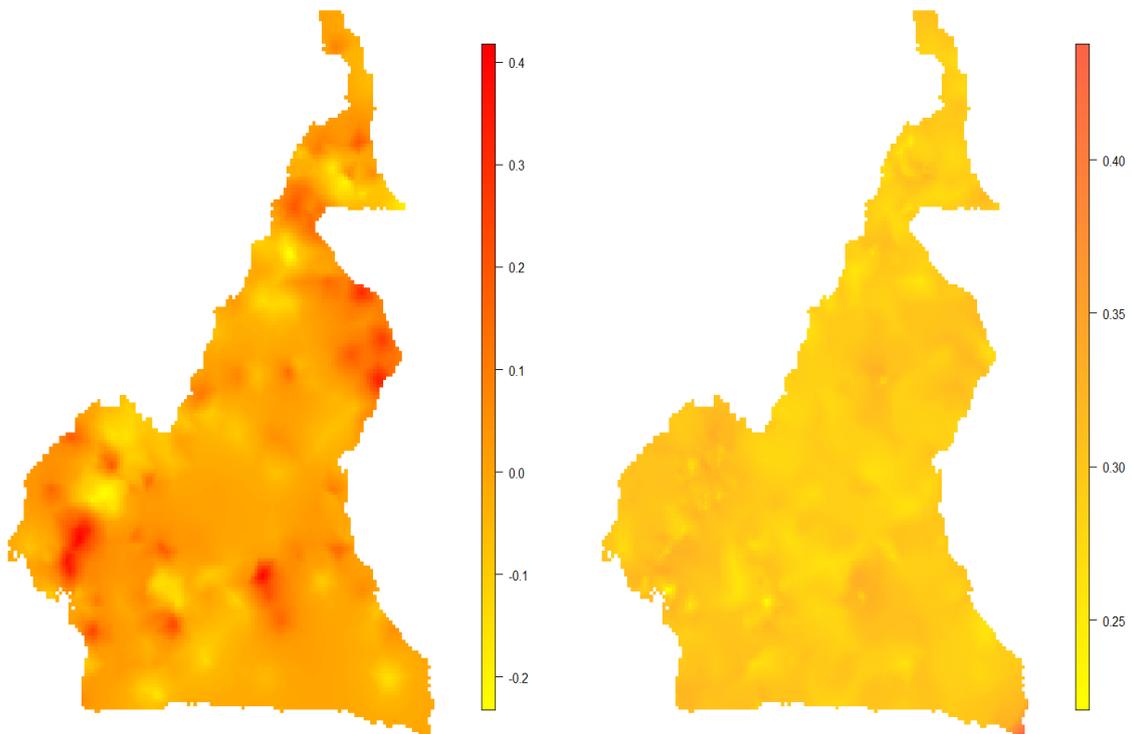
Figure 6.9 : Posterior point estimates of prevalence spatial random effects (left) and its standard deviation (right) of the Bernoulli model on children less than five years old group (without neonates).

6.4 Discussion

Our study was the first of its kind to assess the short-term effects of seasonal variations in malaria prevalence on under-five mortality risks in Cameroon. Our study period ran from January 2007 to December 2011. This preceded the period of intensification of malaria prevention and treatment interventions in Cameroon (Koenker et al., 2018; WHO, 2017c). The 2014 Cameroon MICS mortality data were combined with the estimates of malaria prevalence obtained during the rainy (MIS) and dry seasons (DHS). Bayesian geostatistical models were fitted to estimate the geographical distribution of mortality risk among infants and children under five years old. We also included socioeconomic and demographic covariates as potential confounding factors to adjust for the effects of malaria parasitaemia risk in all our models.

We split individual mortality data into different lines to differentiate dry season from the rainy season of each year. An abundance of zero was observed in the vital status variable and that motivated the use of zero-inflated Bernoulli (ZIB) models. However, in all groups the deviance information criteria and posterior estimates of regression coefficients obtained from the Bernoulli models were close to those produced by the corresponding ZIB models. However the spatial range was smaller in Bernoulli models among under-five year. Thus, the hypothesis of ZIB was not supported by the data.

The effect of malaria risk on under-five mortality was more statistically important when the neonates were excluded from mortality data. These results could confirm the relatively low contribution of malaria on neonates in Cameroon. This is consistent with the results of some cohort studies that identified complications like infections, respiratory distress and asphyxia occurring during pre-term birth as major causes of neonatal mortality. Spatial random effects among under-five year identified hotspots in the North and the Center regions that were similar to those captured by the work of Papaioanou et al (2019). However they showed the

influence of the combination malaria-anemia on under-five mortality in sub-Saharan Africa and conclude that the malaria prevalence alone was not able to capture a statistically important relationship with under-five mortality (Liu et al., 2015; Ndombo et al., 2017; Papaioannou et al., 2019).

Compared to first births, those with three-year intervals between births were less exposed to early mortality. In contrast, the risk of mortality increased among those with less than two-year intervals between births. This result could support the influence of family planning and child spacing in reducing mortality risks. Children whose mothers were aged from 25 to 34 years old were less likely to die than the others. A gradient of odds ratio was also observed on the mother's year of birth. Mortality risks seemed to decline gradually from older mothers to the new generations, and it was only statistically important among infants. This result could be due to short-term period used in our analysis. Mothers' education level was statistically important only without neonates, and regardless of age groups the socioeconomic status was not statistically important. However, a decreasing gradient of odds ratio was observed in categories of socioeconomic status among children under-five (Table 6.1, Table 6.2 and Table 6.3).

Improved sanitation was statistically important in the under-five group (excluding the neonates). It was of borderline importance for the group of infants. This result indicates the contribution of infections associated with sanitation in mortality risk in children aged one to five years. Having safe or unsafe drinking water was of no statistical importance.

The mortality risk of children aged less than five years was higher during the dry season. In the Sahel zones, a substantial reduction of water and food is observed in the dry season, at the same time, an increase of dust winds is occurring. These weather modifications are favorable for malnutrition, anemia, diarrhea, respiratory infections and meningitis among children (Einterz and Bates, 2011; Agier et al., 2013; Libwea et al., 2019). This could explain the high

seasonal variation captured by the spatially varying coefficient model among infants in the northern regions compared to others. A cross-sectional study showed a synergistic relationship between malaria, anemia and malnutrition in the northern part of Cameroon (Sakwe et al., 2019).

The place of residence, household socioeconomic status and improved quality of water were of borderline statistical importance in children below five years. Although children living in urban areas as well as those who benefited from improved water were less likely than others to die prematurely. In developing countries like Cameroon, access to health care remains a challenge to many households due to financial and socio-cultural barriers, and people living in rural areas were more affected (Horowitz et al., 2000; Saeed et al., 2016). The relative decrease of odds ratio in socioeconomic status could also partly explain this situation.

The percentage of households with at least one ITN and the land surface temperature during the day were not of statistical importance. From 2007 to 2011, malaria prevention interventions were focused on the free distribution of ITN to pregnant women and to children under five. This result suggests a limit of this strategy to reduce the under-five mortality rate at the national level.

Spatial variations were captured using a Gaussian process on cluster-level random effects. Surfaces having high or low spatial variations were identified among infant, and under-five children in the East region parts of North and Adamawa regions that border the Central African Republic. Additional patterns were noticed in the rural areas of Centre, Far-north, Littoral, South-West, West, and North-West regions. Spatial random effect patterns captured in the Far-North, Centre, North and Adamawa regions were positives and covered a substantial geographic surface. This could be an indication of abnormally high mortality risks among under-five year children that live in rural zones and villages of Centre, Adamawa and North regions. Our maps corroborate those produced by Golding et al(Golding et al., 2017).

In the infant group, estimates of spatial random effects were negative in the bigger and urbanized cities that were at the same time regional capitals. Examples are Yaoundé (the capital city of Centre region), and Bafoussam (the capital city of the West region). Such results were aligned with the 2014 Cameroon MICS estimates at the regional level (INS, Minsante et UNICEF, 2015).

Logistic regression models assume that the mortality risk remains the same within the five years, the relatively low number of death cases and the presence of time related covariates have made the use of survival model inconsistent.

Our study analysed a period during which coverage of malaria interventions was relatively low. In addition, following the scarcity of reliable data on health interventions at the same cluster locations, we used coverage indicator data from a single household survey and assumed that their estimates did not vary throughout the study period. This hypothesis may have a potential influence on the model parameter estimates.

6.5 Conclusion

In a malaria-endemic country like Cameroon that is characterized by heterogeneous ecological areas, and where the geographic distribution of malaria risk is influenced by seasonal changes, it is advisable to consider malaria parasitological data collected at different transmission seasons. Neonates should be excluded in the assessment of the relationship between malaria prevalence and mortality among children below five years old. Continuous surveillance of household during dry and rainy seasons could be helpful to improve the assessment of the relationship between mortality and malaria risks, but also to enable the disease control programmes to identify specific areas in which seasonal interventions may reduce the under-five mortality rate.

Acknowledgments

6.6 Appendix

Model 1: formulation of relationship between climatic factors and parasite risk for each survey DHS and MIS.

Let us consider cluster s_i, $(i = 1, ..., N)$ in which y_i represents the number of children tested positive for malaria parasitaemia among the n_i screened. In the Bayesian inference, we assume that Y_i has arisen from a binomial distribution such as $Y_i|n_i, p_i \sim Bin(n_i, p_i)$, and where the p_i corresponds to the malaria parasite risk at cluster s_i. At the logit scale, a relationship between climatic factors $X_i = \left(X_{i0} = 1, X_{i1}, X_{i2}, ... X_{ip}\right)$, malaria parasite risk and spatial random effect U_i were defined as $\text{logit}(p_i) = \sum_{k=0}^{p} \beta_k X_{ik} + U_i$. In another way, we have assumed that U_i arise from a multivariate Gaussian distribution i.e. $U \sim MVN(\mathbf{0}, \Omega)$, where $U = (U_1, U_2, ..., U_N)^T$ and Ω the variance-covariance matrix NxN, so that $\Omega_{ij} = \sigma^2 \exp(-\rho d_{ij})$, d_{ij} is the Euclidian distance between cluster s_i and s_j, σ^2 is the spatial variance also called the partial sill and ρ the smoothing parameter to control correlation. We have used a non-informative prior for β_j defined such as $\beta_j \sim N(0,100)$.

Prediction at mortality locations (MICS) using rainy season (MIS).

MICS and MIS samples were randomly selected from DHS. However, the MIS size sample was small than the MICS, we had to consider the presence of clusters s'_j $(j \in 1, ..., N)$ where the parasitaemia data were not collected during the rainy season. The parasitaemia risk at unobserved s'_j location was obtained as $\text{logit}\left(p_j\right) = Z(s'_j) = \sum_{k=0}^{p} \beta_k X'_{jk} + U'(s'_j)$, and conditional to the spatial process and regression parameters estimates by MIS data, the distribution of $Z\left(s'_j\right)|\beta, U(si) \sim N(\beta X_j + \Omega_{ss'}\Omega_s^{-1}U, \Omega_{s'} - \Omega_{s's}\Omega_{s'}^{-1}\Omega_{ss'}$, where U' arises from an isotropic spatial Gaussian process similar to the previous ones described above, but with different spatial variance and range.

Model 2: formulation of relationship between mortality risk and prevalence of parasitaemia.

Let us consider our set of cluster locations $(s_1, s_2, s_3, \ldots, s_N)$, we divide the five the period into parts of six months such as any child i living in a cluster j, can be observed for a maximum of twelve time period (t=1…12) before having 6 years. For a child j, located at cluster s_i and observed at time t. We assumed that a death status $Y_i(s_j, t)$ arise from a Zero inflated Binomial distribution $Y_i(s_j, t) \sim ZIBern(P_i(s_j, t), \theta)$, with θ the inflated zero probability, and π the Bernoulli probability distribution.

$$\pi\big(Y_i(s_j, t) = j\big) = \begin{cases} \theta + (1 - \theta)\pi\big(Y_i(s_j, t) = 0\big) \ if \ Y_i(s_j, t) = 0 \\ (1 - \theta)\pi\big(Y_i(s_j, t) = 1\big) \qquad if \ Y_i(s_j, t) = 1 \end{cases}$$

$E\left(Y_i(s_j, t)\right) = P_i(s_j, t)$, then we also considered a Bayesian logistic regression with a link function: $logit(P_i(s_j, t)) = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ijt}^k + U(s_j)$ (1). The value X_{ijt}^p comes from the p covariates climatic factors. The vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2 \ldots \beta_p)$ of regression coefficient used to calculate the specific odds ratio (OR) exp (β_l), *l=1…p*. The spatial random effect that captures cluster-specific frailty and correlation in mortality was introduced in the model so that $U = (U(s_1), U(s_2), \ldots, U(s_N))^T \sim MVN(0, \Sigma^{(1)})$ and the covariance structure was chosen among Matérn functions family as $\Sigma^{(1)}{}_{s_i, s_j} = \sigma_u^2 \exp(-\rho_u \|s_i - s_j\|)$, $\|.\|$ is the Euclidean distance. Prior of the regression coefficients was assigned using normal and gamma distributions as $\beta_k \sim N(0, \tau_k), \tau_k \sim \text{Inv} - \text{Gamma}(0.01, 0.01)$. We had also performed a spatially varying coefficient regressions to assess the seasonal effect at high-resolution, then the equation (1) was slightly modified such as : $logit(P_i(s_j, t)) = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ijt}^k + \beta_{s_j} *$ $W(s_j) + U(s_j)$ (2), where The spatial random effect that captures cluster-specific seasonal

effect was introduced in the model such as $W(s_j)$, where $W = (W(s_1), W(s_2), \ldots, W(s_N))^T \sim MVN(0, \Sigma^{(2)})$ and the covariance structure was chosen among Matérn functions family such as $\Sigma^{(2)}{}_{s_i,s_j} = \sigma_w^2 \exp(-\rho_w \|s_i - s_j\|) \, d_{i,j}$ is the Euclidean distance. The range parameters ρ_u, ρ_w that monitore the spatial correlation and σ_w^2, σ_u^2 the spatial variances.

Variable selection of covariates

Factors identified as important were involved in model fitting. Bayesian variable was applied on the health interventions, socioeconomic factors (confounding excluded) to select the best predictors. A categorical indicator I_k related to each candidate predictor and that can take value between 0 and 1 to indicate the inclusion or exclusion. When $I_k = 0$ correspond to the exclusion of factor. A multinomial distribution prior was assumed for I_k with probability mass function $\prod_{l=0}^{1} \pi_l{}^{\delta_l(I_k)}$, where π_l is the inclusion probability of each form ($l = 0, 1$). $\delta_k(.)$ the Dirac function such as $\delta_l(I_k) = \begin{cases} 1 \ if \ I_k = l \\ 0 \ if \ I_k \neq l \end{cases}$. In the model, regression coefficients were based on spike and lab approach in which β_k was built as a mixture of normal prior distribution i.e. $\beta_k \sim \delta_l(I_k)N(0, \tau_k^2) + (1 - \delta_l(I_k))N(0, c\tau_k^2)$. The constant c was fixed at 100 000 to shrink the β_k close to zero when the variable was excluded. The inclusion probabilities were constructed using a Bernoulli distribution as $I_k \sim Bern(0.5)$.

Table 6.4: Bayesian geostatistical model-based estimates (posterior median and 95% BCI) obtained from infant mortality risk analysis with and without the neonates (ZIB and Bernoulli models).

| Parameters | Infants (without neonates) | | | | Infants (including neonates) | | | |
| | Zero-inflated Bernoulli | | Bernoulli | | Zero-inflated Bernoulli | | Bernoulli | |
	Median	OR(95%BCI)	Median	OR (95%BCI)	Median	OR (95%BCI)	Median	OR(95%BCI)
Malaria risk	1.82	(0.98, 3.35)	1.76	(0.97, 3.16)	2.99*	(1.28, 7.05)	2.86*	(1.38, 5.83)
Age groups								
< 6 months	1		1		1		1	
others	0.5*	(0.4, 0.63)	0.51*	(0.41, 0.63)	1.82*	(1.32, 2.54)	1.67*	(1.26, 2.22)
Gender								
Male	1		1		1		1	
Female	0.78*	(0.63, 0.96)	0.78*	(0.64, 0.96)	0.77	(0.56, 1.05)	0.8	(0.6, 1.05)
Mother age at birth								
<24	1		1		1		1	
24-35	0.52*	(0.36, 0.75)	0.53*	(0.37, 0.76)	0.57*	(0.33, 0.97)	0.59*	(0.36, 0.95)
>35	0.77	(0.46, 1.26)	0.77	(0.47, 1.25)	0.91	(0.42, 1.94)	0.89	(0.46, 1.73)
Place of residence								
Rural	1		1		1		1	
Urban	0.9	(0.62, 1.27)	0.91	(0.64, 1.27)	0.89	(0.53, 1.47)	0.94	(0.59, 1.47)
Preceding birth interval								
Firstborn	1		1		1		1	
< 2 years	1.78*	(1.28, 2.49)	1.76*	(1.28, 2.44)	1.93*	(1.16, 3.26)	1.78*	(1.13, 2.82)
2 years	0.8	(0.56, 1.14)	0.8	(0.57, 1.14)	1.02	(0.6, 1.74)	1.08	(0.67, 1.74)
3 years	0.54*	(0.33, 0.85)	0.55*	(0.34, 0.86)	0.61	(0.29, 1.2)	0.68	(0.35, 1.26)
=> 4 years	0.58*	(0.37, 0.9)	0.59*	(0.38, 0.9)	0.62	(0.31, 1.2)	0.68	(0.36, 1.24)

| Parameters | Infants (without neonates) | | | | Infants (including neonates) | | | |
| | Zero-inflated Bernoulli | | Bernoulli | | Zero-inflated Bernoulli | | Bernoulli | |
	Median	OR(95%BCI)	Median	OR (95%BCI)	Median	OR (95%BCI)	Median	OR(95%BCI)
Socio ecomomic status								
Poor	1		1		1		1	
Middle	0.88	(0.64, 1.2)	0.87	(0.64, 1.18)	0.7	(0.43, 1.1)	0.72	(0.47, 1.08)
Rich	0.79	(0.53, 1.18)	0.8	(0.54, 1.18)	0.75	(0.41, 1.34)	0.74	(0.43, 1.26)
Mother education								
No formal education	1		1		1		1	
Primary	0.81	(0.62, 1.05)	0.81	(0.63, 1.04)	0.72	(0.49, 1.05)	0.76	(0.55, 1.05)
Above secondary	0.81	(0.57, 1.13)	0.81	(0.59, 1.13)	0.57*	(0.34, 0.94)	0.62*	(0.39, 0.96)
Improved sanitation								
No	1		1		1		1	
Yes	1.02	(0.62, 1.61)	1.01	(0.62, 1.58)	1.03	(0.47, 2.07)	1.03	(0.5, 1.95)
Improved water source								
No	1		1		1		1	
Yes	0.88	(0.65, 1.2)	0.89	(0.66, 1.19)	1.02	(0.65, 1.6)	1.03	(0.68, 1.54)
Multiple births								
Twin	1		1		1		1	
Single	0.19*	(0.13, 0.27)	0.2*	(0.14, 0.27)	0.47	(0.22, 1.05)	0.55	(0.3, 1.12)
Time								
2007	1		1		1		1	
2008	1.12	(0.81, 1.56)	1.13	(0.82, 1.55)	1.16	(0.7, 1.93)	1.12	(0.72, 1.75)
2009	0.93	(0.66, 1.31)	0.93	(0.67, 1.3)	0.89	(0.53, 1.49)	0.89	(0.56, 1.42)
2010	1.07	(0.77, 1.49)	1.07	(0.78, 1.48)	1.01	(0.6, 1.69)	0.99	(0.63, 1.56)
2011	1.02	(0.73, 1.43)	1.01	(0.72, 1.41)	1.16	(0.7, 1.92)	1.12	(0.72, 1.76)
Season								
Rainy	1		1		1		1	

Parameters	Infants (without neonates)				Infants (including neonates)			
	Zero-inflated Bernoulli		Bernoulli		Zero-inflated Bernoulli		Bernoulli	
	Median	OR(95%BCI)	Median	OR (95%BCI)	Median	OR (95%BCI)	Median	OR(95%BCI)
Dry	1.02	(0.8, 1.29)	1.01	(0.8, 1.28)	0.91	(0.64, 1.3)	0.93	(0.67, 1.28)
Mother's year of birth								
<1980	1		1		1		1	
1980-1985	0.83	(0.61, 1.12)	0.83	(0.62, 1.12)	0.93	(0.58, 1.48)	0.98	(0.65, 1.48)
>1985	0.59*	(0.41, 0.84)	0.6*	(0.42, 0.85)	0.72	(0.42, 1.22)	0.76	(0.46, 1.22)
Latitude	1.14*	(1.04, 1.27)	1.13*	(1.03, 1.26)	1.19*	(1.02, 1.43)	1.16*	(1.02, 1.31)
Longitude	0.98	(0.86, 1.1)	0.98	(0.86, 1.1)	0.89	(0.7, 1.07)	0.91	(0.79, 1.03)
Land surface temperature day	0.99	(0.96, 1.02)	0.99	(0.96, 1.02)	0.99	(0.95, 1.04)	0.99	(0.95, 1.04)
Proportion households with at least one ITN	0.61	(0.29, 1.32)	0.61	(0.3, 1.28)	0.38	(0.13, 1.12)	0.42	(0.17, 1.07)
Zero inflated parameter	0.59	(0.15, 0.84)			0.23	(0.005, 0.84)		
Spatial parameters								
Varince	0.17	(0.04, 1.3)	0.18	(0.04, 1.87)	1.06	(0.61, 2.82)	0.62	(0.04, 31.12)
Range	0.67	(0.14, 4.29)	0.73	(0.17, 6.89)	1.33	(0.44, 5.05)	1.32	(0.11, 31.42)
DIC	3520.94		3523.38		2122.25		2123.96	

*: Statistically important
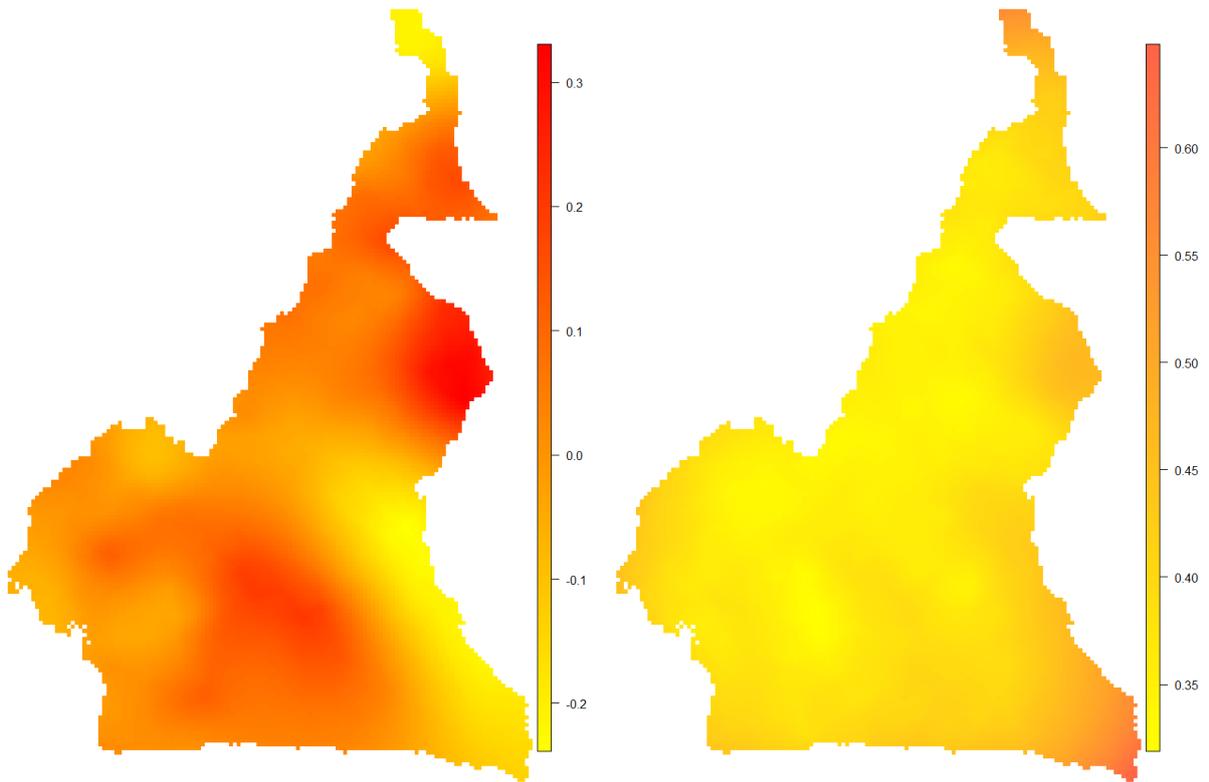
Figure 6.10: Posterior point estimates of spatial random effect (left) and its standard deviation (right) of Zero-inflated Bernoulli model on children less than one-year group (without neonates).
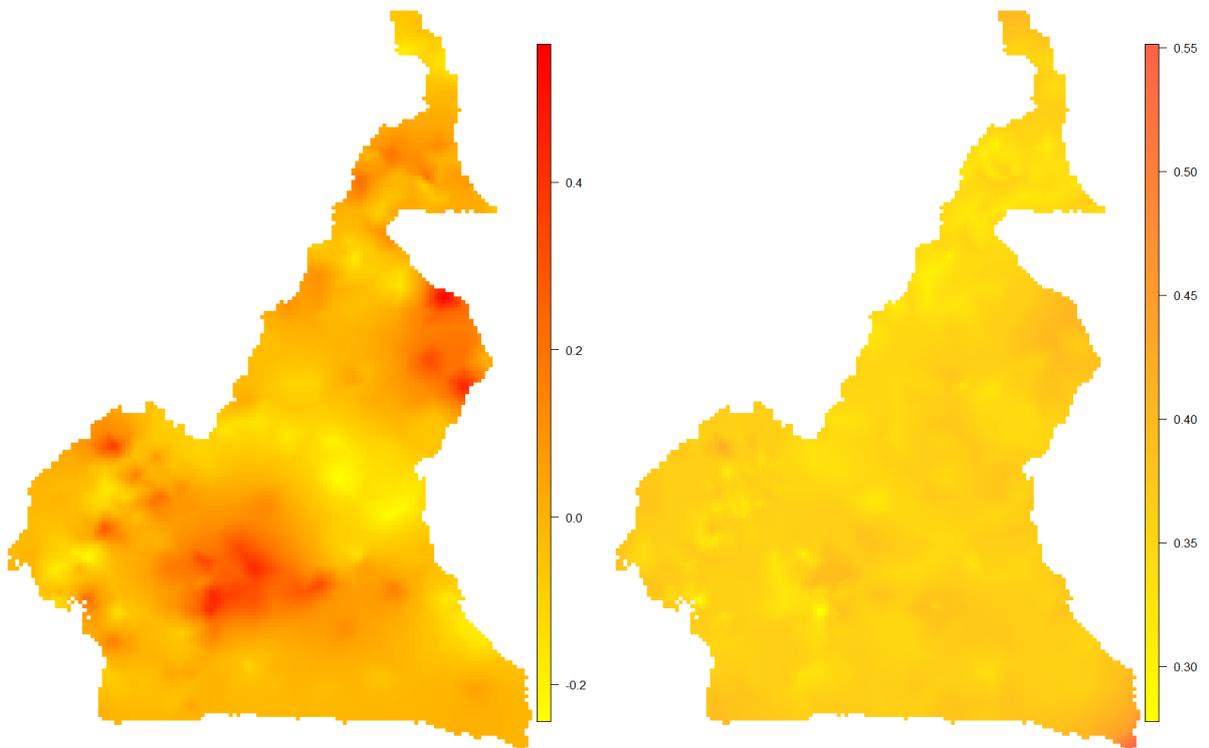
Figure 6.11: Posterior point estimates of spatial random effect (left) and its standard deviation (right) of Zero-inflated Bernoulli model on children less than one-year-old group (including neonates).

Table 6.5: Bayesian geostatistical model-based estimates (posterior median and 95% BCI) obtained from under-five mortality risk analysis with and without the neonates (ZIB and Bernoulli models).

| Parameters | Under five 5 (including neonates) | | | | Under five 5 (without neonates) | | | |
| | Zero-inflated Bernoulli | | Bernoulli | | Zero-inflated Bernoulli | | Bernoulli | |
	Median	OR(95%BCI)	Median	OR(95%BCI)	Median	OR(95%BCI)	Median	OR(95%BCI)
Malaria risk	1.58*	(1.01, 2.45)	1.57*	(1.03, 2.38)	2.02*	(1.18, 3.41)	1.98*	(1.21, 3.21)
Age groups								
< 6 months	1		1		1		1	
6 months-1 year	0.52*	(0.42, 0.65)	0.53*	(0.42, 0.65)	1.84*	(1.37, 2.49)	1.77*	(1.34, 2.35)
1- 2 years	0.31*	(0.25, 0.38)	0.31*	(0.25, 0.38)	1.06	(0.8, 1.41)	1.05	(0.8, 1.39)
2-3 years	0.21*	(0.16, 0.26)	0.21*	(0.16, 0.26)	0.7*	(0.51, 0.96)	0.71*	(0.52, 0.96)
3-4 years	0.12*	(0.09, 0.16)	0.12*	(0.09, 0.16)	0.41*	(0.28, 0.59)	0.42*	(0.29, 0.6)
4-5 years	0.06*	(0.04, 0.09)	0.06*	(0.04, 0.09)	0.22*	(0.15, 0.33)	0.23*	(0.15, 0.34)
Gender								
Male	1		1		1		1	
Female	0.86	(0.74, 1)	0.86	(0.74, 1)	0.89	(0.74, 1.07)	0.9	(0.76, 1.07)
Mother age at birth								
<24	1		1		1		1	
24-35	0.73*	(0.55, 0.95)	0.72*	(0.55, 0.94)	0.88	(0.63, 1.21)	0.87	(0.64, 1.19)
>35	1.17	(0.81, 1.69)	1.15	(0.8, 1.66)	1.54	(0.98, 2.42)	1.49	(0.97, 2.28)
Place of residence								
Rural	1		1		1		1	
Urban	0.79	(0.61, 1.03)	0.8	(0.62, 1.03)	0.75	(0.54, 1.02)	0.77	(0.57, 1.04)
Preceding birth interval								

Parameters	Under five 5 (including neonates)				Under five 5 (without neonates)			
	Zero-inflated Bernoulli		Bernoulli		Zero-inflated Bernoulli		Bernoulli	
	Median	OR(95%BCI)	Median	OR(95%BCI)	Median	OR(95%BCI)	Median	OR(95%BCI)
Firstborn	1		1		1		1	
< 2 years	1.66*	(1.3, 2.12)	1.66*	(1.3, 2.12)	1.65*	(1.23, 2.24)	1.63*	(1.22, 2.17)
2 years	1	(0.77, 1.29)	1	(0.77, 1.29)	1.19	(0.88, 1.62)	1.2	(0.89, 1.61)
3 years	0.71*	(0.51, 0.98)	0.71*	(0.51, 0.98)	0.82	(0.55, 1.21)	0.83	(0.57, 1.21)
=> 4 years	0.59*	(0.42, 0.82)	0.59*	(0.42, 0.82)	0.61*	(0.4, 0.92)	0.62*	(0.41, 0.92)
Socio ecomomic status								
Poor	1		1		1		1	
Middle	0.86	(0.69, 1.08)	0.85	(0.68, 1.07)	0.8	(0.6, 1.04)	0.79	(0.61, 1.03)
Rich	0.81	(0.6, 1.09)	0.81	(0.6, 1.08)	0.8	(0.56, 1.15)	0.8	(0.56, 1.12)
Mother education								
No formal education	1		1		1		1	
Primary	0.86	(0.71, 1.04)	0.84	(0.7, 1.02)	0.86	(0.69, 1.08)	0.85	(0.69, 1.06)
Above secondary	0.8	(0.62, 1.03)	0.78	(0.61, 1.004)	0.7*	(0.52, 0.95)	0.7*	(0.52, 0.941)
Improved sanitation								
No	1		1		1		1	
Yes	0.72	(0.46, 1.07)	0.72	(0.47, 1.06)	0.58*	(0.32, 0.97)	0.58*	(0.32, 0.97)
Improved water source								
No	1		1		1		1	
Yes	0.86	(0.68, 1.08)	0.85	(0.68, 1.07)	0.89	(0.67, 1.17)	0.89	(0.68, 1.16)
Multiple births								
Twin	1		1		1		1	
Single	0.28*	(0.21, 0.37)	0.28*	(0.22, 0.37)	0.51*	(0.33, 0.8)	0.54*	(0.37, 0.84)
Time								
2007	1		1		1		1	
2008	1.14	(0.89, 1.46)	1.14	(0.9, 1.46)	1.14	(0.85, 1.54)	1.13	(0.85, 1.5)
2009	1	(0.78, 1.28)	1	(0.78, 1.28)	0.99	(0.73, 1.34)	0.99	(0.74, 1.32)

Parameters	Under five 5 (including neonates)				Under five 5 (without neonates)			
	Zero-inflated Bernoulli		Bernoulli		Zero-inflated Bernoulli		Bernoulli	
	Median	OR(95%BCI)	Median	OR(95%BCI)	Median	OR(95%BCI)	Median	OR(95%BCI)
2010	1.12	(0.88, 1.43)	1.12	(0.88, 1.43)	1.09	(0.81, 1.47)	1.08	(0.81, 1.43)
2011	0.94	(0.73, 1.21)	0.94	(0.73, 1.21)	0.93	(0.69, 1.26)	0.94	(0.7, 1.25)
Season								
Rainy	1		1		1		1	
Dry	1.23*	(1.03, 1.47)	1.23*	(1.03, 1.46)	1.28*	(1.04, 1.58)	1.27*	(1.04, 1.56)
Mother's year of birth								
<1980	1		1		1		1	
1980-1985	0.95	(0.77, 1.19)	0.95	(0.77, 1.18)	1.06	(0.82, 1.38)	1.07	(0.83, 1.37)
>1985	0.88	(0.67, 1.14)	0.88	(0.67, 1.14)	1.15	(0.83, 1.59)	1.14	(0.83, 1.55)
Latitude	1.07	(0.99, 1.14)	1.07*	(1.001, 1.14)	1.07	(0.98, 1.16)	1.07	(0.98, 1.15)
Longitude	1.01	(0.94, 1.1)	1.01	(0.94, 1.09)	0.99	(0.9, 1.08)	0.98	(0.91, 1.07)
Land surface temperature day	0.99	(0.97, 1.02)	0.99	(0.97, 1.02)	0.99	(0.97, 1.03)	0.99	(0.97, 1.03)
Proportion households with at least one ITN	0.79	(0.46, 1.38)	0.8	(0.47, 1.36)	0.76	(0.39, 1.47)	0.79	(0.42, 1.48)
Zero inflated parameter	0.71	(0.60, 0.80)			0.074	(0.005, 0.18)		
Spatial parameters								
Variance	0.5	(0.15, 3.83)	0.25	(0.04, 2.85)	0.33	(0.29, 0.37)	0.14	(0.09, 0.28)
Range	1.04	(0.38, 5.23)	0.56	(0.12, 3.92)	1.11	(0.46, 2.14)	0.37	(0.23, 0.53)
DIC	7112.15		7113.3		5672.02		5671.28	

*: Statistically important
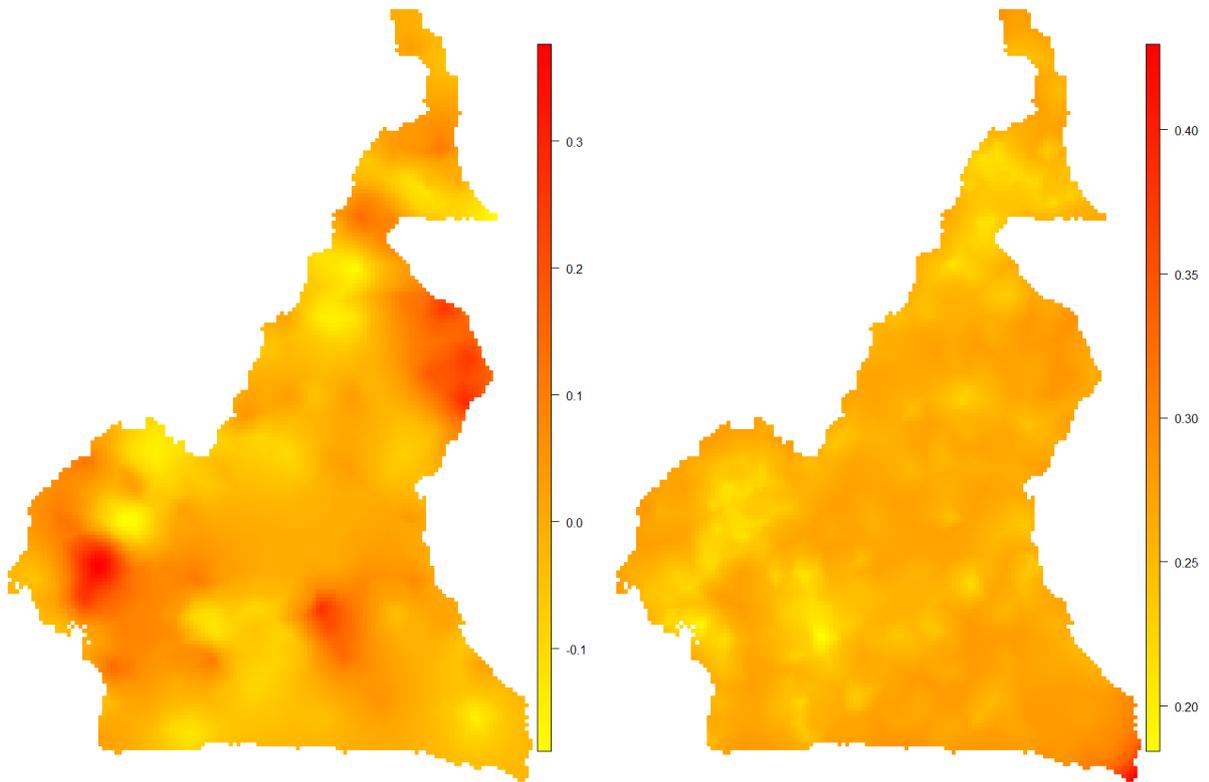
Figure 6.12: Posterior point estimates of spatial random effect (left) and its standard deviation (right) of Zero-inflated Bernoulli model on children less than five years old group (without neonates).
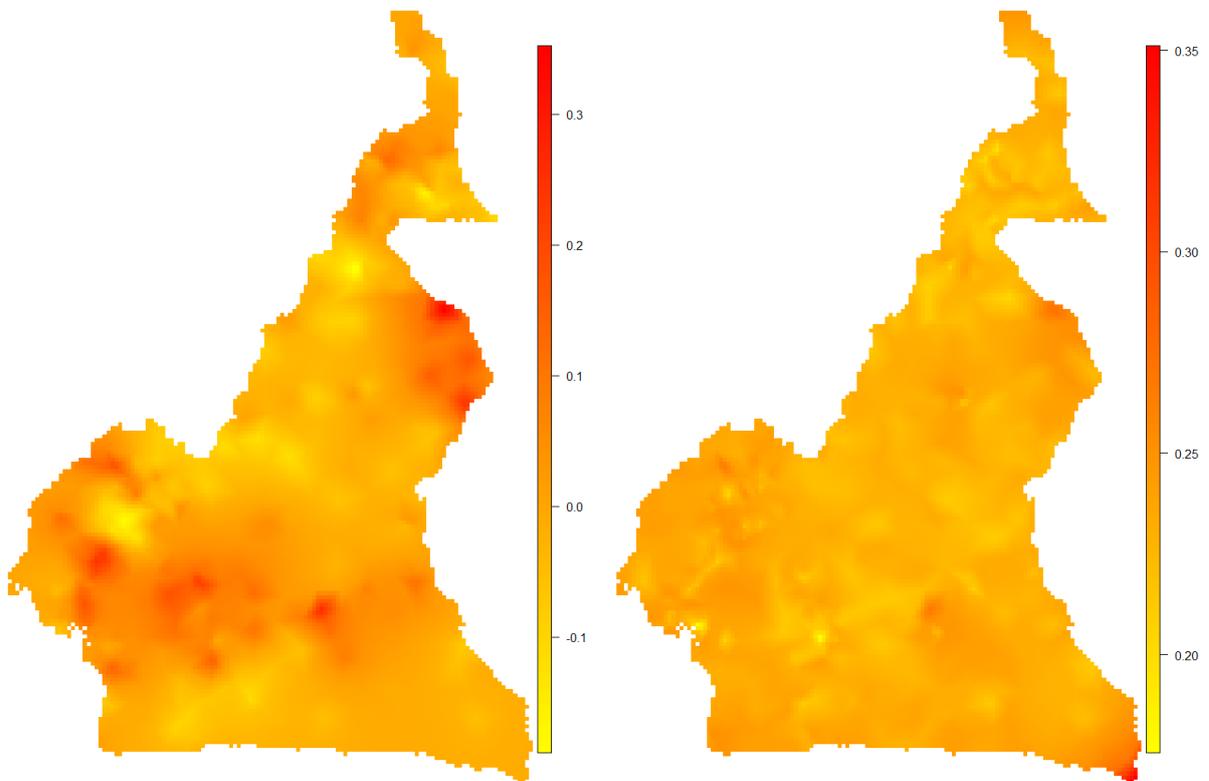

Figure 6.13: Posterior point estimates of spatial random effect (left) and its standard deviation (right) of Zero-inflated Bernoulli model on children less than five years old group (including neonates).

Chapter 7 Discussion and perspectives

The main contributions of the PhD thesis are the evaluation and further development of the malaria survey design methodology related to the timing of the survey, the influence of jittering in the coordinates of the reported cluster locations, the selection of the survey locations and the seasonal monitoring of malaria data. Furthermore, the ability of the survey data to estimate the malaria-related deaths was assessed and estimates of the effects of malaria interventions based on data obtained from household surveys and from HMIS were compared. Our analyses employed data collected by the standardized national and representative household surveys generally conducted to evaluate progress in disease control and socioeconomic conditions in low-resource countries. In particular, the thesis used Demographic and Health Survey (DHS), Malaria Indicators Survey (MIS) carried out in Cameroon in 2011, and the Multiple Cluster Indicators Survey (MICS) conducted in the country in 2014. Although DHS and MIS were conducted in the same year, MIS covered the rainy season while DHS covered the dry one. These surveys were carried out at the same cluster locations. Therefore, it was an interesting opportunity to avoid the cluster misalignment effect on model parameter estimates. Parasitological test results obtained from children aged less than five years using rapid malaria diagnostic tests were used to assess malaria risks in the country. Malaria incidence data were extracted from the national malaria database. Malaria cases tested and diagnosed by health facilities were used in this work to assess the distribution of malaria incidence among groups of persons less than five years and those above five years. The climatic and environmental conditions of each cluster were approximated using remote sensing data and their geographic coordinates.

7.1 Malaria seasonality and survey design

Our findings showed that survey timing might affect the distribution of disease risks in a country with different epidemiological facets. Although DHS remains the major national survey for many low-resource countries with regards to the sample size and spatial coverage, DHS data were not able to capture the entire malaria risk structure of the country. Even though MIS was conducted during the rainy season, it did not capture the overall disease risk pattern. This finding was an indication of the complexities involved in assessing the disease's risks with a single survey in a country with various ecological zones.

Altitude and vegetation coverage have been clearly identified as the main factors of malaria transmission at the national level. Disease risk was positively associated with vegetation coverage and negatively with altitude. In addition, the distance to water bodies was only important during the rainy season.

Malaria incidence collected passively on a routine basis by health facilities has identified the same areas already captured by the DHS and MIS. However, malaria incidence maps were less accurate than household survey maps. This has already been observed in the analysis of malaria incidence in Uganda (Ssempiira et al., 2018). This is an indication of the improvement achieved by the national health management information system during recent years. Those data could be used to monitor the disease risk distribution between household surveys, but health data could be improved to include georeferenced data such as village locations of patients and intervention indicators.

The coastal and northern parts of the country were identified as malaria-sensitive areas during the rainy season. Lack of adequate infrastructure to drain water during the rainy season in the coastal and north regions may explain these changes. Malaria transmission was relatively stable in the eastern part where vegetation coverage was dense. Geostatistical models attributed lower exposure to malaria among populations in the western and Adamawa

highlands, especially during the rainy season, probably because of the presence of mountain ranges that provide high altitudes and cold temperatures. Malaria prevalence was more statistically important for under-5 mortality when adjusted for interventions and seasonal effects, excluding neonatal deaths. Our results suggested that the dry season was positively associated with under-5 mortality risk at the national level and particularly in the northern and Sahelian regions. This situation was likely related to increasing malnutrition, anemia and other associated infection risk of many deadly diseases as diarrhea, meningitis and other respiratory complications.

7.2 Selection of cluster locations and jittering of reported survey coordinates

DHS apply high standards to protect the medical results of the respondents in households interviewed through the random displacement of the reported cluster locations. This modification of cluster coordinates complies with the needs to avoid any potential social stigma and marginalization, as surveys collect data not only on malaria but on other more sensitive health outcomes such as the *immunodeficiency virus* (HIV). The thesis improved our understanding of the jittering influence on the geostatistical model-based malaria risk predictions. It was shown that in Cameroon, the miss-placement of cluster coordinates has a moderate influence on the malaria risk maps, probably because the main climatic predictors such as vegetation index and altitude were relatively stable within the jittering buffer. The relationship between the disease risk and the selected climatic, environmental factors, and intervention indicators were often modified. However, the direction of their association with disease risk was still aligned with scientific literature. The geographic distribution of the uncertainty and the Bayesian confidence intervals were also influenced by the shifting of cluster locations.

Regarding the selection of cluster locations, an improved adaptive Bayesian geostatistical design was proposed by integrating predictive risk uncertainty, human population availability,

and objectives of control programme into the selection of new clusters. Our contribution has helped identifying additional locations where future surveys might focus on understanding the reasons of their abnormally high or low malaria risks. These cluster locations should be analyzed to improve our understanding of malaria transmission and intervention effects. The modified algorithm can be applied to any environmental disease and could then be of interest in a national context of integrated vector surveillance. Identification of cluster locations could be made for several neglected diseases and the shared spatial areas highlighted for further investigations. Such collaborative strategy can lead to resource efficiency and facilitate the improvement of human resources.

7.3 The effects of malaria interventions based on survey and incidence data

The effects of malaria intervention were assessed by including in our models the malaria indicators designed for this purpose by the international Roll Back Malaria initiative and World health organization (RBM Monitoring and Evaluation Group, 2005; Roll Back Malaria, USAID, CDC, UNICEF, WHO, 2013). Bayesian variable selection was always performed to identify the most important indicators to include in the final model. The coverage indicators concerning insecticide-treated nets (ITNs) were mostly selected by the DHS and MIS surveys. Particularly, the proportion of households with at least one ITN per two persons and the percentage of population access to an ITN in their household within a cluster were selected and negatively associated with the disease risk, in MIS (rainy) and DHS (dry) respectively. The ITN coverage indicator identified using DHS data incorporates the effective use of ITN and it was less associated to the malaria risk than the one selected with MIS data. This result showed the importance of ITN presence at household's level during the rainy season. Out of the rainy season, DHS had identified the proportion of fever cases that received a recommended artemisinin combination therapy (ACT) during the last two weeks as

important. It was positively associated with malaria risk among children. This result may be an indication of the ACT use among children suffering from fever.

The proportion of households with at least one ITN per two persons within a cluster was statistically important and negatively associated to confirmed malaria cases. Increase in the coverage of intervention influenced the disease trends. In the under-five mortality analysis, the possession of ITN was negatively associated to under-five mortality but the relation was not statistically important, may be an indication of the remaining efforts to improve the effective use of ITNs.

The following demographic and socioeconomic factors were statistically important to understand the disease risk as well as the under-five mortality: age, education of mothers, and household wealth index. Children aged less than one year or living in households with high wealth index categories as well as children with mothers who attended at least secondary school were less affected than the others.

7.4 Limitations and extensions

Malaria parasite prevalence among children was used as a proxy of malaria risk across the country. The disease risk within a cluster was assumed to be similarly across households, thus individual and familial biological assets and the presence of an asymptomatic population in children group were assumed to be negligible. The effects of climatic factors were assessed only at the national level using a stationary and isotropic Gaussian process.

Different models performed at the regional level should be used to improve malaria risk maps and could help to identify important climatic factors at the regional level. This approach already designed and applied in Mali could exhibit other high spot areas (Giardina et al., 2016).

ITN coverage indicators were calculated without adjusting the quality and duration of ITN. Those data could be included in models to improve the assessment of intervention effects on

malaria risk. Furthermore, surveillance data on malaria vector characteristics and capabilities as the entomological inoculation rate was not available in these representative household surveys. The relationship between malaria risk and climatic factors could be adjusted by such indicators.

Our work could be extended to combine the DHS jittering approach with Bayesian models to identify the subsets of appropriate shifted locations that preserve the true relationship between disease risk estimates, climatic factors, and intervention indicators. Cameroon has conducted in 2018 another DHS it could be an opportunity to evaluate the survey design and to assess effects of interventions on malaria risk and under-five mortality.

Conclusion

This thesis improved our understanding of the effects of survey timing, jittering and seasons on the assessment of disease burden and proposed methodology to optimize the selection of survey sampling locations. The results are useful to National Malaria Control programs and to researchers involved in the spatial analyses of DHS, MICS, and MIS data. In recent decades, efforts were made to design and introduce standardized surveys and collection tools that ensure data reliability and comparability in low-resource countries. The thesis identified the need to improve the temporal and spatial coverage of household surveys, but also to strengthening the HMIS by collecting georeferenced data and to combine the analysis of incidence with survey data. The choice of relevant cluster locations and a continuous collection of household data seem like another requirement for countries that aims to achieve malaria elimination and eradication. The introduction of a health demographic surveillance system for malaria-endemic countries could be helpful.

.

Bibliography

Adigun, A.B., Gajere, E.N., Oresanya, O., Vounatsou, P., 2015. Malaria risk in Nigeria: Bayesian geostatistical modelling of 2010 malaria indicator survey data. Malaria Journal 14. https://doi.org/10.1186/s12936-015-0683-6

Agier, L., Deroubaix, A., Martiny, N., Yaka, P., Djibo, A., Broutin, H., 2013. Seasonality of meningitis in Africa and climate forcing: aerosols stand out. J R Soc Interface 10. https://doi.org/10.1098/rsif.2012.0814

Alegana, V.A., Atkinson, P.M., Wright, J.A., Kamwi, R., Uusiku, P., Katokele, S., Snow, R.W., Noor, A.M., 2013. Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial-temporal models. Spat Spatiotemporal Epidemiol 7, 25–36. https://doi.org/10.1016/j.sste.2013.09.001

Antonio-Nkondjio, C., Kerah, C.H., Simard, F., Awono-Ambene, P., Chouaibou, M., Tchuinkam, T., Fontenille, D., 2006. Complexity of the malaria vectorial system in Cameroon: contribution of secondary vectors to malaria transmission. J. Med. Entomol. 43, 1215–1221.

Antonio-Nkondjio, C., Ndo, C., Njiokou, F., Bigoga, J.D., Awono-Ambene, P., Etang, J., Ekobo, A.S., Wondji, C.S., 2019. Review of malaria situation in Cameroon: technical viewpoint on challenges and prospects for disease elimination. Parasites & Vectors 12, 501. https://doi.org/10.1186/s13071-019-3753-8

Apinjoh, T.O., Anchang-Kimbi, J.K., Mugri, R.N., Tangoh, D.A., Nyingchu, R.V., Chi, H.F., Tata, R.B., Njumkeng, C., Njua-Yafi, C., Achidi, E.A., 2015. The Effect of Insecticide Treated Nets (ITNs) on Plasmodium falciparum Infection in Rural and Semi-Urban Communities in the South West Region of Cameroon. PLoS One 10. https://doi.org/10.1371/journal.pone.0116300

Arab, A., Jackson, M.C., Kongoli, C., 2014. Modelling the effects of weather and climate on malaria distributions in West Africa. Malar J 13, 126. https://doi.org/10.1186/1475-2875-13-126

Atieli, H.E., Zhou, G., Afrane, Y., Lee, M.-C., Mwanzo, I., Githeko, A.K., Yan, G., 2011. Insecticide-treated net (ITN) ownership, usage, and malaria transmission in the highlands of western Kenya. Parasit Vectors 4, 113. https://doi.org/10.1186/1756-3305-4-113

Ayala, D., Costantini, C., Ose, K., Kamdem, G.C., Antonio-Nkondjio, C., Agbor, J.-P., Awono-Ambene, P., Fontenille, D., Simard, F., 2009. Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. Malar J 8, 307. https://doi.org/10.1186/1475-2875-8-307

Barbieri, M.M., Berger, J.O., 2004. Optimal predictive model selection. Ann. Statist. 32, 870–897. https://doi.org/10.1214/009053604000000238

Berg, P., Moseley, C., Haerter, J.O., 2013. Strong increase in convective precipitation in response to higher temperatures. Nature Geoscience 6, 181–185. https://doi.org/10.1038/ngeo1731

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., Songini, M., 1995. Bayesian analysis of space—time variation in disease risk. Statistics in Medicine 14, 2433–2443. https://doi.org/10.1002/sim.4780142112

Berzosa, P., de Lucio, A., Romay-Barja, M., Herrador, Z., González, V., García, L., Fernández-Martínez, A., Santana-Morales, M., Ncogo, P., Valladares, B., Riloha, M., Benito, A., 2018. Comparison of three diagnostic methods (microscopy, RDT, and PCR) for the detection of malaria parasites in representative samples from Equatorial Guinea. Malaria Journal 17, 333. https://doi.org/10.1186/s12936-018-2481-4

Besag, J., Green, P.J., 1993. Spatial Statistics and Bayesian Computation. Journal of the Royal Statistical Society. Series B (Methodological) 55, 25–37.

Bigoga, J.D., Manga, L., Titanji, V.P., Coetzee, M., Leke, R.G., 2007. Malaria vectors and transmission dynamics in coastal south-western Cameroon. Malaria Journal 6, 5. https://doi.org/10.1186/1475-2875-6-5

Bonaventure Tientche, Damian Nota Anong, Smith Asaah, Jerome Fru-Cho,Theresa K. Nkuo-Akenji, Damian Nota Anong, 2016. Asymptomatic malaria parasitaemia in school children of E kondo Titi sub-division, Cameroon. International Research Journal of Public and Environmental Health 3 (8), 182–190. http://dx.doi.org/10.15739/irjpeh.16.024

Bousema, T., Stevenson, J., Baidjoe, A., Stresman, G., Griffin, J.T., Kleinschmidt, I., Remarque, E.J., Vulule, J., Bayoh, N., Laserson, K., Desai, M., Sauerwein, R., Drakeley, C., Cox, J., 2013. The impact of hotspot-targeted interventions on malaria transmission: study protocol for a cluster-randomized controlled trial. Trials 14, 36. https://doi.org/10.1186/1745-6215-14-36

Boussougou-Sambe, S.T., Awono-Ambene, P., Tasse, G.C.T., Etang, J., Binyang, J.A., Nouage, L.D., Wamba, G., Enyong, P., Fokam, E.B., 2017. Physical integrity and residual bio-efficacy of used LLINs in three cities of the South-West region of Cameroon 4 years after the first national mass-distribution campaign. Malar J 16. https://doi.org/10.1186/s12936-017-1690-6

Bowen, H.L., 2013. Impact of a mass media campaign on bed net use in Cameroon. Malar. J. 12, 36. https://doi.org/10.1186/1475-2875-12-36

Briët, O.J.T., Amerasinghe, P.H., Vounatsou, P., 2013. Generalized Seasonal Autoregressive Integrated Moving Average Models for Count Data with Application to Malaria Time Series with Low Case Numbers. PLOS ONE 8, e65761. https://doi.org/10.1371/journal.pone.0065761

BUCREP, 2011. Projections demographiques. Ministere de l'Economie, Amenagement et de la Planification.

BUCREP, 2010. La population du Cameroun en 2010.

BUCREP, 2005. Projections Demographiques, 3eme RGPH.

Burgert CR, Colston J, Roy T, Zachary B., 2013. Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys (No. DHS Spatial Analysis Report No. 7). USAID, ICF International, Calverton, Maryland, USA.

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A.M., Morse, A.P., Colón-González, F.J., Stenlund, H., Martens, P., Lloyd, S.J., 2014. Impact of climate change on global malaria distribution. Proc Natl Acad Sci U S A 111, 3286–3291. https://doi.org/10.1073/pnas.1302089111

Chapman Hall CRC, 2010. Spatial Aggregation and the Ecological Fallacy. Chapman Hall CRC Handb Mod Stat Methods 2010, 541–558. https://doi.org/10.1201/9781420072884-c30

Christiansen-Jucht, C., Parham, P.E., Saddler, A., Koella, J.C., Basáñez, M.-G., 2014. Temperature during larval development and adult maintenance influences the survival of Anopheles gambiae s.s. Parasites & Vectors 7, 489. https://doi.org/10.1186/s13071-014-0489-3

Cibulskis, R.E., Aregawi, M., Williams, R., Otten, M., Dye, C., 2011. Worldwide Incidence of Malaria in 2009: Estimates, Time Trends, and a Critique of Methods. PLOS Medicine 8, e1001142. https://doi.org/10.1371/journal.pmed.1001142

Consultative Group on Health Systems and Operational, 2011. A Research Agenda for Malaria Eradication: Health Systems and Operational Research. PLOS Medicine 8, e1000397. https://doi.org/10.1371/journal.pmed.1000397

Cressie, N., 1993. Statistics for Spatial Data, Revised Edition. ed. Wiley Series in Probability and Statistics. https://doi.org/10.1002/9781119115151

Cressie, N., Kornak, J., 2003. Spatial Statistics in the Presence of Location Error with an Application to Remote Sensing of the Environment. Statist. Sci. 18, 436–456. https://doi.org/10.1214/ss/1081443228

Dalrymple, U., Arambepola, R., Gething, P.W., Cameron, E., 2018. How long do rapid diagnostic tests remain positive after anti-malarial treatment? Malaria Journal 17, 228. https://doi.org/10.1186/s12936-018-2371-9

Diboulo, E., Sié, A., Diadier, D.A., Voules, D.A.K., Yé, Y., Vounatsou, P., 2015. Bayesian variable selection in modelling geographical heterogeneity in malaria transmission from sparse data: an application to Nouna Health and Demographic Surveillance System (HDSS) data, Burkina Faso. Parasites & Vectors 8, 118. https://doi.org/10.1186/s13071-015-0679-7

Diboulo, E., Sié, A., Vounatsou, P., 2016. Assessing the effects of malaria interventions on the geographical distribution of parasitaemia risk in Burkina Faso. Malaria Journal 15. https://doi.org/10.1186/s12936-016-1282-x

Diggle P. J., Tawn J. A., Moyeed R. A., 2002. Model-based geostatistics. Journal of the Royal Statistical Society: Series C (Applied Statistics) 47, 299–350. https://doi.org/10.1111/1467-9876.00113

Diggle, P., Lophaven, S., 2006. Bayesian Geostatistical Design. Scandinavian Journal of Statistics 33, 53–64.

Diggle, P.J., Menezes, R., Su, T., 2010. Geostatistical Inference under Preferential Sampling. Journal of the Royal Statistical Society. Series C (Applied Statistics) 59, 191–232.

Diggle, P.J., Moraga, P., Rowlingson, B., Taylor, B.M., 2013. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. Statist. Sci. 28, 542–563. https://doi.org/10.1214/13-STS441

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. Journal of the Royal Statistical Society: Series C (Applied Statistics) 47, 299–350. https://doi.org/10.1111/1467-9876.00113

Djaman, J.A., Olefongo, D., Ako, A.B., Roman, J., Ngane, V.F., Basco, L.K., Tahar, R., 2017. Molecular Epidemiology of Malaria in Cameroon and Côte d'Ivoire. XXXI. Kelch 13 Propeller Sequences in Plasmodium falciparum Isolates before and after Implementation of Artemisinin-Based Combination Therapy. Am. J. Trop. Med. Hyg. 97, 222–224. https://doi.org/10.4269/ajtmh.16-0889

Dlamini, S.N., Beloconi, A., Mabaso, S., Vounatsou, P., Impouma, B., Fall, I.S., 2019. Review of remotely sensed data products for disease mapping and epidemiology. Remote Sensing Applications: Society and Environment 14, 108–118. https://doi.org/10.1016/j.rsase.2019.02.005

Doolan, D.L., Dobaño, C., Baird, J.K., 2009. Acquired Immunity to Malaria. Clin. Microbiol. Rev. 22, 13–36. https://doi.org/10.1128/CMR.00025-08

Dwyer-Lindgren, L., Squires, E.R., Teeple, S., Ikilezi, G., Allen Roberts, D., Colombara, D.V., Allen, S.K., Kamande, S.M., Graetz, N., Flaxman, A.D., El Bcheraoui, C., Asbjornsdottir, K., Asiimwe, G., Augusto, Â., Augusto, O., Chilundo, B., De Schacht, C., Gimbel, S., Kamya, C., Namugaya, F., Masiye, F., Mauieia, C., Miangotar, Y., Mimche, H., Sabonete, A., Sarma, H., Sherr, K., Simuyemba, M., Sinyangwe, A.C., Uddin, J., Wagenaar, B.H., Lim, S.S., 2018. Small area estimation of under-5 mortality in Bangladesh, Cameroon, Chad, Mozambique, Uganda, and Zambia using spatially misaligned data. Population Health Metrics 16, 13. https://doi.org/10.1186/s12963-018-0171-7

Edlund, S., Davis, M., Douglas, J.V., Kershenbaum, A., Waraporn, N., Lessler, J., Kaufman, J.H., 2012. A global model of malaria climate sensitivity: comparing malaria response to historic climate data based on simulation and officially reported malaria incidence. Malaria Journal 11, 331. https://doi.org/10.1186/1475-2875-11-331

Einterz, E.M., Bates, M., 2011. Causes and circumstances of death in a district hospital in northern Cameroon, 1993-2009. Rural Remote Health 11, 1623.

Elias T. Krainski et al, 2018. Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. CRC Press Book.

ESRI, 2013. ESRI 2013. ArcGIS Desktop. Release 10.2.1.

Fançony, C., Sebastião, Y.V., Pires, J.E., Gamboa, D., Nery, S.V., 2013. Performance of microscopy and RDTs in the context of a malaria prevalence survey in Angola: a comparison using PCR as the gold standard. Malaria Journal 12, 284. https://doi.org/10.1186/1475-2875-12-284

Fowkes, F.J.I., Boeuf, P., Beeson, J.G., 2016. Immunity to malaria in an era of declining malaria transmission. Parasitology 143, 139–153. https://doi.org/10.1017/S0031182015001249

Frey, C., Traoré, C., De Allegri, M., Kouyaté, B., Müller, O., 2006. Compliance of young children with ITN protection in rural Burkina Faso. Malar J 5, 70. https://doi.org/10.1186/1475-2875-5-70

Fru-Cho, J., Bumah, V.V., Safeukui, I., Nkuo-Akenji, T., Titanji, V.P., Haldar, K., 2014. Molecular typing reveals substantial Plasmodium vivax infection in asymptomatic adults in a rural area of Cameroon. Malaria Journal 13, 170. https://doi.org/10.1186/1475-2875-13-170

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association 85, 398. https://doi.org/10.2307/2289776

Gelfand, A.E., Vounatsou, P., 2003. Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics 4, 11–15. https://doi.org/10.1093/biostatistics/4.1.11

Gemperli, A., Sogoba, N., Fondjo, E., Mabaso, M., Bagayoko, M., Briët, O.J.T., Anderegg, D., Liebe, J., Smith, T., Vounatsou, P., 2006. Mapping malaria transmission in West and Central Africa. Tropical Medicine & International Health 11, 1032–1046. https://doi.org/10.1111/j.1365-3156.2006.01640.x

Gemperli, A., Vounatsou, P., Kleinschmidt, I., Bagayoko, M., Lengeler, C., Smith, T., 2004. Spatial Patterns of Infant Mortality in Mali: The Effect of Malaria Endemicity. Am J Epidemiol 159, 64–72. https://doi.org/10.1093/aje/kwh001

George Casella and Roger L Berger, 2002. Statistical inference. Duxury Thomson Learning, California.

Gething, P.W., Casey, D.C., Weiss, D.J., Bisanzio, D., Bhatt, S., Cameron, E., Battle, K.E., Dalrymple, U., Rozier, J., Rao, P.C., Kutz, M.J., Barber, R.M., Huynh, C., Shackelford, K.A., Coates, M.M., Nguyen, G., Fraser, M.S., Kulikoff, R., Wang, H., Naghavi, M., Smith, D.L., Murray, C.J.L., Hay, S.I., Lim, S.S., 2016. Mapping Plasmodium falciparum Mortality in Africa between 1990 and 2015. New England Journal of Medicine 375, 2435–2445. https://doi.org/10.1056/NEJMoa1606701

Gething, P.W., Patil, A.P., Smith, D.L., Guerra, C.A., Elyazar, I.R., Johnston, G.L., Tatem, A.J., Hay, S.I., 2011. A new world malaria map: Plasmodium falciparum endemicity in 2010. Malaria Journal 10, 378. https://doi.org/10.1186/1475-2875-10-378

Geweke, J.F., 1991. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (No. 148), Staff Report. Federal Reserve Bank of Minneapolis.

GFATM, 2018. Cameroon approuved proposals for malaria [WWW Document]. URL https://www.theglobalfund.org/en/portfolio/country/?loc=CMR&k=abd20e6f-1d8d-4fd0-964e-f429e276d3ac (accessed 4.17.18).

Giardina, F., Gosoniu, L., Konate, L., Diouf, M.B., Perry, R., Gaye, O., Faye, O., Vounatsou, P., 2012. Estimating the Burden of Malaria in Senegal: Bayesian Zero-Inflated Binomial Geostatistical Modeling of the MIS 2008 Data. PLoS ONE 7, e32625. https://doi.org/10.1371/journal.pone.0032625

Giardina, F., Kasasa, S., Sié, A., Utzinger, J., Tanner, M., Vounatsou, P., 2014. Effects of vector-control interventions on changes in risk of malaria parasitaemia in sub-Saharan Africa: a spatial and temporal analysis. Lancet Glob Health 2, e601-615. https://doi.org/10.1016/S2214-109X(14)70300-6

Giardina, F., Sogoba, N., Vounatsou, P., Sogoba, N., Vounatsou, P., 2016. Handbook of Spatial Epidemiology, in: Bayesian Variable Selection in Semiparametric and Nonstationary Geostatistical Models: An Application to Mapping Malaria Risk in Mali. Chapman and Hall/CRC. https://doi.org/10.1201/b19470-31

Giorgi, E., Diggle, P.J., Snow, R.W., Noor, A.M., 2018. Geostatistical methods for disease mapping and visualization using data from spatio-temporally referenced prevalence surveys. arXiv:1802.06359 [stat].

Golassa, L., Enweji, N., Erko, B., Aseffa, A., Swedberg, G., 2013. Detection of a substantial number of sub-microscopic Plasmodium falciparum infections by polymerase chain reaction: a potential threat to malaria control and diagnosis in Ethiopia. Malaria Journal 12, 352. https://doi.org/10.1186/1475-2875-12-352

Golding, N., Burstein, R., Longbottom, J., Browne, A.J., Fullman, N., Osgood-Zimmerman, A., Earl, L., Bhatt, S., Cameron, E., Casey, D.C., Dwyer-Lindgren, L., Farag, T.H., Flaxman, A.D., Fraser, M.S., Gething, P.W., Gibson, H.S., Graetz, N., Krause, L.K., Kulikoff, X.R., Lim, S.S., Mappin, B., Morozoff, C., Reiner, R.C., Sligar, A., Smith, D.L., Wang, H., Weiss, D.J., Murray, C.J.L., Moyes, C.L., Hay, S.I., 2017. Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals. Lancet 390, 2171–2182. https://doi.org/10.1016/S0140-6736(17)31758-0

Gosoniu, L., Msengwa, A., Lengeler, C., Vounatsou, P., 2012. Spatially Explicit Burden Estimates of Malaria in Tanzania: Bayesian Geostatistical Modeling of the Malaria Indicator Survey Data. PLOS ONE 7, e23966. https://doi.org/10.1371/journal.pone.0023966

Gosoniu, L., Veta, A.M., Vounatsou, P., 2010. Bayesian Geostatistical Modeling of Malaria Indicator Survey Data in Angola. PLoS ONE 5, e9322. https://doi.org/10.1371/journal.pone.0009322

Green, P.J., Łatuszyński, K., Pereyra, M., Robert, C.P., 2015. Bayesian computation: a summary of the current state, and samples backwards and forwards. Stat Comput 25, 835–862. https://doi.org/10.1007/s11222-015-9574-5

Groenigen, V., W, J., Stein, A., 1998. Constrained Optimization of Spatial Sampling using Continuous Simulated Annealing. Journal of Environmental Quality 27, 1078–1086. https://doi.org/10.2134/jeq1998.00472425002700050013x

Gryparis, A., Paciorek, C.J., Zeka, A., Schwartz, J., Coull, B.A., 2009. Measurement error caused by spatial misalignment in environmental epidemiology. Biostatistics 10, 258–274. https://doi.org/10.1093/biostatistics/kxn033

Guo, C., Yang, L., Ou, C.-Q., Li, L., Zhuang, Y., Yang, J., Zhou, Y.-X., Qian, J., Chen, P.-Y., Liu, Q.-Y., 2015. Malaria incidence from 2005–2013 and its associations with meteorological factors in Guangdong, China. Malaria Journal 14, 116. https://doi.org/10.1186/s12936-015-0630-6

Haerter, J.O., Berg, P., Hagemann, S., 2018. Heavy rain intensity distributions on varying time scales and at different temperatures. Journal of Geophysical Research: Atmospheres 115. https://doi.org/10.1029/2009JD013384

Handcock, M.S., Stein, M.L., 1993. A Bayesian Analysis of Kriging. Technometrics 35, 403–410. https://doi.org/10.2307/1270273

Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R.F., Brooker, S., Smith, D.L., Moyeed, R.A., Snow, R.W., 2009. A World Malaria Map: Plasmodium falciparum Endemicity in 2007. PLoS Medicine 6, e1000048. https://doi.org/10.1371/journal.pmed.1000048

Head, M.G., Goss, S., Gelister, Y., Alegana, V., Brown, R.J., Clarke, S.C., Fitchett, J.R.A., Atun, R., Scott, J.A.G., Newell, M.-L., Padmadas, S.S., Tatem, A.J., 2017. Global funding trends for malaria research in sub-Saharan Africa: a systematic analysis. Lancet Glob Health 5, e772–e781. https://doi.org/10.1016/S2214-109X(17)30245-0

Heckmann, T., Gegg, K., Gegg, A., Becht, M., 2014. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. Natural Hazards and Earth System Science 14, 259–278. https://doi.org/10.5194/nhess-14-259-2014

Hoerl, A.E., Kennard, R.W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12, 55–67. https://doi.org/10.1080/00401706.1970.10488634

Horowitz, C.R., Davis, M.H., Palermo, A.-G.S., Vladeck, B.C., 2000. Approaches to Eliminating Sociocultural Disparities in Health. Health Care Financ Rev 21, 57–74.

Hou, J., Du, L., Liu, K., Hu, Y., Zhu, Y., 2018. Characteristics of vegetation activity and its responses to climate change in desert/grassland biome transition zones in the last 30 years based on GIMMS3g. Theor Appl Climatol 1–14. https://doi.org/10.1007/s00704-018-2527-0

Hsiang, T.C., 1975. A Bayesian View on Ridge Regression. Journal of the Royal Statistical Society. Series D (The Statistician) 24, 267–268. https://doi.org/10.2307/2987923

Ikeda, T., Behera, S.K., Morioka, Y., Minakawa, N., Hashizume, M., Tsuzuki, A., Maharaj, R., Kruger, P., 2017. Seasonally lagged effects of climatic factors on malaria incidence in South Africa. Scientific Reports 7. https://doi.org/10.1038/s41598-017-02680-6

INS, 2014. Profil de pauvreté monétaire (Rapport principal de l'ECAM4). INS, Yaounde.

INS, 2013. Annuaire Statistique du Cameroun 2013. Institut National de la Statistique, Yaoundé/Cameroun.

INS, M., International ICF, 2012. Cameroun Enquête Démographique et de Santé et à Indicateurs Multiples (EDS-MICS) 2011. Yaounde.

INS, Minsante et UNICEF, 2015. Cameroun - Enquête par Grappes à Indicateurs Multiples 2014 [WWW Document]. URL http://microdata.worldbank.org/index.php/catalog/2895 (accessed 1.29.18).

Ishwaran, H., Rao, J.S., 2005. Spike and slab variable selection: Frequentist and Bayesian strategies. The Annals of Statistics 33, 730–773. https://doi.org/10.1214/009053604000001147

John Weier and Herring David, 2000. Measuring Vegetation (NDVI, EVI) [WWW Document]. URL https://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring_vegetation_1.php

Kabaghe, A.N., Chipeta, M.G., McCann, R.S., Phiri, K.S., Vugt, M. van, Takken, W., Diggle, P., Terlouw, A.D., 2017. Adaptive geostatistical sampling enables efficient identification of malaria hotspots in repeated cross-sectional surveys in rural Malawi. PLOS ONE 12, e0172266. https://doi.org/10.1371/journal.pone.0172266

Kar, N.P., Kumar, A., Singh, O.P., Carlton, J.M., Nanda, N., 2014. A review of malaria transmission dynamics in forest ecosystems. Parasites & Vectors 7, 265. https://doi.org/10.1186/1756-3305-7-265

Kazembe, L.N., Appleton, C.C., Kleinschmidt, I., 2007. Spatial analysis of the relationship between early childhood mortality and malaria endemicity in Malawi. Geospat Health 2, 41–50. https://doi.org/10.4081/gh.2007.253

Khagayi, S., Desai, M., Amek, N., Were, V., Onyango, E.D., Odero, C., Otieno, K., Bigogo, G., Munga, S., Odhiambo, F., Hamel, M.J., Kariuki, S., Samuels, A.M., Slutsker, L., Gimnig, J., Vounatsou, P., 2019. Modelling the relationship between malaria prevalence as a measure of transmission and mortality across age groups. Malaria Journal 18, 247. https://doi.org/10.1186/s12936-019-2869-9

Kibret, S., Lautze, J., McCartney, M., Wilson, G.G., Nhamo, L., 2015. Malaria impact of large dams in sub-Saharan Africa: maps, estimates and predictions. Malaria Journal 14, 339. https://doi.org/10.1186/s12936-015-0873-2

Kimbi, H.K., Sumbele, I.U., Nweboh, M., Anchang-Kimbi, J.K., Lum, E., Nana, Y., Ndip, L.M., Njom, H., Lehman, L.G., 2013. Malaria and haematologic parameters of pupils at different altitudes along the slope of Mount Cameroon: a cross-sectional study. Malaria Journal 12, 193. https://doi.org/10.1186/1475-2875-12-193

Kipruto, E.K., Ochieng, A.O., Anyona, D.N., Mbalanya, M., Mutua, E.N., Onguru, D., Nyamongo, I.K., Estambale, B.B.A., 2017. Effect of climatic variability on malaria trends in Baringo County, Kenya. Malaria Journal 16, 220. https://doi.org/10.1186/s12936-017-1848-2

Kleinschmidt, I., Bradley, J., Knox, T.B., Mnzava, A.P., Kafy, H.T., Mbogo, C., Ismail, B.A., Bigoga, J.D., Adechoubou, A., Raghavendra, K., Cook, J., Malik, E.M., Nkuni, Z.J., Macdonald, M., Bayoh, N., Ochomo, E., Fondjo, E., Awono-Ambene, H.P., Etang, J., Akogbeto, M., Bhatt, R.M., Chourasia, M.K., Swain, D.K., Kinyari, T., Subramaniam, K., Massougbodji, A., Okê-Sopoh, M., Ogouyemi-Hounto, A., Kouambeng, C., Abdin, M.S., West, P., Elmardi, K., Cornelie, S., Corbel, V., Valecha, N., Mathenge, E., Kamau, L., Lines, J., Donnelly, M.J., 2018. Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: a WHO-coordinated, prospective, international, observational cohort study. The Lancet Infectious Diseases 18, 640–649. https://doi.org/10.1016/S1473-3099(18)30172-5

Koenker, H., Arnold, F., Ba, F., Cisse, M., Diouf, L., Eckert, E., Erskine, M., Florey, L., Fotheringham, M., Gerberg, L., Lengeler, C., Lynch, M., Mnzava, A., Nasr, S., Ndiop, M., Poyer, S., Renshaw, M., Shargie, E., Taylor, C., Thwing, J., Van Hulle, S., Ye, Y., Yukich, J., Kilian, A., 2018. Assessing whether universal coverage with insecticide-treated nets has been achieved: is the right indicator being used? Malar J 17. https://doi.org/10.1186/s12936-018-2505-0

Korenromp, E., Mahiané, G., Hamilton, M., Pretorius, C., Cibulskis, R., Lauer, J., Smith, T.A., Briët, O.J.T., 2016. Malaria intervention scale-up in Africa: effectiveness predictions for health programme planning tools, based on dynamic transmission modelling. Malaria Journal 15, 417. https://doi.org/10.1186/s12936-016-1461-9

Korenromp, E.L., Hosseini, M., Newman, R.D., Cibulskis, R.E., 2013. Progress towards malaria control targets in relation to national malaria programme funding. Malar J 12, 18. https://doi.org/10.1186/1475-2875-12-18

Krnjajić, M., Draper, D., 2014. Bayesian model comparison: Log scores and DIC. Statistics & Probability Letters 88, 9–14. https://doi.org/10.1016/j.spl.2014.01.023

Kwan, M.-P., 2012. The Uncertain Geographic Context Problem. Annals of the Association of American Geographers 102, 958–968. https://doi.org/10.1080/00045608.2012.687349

Kyriakidis, P.C., Dungan, J.L., 2001. A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions. Environmental and Ecological Statistics 8, 311–330. https://doi.org/10.1023/A:1012778302005

Laguna, F., Grillet, M.E., León, J.R., Ludeña, C., 2017. Modelling malaria incidence by an autoregressive distributed lag model with spatial component. Spatial and Spatio-temporal Epidemiology 22, 27–37. https://doi.org/10.1016/j.sste.2017.05.001

Lark, R.M., 2002. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. Geoderma 105, 49–80. https://doi.org/10.1016/S0016-7061(01)00092-1

Larsen, D.A., Hutchinson, P., Bennett, A., Yukich, J., Anglewicz, P., Keating, J., Eisele, T.P., 2014. Community Coverage with Insecticide-Treated Mosquito Nets and Observed Associations with All-Cause Child Mortality and Malaria Parasite Infections. Am J Trop Med Hyg 91, 950–958. https://doi.org/10.4269/ajtmh.14-0318

Lelisa, K., Asale, A., Taye, B., Emana, D., Yewhalaw, D., 2017. Anopheline mosquitoes behaviour and entomological monitoring in southwestern Ethiopia. J Vector Borne Dis 54, 240–248.

Leonard Amekudzi, and al, 2014. The impact of climate change on malaria in coastal Ghana [WWW Document]. ResearchGate. URL https://www.researchgate.net/publication/280087746_The_impact_of_climate_change_on_malaria_in_coastal_Ghana (accessed 7.28.17).

Li, D., Zhang, J., Wu, H., 2012. Spatial data quality and beyond. International Journal of Geographical Information Science 26, 2277–2290. https://doi.org/10.1080/13658816.2012.719625

Li, Y., Brown, P., Gesink, D.C., Rue, H., 2012. Log Gaussian Cox processes and spatially aggregated disease incidence data   , Log Gaussian Cox processes and spatially aggregated disease incidence data. Stat Methods Med Res 21, 479–507. https://doi.org/10.1177/0962280212446326

Libwea, J.N., Kingue, S.R.B., Ashukem, N.T., Kobela, M., Boula, A., Sinata, K.-S., Ndombo, P.K., 2019. Assessing the causes of under-five mortality and proportion associated with pneumococcal diseases in Cameroon. A case-finding retrospective observational study: 2006–2012. PLOS ONE 14, e0212939. https://doi.org/10.1371/journal.pone.0212939

Lindén Andreas, Mäntyniemi Samu, 2011. Using the negative binomial distribution to model overdispersion in ecological count data. Ecology 92, 1414–1421. https://doi.org/10.1890/10-1831.1

Lindgren, F., Rue, H., 2015. Bayesian Spatial Modelling with R - INLA. Journal of Statistical Software 63. https://doi.org/10.18637/jss.v063.i19

Liu, L., Oza, S., Hogan, D., Perin, J., Rudan, I., Lawn, J.E., Cousens, S., Mathers, C., Black, R.E., 2015. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. The Lancet 385, 430–440. https://doi.org/10.1016/S0140-6736(14)61698-6

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing 10, 325–337. https://doi.org/10.1023/A:1008929526011

MacLehose, R.F., Hamra, G.B., 2014. Applications of Bayesian Methods to Epidemiologic Research. Curr Epidemiol Rep 1, 103–109. https://doi.org/10.1007/s40471-014-0019-z

Massoda Tonye, S.G., Kouambeng, C., Wounang, R., Vounatsou, P., 2018. Challenges of DHS and MIS to capture the entire pattern of malaria parasite risk and intervention effects in countries with different ecological zones: the case of Cameroon. Malaria Journal 17, 156. https://doi.org/10.1186/s12936-018-2284-7

Matern, B., 1986. Spatial Variation, 2nd ed, Lecture Notes in Statistics. Springer-Verlag, New York.

Mbakop, L.R., Awono-Ambene, P.H., Mandeng, S.E., Ekoko, W.E., Fesuh, B.N., Antonio-Nkondjio, C., Toto, J.-C., Nwane, P., Fomena, A., Etang, J., 2019. Malaria Transmission around the Memve'ele Hydroelectric Dam in South Cameroon: A Combined Retrospective and Prospective Study, 2000–2016. Int J Environ Res Public Health 16. https://doi.org/10.3390/ijerph16091618

Mcbratney, A., Webster, R., Burgess, T.M., 1981. The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables—II. Computers & Geosciences 7, 331–334. https://doi.org/10.1016/0098-3004(81)90077-7

Menkin-Smith, L., Winders, W.T., 2019. Malaria (Plasmodium Vivax), in: StatPearls. StatPearls Publishing, Treasure Island (FL).

Ministere de la sante publique, 2011. Plan National de Developpement Sanitaire (PNDS) 2011-2015.

Ministry of Economy, Planning and regional development, 2009. Cameroon Vision 2035.

Ministry of Public Health, 2010. Scaling up simple malaria treatment with combination of artesunate-amadiaquine [WWW Document]. URL https://www.theglobalfund.org/en/portfolio/country/grant/?k=ecd945a0-f74f-49bb-a379-64991f360475&grant=CMR-910-G07-M (accessed 4.9.18).

Minsante, INS, 2012. Enquête sur les indicateurs du paludisme.

Minsante, INS and UNFPA, 2016. Projections demographiques et estimations des cibles prioritaires des differents programmes et interventions de sante.

MoH, 2001. Sectorial Health Strategic Plan 2001-2015.

Moraga, P., Lawson, A.B., 2012. Gaussian component mixtures and CAR models in Bayesian disease mapping. Computational Statistics & Data Analysis 56, 1417–1433. https://doi.org/10.1016/j.csda.2011.11.011

Motta, P., Porphyre, T., Handel, I., Hamman, S.M., Ngu Ngwa, V., Tanya, V., Morgan, K., Christley, R., Bronsvoort, B.M. deC, 2017. Implications of the cattle trade network in Cameroon for regional disease prevention and control. Sci Rep 7, 43932. https://doi.org/10.1038/srep43932

Munier, A., Njanpop-Lafourcade, B.-M., Sauvageot, D., Mhlanga, R.B., Heyerdahl, L., Nadri, J., Wood, R., Ouedraogo, I., Blake, A., Akilimali Mukelenge, L., Anné, J.-C.B., Banla Kere, A., Dempouo, L., Keita, S., Langa, J.P.M., Makumbi, I., Mwakapeje, E.R., Njeru, I.J., Ojo, O.E., Phiri, I., Pezzoli, L., Gessner, B.D., Mengel, M., 2017. The African cholera surveillance network (Africhol) consortium meeting, 10–11 June 2015, Lomé, Togo. BMC Proceedings 11, 2. https://doi.org/10.1186/s12919-016-0068-z

Musenge, E., Chirwa, T.F., Kahn, K., Vounatsou, P., 2013. 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 Geoinf 22, 86–98. https://doi.org/10.1016/j.jag.2012.04.001

Nahum, A., Erhart, A., Mayé, A., Ahounou, D., van Overmeir, C., Menten, J., van Loen, H., Akogbeto, M., Coosemans, M., Massougbodji, A., D'Alessandro, U., 2010. Malaria incidence and prevalence among children living in a peri-urban area on the coast of benin, west Africa: a longitudinal study. Am. J. Trop. Med. Hyg. 83, 465–473. https://doi.org/10.4269/ajtmh.2010.09-0611

Nambuusi, B.B., Ssempiira, J., Makumbi, F.E., Kasasa, S., Vounatsou, P., 2019. The effects and contribution of childhood diseases on the geographical distribution of all-cause under-five mortality in Uganda. Parasite Epidemiol Control 5. https://doi.org/10.1016/j.parepi.2019.e00089

National Institute of Statistics, 2017. Cameroon 2014 MICS key findings report released - unicef mics.

National Institute of Statistics, 2013. Enquête post campagne sur l'utilisation des moustiquaires imprégnées d'insecticide a longue durée d'action. Yaounde.

National Institute of Statistics, 2011. The DHS Program - Cameroon: Standard DHS, 2011 [WWW Document]. URL https://dhsprogram.com/what-we-do/survey/survey-display-337.cfm (accessed 1.29.18).

Ndombo, P.K., Ekei, Q.M., Tochie, J.N., Temgoua, M.N., Angong, F.T.E., Ntock, F.N., Mbuagbaw, L., 2017. A cohort analysis of neonatal hospital mortality rate and predictors of neonatal mortality in a sub-urban hospital of Cameroon. Italian Journal of Pediatrics 43, 52. https://doi.org/10.1186/s13052-017-0369-5

Noor, A.M., ElMardi, K.A., Abdelgader, T.M., Patil, A.P., Amine, A.A.A., Bakhiet, S., Mukhtar, M.M., Snow, R.W., 2012. Malaria Risk Mapping for Control in the Republic of Sudan. American Journal of Tropical Medicine and Hygiene 87, 1012–1021. https://doi.org/10.4269/ajtmh.2012.12-0390

O'Hara, R.B., Sillanpää, M.J., 2009. A review of Bayesian variable selection methods: what, how and which. Bayesian Anal. 4, 85–117. https://doi.org/10.1214/09-BA403

Okuneye, K., Eikenberry, S.E., Gumel, A.B., 2019. Weather-driven malaria transmission model with gonotrophic and sporogonic cycles. Journal of Biological Dynamics 0, 1–37. https://doi.org/10.1080/17513758.2019.1570363

Organisation mondiale de la santé, 2013. Recruscence des cas de paludisme dans l'Extrême Nord du Cameroun.

Pandey, P.K., Das, L., Jhajharia, D., Pandey, V., 2018. Modelling of interdependence between rainfall and temperature using copula. Model. Earth Syst. Environ. 4, 867–879. https://doi.org/10.1007/s40808-018-0454-9

Papaioannou, I., Utzinger, J., Vounatsou, P., 2019. Malaria-anemia comorbidity prevalence as a measure of malaria-related deaths in sub-Saharan Africa. Sci Rep 9, 11323. https://doi.org/10.1038/s41598-019-47614-6

Perez-Haydrich, C., Warren, J.L., Burgert, C.R., Emch, M.E., 2013. Guidelines on the use of DHS GPS data.

PNLP, 2016. Rapport D'activites 2016 du Programme National de Lutte Contre Le Paludisme. Minsante, Yaoundé/Cameroun.

PNLP, 2014a. Rapport de lutte contre le paludisme 2014.

PNLP, 2014b. Plan Strategique National de Lutte contre le Paludisme 2014-2018.

PNLP, 2013. Rapport D'activites 2013 du Programme National de Lutte Contre Le Paludisme. Ministere de la sante publique.

PNLP, 2012. Rapport D'activites 2012 du Programme National de Lutte Contre Le Paludisme. Ministere de la sante publique.

PNLP, 2011. Rapport D'activites 2011 du Programme National de Lutte Contre Le Paludisme. Ministere de la sante publique.

PNLP, M., 2007. Profil entomologique du paludisme au Cameroun.

PNLP, Minsante, 2013. Plan de suivi-evaluation du plan strategique national de lutte contre le paludisme 2011-2015.

PNLP, Minsante, M., 2020. Rapport d'activites 2019 de lutte contre le paludisme. Yaoundé/Cameroun.

Polson, N.G., Scott, J.G., 2011. Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction *. Oxford University Press.

Pothin, E., Ferguson, N.M., Drakeley, C.J., Ghani, A.C., 2016. Estimating malaria transmission intensity from Plasmodium falciparum serological data using antibody density models. Malaria Journal 15, 79. https://doi.org/10.1186/s12936-016-1121-0

Programme National de Lutte contre le Paludisme-Cameroun., 2007. Profil entomologique du paludisme au Cameroun.

R Core Team, 2016. R: A Language and Environment for Statistical Computing [WWW Document]. R Foundation for Statistical Computing. URL https://www.R-project.org/

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

RBM et al, 2008. The global malaria action plan.

RBM Monitoring and Evaluation Group, 2005. Malaria indicator survey : basic documentation for survey design and implementation. World Health Organization.

Ribeiro, D., 2017. Model-based Geostatistics | Peter Diggle | Springer, Springer Series in Statistics. ed. Springer-Verlag New York.

Rich, S.M., Ayala, F.J., 2006. Evolutionary Origins of Human Malaria Parasites, in: Dronamraju, K.R., Arese, P. (Eds.), Malaria: Genetic and Evolutionary Aspects, Emerging Infectious Diseases of the 21st Century. Springer US, Boston, MA, pp. 125–146. https://doi.org/10.1007/0-387-28295-5_6

Roll Back Malaria, USAID, CDC, UNICEF, WHO, 2013. Household survey indicators for malaria control.

Royle, J.A., Nychka, D., 1998. An algorithm for the construction of spatial coverage designs with implementation in SPLUS. Computers & Geosciences 24, 479–488. https://doi.org/10.1016/S0098-3004(98)00020-X

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71, 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Russell, L.B., Gold, M.R., Siegel, J.E., Daniels, N., Weinstein, M.C., 1996. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. JAMA 276, 1172–1177.

Russo, D., 1984. Design of an Optimal Sampling Network for Estimating the Variogram 1. Soil Science Society of America Journal 48, 708–716. https://doi.org/10.2136/sssaj1984.03615995004800040003x

Saeed, B.I.I., Yawson, A.E., Nguah, S., Agyei-Baffour, P., Emmanuel, N., Ayesu, E., 2016. Effect of socio-economic factors in utilization of different healthcare services among older adult men and women in Ghana. BMC Health Services Research 16, 390. https://doi.org/10.1186/s12913-016-1661-6

Sakwe, N., Bigoga, J., Ngondi, J., Njeambosay, B., Esemu, L., Kouambeng, C., Nyonglema, P., Seumen, C., Gouado, I., Oben, J., 2019. Relationship between malaria, anaemia, nutritional and socio-economic status amongst under-ten children, in the North Region of Cameroon: A cross-sectional assessment. PLOS ONE 14, e0218442. https://doi.org/10.1371/journal.pone.0218442

Samadoulougou, S., Maheu-Giroux, M., Kirakoya-Samadoulougou, F., De Keukeleire, M., Castro, M.C., Robert, A., 2014. Multilevel and geo-statistical modeling of malaria risk in children of Burkina Faso. Parasit Vectors 7. https://doi.org/10.1186/1756-3305-7-350

Scheme, W., 2013. Guidelines for laboratory and field-testing of long-lasting insecticidal nets. World Health Organization.

Schmidt, M., Klein, D., Conrad, C., Dech, S., Paeth, H., 2014. On the relationship between vegetation and climate in tropical and northern Africa. Theor Appl Climatol 115, 341–353. https://doi.org/10.1007/s00704-013-0900-6

Schrödle, B., Held, L., 2011. Spatio-temporal disease mapping using INLA. Environmetrics 22, 725–734. https://doi.org/10.1002/env.1065

Sharma, V.P., 2006. Vector Genetics in Malaria Control, in: Dronamraju, K.R., Arese, P. (Eds.), Malaria: Genetic and Evolutionary Aspects, Emerging Infectious Diseases of the 21st Century. Springer US, Boston, MA, pp. 147–167. https://doi.org/10.1007/0-387-28295-5_7

Siwal, N., Singh, U.S., Dash, M., Kar, S., Rani, S., Rawal, C., Singh, R., Anvikar, A.R., Pande, V., Das, A., 2018. Malaria diagnosis by PCR revealed differential distribution of mono and mixed species infections by Plasmodium falciparum and P. vivax in India. PLoS ONE 13, e0193046. https://doi.org/10.1371/journal.pone.0193046

Skinner, C.J., 2016. Probability Proportional to Size (PPS) Sampling, in: Wiley StatsRef: Statistics Reference Online. American Cancer Society, pp. 1–5. https://doi.org/10.1002/9781118445112.stat03346.pub2

Snow, R.W., Kibuchi, E., Karuri, S.W., Sang, G., Gitonga, C.W., Mwandawiro, C., Bejon, P., Noor, A.M., 2015. Changing Malaria Prevalence on the Kenyan Coast since 1974: Climate, Drugs and Vector Control. PLOS ONE 10, e0128792. https://doi.org/10.1371/journal.pone.0128792

Ssempiira, J., Kissa, J., Nambuusi, B., Kyozira, C., Rutazaana, D., Mukooyo, E., Opigo, J., Makumbi, F., Kasasa, S., Vounatsou, P., 2018. The effect of case management and vector-control interventions on space–time patterns of malaria incidence in Uganda. Malaria Journal 17, 162. https://doi.org/10.1186/s12936-018-2312-7

Ssempiira, J., Nambuusi, B., Kissa, J., Agaba, B., Makumbi, F., Kasasa, S., Vounatsou, P., 2017. Geostatistical modelling of malaria indicator survey data to assess the effects of interventions on the geographical distribution of malaria prevalence in children less than 5 years in Uganda. PLOS ONE 12, e0174948. https://doi.org/10.1371/journal.pone.0174948

Stresman, G.H., 2010. Beyond temperature and precipitation: Ecological risk factors that modify malaria transmission. Acta Tropica 116, 167–172. https://doi.org/10.1016/j.actatropica.2010.08.005

Sturtz, S., Ligges, U., Gelman, A., 2005. R2WinBUGS : A Package for Running WinBUGS from R. Journal of Statistical Software 12. https://doi.org/10.18637/jss.v012.i03

Swiss TPH et al, 1996. Mapping Malaria Risk in Africa (MARA) [WWW Document]. URL https://www.mara-database.org//login.html (accessed 10.1.19).

Tabue, R.N., Awono-Ambene, P., Etang, J., Atangana, J., C, A.-N., Toto, J.C., Patchoke, S., Leke, R.G.F., Fondjo, E., Mnzava, A.P., Knox, T.B., Tougordi, A., Donnelly, M.J., Bigoga, J.D., 2017. Role of Anopheles (Cellia) rufipes (Gough, 1910) and other local anophelines in human malaria transmission in the northern savannah of Cameroon: a cross-sectional survey. Parasites & Vectors 10, 22. https://doi.org/10.1186/s13071-016-1933-3

Tabue, R.N., Nem, T., Atangana, J., Bigoga, J.D., Patchoke, S., Tchouine, F., Fodjo, B.Y., Leke, R.G., Fondjo, E., 2014. Anopheles ziemanni a locally important malaria vector in Ndop health district, north west region of Cameroon. Parasites & Vectors 7, 262. https://doi.org/10.1186/1756-3305-7-262

Tandi, T.E., Cho, Y., Akam, A.J.-C., Afoh, C.O., Ryu, S.H., Choi, M.S., Kim, K., Choi, J.W., 2015. Cameroon public health sector: shortage and inequalities in geographic distribution of health personnel. Int J Equity Health 14, 43. https://doi.org/10.1186/s12939-015-0172-0

Tanser, F.C., Sharp, B., le Sueur, D., 2003. Potential effect of climate change on malaria transmission in Africa. The Lancet 362, 1792–1798. https://doi.org/10.1016/S0140-6736(03)14898-2

Tatem, A.J., Garcia, A.J., Snow, R.W., Noor, A.M., Gaughan, A.E., Gilbert, M., Linard, C., 2013. Millennium development health metrics: where do Africa's children and women of childbearing age live? Population Health Metrics 11, 11. https://doi.org/10.1186/1478-7954-11-11

Tchinda, V.H.M., Socpa, A., Keundo, A.A., Zeukeng, F., Seumen, C.T., Leke, R.G.F., Moyou, R.S., 2012. Factors associated to bed net use in Cameroon: a retrospective study in Mfou health district in the Centre Region. Pan Afr Med J 12.

Tebeu, P.-M., Halle-Ekane, G., Da Itambi, M., Mbu, R.E., Mawamba, Y., Fomulu, J.N., 2015. Maternal mortality in Cameroon: a university teaching hospital report. Pan Afr Med J 21. https://doi.org/10.11604/pamj.2015.21.16.3912

The Global Fund, 2006. The Global Fund annual report 2006 [WWW Document]. URL https://www.theglobalfund.org/media/1332/corporate_2006annual_report_en.pdf?u=6371532 79440000000 (accessed 5.17.19).

The World Bank, 2016. World Development Indicators 2016.

The World Bank, 2016. Country Economic Memorandum.

Thomas, C.J., Cross, D.E., Bøgh, C., 2013. Landscape Movements of Anopheles gambiae Malaria Vector Mosquitoes in Rural Gambia. PLOS ONE 8, e68679. https://doi.org/10.1371/journal.pone.0068679

Travis, P., Bennett, S., Haines, A., Pang, T., Bhutta, Z., Hyder, A.A., Pielemeier, N.R., Mills, A., Evans, T., 2004. Overcoming health-systems constraints to achieve the Millennium Development Goals. The Lancet 364, 900–906. https://doi.org/10.1016/S0140-6736(04)16987-0

Tusting, L.S., Bottomley, C., Gibson, H., Kleinschmidt, I., Tatem, A.J., Lindsay, S.W., Gething, P.W., 2017. Housing Improvements and Malaria Risk in Sub-Saharan Africa: A Multi-Country Analysis of Survey Data. PLoS Med 14. https://doi.org/10.1371/journal.pmed.1002234

UNDP, 2015. Human Development Report 2015.

UNICEF, 1995. A pratical handbook for Multiple Indicator Cluster Survey (MICS).

United Nations, 2015. Sustainable Development Goals .:. Sustainable Development Knowledge Platform [WWW Document]. URL https://sustainabledevelopment.un.org/?menu=1300 (accessed 5.17.19).

University of Southampton, 2013. WorldPop datasets [WWW Document]. URL https://www.worldpop.org/focus_areas

University of Columbia, 1997. Socioeconomic Data and Applications Center (sedac) [WWW Document]. URL http://sedac.ciesin.columbia.edu/data/set/grump-v1-population/data-download

U.S. Geological Survey, 2018. MODIS Products Table | LP DAAC :: NASA Land Data Products and Services [WWW Document]. URL https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table (accessed 3.5.18).

U.S. Geological Survey, 2017. MODIS Products Table [WWW Document]. URL https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table

U.S. Geological Survey, 1985. FEWSNET Data Portals [WWW Document]. URL https://earlywarning.usgs.gov/fews/datadownloads/Continental%20Africa/Dekadal%20RFE

USAID, 1997. DHS Analytical reports.

Van de Pas, R., Hill, P.S., Hammonds, R., Ooms, G., Forman, L., Waris, A., Brolan, C.E., McKee, M., Sridhar, D., 2017. Global health governance in the sustainable development goals: Is it grounded in the right to health? Global Chall 1, 47–60. https://doi.org/10.1002/gch2.1022

Wanzirah, H., Tusting, L.S., Arinaitwe, E., Katureebe, A., Maxwell, K., Rek, J., Bottomley, C., Staedke, S.G., Kamya, M., Dorsey, G., Lindsay, S.W., 2015. Mind the Gap: House Structure and the Risk of Malaria in Uganda. PLoS One 10. https://doi.org/10.1371/journal.pone.0117396

Warren, A.E., Wyss, K., Shakarishvili, G., Atun, R., de Savigny, D., 2013. Global health initiative investments and health systems strengthening: a content analysis of global fund investments. Global Health 9, 30. https://doi.org/10.1186/1744-8603-9-30

Warren, J.L., Perez-Heydrich, C., Burgert, C.R., Emch, M.E., 2016a. Influence of Demographic and Health Survey Point Displacements on Point-in-Polygon Analyses. Spat Demogr 4, 117–133. https://doi.org/10.1007/s40980-015-0015-z

Warren, J.L., Perez-Heydrich, C., Burgert, C.R., Emch, M.E., 2016b. Influence of Demographic and Health Survey Point Displacements on Distance-Based Analyses. Spat Demogr 4, 155–173. https://doi.org/10.1007/s40980-015-0014-0

White, N.J., Pukrittayakamee, S., Hien, T.T., Faiz, M.A., Mokuolu, O.A., Dondorp, A.M., 2014. Malaria. The Lancet 383, 723–735. https://doi.org/10.1016/S0140-6736(13)60024-0

WHO, 2019. World malaria report 2019 [WWW Document]. URL https://www.who.int/publications-detail-redirect/9789241565721 (accessed 11.1.20).

WHO, 2018a. World Malaria Report 2018 [WWW Document]. WHO. URL http://www.who.int/malaria/publications/world-malaria-report-2018/en/ (accessed 7.25.19).

WHO, 2018b. Malaria surveillance, monitoring & evaluation: a reference manual. World health Organization Press, Geneva, Switzerland.

WHO, 2017a. World malaria report 2017.

WHO, 2017b. OMS Cameroun : Rapport annuel 2017. Yaounde.

WHO, 2017c. WHO | Achieving and maintaining universal coverage with long-lasting insecticidal nets for malaria control.

WHO, 2016. Global technical strategy for malaria 2016-2030.

WHO, 2015a. Treatment of Plasmodium vivax, P. ovale, P. malariae and P. knowlesi infections. World Health Organization.

WHO, 2015b. Guidelines for the treatment of malaria - Third edition.

WHO, 2015c. World Malaria Report 2015.

WHO, 2013a. Seasonal malaria chemoprevention with sulfadoxine-pyrimethamine plus amodiaquine in children: A field guide [WWW Document]. WHO. URL https://www.who.int/malaria/publications/atoz/9789241504737/en/ (accessed 7.31.19).

WHO, 2013b. WHO policy brief for the implementation of intermittent preventive treatment of malaria in pregnancy using sulfadoxine-pyrimethamine (IPTp-SP) [WWW Document]. WHO. URL https://www.who.int/malaria/publications/atoz/policy_brief_iptp_sp_policy_recommendation/en/ (accessed 7.31.19).

William G. Cochran, 1977. Sampling Techniques, 3rd Edition.

World Health Organization, 2017. A framework for malaria elimination.

World Health Organization, 2016. Eliminating malaria (No. WHO/HTM/GMP/2016.28).

World Health Organization, 2009. Global Health Risks - WHO - OMS -. Geneva, Switzerland.

Zacarias, O.P., Andersson, M., 2011. Spatial and temporal patterns of malaria incidence in Mozambique. Malaria Journal 10, 189. https://doi.org/10.1186/1475-2875-10-189

Zelman, B., Melgar, M., Larson, E., Phillips, A., Shretta, R., 2016. Global fund financing to the 34 malaria-eliminating countries under the new funding model 2014–2017: an analysis of national allocations and regional grants. Malar J 15. https://doi.org/10.1186/s12936-016-1171-3

Zhao, X., Chen, F., Feng, Z., Li, X., Zhou, X.-H., 2014. The temporal lagged association between meteorological factors and malaria in 30 counties in south-west China: a multilevel distributed lag non-linear analysis. Malar J 13, 57. https://doi.org/10.1186/1475-2875-13-57

Zidek, J.V., Shaddick, G., Taylor, C.G., 2014. Reducing estimation bias in adaptively changing monitoring networks with preferential site selection. The Annals of Applied Statistics 8, 1640–1670. https://doi.org/10.1214/14-AOAS745

Curriculum vitae

Date and place of birth	March 19, 1982 in Nanga Eboko, Cameroon
Nationality	Cameroonian
E-mail	massoda.salomon@gmail.com,
Address	Byfangweg 6, 4051 Basel, Switzerland

EDUCATION

2016-2020	PhD thesis " Survey design effects on geostatistical model-based estimates of malaria burden and effects of control interventions" (2020) under the supervision of Dr. P. Vounatsou and Prof. Dr. J. Utzinger at the Swiss Tropical Institute, University of Basel, Basel, Switzerland.
2014-2015	Master of Advanced Studies in Public Health (MPH), Pasteur-CNAM School of Public Health, Paris, France. Thesis: Detection and forecasting of malaria outbreak in Cameroon
2008-2010	Master of Science (M.Sc.) in Statistic and Econometric at the University of Toulouse 1 Capitole, France. Thesis: Analysis of malaria transmission in Cameroon: approach based on morbidity data.
2004-2008	Diploma of Engineer in Applied Statistic at the Regional Institute of Statistic and Economy of Yaoundé, Cameroon.
2003-2005	Master of Science (M.Sc.) in Computer Science, University of Yaoundé 1, Cameroon (without Thesis).
2000-2003	Bachelor (B.Sc) in Computer Science, University of Yaoundé 1, Cameroon.

PROFESSIONAL EXPERIENCE AND TEACHINGS

2010-2016	Biostatistician National Malaria Control Program, Cameroon.
2010-2014	Assistant in computer science courses at Regional Institute of Statistic and Economy of Yaoundé, Cameroon.

Massoda Tonye, Salomon G., Celestin Kouambeng, Romain Wounang, and Penelope Vounatsou. 2018. "Challenges of DHS and MIS to Capture the Entire Pattern of Malaria Parasite Risk and Intervention Effects in Countries with Different Ecological Zones: The Case of Cameroon." Malaria Journal 17: 156.

G. Nko'Ayissi, W. Mbacham, **G. Massoda**. Morbidité du paludisme dans les faciès camerounais : analyse des données de surveillance de 36 mois sur les groupes à risque. Médecine d'Afrique Noire 5909 - Août/sept. 2012 - pages 421-429