



Spatial Modelling of Malaria Incidence and its Risk Factors in Namibia

by:

Remember Ndahalashili

Katale

214044866

*A thesis submitted in partial fulfilment of the requirements
for the degree of Master of Science in Applied Statistics*

in the

Department of Mathematics and Statistics
Faculty of Health and Applied Sciences
Namibia University of Science and Technology

Supervisors:

Main supervisor: Dr Dibaba Bayisa Gemechu

Co-Supervisors:

Mr Johannes Swartz and
Mr Etuhole Moshili Mwahi

May 2022

DECLARATION

I, Remember Ndahalashili Katale, hereby declare that this study "*Spatial Modelling of Malaria Incidence and its Risk Factors in Namibia*" in partial fulfillment for master's degree in Applied Statistics is my own original work. Without the prior written Permission of the author or NUST. This report may not be copied, stored in any retrieval system or transmitted in any form or by any means (electronic, recording, photocopying, mechanical, or any other method).

I, Remember Ndahalashili Katale, grant NUST permission to reprint this thesis in whole or in part, in any style or format NUST deems suitable.

Ms R.N. Katale

Author

Signature.....*R.Katale*.....

Date.....12 *May* 2022.....

CERTIFICATION

This is to certify that this thesis titled *“Spatial Modelling of Malaria Incidence and its Risk Factors in Namibia”* was undertaken by **Remember Ndahalashili Katale** at the department of Mathematics and Statistics, Namibia Universtiy of Science and Technology in partial fulfillment for the requirements of the award of the Master degree in Applied Statistics.

Dr D.B. Gemechu

Supervisor

Signature..... 

Date..... 30/05/2022

Mr J. Swartz

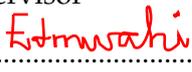
First co-supervisor

Signature..... 

Date..... 30/05/2022

Mr E.M. Mwahi

Second co-supervisor

Signature..... 

Date..... 30/05/2022

Mr B.E Obabueki

Head of Department

Signature..... 

Date..... 30/5/2022

Prof L. Kazembe

External Examiner

Signature..... 

Date..... 31/05/2022

DEDICATION

I dedicate this thesis to my Mother Ulalia Mumangeni and my late father Natangwe Ndakoneka Katale. I thank you for all you have done for me. You were my inspiration through out this study. I thank you very much for all the efforts.

Abstract

Malaria distribution is known to be geographical and temporal heterogeneous, with cases fluctuating across space and time, and climatic conditions are largely connected with regard to malaria occurrence, both temporally and spatially. Millions of dollars have been spent on malaria control in Namibia to achieve the goal of reducing malaria incidence from 13 to less than 1 malaria case per 1000 population in 2016 and becoming malaria free by 2020. However, malaria still remains a major public health challenge in Namibia, primarily in the Kavango West and East, Ohangwena, and Zambezi regions. The primary purpose of this research was to fit a spatial model to profile spatial variation in malaria incidence (MoHSS) and to investigate possible associations between disease risk and environmental factors in these areas. To explain disease trends, identify malaria risk factors, and locate malaria hotspots, the INLA package in R software was used to fit a range of models, including non-spatial, spatial, and spatio-temporal models. Malaria data for 2018 to 2020 were obtained from the Ministry of Health and Social Services, while monthly weather data for 2018 to 2020 were obtained from SASSCAL, and population estimates for each constituency were used to project the population for 2018 to 2020. Since more than 96% of the 2018-2019 reported malaria cases were from the Kavango East and West, Zambezi, and Ohangwena regions, and more than 80% in 2020, this study was restricted to those areas. A hierarchical Bayesian CAR model was fitted to these datasets to investigate climatic and other related factors that could explain the spatial/temporal variation of malaria infection in Namibia. Average rainfall received on an annual basis and maximum temperature were found to have a significant spatial and temporal variation on malaria infection. Every mm increase in annual rainfall in a specific constituency in each year increases annual mean malaria cases by 0.6% in that constituency. Also, for every one °C increase in annual maximum temperature in a certain constituency, it will increase the annual mean cases of malaria by

0.6%. The posterior means of the time main effect (year t) showed a visible slightly increasing global trend from 2018 to 2020. Constituencies in the Kavango outskirts East and West regions revealed a high spatial risk and posterior relative risk (RR: 1.57 to 1.78). Both unstructured random effects (spatial and temporal) as well as temporal structured random effects revealed a significant variation of malaria. Future studies should consider examining all possible putative sources of malaria transmission including travel histories and networks, and treatment seeking behavior and should mostly focus on finding and mapping potential anopheles mosquito habitat that was missed in this study due to a lack of information in the datasets on anopheles mosquito breeding locations (e.g., irrigated agriculture).

ACKNOWLEDGEMENTS

First, I would like to thank God for making this possible and accompanying me this far. Secondly, I would like to register my gratitude to my supervisor Dr Dibaba Bayisa Gemechu for his excellent guidance, patience and for being supportive throughout this study. His encouragement has been a driving force during the period of this study, and he has been my light throughout this project. Thank you for taking the time to listen to my issues. Your advice and coaching have been greatly appreciated. It has been an honor to be mentored by you, and I aim to make you proud by keeping the scientific and personal principles you instilled in me.

I am also grateful to the NUST Registrar's Office for sponsoring my first three semesters of my master's degree while serving their office as an assistant to the Registrar (Statistician).

To Mr Andrias Shipanga, Mrs Natalia Tulimuwo Mungoba, and Mr Vellem Mayer, all from MoHSS, I am grateful for your patience, even though I had many unsettling situations, and for the help you provided me in order for me to complete my research.

Special thanks go to my loving mother Ulalia Mumangeni and my late father Natangwe Ndakoneka Katale for all the support, courage and hope they gave me during difficult times and making sure that i persevered up to the end.

CONTENTS

DECLARATION	ii
CERTIFICATION	iii
Dedication	iv
Abstract	v
Acknowledgements	vii
1 INTRODUCTION	1
1.1 Background	1
1.2 Problem statement	5
1.3 Objectives of the study	7
1.3.1 Main objective	7
1.3.2 The specific objectives	7
1.4 The significance of the study	7
1.5 Thesis structure	8
2 LITERATURE REVIEW	9
2.1 Introduction	9
2.2 Review of related studies	10
2.3 Types of spatial data	12
2.4 Review on statistical methods of spatial data analysis	14
2.4.1 Cluster analysis	14
2.4.2 Spatial autocorrelation	15
2.5 Generalized Linear Models	19
2.5.1 Introduction	19
2.5.2 Generalized Linear Mixed Models (GLMMs)	20
2.5.3 Conditional Autoregressive Regression models (CARs)	22

2.5.4	Bayesian Hierarchical approach	23
2.5.5	The Likelihood Function	23
2.5.6	Prior Distributions	24
2.5.7	Posterior Distributions	27
2.5.8	Integrated Nested Laplace approximation (INLA)	27
3	DATA AND METHODOLOGY	29
3.1	Study area	29
3.2	Study design	29
3.3	Data	31
3.3.1	Population at risk	31
3.3.2	Malaria data	33
3.3.3	Environmental/weather /climatic data of the constituency (2017 - 2020)	37
3.4	Ethical considerations	38
3.5	Data management	38
3.5.1	Data cleaning	38
3.5.2	Standardised Incidence Ratio (SIR)	39
3.6	Model building for count data	40
3.6.1	Poisson regression model (non-spatial model)	41
3.6.2	Negative binomial regression	43
3.6.3	Model specification (spatial model and spatial temporal)	45
3.7	Parameters estimation, spatial autocorrelation, and best model selection criteria	54
3.8	Model selection criterion	56
3.8.1	Akaike Information Criterion (AIC)	56
3.8.2	Bayesian Information Criterion (BIC)	57
3.8.3	Deviance Information Criterion (DIC)	57
4	RESULTS	59
4.1	Description and distribution of 2018 to 2020 malaria incidence rate in Kavango E and W, Zambezi and Ohangwena region	59

4.2	Non-spatial examination of malaria risk variables in Namibia (Non-spatial modelling)	68
4.3	Modeling spatial patterns of malaria	74
4.3.1	Results of the spatial model	78
4.4	Modelling spatial and time patterns of malaria	83
4.4.1	Results of the spatio-temporal model	85
5	DISCUSSIONS, CONCLUSIONS AND RECOMMENDATIONS	93
5.1	Discussions	93
5.2	Conclusions	97
5.3	Recommendations	99
	Bibliography	100
A	R code	107
A.1	non-spatial and spatial Analysis	107
A.2	spatio - temporal Analysis	120
B	Malaria active case detection questionnaire	126
C	Additional marerial	127
C.1	Maps for the spatio-temporal model without added covariaariates effects	127

LIST OF FIGURES

1.1	Namibia malaria transmission map (MoHSS, 2014)	3
3.1	Map of the 4 regions (Kavango West and East, Zambezi and Ohangwena)	32
3.2	Map of all constituencies in the 4 regions	33
4.1	Trend analysis of malaria (2018-2020)	60
4.2	Constituency names in the 4 regions map	62
4.3	Malaria Incidence rate per 1000 population at constituency level	63
4.4	Annual average malaria incidence rate per 1000 population per season	65
4.5	Malaria risk map (unpredicted)	67
4.6	Neighbourhoods at constituency level	75
4.7	Moran’s I scatter plot	76
4.8	Local Moran’s I statistics map (2018 -2020).	77
4.9	Getis - Ord Gi statistic cluster map I significant cluster map (2018 - 2020)	77
4.10	Local Moran’s I significant cluster map (2018 - 2020)	78
4.11	Global linear temporal trend for malaria (spatio temporal model with added covariates)	87
4.12	Spatial main effect and differential temporal maps of the spatiotemporal model 3 (posterior mean obtained using a random walk of order 2)	88
4.13	Posterior mean for malaria incidence (Nonspatially or Temporally interaction) of the non-parametric spatio-temporal model using RW2 with added covariate effect	90
4.14	Posterior mean for malaria incidence (Temporally Structured interaction) of the non-parametric spatio-temporal model using RW2 with added covariate effect	91

4.15	Posterior mean for malaria incidence (Spatially Structured interaction) of the non-parametric spatio-temporal model using RW2 with added covariate effect	92
B.1	Malaria active case detection questionnaire	126
C.1	Spatial main effect and differential temporal maps of the spatiotemporal model without added covariates (posterior mean obtained using a random walk of order 2)	127
C.2	Posterior mean for malaria of the nonspatially or temporal interaction model without added covariates using RW2	128
C.3	Posterior mean for malaria of the temporal structured interaction model without added covariates using random walk of order 2	129
C.4	Posterior mean for malaria of the spatial structured interaction model without added covariates using random walk of order 2	130

LIST OF TABLES

1.1	OPD and IPD reported cases and death cases for Malaria in Namibia from 2001 to 2020 (MoHSS, 2018)	4
1.2	Reported OPD malaria cases by region, 2010 - 2018	6
3.1	Total cases of malaria per year per region and health districts (2018 – 2020)	35
3.2	Sample size for Individual dataset 2017 to 2019	37
3.3	List of merged datasets	38
3.4	List of interaction considered in the study	51
3.5	List of Analysis performed in the study	56
4.1	Total number of people suspected to have malaria and tested positive per region per season	61
4.2	Summary of the 11 constituency that revealed the highest malaria incidence rate per 1000 population in 2018, 2019, and 2020	64
4.3	Summary of computed malaria incidence per year	66
4.4	Variable considered for non-spatial Negative Binomial regression model (aggregated data 2018-2020)	69
4.5	Variable considered for non-spatial Negative Binomial regression model (Individual data 2017-2019)	69
4.6	Comparisons of the non-spatial implemented models (2018 – 2020 aggregated data)	70
4.7	Poisson and Negative Binomial regression non-spatial model (aggregated data 2018 -2020)	71
4.8	Negative Binomial regression non-spatial model (Individual data 2017–2019)	73

4.9	Comparisons of the spatial implemented models (2018 – 2020 aggregated data)	80
4.10	Pearson’s correlation between sum of positive tested malaria cases with human population density as well as climatic variables included in the spatial model (aggregated data 2018 – 2020)	80
4.11	log scale Parameter for the Negative Binomial BYM model (aggregated data 2018 - 2020)	81
4.12	Comparison of the two models (Individual data 2017-2019)	83
4.13	log scale Parameter for the Bayesian Negative Binomial BYM model (Individual data 2017 - 2019)	84
4.14	Comparison of the spatio-temporal models (aggregated data 2018-2020)	86
4.15	Log scale parameter for the Bayesian Negative Binomial (BYM) unstructured interactive spatio–temporal model with added significant climatic variables from spatial model using 2018-2020 aggregated dataset	86

LIST OF ABBREVIATIONS

ACD	Active Case Detection
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BYM	Besag-York-Mollié model
CAR	Conditional Auto-Regressive
CSR	Complete Spatial Randomness
DIC	Deviance Information Criterion
DHIS2	District Health Information Software 2
EA	Enumeration Area
GIS	Geographic Information Systems
GLM	Generalized Linear Model
GOF	Goodness Fit
GLMM	Generalized Linear Mixed Model
ICAR	Intrinsic Conditional Auto-Regressive
IPD	Inpatient Diagnosis
LMM	Linear Model
LMM	Linear Mixed Model
MAWF	Ministry of Agriculture, Water and Forestry
MCH	Maternal and Child Health
MIS	Malaria Indicator Survey
MoHSS	Ministry of Health and Social Services
OPD	Outpatient Diagnosis
PCR	Polymerase Chain Reaction
PHS	Public Health Sectors
RDT	Rapid Diagnosis Test
SASSCAL	The Southern African Science Service Centre for

STAR Structured Additive Regression
WHO World Health Organization

CHAPTER 1

INTRODUCTION

1.1 Background

Malaria is the most lethal protozoan illness, killing over 750,000 people in Africa each year, the majority of whom are children and pregnant women, with 435,000 deaths occurring among children below the age of five (Greenwood, 2012; Hussein, 2020; WHO, 2021; Tseha, 2021) and every two minutes, a child under the age of five dies as a result of this disease (Organization, 2016).

Malaria affects more than half of the world's population (Hamza et al., 2015) and an estimated 3.2 billion people globally are still at risk of contracting malaria (WHO, 2016). The World Health Organization reports that roughly 214 million new cases of malaria were reported in 2014, resulting in 428,000 fatalities worldwide. In 2015, the WHO African Region reported 88% of malaria cases, the WHO South-East Region reported 10%, and the WHO Eastern Mediterranean Region recorded 2% (WHO, 2016). In 2018, there were around 228 million cases of malaria and 405,000 deaths worldwide, with Africa having the largest number of cases and fatalities (Hussein et al., 2020). Malaria killed an estimated 409 000 million people worldwide in 2019, with African countries bearing the brunt of the toll (Aborode, 2021). Young children (children under the age of five) are the most prone to malaria, accounting for 67% (274 000) of all malaria fatalities globally in 2019. The WHO's African Region bears a disproportionately substantial amount of the world's malaria incidence. In 2019, the region was responsible for 94% of malaria infections and deaths (Abel, 2021).

Malaria is also one of the top ten primary causes of illness and mortality in the majority of Sub-Saharan African nations, including Namibia (WHO, 2016). According to the World Health Organization (WHO) estimates from 2018, there were around 228 million cases of malaria and 405,000 deaths worldwide, with Africa having the largest number of cases and fatalities (Hussein et al., 2020). Malaria killed an estimated 229 million people worldwide in 2019, with African countries bearing the brunt of the toll (Aborode, 2021). In 2019, there were an estimated 229 million cases of malaria all across the world, with an estimated 409 000 malaria fatalities (WHO, 2021). Young children (those under the age of five) are the most susceptible to malaria accounting for 67% (274 000) of all malaria fatalities globally in 2019. The WHO's African Region bears a disproportionately substantial amount of the global malaria burden. In 2019, the region was responsible for 94% of malaria infections and deaths (Abel, 2021). This is due to climate ideal for the Anopheles mosquitoes that transmit malaria parasites e.g., warm temperatures, humid conditions, and high rainfall. The sickness is the result of parasites that are spread to humans through the bites of infected female Anopheles mosquitos (Turner et al., 2013). Malaria affects both men and women of all ages, with children under the age of five and pregnant women being the most vulnerable (Lamb, 2012).

Furthermore, malaria's economic impact in Africa was estimated to cost approximately \$ 12 billion per year, which has a negative impact on business because the disease is responsible for employee absenteeism, increases in healthcare expenditure/spending, and decreases in productivity, all of which can have a negative impact on a company's reputation (Worrall, Rietveld, and Delacollette, 2004). Malaria is also to blame for the consequences of poor economic growth, which contributes to an increase in malaria cases, mortality, and morbidity rates in Namibia, particularly in the north (Union, 2006).

When it comes to country malaria transmission risk, Namibia is divided into three zones and among the three zones, zone 1 and 2 are still at risk of malaria where zone 1 is classified to be areas with moderate/high malaria transmission, zone 2 is classified to be areas with low malaria transmission and zone 3 is classified to

12 045 to 66 141 before decreasing to 36 451 in 2018 (Table 1.1) and this information was retrieved from the malaria annual report of 2018.

TABLE 1.1: OPD and IPD reported cases and death cases for Malaria in Namibia from 2001 to 2020 (MoHSS, 2018)

YEAR	OPD REPORTED CASES	IPD REPORTED CASES	DEATH CASES
2001	521067	41636	1747
2002	439760	23984	1030
2003	418146	20295	1094
2004	559324	36043	1734
2005	396579	23339	1137
2006	319676	27690	612
2007	102381	4242	181
2008	119711	4907	174
2009	70807	1864	64
2010	22359	1505	45
2011	15774	984	36
2012	3163	50	4
2013	4727	102	21
2014	16128	787	61
2015	12045	561	43
2016	24879	94	94
2017	66141	1969	331
2018	36451	1754	82
2019	2154	836	7
2020	11539	968	40

OPD: Outpatient Diagnosis

IPD: Inpatient Diagnosis

From 2014 to 2018, the Kavango West and East regions reported the highest number of malaria cases in Namibia, followed by the Zambezi region and the Ohangwena region (Haiyambo et al., 2019) (Table 1.1). Furthermore, Out of 68 110, and 38 205 malaria cases reported in Namibia in 2017 and 2018 respectively, the same four regions contributed about 96% of the total cases reported, with Kavango leading with 81%, followed by Zambezi with 10%, and Ohangwena with 5%, with the remaining regions accounting for only 4% of the total cases reported. Still, from the descriptive statistics of the total malaria cases reported in 2019 and 2020, the four contributed more than 90% of the total reported cases (MoHSS, 2019) (Table 1.2). Moreover, looking at the trend analysis of malaria cases reported from 2001 to 2020, it is clear that more research on malaria is needed, as the total number of malaria cases reported

annually continues to rise and fall (Table 1.1).

1.2 Problem statement

Malaria transmission remains unstable in most of the high and moderate endemic malaria nations where climatic factors are known to be mostly associated with malaria incidence from temporal and spatial perspective (Alemu et al., 2011). Namibia is among the countries that intend to achieve the third Sustainable Development Goal (SDG 3), which calls for the end/elimination of the malaria outbreak by 2030. Malaria has been eradicated in several countries. However, over the previous decade, most African countries, including Namibia, have seen an increase in malaria cases (Newby et al., 2016; Feachem et al., 2019). Millions of dollars have been spent on controlling malaria in Namibia so that the goal of reducing malaria incidence from 13 to less than 1 malaria case per 1000 population by 2016 and be malaria free by 2020 could be achieved. In 2017, the country needed N\$ 1.2 billion to eliminate malaria and the government was able to commit 65% of the funding. However, malaria remains a major public health concern in Namibia, mostly in Kavango West and East, Ohangwena and Zambezi region (Table 1.2), although it is preventable and curable. Complete datasets on malaria cases and environmental/climatic data that explain seasonal characteristics are being recorded. However, these datasets were never fully merged for them to be analysed and, to the best of our knowledge, no in-depth studies on spatial modelling of malaria incidences have been done considering a bigger sample size of a population at risk using climatic variables to estimate accurate overall malaria incidence at constituency level.

The distribution of malaria is known to have a spatial and temporal heterogeneity where cases vary through space and time, and most of the time climatic factors are known to be mostly associated with malaria incidence from temporal and spatial perspectives. Hence, it is important to describe the geographic variation of malaria risk, identify possible risk factors/covariates that might explain spatial variation, identify high-risk malaria significant areas (hot spot areas) through the Bayesian

conditional autoregression approach so that appropriate actions may be taken, and to assess health inequalities for a better allocation of health care resources.

TABLE 1.2: Reported OPD malaria cases by region, 2010 - 2018

Year	Zambezi	Kavango E& W	Omusati	Oshana	Oshana	Oshikoto	Kunene	Otjozondjupa	omaheke	Karas	Khomas	Erongo	Hardap	Namibia
2010	2929	13258	1505	68	3078	747	250	58	19	10	337	94	6	22359
2011	2434	11546	729	132	441	152	138	51	37	7	69	36	2	15774
2012	1274	948	410	49	196	67	97	45	12	2	42	18	3	3163
2013	2564	1454	161	48	236	43	35	85	4	14	34	43	6	4727
2014	1549	12959	477	112	505	158	38	221	58	0	21	11	19	16128
2015	479	10513	378	56	280	99	44	79	28	6	45	30	8	12045
2016	3154	14980	1584	293	3091	873	265	266	97	36	175	49	16	24879
2017	5085	52044	1470	331	5095	1057	171	437	130	24	211	70	16	66141
2018	3629	29617	152	131	1706	699	53	318	52	15	0	60	19	36451

Legends

High Cases	
Moderate Cases	
Low Cases	

1.3 Objectives of the study

1.3.1 Main objective

The primary goal of this research was fitting a spatial and spatio-temporal model that examines the effects of climatic variables and determine the spatial, and temporal pattern of malaria as well as risk factors/covariates that might explain spatial variation malaria infection in North Namibia.

1.3.2 The specific objectives

The study's key objectives were as follows:

- (i) To explore covariates factors associated with malaria in North Namibia;
- (ii) To simulate the link between malaria incidence and risk variables in North Namibia;
- (iii) To ascertain the spatial and temporal pattern of malaria as well as risk covariates that might explain the spatial/temporal variation of malaria infection in North Namibia; and
- (iv) To identify and map high-risk constituencies in North Namibia (hot spot areas).

1.4 The significance of the study

This study made use of areal/lattice malaria data that has a spatial structure, hence it employed Bayesian hierarchical Conditional Auto-Regressive (CAR) models introduced by Besag (Besag, 1974). These techniques anticipate risk in places where data are not collected while flattening variability in areas where the population is small (Gelfand and Vounatsou, 2003). The study fitted the Besag York Mollié (BYM) model which is a lognormal Poisson model which incorporates both an Intrinsic Conditional Auto-Regressive (ICAR) component (subclass of CAR models) for spatial smoothing, and a conventional random-effects component for non-spatial heterogeneity (Morris et al., 2019). Hence, results of the study may help public

health experts and policy makers to easily recognize the geographical distribution of malaria, identify risk hazards related to the disease, forecast epidemics in regions where such dangers exist, and identify vulnerable populations (malaria hot spot areas) for better guidance on monitoring and planning resources needs at all levels of health care and designing appropriate interventions to areas or communities deserving closer inspection by carefully using the available limited resources. These could help Namibia meeting Sustainable Development Goal 3 (SDG3) that aimed at ending malaria by 2030.

1.5 Thesis structure

This thesis consists of five sections. The first chapter provides a summary of the malaria history before delving into the study's problem statement, aims, and significance. The second chapter discusses prior work on malaria modeling as well as other techniques to disease mapping, such as malaria. The third chapter discusses the study technique, starting with an area study and study design, then moving on to the study population and sample size, data, and finally model description. We started with descriptive findings in Chapter 4, then moved on to model results (non-spatial, spatial, and spatio-temporal outcomes), and finally to results discussion. The conclusion and suggestions section of the study (Chapter 5) is then transcribed. Appendix A contains the R code used in the study.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Previously, it was thought that malaria was caused by stale air (Hempelmann and Krafts, 2013). In 1880, Alphonse Laveran, who was looking for a bacterial cause of malaria, discovered the blood parasites that cause malaria. Malaria is a parasitic disease caused by the protozoan Plasmodium, and it is still an endemic and significant public health issue in many countries, including Namibia. Plasmodium falciparum, Plasmodium vivax, Plasmodium knowlesi, and Plasmodium malariae are four Plasmodium species that cause malaria in humans, with Plasmodium falciparum being the deadliest (Parham and Michael, 2010). According to previous reports, malaria prevention and control measures have resulted in a 29% reduction in global morbidity and mortality rates since 2010, but millions of malaria cases are still reported annually in Africa (Parham and Michael, 2010).

According to the literature, malaria risk infection groups primarily include rural populations, people living near bodies of water (mosquito breeding sites), people living at low altitudes, poor and less educated people, and people working in farming and fishing areas, particularly pregnant women, breast-feeding women, young children (less than 5 years), and HIV-infected people, mostly those who are not on treatment (National Malaria Elimination Strategic Plan 2017-2021, 2017).

The reproduction of the anopheles mosquito depends on environmental factors like rainfall, temperature, and humidity in connection with vegetation cover and hydrology, especially water bodies (Sipe and Dale, 2003). The malaria parasite cannot

develop below 18 °C and over 40 °C hence the highest percentage of vectors that survive the incubation period are found at temperatures ranging from 28 °C to 32 °C (Alemu et al., 2011). High rainfall has an impact on mosquito breeding sites, but too much rain can wreak havoc on breeding grounds (Parham and Michael, 2010). They will prefer an environment with tall grasses, weeds, wet ground, and forest (Hardy et al., 2019). Mosquitoes prefer low-altitude environments because of the high temperature and humidity (Tuyishimire et al., 2016). Furthermore, environmental changes such as deforestation, vegetation clearance for crop production and marshland conversion, and people living near mosquito breeding areas are thought to be at high risk of contracting malaria (Tuyishimire et al., 2016).

Malaria patients experience symptoms such as fever, nonstop headache, sweats, muscle aches, and chills, as well as signs such as vomiting, yellowing of the skin and eyes, diarrhea, bleeding problems, shock, kidney, and liver failure. Malaria weakens the body to the point where an employee cannot go to work at times, and the disease can lead to death if a patient is not treated promptly (McKenna, 2008.)

2.2 Review of related studies

Disease mapping research has become a common application method used by biostatisticians, medical demographers, and epidemiologists to understand the geographical distribution of a disease, which is typically analysed in the formulation of a Bayesian hierarchical model with the primary goal of investigating the spatial distribution of malaria.

Several studies on malaria spatial modeling have been conducted in African countries using various approaches in the analysis of malaria spatial variation and its relationship with environmental, socio-demographic, and economic factors. The different types of analysis considered included spatial analysis, and Bayesian spatio-temporal analysis. These employed different methods/tools of analysis such as Global and local Moran I statistics, Point pattern analysis, SaTScan Technique, and

Getis Ordi (G_i^*) spatial statistics. Some reseahers have applied Bayesian hierarchical approach using the Markov chain Monte Carlo (MCMC) method in Win BUGS programme or the INLA package in R software. Some other methods included Bayesian hierarchical conditional auto regression (CAR) model in Win BUGS software , Bayesian hierarchical generalised linear mixed model in the INLA package in R software, Bayesian Poisson regression approach, GIS-based spatial modeling techniques, Leroux model, Dean Model, Besag-York-Mollié (BYM) model, Bayesx model, BYSTAR model and many more (Sipe, 2003; Guerra et al., 2006; Kazembe, 2007; Victor, 2009; Tuyishimire 2016; Joao, 2018).

The majority of the epidemiological studies on spatial modeling of malaria risk factors conducted in Africa focused on the ecology of the vector (Turner et al., 2013), the effectiveness of control measures (Guerra, Snow, and Hay, 2006), and spatial disease modeling (Guerra, Snow, and Hay, 2006). Ealier studies have revealed that malaria is influenced by three factors: environmental, demographic, and economic. Some of the previous are presented as follows.

Tuyishimire (2016) conducted a study on spatial modeling of malaria risk factors in Ruhuha sector in the east of Rwanda using Getis and Ord spatial statistics to simulate malaria risk factors geographically in largely rural areas of Southeastern Rwanda. The author's results revealed that malaria prevalence increases were associated with the proximity to irrigated farmland, housing quality and household size where lower housing quality for instance mud houses, earth floor and unburnt brick walls were associated with a high risk of malaria infection. The author concluded that irrigated farmlands are the main anopheles mosquito breeding sites in Rwanda and suggested that people should not only live far away from irrigated farmlands, which are thought to be the primary anopheles mosquito breeding places, but housing quality should also be addressed.

Ferrao (2018) also studied mapping and modeling malaria risk in Mozambique by using GIS-based spatial modeling techniques. The findings of the author showed

that distance to water bodies, mean temperature, precipitation, altitude, slope, distance to the road, land use and land cover, and population density were associated with malaria risk in Mozambique.

Another study on spatial modeling and risk factors of malaria incidence in northern Malawi was done by Kazembe (2007). The author applied a Bayesian Poisson regression model assuming different spatial structures to model malaria risk factors in Malawi. Altitude and precipitation were found to be associated with malaria risk in northern Malawi. The author's results revealed that the geographical variation in malaria risk was caused by a combination of observed and unseen environmental factors and highlighted the overall effect of these factors using a map (Kazembe, 2007).

2.3 Types of spatial data

The application of data or information to determine the geographic proximity of the earth's features and boundaries to represent an object's location, size, and shape, such as lakes, flat ground, and township, is referred to as spatial data analysis. Spatial data can be mapped, normalcy can be stored as coordinates, and topology can be stored. There are three types of spatial data (DiMaggio, 2012):

I. Spatial point processes (Point data)

A point data set is a discrete unit of information that is typically derived from measurement and can be represented numerically or graphically. This means that the spatial point process is a d -dimensional random pattern of points (where $d = 2$ or $d = 3$ in applications) (DiMaggio, 2012).

Point processes are assumed to follow a probability distribution and they are frequently described as Poisson processes. Spatial points can be associated with covariates where analyses may include assessing the role of covariates in finding intensity or controlling for covariates effects when assessing the interaction between points

(Morris et al., 2019). As a result, a point process is used for the analysis of observed point patterns, where the points denote the locations of some research object, such as disease cases.

II. Areal (Lattice) data

These data are associated with population surveys such as census and health statistics, and were originally referred to individuals located in specific points of space to submit information on the size, distribution, composition, and other social and economic characteristics of the population, as well as household amenities and housing circumstances (Stevenson, 2003).

Lattice data includes socioeconomic data pertaining to administrative regions (e.g., census, administration), climate model predictions as aggregates across regions, and health data pertaining to hospital wards. Lattice data is used to calculate proportions or risks based on count data (Anselin, 1998).

Exploratory methods for point data such as Kernel density estimation, can be used to investigate the process under investigation, whereas different methods for larger units, such as standardized mortality ratio (SMR), Bayesian smoothing, and autocorrelation statistics, can be used (Wilesmith et al., 2003).

III. Continuous data (Geo-statistical data)

Continuous data consists of point samples drawn from a continuous spatial distribution such as temperature readings taken at various point locations. They typically observe a complete collection of data points generated on data interpolated to unseen points on a continuous surface in a regular or irregular manner as a function of distance (Wilesmith et al., 2003).

Furthermore, this data is estimated from a set of field samples that can be distributed regularly or irregularly and results from natural resources such as air pollution,

rainfall, soil mineral concentrations, humidity, and variables that can be measured at all possible locations (Wilesmith et al., 2003).

2.4 Review on statistical methods of spatial data analysis

2.4.1 Cluster analysis

A set of objects is grouped into clusters in cluster analysis so that an object in one group is homogeneous. Cluster analysis is divided into two categories: global cluster analysis, which assumes that the risk surface is clustered, implying that areas of similar elevated risk exist, and local cluster analysis, which provides enough information about where clusters exist. These are indicators that show local patterns and measure local instabilities and are useful for identifying specific clusters in a data set (Gangnon and Clayton, 2003). In the absence of actual examination using different clustering scenarios such as hot spot clustering, local clustering assumes the presence of spatial clustering.

Due to the assumption of zero neighbourhood criteria, a hotspot analysis is defined as an area/location where people have a higher-than-average risk of victimisation. In other words, it is an area or location that possesses a higher-than-average incidence of the event under consideration in a cluster. Furthermore, hotspot analysis utilizes vectors to identify statistically significant hotspots (high) and cold spots (low) in data by forming polygons or converging points that are close to one another based on a measured distance. Moran's and Getis – Ord General (Getis – Ord G_i^*) are two methods used in hotspot analysis to retain values such as z-score and, when combined, will indicate whether or not clustering exists in the data (Achu and Rose, 2016). Getis –Ord G_i^* examines each feature in the dataset, and not all high-value features can be statistically significant hotspots.

A feature with a substantial hotspot is indicated by a high z-score and a low p-value, whereas a feature with a low negative z-score and a low p-value indicates a significant cold spot. The more intensive the clustering, the higher or lower the z-score, and if the z-score is close to 0, no spatial clustering exists (Zuur, Ieno and

Saveliew, 2017).

The Getis –Ord local statistics are given as follows:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij}x_j - \bar{X} \sum_{j=1}^n w_{ij}}{s \sqrt{\frac{n \sum_{j=1}^n w_{ij}^2 - (\sum_{j=1}^n w_{ij})^2}{n-1}}}, \quad (2.1)$$

where x_j is the attribute value for feature j , w_{ij} is the special weight between features I and j , n is the total number of features, \bar{X} is the mean of the x variable, and s is the standard deviation.

2.4.2 Spatial autocorrelation

The correlation of a variable with itself over space is quantified by spatial autocorrelation. To estimate the spatial effect of the region being studied (e.g., region/constituency), the neighbourhood information is computed using the boundaries of that constituency. A neighbouring structure at the constituency level, as well as an adjacency matrix and a contiguity weight matrix (W) with elements of 1 and 0, are created with varying distance lags.

$$W = \begin{pmatrix} 1 & w_{11} & w_{12} & w_{13} & \dots & w_{1k} \\ 1 & w_{21} & w_{22} & w_{23} & \dots & w_{2k} \\ 1 & w_{31} & w_{32} & w_{33} & \dots & w_{3k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & w_{n1} & w_{n2} & w_{n3} & \dots & w_{nk} \end{pmatrix}, \quad (2.2)$$

where k is the number of locations (constituencies) under the study, W_{ij} is the element of W representing a weight for $(i, j)^{th}$ constituency $i, j = 1, 2, 3, \dots, n$ and W is

constructed using the following 3 description types (Dessie, 2017).

- (i) The rook continuity description considers common edge shared objects by letting $W_{ij} = 1$ for constituencies that share common edge with the constituency of interest otherwise $W_{ij} = 0$.
- (ii) Descriptions of the contiguousness of the bishop consider objects that share a common vertex as neighbours, $W_{ij} = 1$ for entities that share common vertex with the region of $W_{ij} = 0$ and
- (iii) A queen contiguousness description integrates both the rook and bishop description as any object sharing either a common edge or common vertex to be considered as a neighbour, $W_{ij} = 1$ for entities that share common edge or vertex with constituency of the interest otherwise $W_{ij} = 0$.

The distance and inverse distance square between the centers of areas or points are used to define the location contiguity matrix in Cartesian space.

$$W_{ij} = \frac{1}{d_{ij}} \quad \text{or} \quad W_{ij} = \frac{1}{d_{ij}^2} \quad \text{or} \quad W_{ij} = \begin{cases} 1, & \text{if } d_{ij} < r \\ 0, & \text{if } d_{ij} > r \end{cases}$$

where d_{ij} is the distance between center of i^{th} and j^{th} location and r is prespecified radius distance.

Spatial autocorrelation is crucial in geographical analysis where Moran's I is a correlation coefficient that measures the data set's overall (global) spatial autocorrelation. Global measures of spatial autocorrelation determine whether or not the data shows spatial autocorrelation against H_0 , as well as the magnitude and direction (positive or negative) of any spatial autocorrelation. If the data is globally spatially correlated, the specific observations that are auto correlated with neighbouring observations of the variable of interest as well as strength and direction must be identified by considering local spatial autocorrelation.

In this case, Moran I scatter plots are plotted so that one can be able to picture the linear correlation between the spatial lag y and Wy . In detail, Wy is plotted against Y and the slope of the regression curve is defined by the Moran's I value where: W is the non-zero elements of the spatial weight matrix and y is the weighted sum or a weighted average of the neighbouring values.

Moran's I

Moran's I test is based on deviations from the mean as a cross-product and is used to detect global spatial autocorrelation. It is computed for n observations on a variable x at various locations and is defined as:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{X})(x_j - \bar{X})}{\sum_i (x_i - \bar{X})^2}, \quad (2.3)$$

where S_0 is the sum of the elements of the weight matrix: $S_0 = \sum_i \sum_j w_{ij}$, w_{ij} are the weight matrix entries, and \bar{X} is the mean of the x variable.

Moran's I is related to, but not the same as a correlation coefficient. In the absence of autocorrelation, it differs from -1 to $+1$, and regardless of the stated weight matrix, the expectation of Moran's I statistic, independent of the provided weight matrix, is $-1/(n-1)$, which decreases to zero as the sample size increases. The normalizing factor $S_0 = n$ for a row-standardized spatial weight matrix, and the statistic is simplified to a spatial cross product to variance ratio. Positive spatial autocorrelation is indicated by Moran's I coefficient greater than $-1/(n-1)$, whereas negative spatial autocorrelation is indicated by Moran's I coefficient less than $-1/(n-1)$. This means that Moran's $I = 1$ value close to 0 indicates perfect positive spatial autocorrelation, whereas Moran's I value close to 0 indicates no spatial autocorrelation (De Jong, Sprenger, and Van Veen, 1984; Assuncao and Reis, 1999; Li, Calder, and Cressie, 2007).

Local indicators of spatial association (LISA)

At each observation, the local indicator of spatial association provides a signal of strong spatial clustering of related values surrounding that observation. The sum of LISA is proportional to global indicators. The following is a definition of local Moran's I statistics.

$$I_i(d) = (x_i - \bar{x}) \sum_{j=1}^n w_{ij}(d) (x_j - \bar{x}), j \neq i, \quad (2.4)$$

where x_i and x_j denote the number of count cases at constituency i and j , respectively, and w_{ij} denotes the spatial weight matrix based on the defined distance lags between constituency i and j (most of the time lags are in kilometers). If the distance between constituency i and constituency j is less than d , $w_{ij}(d) = 1$; otherwise, $w_{ij}(d) = 0$.

Geary's C

Geary's C statistic (Geary, 1954) is calculated by comparing the variations in response to each other's observations. See the following equation:

$$C = \frac{n-1}{2S_0} \frac{\sum_i \sum_j w_{ij} (x_i - x_j)^2}{\sum_i (x_i - \bar{X})^2} \quad (2.5)$$

Geary's C spans from 0 (maximal positive autocorrelation) to a positive number when there is a lot of negative autocorrelation. In the absence of autocorrelation and regardless of the weight matrix provided, its expectation is 1 (Sokal and Oden,

1978). If Geary's C is smaller than just 1, it suggests that there is positive spatial autocorrelation (De Jong, Sprenger, and Van Veen, 1984).

2.5 Generalized Linear Models

2.5.1 Introduction

Generalized linear models (GLMs) in statistics are a versatile generalisation of ordinary linear regression that permits response variables to have error distribution models that are not the normal distribution. The GLM improves on linear regression by allowing the linear model to be connected to the response variable through a link function, and the variance of each measurement is a function of its projected value (Dey, Ghosh, and Mallick, 2000; Guisan, Edwards Jr, and Hastie, 2002; Hedeker, 2005; Hardin et al., 2007).

In GLM, each dependent variable's outcome y is considered to be generated by a specific distribution in an exponential family e.g., Normal, Binomial, Poisson and gamma distribution where in both the mean, μ , of the distribution depending on the independent factors, x (Dey, Ghosh, and Mallick, 2000; Guisan, Edwards Jr, and Hastie, 2002; Hedeker, 2005; Hardin et al., 2007). The exponential family:

$$f(y_i|\theta_i, \varnothing, w_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\varnothing}\right) w_i + c(y_i, \varnothing, w_i), \quad (2.6)$$

where θ_i is the natural parameter of the exponential family, \varnothing is a dispersion parameter shared by all observations, w_i is a weight, and $b(\cdot)$ and $c(\cdot)$ are functions that vary depending on the exponential family.

The expectation $E(y_i|u_i, \gamma) = \mu_i$ is linked to the linear predictor η_i by $\mu_i = g(\eta_i)$ and $\eta_i = \mu_i' \gamma$, where g is a known response function and γ are unknown regression

coefficients (Belitz et al., 2009; Mwahi, 2014).

For example, let consider a Poisson regression model that expresses explanatory/covariates that have a statistically significant effect on the response variable. This model predicts dependent variable (y_i) where y_i is the count of people that tested malaria positive in constituency i , and it consists of one or more independent variables and the measure of the effects of the given predictors is expressed in terms of incidence rate.

Let $Y_i \sim \text{poisson}(\mu_i)$, where μ_i is the expected count of y_i . The general Poisson model is given by:

$$\log E(y_i) = \log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}, \quad (2.7)$$

where β_0 is the intercept, $\beta_1, \beta_2, \dots, \beta_k$ are parameters and x_1, x_2, \dots, x_k are explanatory variables (fixed effects) that are included in the model which sometime is known as covariates and $\log E(y_i)$ is the mean expected value of the outcome being a case for subject i .

Exponentiating eq. 3.1 we obtain a multiplicative model for the mean itself:

$$E(y_i|x_i) = \mu_i = e^{x_i^T \beta}, \quad (2.8)$$

where i is an observation, x_i^T is a set of independent variables including an intercept, $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^T$ are a set of coefficients.

2.5.2 Generalized Linear Mixed Models (GLMMs)

Generalized linear mixed models (GLMMs) are a type of linear mixed model that has been extended. GLMMs can also be thought of as a generalized linear model (e.g., logistic regression) that estimate both fixed and random effects (hence mixed models) and are especially useful when the dependent variable is binary, ordinal,

count or quantitative but not normally distributed. The general form of the linear mixed model (in matrix notation) is given as:

$$y = X\beta + Zu + \epsilon, \quad (2.9)$$

where y is a $N \times 1$ column vector, the outcome variable; X is a $N \times p$ matrix of the p predictor variables; β is a $p \times 1$ column vector of the fixed effects regression coefficients (the β s); Z is the $N \times q$ design matrix for the q random effects (the random complement to the fixed X); μ is a $q \times 1$ vector of the random effects (the random complement to the fixed β); and ϵ is a $N \times 1$ column vector of the residuals, that part of y that is not explained by the model, $X\beta + Zu$.

The difference between Linear Mixed Model (LMMs) and Generalized Linear Mixed Model (GLMMs) is that in GLMMs, aside from gaussian distributions, the response variables might come from any distribution. In addition, rather than modeling the responses directly, a link function, such as a log link, is sometimes used.

Let the linear predictor, η , be the combination of the fixed and random effects excluding the residuals.

$$\eta = X\beta + Z\gamma, \quad (2.10)$$

The generic link function is called $g(\cdot)$, the link function relates the outcome y to the linear predictor η . Hence, the conditional expectation of y (conditional because it is the expected value depending on the level of the predictors) is denoted as: $g(E(y)) = \eta$. The expectation of y can also be modelled: $E(y) = g(\eta) = \mu$, with $y = g(\eta) + \epsilon$.

In simple notation, adding random effects to the Poisson regression model Eq.(2.7), the model is rewritten as follows:

$$\ln(E(y_i|\mu_{it})) = x_{it}^T \beta = \beta_0 + \sum_{k=1}^k \beta_k x_{kit} + \mu_i, \quad (2.11)$$

where y_{it} is the count of people tested malaria positive in constituency i at time t , μ_{it} is the random effect for constituency i assumed to follow a normal distribution with mean 0, and variance (σ^2) and each independent variable, k , is multiplied by a coefficient β_k and is added to a constant, β_0 . In more familiar notation, $\ln(E(y_{it}|\mu_{it})) = \ln(\mu_{it}) = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \dots + \beta_k x_{kit} + \mu_{it}$. The parameters $(\beta_0, \beta_1, \dots, \beta_k)$ are estimated by maximum likelihood estimators (MLEs) using iterative algorithms such as Newton-Raphson (NR) and iteratively re-weighted least square and the goodness fit of the model is checked using Pearson chi-square statistic χ^2 .

2.5.3 Conditional Autoregressive Regression models (CARs)

Generally, GRMs do not account for non-linear effects of the covariates as well as the spatial and temporal structure of the data. Therefore, Conditional Autoregressive Regression models (CARs) are considered. CAR models are the most often used spatial analytic models for describing spatial autocorrelation in data relating to a group of non-overlapping areal units, assessing how quantities of interest fluctuate with factors, and locating hot spot clusters. These models are often stated in a hierarchical Bayesian framework, with Markov chain Monte Carlo (MCMC) simulation used for inference or Integrated Nested Laplace approximation (INLA) (Lee and Kim, 2008).

In spatial analysis of disease mapping, in a hierarchical Bayesian model, CAR models are quantified as a prior distribution for a series of random effects to study the patterns of diseases (Lee and Mitchell, 2013).

2.5.4 Bayesian Hierarchical approach

Fully Bayesian in disease mapping is important because fully bayesian posterior standard deviations of small-area relative risks are more reflective of the uncertainty associated with the relative risk estimation (Ugarte, Militino, and Goicoa, 2008). The Bayesian approach provides samples of the entire posterior distribution of incidence rates or relative rates for each area by providing more information than a single point estimate where all parameters are allocated to cope with their likely volatility prior to distribution and this can be achieved through Markov chain Monte Carlo (MCMC) or Integrated Nested Laplace approximation (INLA).

The parameters in Bayesian modeling have distributions that regulate their shape and are specified by the investigator based on their prior beliefs about their behavior. (Lawson, 2018).

2.5.5 The Likelihood Function

The likelihood function describes the observed data's joint probability as a function of the statistical model's parameters (Luerken, 2009).

Let $y_i, i = 1, 2, 3, \dots, n$ be a random variable with probability density function $f(y_i|\theta)$, where $\theta = \theta_1, \theta_2, \dots, \theta_p$ is a p is a length vector of relative risk parameters, then the likelihood of y_i is defined as:

$$F(y|\theta) = \sum_{i=1}^n f(y_i|\theta), \quad (2.12)$$

with the assumption that $y = y_1, y_2, \dots, y_m$ given θ are independent, making it possible to take the product of individual contribution of $f(y_i|\theta)$ (Lawson, 2018).

Poisson counts Likelihood

Poisson regression is used to model response variables (Y-values) for small area count data. Poisson explains which explanatory variables have a statistically significant effect on the response variable. The model assume that count variable y_i have a mean μ_i and is independently distributed as $y_i \sim \text{poisson}(\mu_i)$.

The likelihood function is denoted as:

$$L(\mu_i) = \prod_{i=1}^n f(x_i; \mu_i) = \prod_{i=1}^n \frac{\mu_i^{x_i} e^{-\mu_i}}{x_i!} = \frac{\mu_i^{\sum_i x_i} e^{-n\mu_i}}{x_1! x_2! x_3! \dots x_n!}. \quad (2.13)$$

The expectation is $E(y_i) = \mu_i = E_i \theta_i$, where E_i is the expected rate for the i^{th} area and θ_i is the relative risk for the i^{th} area.

2.5.6 Prior Distributions

Generally prior distribution represents belief about the true value of a parameter θ before observing the data y_i . For example, given parameter θ , the prior distribution can be denoted by $p(\theta)$, while for a parameter vector v , the joint distribution is denoted as $p(v)$. There are different types of priors:

Uniformed priors

Uniformed priors describe that there is no prior knowledge. The distribution of an uninformed prior is believed to add no information to Bayesian inference. For example, a Uniform(0,1) distribution might be regarded an uninformed prior when estimating a probability since it states that before collecting any data, we consider every feasible value for the true probability to be equally likely.

Conjugate priors

A parametric distribution that can be easily updated. The prior in models with conjugate priors is the same shape as the likelihood. If the likelihood is a Gaussian with known precision, then the conjugate prior on the mean is also a Gaussian. This

ensures that the mean's posterior distribution is also Gaussian.

For example, let y_i be a set of observations $\{y_i\}_{i=1}^n$ that follow a Gaussian distribution:

$$y_i | \mu, \tau \sim N(\mu, \tau), \quad (2.14)$$

,

where μ is the unknown mean and τ is the known precision.

Then the prior on μ can be a Normal distribution with mean μ_0 and precision τ_0 :

$$\mu \sim N(\mu_0, \tau_0). \quad (2.15)$$

The posterior distribution of $\mu | \text{data } y \sim N(\mu_1, \tau_1)$,

where

$$\mu_1 = \mu_0 \frac{\tau_0}{\tau_0 + \tau_n} + \bar{y} \frac{\tau_n}{\tau_0 + \tau_n}$$

and,

$$\tau_1 = \tau_0 + \tau_n.$$

Subjective priors

In subjective priors, a distribution is constructed from an expert's opinion to describes the informed opinion of the value of a parameter prior to the collection of data.

Improper priors

Improper priors distribution does not normalize to unity. If prior distributions are given an improper uniform prior, $p(\theta) \propto 1$, then the posterior distribution is proportional to the likelihood,

$$p(\theta|y) \propto p(y|\theta)p(\theta) \propto p(y|\theta) \quad (2.16)$$

yet, a prior distribution $p(\theta)$ is an improper when it is not a probability distribution, meaning,

$$\int p(\theta)d\theta = \infty \quad (2.17)$$

Perhaps the most common improper distribution is an unbounded uniform distribution, $p(\theta) \propto 1$ for $-\infty < \theta < \infty$. These priors can be used, because in some cases, the posterior distribution can still be proper even if the prior is not.

Informed priors

Informed priors are a description of the level of knowledge you have. The distribution of an informed prior provides information to the Bayesian inference. It's either the outcome of a previous statistical study of other data that provided you with knowledge about the parameter (example) or it was built from an expert's estimate of the parameter. These priors can be modeled in a number of different ways. If the parameter values produce a distribution that differs from an uninformed prior, the conjugate prior is an informed prior. For example, a $Beta(1,1)$ distribution is usually considered an uninformed prior when estimating a binomial probability because it assigns equal weight to all values of p between 0 and 1. Thus, a $Beta(4,2)$ distribution, for example, is an informed prior because its shape is different: it peaks at 0.75.

2.5.7 Posterior Distributions

The posterior distribution, also known as the posterior, is the conditional distribution of a set of unknown parameters, latent variables, or otherwise missing variables of interest in Bayesian research. The posterior distribution uses the current data to update previous information about that parameter, which is referred to as a prior. Essentially, the prior distribution gives information about the parameter based on prior beliefs or assumptions, whereas the probability uses facts to provide information. The product of the likelihood and the prior distribution is then called the posterior distribution where Bayes' theorem is used to derive a posterior distribution ($p(\theta|x)$),

$$P(\theta|x) = \frac{p(x|\theta)P(\theta)}{p(x)} = \frac{p(x|\theta)P(\theta)}{\int P(x|\theta)p(\theta)d(\theta)}, \quad (2.18)$$

where θ is the unknown parameter(s) and x is the current data, $p(x|\theta)$ is the probability of the data given the parameter $P(\theta|x)$ is the likelihood $L(\theta|x)$, The prior distribution, $p(\theta)$, is user specified to represent prior knowledge about the unknown parameter(s), and the last piece of Bayes' theorem, the marginal distribution of data, $p(x)$, is computed using the likelihood and the prior.

2.5.8 Integrated Nested Laplace approximation (INLA)

According to the literature, the most often used approach for implementing Approximate Bayesian Inference is Integrated Nested Laplace Approximation (INLA) (Akerkar, Martino and Rune, 2020). Although the INLA methodology focuses on models defined as Latent Gaussian Markov random fields (GMRF), it incorporates a vast family of models that are employed in reality (Akerkar, Martino, and Rune, 2020).

A Latent GMRF model is a hierarchical model in which a distributional for the observables y is found at the first stage given some latent parameters η and, optionally, some extra parameters θ_1 .

$$\pi(y|\eta, \theta_i) = \prod_j \pi(y_j|\eta_j, \theta_i) \quad (2.19)$$

The latent parameters η are a subset of the wider latent random field x , which is the second level of our hierarchical model. The latent field x is modelled as a GMRF with the assistance of a precise matrix Q that varies with several hyperparameters θ_2 .

$$\pi(x|\theta_2) \propto \exp \left\{ -\frac{1}{2} (x - u)^T Q (x - u) \right\} \quad (2.20)$$

(Akerkar, Martino, and Rue, 2010; Gómez-Rubio, 2020)

The prior distribution for the hyperparameters $\theta = (\theta_1, \theta_2)$ is the model's third and final stage.

The INLA method presents a method for quick Bayesian inference utilising realistic approximations to $\pi(\theta|y)$ and $\pi(x_i|y), i = 0, \dots, n$ the hyperparameter marginal posterior densities and the latent variable posterior marginal densities.

CHAPTER 3

DATA AND METHODOLOGY

3.1 Study area

Namibia has 14 regions 35 health districts and 122 constituencies in total with malaria present in Northern regions mainly Kavango West and East, Zambezi and Ohangwena Region and these four regions consist only 34 of the 122 constituencies and hundreds of clinics that keep records of individuals that tested malaria that could be aggregated at constituency level (Maina et al., 2019).

3.2 Study design

After discovering that malaria elimination demands a robust surveillance system for early detection of malaria infections and allow for a quick and efficient reaction, the World Health Organization and Global Fund promoted the use of a health information system, and the majority of developing countries, including Namibia, have implemented the District Health Information Software (DHIS), which records information on all people who have tested malaria positive (Dehnavieh et al., 2019). This program is a free and open-source platform for everyday health information management, with a primary focus on creating health statistics.

Although Polymerase Chain Reaction (PCR) is most useful for confirming the species of malaria parasite after the diagnosis has been established by either smear microscopy or Rapid Diagnostic Test (RDT). This study has used secondary malaria data collected through a passive surveillance case detection that was carried out among people who visited any of Namibia's health facilities reporting

fever/suspected of having malaria and were tested for malaria using the Rapid Diagnostic Test (RDT) screening and this might have effect on the accuracy of confirmed cases. All individuals who tested positive as per RDT test were given a Rapid Case Notification form (Malaria Case Investigation form) to fill out, which contains more information about the patient, and the aggregated information of people who tested malaria positive on a daily basis was recorded in the weekly surveillance as aggregated data.

The form provided information such as the date the patient tested positive for malaria, the region, the health district, the health facility center, the patient's home village/town, the patient's age, gender, pregnant status, occupation, malaria symptoms, travel history, use of mosquito nets, home sprayed in the previous 12 months, case classification (i.e., local or imported/non-local), nationality, and many more.

The filled-out forms are then transferred from health facility centers to health districts for active surveillance. Malaria positive people were followed up at their homes one week after having tested positive for malaria so that the geo-coordinates of breeding sites areas could be recorded. All members of each household and their immediate neighbours within a 100-meter radius were also tested for malaria.

Malaria-positive individuals were treated immediately, and those who were seriously ill (malaria severe) were transported to the nearest hospital. Some in-depth investigation was carried out in villages where more people tested positive during the tracing, and they were also given the active malaria form to fill out. During household visits, some additional vector control research was conducted at the community level. The information of all people who tested positive for malaria using RDT was then entered into an electronic database known as District Health Information Software 2 (DHIS2) and this is kept as confidential data for future use.

Therefore, a cross sectional study was employed using two different datasets: 2018 to 2020 malaria weekly surveillance data (complete aggregated dataset) and 2017 to 2020 individual records of malaria confirmed cases by Rapid Diagnostic Test (RDT)

in the four regions that were entered on the DHIS system at the time (incomplete dataset). The Ministry of Health and Social Services provided both datasets.

3.3 Data

Malaria has been linked to climatic factors such as temperature, rainfall, wind speed, humidity, altitude, and soil wetness in studies conducted in some African and Asian countries (Kazembe, 2007; Wu et al., 2016; Ferrao et al., 2018; Umer et al., 2019). As a result, the study looked at both climatic variables, as well as some other potential factors based on the literature, such as population density, age, gender, occupation, use of mosquito bed nets, and material used for house walls (i.e. sticks, grass, mud, cow dung, wood, brick, cement, corrugated iron sheets, etc.) and roof material (i.e. grass, sticks, corrugated iron sheets, asbestos) by utilising a combination of datasets from various data sources that comprised 2018 to 2020 malaria weekly surveillance data (aggregated complete dataset), 2017 to 2019 individual malaria tested positive by RTD dataset, 2017 to 2020 monthly meteorological/weather/climate data, 2017 to 2020 projected population, and Namibia shape files for the constituencies. The following is a description of the dataset used in the study.

3.3.1 Population at risk

The Ministry of Health and Social Services' malaria annual report for 2018/2019, indicates that more cases of malaria were recorded in the four regions (Kavango East and West, Zambezi, and Ohangwena), as they had been for years. As a result, and to avoid too many errors, since the surveillance dataset recorded malaria case per clinic not as per constituency, the study primarily targeted the population living in the four northern risk malaria transition constituencies (Kavango East & West, Zambezi and Ohangwena region). These include approximately 645 518 people in 2018, 657 589 people in 2019, and 669 820 people in 2020.

These 4 regions are made up of 34 constituencies: 14 in each of the two Kavango regions, 12 in Ohangwena, and 8 in the Zambezi region. All of these constituencies

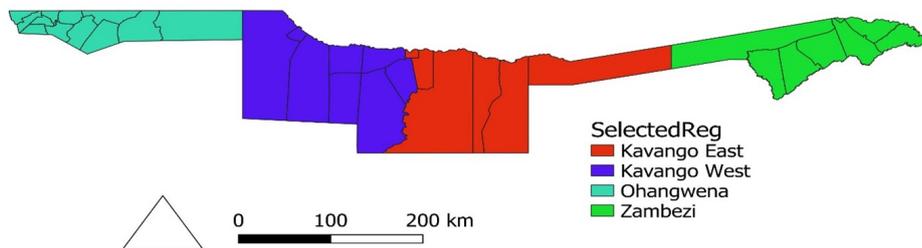


FIGURE 3.1: Map of the 4 regions (Kavango West and East, Zambezi and Ohangwena)

cover an area of 120 110 square kilometers, with Mashare being the largest at 9104 square kilometers and Rundu Urban being the smallest at 18 square kilometers.

More specifically, in Kavango West and East regions, there are 4 health districts with 29 health facilities (1 hospital, 2 health centers and 26 primary health care clinics) while in Ohangwena region there are 3 health districts with 36 health facilities centers (3 hospitals, 2 health centers and 32 primary health care clinics and 144 outreach services points) and in Zambezi region there is one health district with a total number of 29 health facilities (1 hospital, 3 health centers and 25 primary health care clinics). The map in Figure 3.2 depicts the 34 constituencies in the 4 regions studied.

Concerns about the potential long-term effects of any disease as well as other demographic trends, have heightened interest in the accuracy of population projections (Reed et al., 2002). As a result, the Namibia annual population growth rate was used to project the population of each constituency from 2012 to 2020 using the 2011 population data from the shapefiles.

(<https://www.statista.com/statistics/510201/population-growth-in-namibia/>)

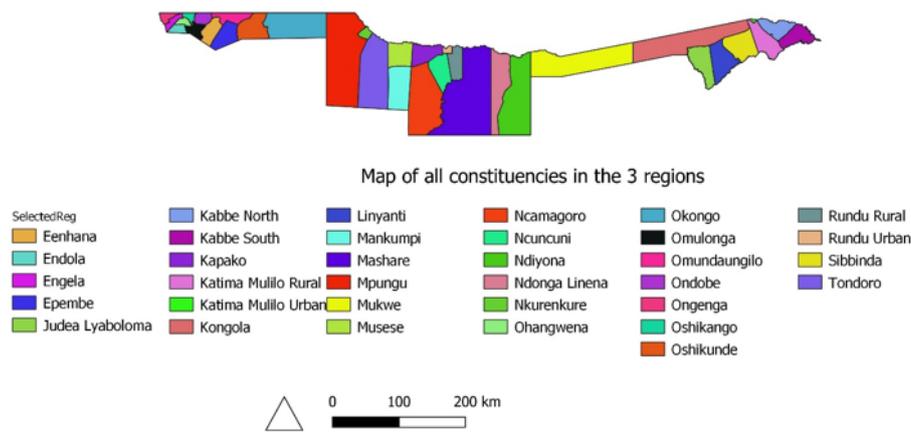


FIGURE 3.2: Map of all constituencies in the 4 regions

3.3.2 Malaria data

Malaria statistics from the Ministry of Health and Social Services from 2010 to 2018 revealed a high number of malaria cases in the 4 Northern regions (Kavango East and West, Zambezi, and Ohangwena) (Table 1.2). According to Table 1.2, Namibia recorded 66 141 malaria cases, 62 204 of which were from the 4 Northern regions (Kavango East and West, Zambezi and Ohangwena). Again, out of 36 451 malaria cases reported in 2018, the same 4 regions contributed 96% (34 952 malaria cases) of the total cases, with Kavango leading the way with 81%, Zambezi with 10 percent, Ohangwena with 5%, and only 4% reported from the other regions.

The sample data included in this study was, therefore, all 2018 – 2020 malaria weekly surveillance recorded cases and 2017 – 2019 individual malaria records that were available on the DHIS2 system of patients from these 4 regions because these are the regions with the highest malaria transmission among all Zone 1 regions. More specifically, the individuals satisfying the following criteria:

- ✓ The individual has visited any of the health service center (i.e., hospital or clinic) located anywhere in the 4 regions from 2017 to 2019
- ✓ The individual suspected to have malaria after visiting the clinic or hospital
- ✓ The individual tested malaria positive through RDT

- ✓ The individual who was willing to complete the Malaria Case Investigation form
- ✓ The individual information was entered on the DHIS2 system

2018 - 2020 complete aggregated malaria dataset(malaria weekly surveillance dataset)

The Ministry of Health and Social Services maintains a summarised statistic (malaria weekly surveillance data/aggregated data) that includes a summary of all malaria daily fever and suspected cases, as well as reported tested malaria positive by RTD cases per clinic. Every day, all people who visit any of the nearest health facility centers reporting fever or suspected of having malaria were recorded in this dataset after testing malaria positive by the Rapid Diagnosis Test (RDT).

This dataset has been identified as a complete malaria dataset. However, because the data is too summarised, the dataset does not provide all detailed information of people who tested positive for malaria, as it does on the active malaria form. This dataset was then obtained in excel format from the MoHSS.

In 2018 – 2020, a total of 44 570 cases of malaria were recorded in the 4 regions, 34 952 in 2018, 2 990 in 2019, and 10 678 in 2020 and this made-up the total sample size of 44 570 (Table 3.1). The following are statistics of all malaria reported cases for the 3 years (2018 - 2020) .

TABLE 3.1: Total cases of malaria per year per region and health districts (2018 – 2020)

Year	Region	Health districts	Malaria cases
2018	Kavango E & W	Andara	4971
		Nankudu	8933
		Nyangana	3 355
		Rundu	9 249
	Total	26298	
	Ohangwena	Eenhana	592
		Engela	615
Okongo		502	
Total	1709		
Zambezi	Katima Mulilo	3402	
Total			31619
2019	Kavango E & W	Andara	122
		Nankudu	374
		Nyangana	3 127
		Rundu	1 268
	Total	1891	
	Ohangwena	Eenhana	183
		Engela	464
Okongo		136	
Total	783		
Zambezi	Katima Mulilo	316	
Total			2 990
2020	Kavango E & W	Andara	1 694
		Nankudu	2 356
		Nyangana	660
		Rundu	1 1 203
	Total	5913	
	Ohangwena	Eenhana	855
		Engela	487
Okongo		791	
Total	2133		
Zambezi	Katima Mulilo	1998	
Total			10035
Grand Total			44644

2017 – 2019 Individual malaria tested positive by RTD dataset

Individuals who tested positive for malaria using a rapid diagnostic test (RDT) were provided an active case malaria detection questionnaire/form to fill out, and this questionnaire is kept at health facility centers for future histories (Smith et al., 2017; McCreesh et al., 2018; Hsiang et al., 2020). According to the literature, the questionnaire provides more information about the individual that may influence malaria, such as age, gender, place of residence, use of mosquito net, occupation, household sprayed in the past 3 months, ownership of mosquito bed nets, slept under mosquito bed net in the past 3 days, and many other factors.

The MoHSS then introduced the DHIS2 tool in 2014, where they began entering the information provided in the active case malaria form for everyone who tested malaria positive at any of Namibia's health facilities, and entering the cases for multiple purposes, including achieving goal 3.3 by 2030. Individual datasets of people who tested positive for malaria using RTD data in the 4 regions of Kavango East and West, Zambezi, and Ohangwena from 2017 to 2018 were then extracted from the DHIS2 system in Microsoft Excel format after permission from the Ministry of Health and Social Services. When people began capturing data on the DHIS2, some forms were missing, resulting in an incomplete individual dataset.

All individuals who tested positive by RDT among people suspected of having malaria from all 8 health districts in the 4 regions (Kavango East and West, Ohangwena, and Zambezi) when they visited any of the health centres from 2017 to 2019 were recorded. The study then included the information from the active case detection form of individuals who tested malaria positive was entered on the DHIS2 system from the 4 regions.

All confirmed and entered 2017 to 2019 malaria cases on DHIS2 system were 55 261 malaria cases and this was considered as the sample for individual data (Table 3.2)

TABLE 3.2: Sample size for Individual dataset 2017 to 2019

Region	Number of malaria cases
Kavango E & W	49905
Ohangwena	5105
Zambezi	251
Grand Total	55261

3.3.3 Environmental/weather /climatic data of the constituency (2017 - 2020)

Weather forecasting is provided by the Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL) Net page "[http://www.sasscalweathernet.org/index.php MIsoCode=NA](http://www.sasscalweathernet.org/index.php?MIsoCode=NA)" and this used to obtain average monthly weather data from 12 weather stations in towns or close towns to the constituencies in the 4 regions with the assistance of the Namibia Meteorological Service Windhoek's climate and data section, whereas constituencies without a weather station used monthly weather data from the nearest weather station. The 12 weather stations were: Alex Muranda Livestock Development Centre, Bagani, Dudukabbe, Hamoye, Kalimbeza, Kanovlei, Mashare, Ngoma, Okalongo, Okashana, Omafo, and Sachinga.

The list of datasets used in the study, as well as the variables of interest from each dataset, are shown in Table 3.3. The study merged all the 4 datasets listed in Table 3.3 into one data to attain all the objectives of the study considering varieties of variables from the 4 listed datasets. Basically, individual data set was analysed to determine some of the non-climatic variables thought to influence malaria through fitting a poisson non-spatial model while malaria aggregated datasets that was merged with climatic as well as population data was used in attaining all the study objectives through fitting a non-spatial, spatial and spatial-temporal models.

TABLE 3.3: List of merged datasets

MERGED DATASET	VARIABLE OF INTEREST
1. 2018 – 2020 Malaria Monthly surveillance data (complete aggregated data)	1. Region and constituency (spatial effect) 2. Year (time effects) 3. Season 4. Total number of people tested malaria positive by RDT in each constituency (dependent variable) 5. Expected cases (offset)
2. Monthly meteorological/weather/climatic data of the constituency (2018-2020)	5. Temperature (average, minimum and maximum) 6. Amount of rainfall recieved 7. Average wind speed 8. Average soil temperature 9. Humidity 10. Average leaf wetness 11. Altitude in meters
3. Population data	12. Population and Human population density per constituency
4. Individual dataset (2017 - 2019) extracted from DHIS2 system	13. Gender 14. Age group 15. Place of residence 16. Type of health facility where the individual tested malaria positive 17. Occupation 18. Slept under mosquito bed net in the past 3 days 19. Home sprayed in the past 12 months

3.4 Ethical considerations

The Namibia University of Science and Technology (NUST) Research Ethical Committee, as well as the Ministry of Health and Social Services Research Ethics Committee, provided ethical approval for the use of RDT-positive malaria cases as secondary data for malaria cases in Namibia’s four most malaria-transmitting regions. The data will be treated confidentially and will not be used for any other purpose other than academic research.

3.5 Data management

3.5.1 Data cleaning

Data cleaning was performed on all datasets prior to analysis. Monthly climatic variables and population of each constituency were aggregated for 2018 - 2020, and the

proportion of some other possible factors from the individual dataset, such as age, gender, material used for wall, roof, and floor, were then computed per constituency (Table 4.4 and 4.5).

Before the analysis, the data was cleaned, variables for both datasets were aggregated per constituency e.g., mean of malaria positive cases (dependent variable), the mean population of each constituency, constituency mean for all climatic variables included in the model as possible factors/covariates (fixed effects), and proportion for the possible factors from the individual dataset e.g., age, gender, material used for wall, roof and floor were computed per constituency (Tables 4.4 and 4.5).

Several models (non-spatial, spatial, and spatio-temporal) were fitted. The spatio-temporal analysis only considered the entire 2018 – 2020 constituency yearly aggregated data. R-statistical software was used for non-spatial, spatial, and spatio-temporal data analysis, with the INLA package being used to explain disease patterns, identify malaria risk factors, and constituencies at high risk (malaria hot-spot). R-code used for the analysis were presented in Appendix A.

3.5.2 Standardised Incidence Ratio (SIR)

Before fitting the models, SIR a straightforward measure of disease risk in a specific population each year was computed. The Standardised Incidence Ratio is used to determine whether the occurrence of any disease is high or low in a relatively small population. SIR values can indicate whether the number of observed cases in each geographic area is higher or lower than predicted based on the population of the community. SIR is 1 if the observed number of instances equals the expected number of cases. The SIR will be more than one if there are more cases than predicted; if there are fewer cases than expected, the SIR will be less than one (Disease Control, Prevention et al., 2008). SIR (Standardized Incidence Ratio) was calculated as follows:

Let y_{ijt} be the observed malaria case in constituency i ($i = 1, 2, 3, \dots, n$), season j ($j = 1, 2, 3$, and 4) and year t ($t = 1, 2$, and 3), then we define SIR as the ratio

of observed cases y_{ijt} to the number of expected cases E_{ijt} in the i^{th} constituency, j season at time t :

$$SIR = RR = \theta_{ijt} = \frac{y_{ijt}}{E_{ijt}} \quad (3.1)$$

With the expected number of malaria cases estimated as the constituency overall mean rate in the year t multiplied by the population of the constituency:

$$E_{ijt} = \text{meanrate}_t \times \text{population}_{ijt} \quad (3.2)$$

In addition, malaria annual incidence rate of each constituency's per 1000 population was also computed by using the projected population: Incidence rate = (number of cases/population at risk) \times 1000

$$IR_{ijt} = \frac{y_{ijt}}{P_{ijt}} \times 1000 \quad (3.3)$$

Small areas, on the other hand, may exhibit extreme SIRs in many situations due to small population sizes or small sample sizes. SIRs may be misleading and insufficiently reliable for reporting in these cases. As a result, it is preferable to estimate disease risk using Bayesian hierarchical models, which allow for information borrowing from neighbouring areas and the incorporation of covariates, resulting in the smoothing or reduction of extreme values (Moraga, 2019).

3.6 Model building for count data

Although count data can be approximated by a normal distribution and reasonably described using a linear model, count data are most typically modeled using a generalised linear model with a Poisson distribution or a negative binomial distribution (GLM). Poisson and negative binomial distributions are discrete probability distributions with two fundamental properties: they only contain zero and positive integers, and the variance is a function of the mean (Gardner, Mulvey, and Shaw, 1995; Sellers and Shmueli, 2010).

This study used the sum and mean of malaria tested positive as the dependent variable, so count models were used to achieve the study's objectives using both individual and aggregated malaria dataset. The following is detailed information on count models.

3.6.1 Poisson regression model (non-spatial model)

Poisson regression model expresses explanatory/covariates that have a statistically significant effect on the response variable. This model is defined such that the observed number of malaria cases (y_{ijt}) in constituency j , season j in year time t , with associated expected counts E_{ijt} could be treated as one realization of poisson random variables with means μ_{ijt} , i.e., $y_{ijt}|E_{ijt}, \theta_{ijt} \sim \text{poisson}(E_{ijt}\theta_{ijt})$, where $\mu_{ijt}(\mu_1, \mu_2, \mu_3, \dots, \mu_n) = E_{ijt} \times \theta_{ijt}$ is a function of the effects of k covariates x_{kijt} as well as as spatial and temporal random effects. Then general Poisson model is given by:

$$E(y_{ijt}|E_{ijt}, \theta_{ijt}) = \ln(\mu_{ijt}) = \exp(x_{ijt}^T \beta + E_{ijt}). \quad (3.4)$$

(Cameron and Trivedi, 2001; Lord et al., 2006; Greene, 2008; Hilbe, 2011; Cameron and Trivedi, 2013). In familiar notation this could be rewritten as:

$$\ln(E(y_{ijt}|E_{ijt}, \theta_{ijt})) = (\mu_{ijt}) = \beta_0 + \beta_1 x_{ijt1} + \beta_2 x_{ijt2} + \beta_3 x_{ijt3} + \dots + \beta_k x_{ijtk} + \text{offset}(E_{ijt}), \quad (3.5)$$

where μ_{ijt} is the expected value of the outcome variable y_{ijt} for subject in this case μ_{ijt} is the expected number of malaria cases, β_0 is an intercept that explain the overall mean incidence, $\beta_1, \beta_2, \dots, \beta_k$ are estimated parameters regression coefficients) and $x_{ij1}, x_{ij2}, x_{ij3} \dots x_{ijk}$, are the explanatory/independent variables with corresponding estimated parameters β_k , and E_{ijt} is the added offset. The parameters $(\beta_0, \beta_1, \dots, \beta_k)$ are estimated by maximum likelihood estimators (MLEs) using iterative algorithms such as Newton-Raphson (NR) and iteratively re-weighted least square and the goodness fit of the model is checked using Pearson chi-square statistic χ^2 .

Interpretation of the Poisson regression parameter

For a binary explanatory variable denoted by an indicator variable, let $x_i = 1$ be the variable in category 1 (if the factor is present) and $x_i = 0$ be the variable in category 0 (if factor is absent) and rate ratio: $RR = \frac{E(y_{ijt}|present)}{E(y_{ijt}|absent)} = e^{\beta_i}$. If x_k is continuous, a one-unit increase will result in a multiplicative effect of e^{β_k} on the rate μ , e^{β_0} is the rate for the observation being studied if all $x_i = 1, 2, 3, \dots, n$ are equal to zero.

Model assumption

- (i) Poisson model assumes that the dependent variable (response) consists of count data (i.e. count y_{ijt}) which is either a zero or greater (nonnegative integer). For one to run a Poisson regression model, the minimum value of y_{ijt} must be 0. In this study y_{ijt} is defined to be the count of people who tested malaria positive by RTD in constituency i at time j and t .

We can test this by computing the expected counts and plot them with the observed counts to see if they are related but in this case one cannot clearly see how these are related. Hence the best way is to use a formal hypothesis test, known as a chi-squared (χ^2) goodness-of-fit test. For instance, if one thinks our data might follow Poisson distribution, the null hypothesis (H_0) can be formulated as alternative Hypothesis (H_1) for example, H_0 : The data follows a Poisson distribution and H_1 : The data does not follow a Poisson distribution. For the goodness-of-fit test to work, all expected frequencies must be ≥ 5 ; to achieve this, adjacent categories can be "pooled" and we reject H_0 if p-value is less than α , level of significance.

- (ii) The information (data) should include at least one or more independent variables that can be measured on a continuous, ordinal, or nominal scale. Ordinary, nominal, or continuous explanatory variables should be used.
- (iii) The observation should be independent. Meaning that the occurrences of a random event in an interval of time are independent. In other words, one

observation cannot provide any information on another observation or to say there is no relationship between the observations in each group or between the groups themselves. This assumption can also be checked by comparing standard model-based errors versus robust errors to see whether there are any significant differences.

Goodness of fit testing

The Wald test statistics could be used to measure the model's quality of fit (named after Abraham Wald). The Wald test assesses statistical parameter constraints by calculating the weighted distance between the unrestricted estimate and its expected value under the null hypothesis, where the weight represents estimate precision.

Wald test statistics is denoted as follows:

$$W = \frac{\hat{\beta}_i - \beta_0}{\widehat{se}\hat{\beta}_i} \quad (3.6)$$

3.6.2 Negative binomial regression

If the data failed the premise of equidispersion, it is usually a good idea to look for "apparent Poisson Overdispersion." As a result, if the assumption is still broken, various changes must be made to the Poisson distribution. model to confirm if it is overdispersed (Hilbe, 2011; Swartout et al., 2015).

- (i) Ensuring that all relevant variables are included in the model;
- (ii) Checking if the data have some outliers;
- (iii) Checking if the model included all relevant interaction terms;
- (iv) Checking if one of the variables needs to be transformed;
- (v) Checking if the data is sufficient for the model or the data is too sparse, for example if there are many gaps present in the data being recorded, and
- (vi) If there are missing values that are not missing at random.

After testing poisson assumptions, the data failed the premise of equidispersion assumption, as a result, Negative binomial regression a variant of Poisson regression that lowers the Poisson model's restrictive constraint that variance equals mean can be utilised (Hilbe and Greene, 2007; Collins, Waititu, and Wanjoya, 2020; Leckie et al., 2020) was considered. Negative Binomial regression a special case of Poisson–gamma mixture assesses the significance of variability in the incidence ratio by modeling Poisson heterogeneity with a gamma distribution and log link function. In this study, the number of cases (y_{ijt}) was assumed to follow a poisson distribution in the negative binomial model, while the mean (μ_{ijt}) follows a Gamma distribution (Cameron and Trivedi, 2001; Lord et al., 2006; Greene, 2008; Hilbe, 2011; Cameron and Trivedi, 2013).

The negative binomial distribution is denoted as follows:

$$p(y) = p(Y = y) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1) \Gamma(\alpha^{-1})} \left(\frac{1}{1 + \alpha\mu} \right)^{\alpha^{-1}} \left(\frac{\alpha\mu}{1 + \alpha\mu} \right)^y, \quad (3.7)$$

where $\mu > 0$ is the mean incidence rate of Y per unit of exposure (e.g., population size area, distance, or time) and $\alpha > 0$ is the heterogeneity parameter.

let $\beta = (\beta_0, \beta_1, \beta_2, \beta_3 \dots \beta_k)^T$ the predictor data is then compiled into a design matrix X , as shown below.:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ 1 & x_{31} & x_{32} & x_{33} & \dots & x_{3k} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \dots & x_{nk} \end{pmatrix}$$

Designating the i^{th} row of x to be x_{ijt} and exponentiating the negative binomial model Eq (3.12), the negative binomial distribution can be rewritten as follows:

$$p(y_{ijt}) = \frac{\Gamma(y_{ijt} + \alpha^{-1})}{\Gamma(y_{ijt} + 1) \Gamma(\alpha^{-1})} \left(\frac{1}{1 + \alpha e^{x_{ijt}\beta}} \right)^{\alpha^{-1}} \left(\frac{\alpha e^{x_{ijt}\beta}}{1 + \alpha e^{x_{ijt}\beta}} \right)^{y_{ijt}}, \quad (3.8)$$

where $i = 1, 2, \dots, n$ and the parameters α and β are estimated using maximum likelihood estimation and the likelihood function is denoted as follows:

$$L(\alpha, \beta) = \prod_{i=1}^n p(y_{ijt}) = \prod_{i=1}^n \frac{\Gamma(y_{ijt} + \alpha^{-1})}{\Gamma(y_{ijt} + 1) \Gamma(\alpha^{-1})} \left(\frac{1}{1 + \alpha e^{x_{ijt}\beta}} \right)^{\alpha^{-1}} \left(\frac{\alpha e^{x_{ijt}\beta}}{1 + \alpha e^{x_{ijt}\beta}} \right)^{y_{ijt}}, \quad (3.9)$$

And the likelihood function is given by:

$$\ln L(\alpha, \beta) = \sum_{i=1}^n \{y_{it} \ln \alpha + y_{it} (x_{it}\beta) - (y_{it} + \alpha^{-1}) \ln(1 + \alpha e^{x_{it}\beta}) + \Gamma(y_{it} + \alpha^{-1}) - \Gamma(y_{it} + 1) - \Gamma(\alpha^{-1})\} \quad (3.10)$$

The value of α and β that maximize $\ln L(\alpha, \beta)$ is the maximum likelihood estimate

The negative binomial model can be expressed as a Poisson-gamma mixture. Let $y_{ijt}|E_{ijt}, \theta_{ijt} \sim \text{Poisson}(\mu_{ijt})$, the Poisson mean μ_{ijt} is structured as follows:

$$E(y_{ijt}|E_{ijt}, \theta_{ijt}) = \mu_{ijt} = \exp(x_{ijt}^T \beta + \sigma \epsilon_{ijt} + E_{ijt}). \quad (3.11)$$

(Cameron and Trivedi, 2001; Lord et al., 2006; Greene, 2008; Hilbe, 2011; Cameron and Trivedi, 2013). In familiar notation this could be rewritten as:

$$\ln(E(y_{ijt}|E_{ijt}, \theta_{ijt})) = \ln(\mu_{ijt}) = \beta_0 + \beta_1 x_{ijt1} + \beta_2 x_{ijt2} + \beta_3 x_{ijt3} + \dots + \beta_k x_{ijt k} + \sigma \epsilon_{ijt} + \text{offset}(E_{ijt}), \quad (3.12)$$

where $\sigma \epsilon_i$ is the disturbance model error that is independent of all covariates, and $\exp(\epsilon_i)$ is assumed to have a gamma distribution with a mean equal to 1 and a smaller variance.

3.6.3 Model specification (spatial model and spatial temporal)

Bayesian hierarchical modelling

Bayesian hierarchical modeling is a statistical model with numerous levels (a hierarchical model) that employs the Bayesian approach to estimate the parameters of the

posterior distribution (Allenby and Rossi, 2006). The hierarchical model is formed by combining the sub-models, and the Bayes' theorem is applied to integrate them with observed data and account for any uncertainty (Allenby and Rossi, 2006). As further evidence on the previous distribution is acquired, the posterior distribution, also known as the updated probability estimate, is the consequence of this integration (Allenby and Rossi, 2006).

Bayesian hierarchical modeling is a statistical model created in several levels (a hierarchical model) that employs the Bayesian approach to estimate the parameters of the posterior distribution (Allenby and Rossi, 2006). The submodels are merged to form a hierarchical model, and the Bayes' theorem is applied to integrate them with observed data and account for all uncertainty. (Allenby and Rossi, 2006). The posterior distribution, also known as the updated probability estimate, is the result of this integration as more data on the previous distribution is collected (Allenby and Rossi, 2006).

Bayesian hierarchical model components

A Bayesian hierarchical model is made up of three parts: the data model, which describes the data distribution given the parameters, the process model, which describes the underlying spatial trend, as well as the parameter model, which describes the prior distribution of the parameters to be estimated (Lesaffre and Lawson, 2012).

Data model: Spatial

Let y_{it} be the observed number of malaria case in each constituency i ($i = 1, 2, 3, \dots, n$) at time t ($t = 1, 2, \text{ and } 3$), E_{it} be the expected cases, and θ_{it} be the relative risk in constituency i at time t .

Then the observed y_{it} in n counties could be treated as one realisation of Negative Binomial random variables, that follows a Negative Binomial distribution with mean μ_{it} which is a function of k covariates (x_{kit}) as well as spatial random effects. Then: $y_{it}|E_{it}, \theta_{it} \sim NB(E_{it}\theta_{it})$. This can be rewritten as: $y_{it} \sim NB(\mu_{it})$

Data model: Spatio-temporal

The Besag, York, and Mollie's (BYM), model developed by Besag, York, and Mollie is a popular hierarchical Bayesian model. When working with area/lattice data, this model (BYM), is known to be the best model for modeling the spatial and temporal effects (Moraga, 2019). As a result, the BYM spatial model was fitted on the 3 years' combined (aggregated) dataset, whereas the BYM spatio-temporal model was fitted on the 2018 to 2020 disaggregated malaria complete data. This model incorporates random effects resulting from unstructured and spatially structured heterogeneity into the log-linear model for relative risk, allowing smoothing relative risks at both the global and local levels.

Let y_{it} be the observed malaria case in each constituency i ($i = 1, 2, 3, \dots, 102$) at time t ($t = 1, 2, \text{ and } 3$), E_{it} be the expected cases in constituency i at time t , $\theta_{it} = \frac{y_{it}}{E_{it}}$ be the relative risk in constituency i at time t .

The observed y_{it} in the 102 counties could be treated as one realisation of Poisson random variables, that follows a Poisson distribution with mean μ_{it} ($\mu_1 \dots \mu_n$) = $E_{it} \times \theta_{it}$ which is a function of k covariates (x_{kit}) as well as spatial and temporal random effects. Then, $y_{it} | E_{it}, \theta_{it} \sim NB(E_{it}\theta_{it})$. This can be rewritten as, $y_{it} \sim NB(\mu_{it})$

Process model and the parameter model

The relative risk structure's underlying structure is described by the process model. We employed the spatial model's spatio-temporal extension (Besag-York-Mollie) model, which is a CAR convolution model with two random effects, one spatially organized area-specific random impact and one unstructured area-specific random influence (Besag, 1991).

I. Spatial model

$$\ln(E(y_{it} | \theta_{it})) = \ln(\mu_i) = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{it3} + \dots + \beta_k x_{itk} + \text{offset}(E_{it}) + \sigma e_{it} + u_i + v_i, \quad (3.13)$$

This model is an extension of non-spatial model Eq. (3.12), where u_i is the spatial structured random effects and, v_i spatially unstructured random effects that account for spatial dependence.

Priors for the spatial random effects were set to follow log gamma distribution with mean = 0, precision = 0.001 that corresponds to large variance because the study used a Negative Binomial model which is a generalisation of the Poisson model, while the default prior assigned to the associated coefficients (and the intercept) was a Gaussian distribution since there was no enough information available to fully specify a precise prior.

II. Spatio-temporal model

The spatial Besag-York-Mollie (BYM) spatial model (Eq. (3.13)) is extended to allow for a temporal component:

$$\ln(E(y_{it} | E_{it}, \theta_{it})) = \ln(\mu_{it}) = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + \dots + \beta_k x_{kit} + \sigma e_{it} + u_i + v_i + \text{Temporal}_t(\beta_t), \quad (3.14)$$

with $t = 1, 2, \text{and } 3$. On $\text{Temporal}_t(\beta_t)$, a parametric or nonparametric structure can be specified to present a parametric trend for the temporal component and nonparametric trend for the temporal component respectively (Blangiardo and Cameletti, 2015).

The linear predictor of parametric trend (parametric model with fixed covariates and time component)

From the spatial Besag-York-Mollie (BYM) spatial model Eq.(3.13), $(\beta + \delta_i) \times t$ denoting the main spatial effect was added to Eq. (3.13) to accommodate for temporal component:

$$\ln(E(y_{it} | E_{it} \theta_{it})) = \ln(\mu_{it}) = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + \dots + \beta_k x_{kit} + \sigma e_{it} + u_i + v_i + (\beta + \delta_i) \times t, \quad (3.15)$$

where β is the main linear trend that represents the global time effect, and δ_i be the differential trend that characterizes the interaction of time and space.

The linear predictor of non-parametric trend:

$$\ln(E(y_{it} | \theta_{it})) = \ln(\mu_{it}) = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + \dots + \beta_k x_{kit} + \sigma e_{it} + u_i + v_i + \gamma_t + \phi_t, \quad (3.16)$$

where γ_t represent the structured temporally effect modelled dynamically using random walk of order 1 or 2 while ϕ_t represent the unstructured temporally effect

III. Full spatio-temporal extension of the spatial Besag-York-Mollie (BYM) model, which is the CAR convolution model with two random effects, one spatially structured and one unstructured area-specific random effect

Full spatio-temporal extension of the spatial Besag-York-Mollie (BYM) model is a extension of Eq.(3.16) where δ_{it} was introduced to allow for a spatial-temporal interaction, which would explain variances in the time trend of malaria cases for various constituencies.

$$\ln(E(y_{it} | \theta_{it})) = \ln(\mu_{it}) = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + \dots + \beta_k x_{kit} + \sigma e_{it} + u_i + v_i + \gamma_t + \phi_t + \delta_{it}, \quad (3.17)$$

where γ_t is a random term representing between time (year) variation and was assumed to be an autoregressive process, u_i is a spatially organised, area-specific

random effect that allows for smoothing between adjacent areas, namely:

$$u_i | u_j \sim N\left(\bar{\mu}_{\delta_i}, \frac{\sigma_u^2}{n_{\delta_i}}\right)$$

with δ_i , be the set of neighbours and n_{δ_i} be the number of neighbours for a specific area i . The unstructured component v_i is modelled using as a Gaussian process.

$$v_i \sim N(0, \sigma_u^2)$$

To account for increased variation in counts caused by unobserved (and geographically unstructured) risk variables, the γ_t term that represents the temporally structured effect was modelled dynamically using random walk of order 2 (*RW of order 2*) defined as:

$$\gamma_t | \gamma_{t-1}, \gamma_{t-2} \sim N(2\gamma_{t-1} + \gamma_{t-2}, \sigma^2).$$

While ϕ_t term is specified using a Gaussian exchangeable prior, which is defined as:

$$\phi_t \sim N\left(0, \frac{1}{\tau_{\phi_t}}\right)$$

To allow time and space to communicate with one another, which explains differences in the time trend of malaria risk for different constituencies, the parameter δ_{it} has a Gaussian distribution with a precision matrix provided by $\tau_{\delta} R_{\delta}$ where τ_{δ} is an unknown scalar and R_{δ} is the structure matrix specifying the type of temporal and/or spatial dependence between the elements of δ .

R_{δ} may be factored as the Kronecker product of the structural matrix of interacting primary effects. The structure matrix can be defined in four ways (Blangiardo and Cameletti, 2015).

The fitted spatio-temporal model considered three different types of interactions (see Table 3.4.)

TABLE 3.4: List of interaction considered in the study

Interaction	Parameter Inter-acting	Rank
I. Nonspatially or Temporally structured interaction	v_i and ϕ_t	nT
II. Temporally structured interaction	v_i and γ_t	$n(T - 2)$ for RW2
III. Spatially structured interaction	ϕ_t and u_i	$(n-1)T$

I. Nonspatially or Temporally structured interaction

This interaction assumes that the unstructured spatial effects (v_i) is interacting with the unstructured temporal effect (ϕ_t) and the structure matrix is denoted as follows:

$$R_\delta = R_u \otimes R_\phi = I \otimes I = I$$

Since we assumed not spatial neither temporal structure on this interaction I, then:

$$\delta_{it} \sim N\left(0, \frac{1}{\tau_\delta}\right)$$

This means that the interaction (I) can be viewed as unobserved independent factors for each constituency and year combination, resulting in no structure (López-Quilez and Munoz, 2009). However, if the model includes spatial and temporal main effects, this interaction effect simply suggests independence in deviations from them (López-Quilez and Munoz, 2009). Due to the main effects, contributions to malaria risk in neighbouring constituencies or in subsequent years (e.g., 2018, 2019, and 2020) can still be highly connected. As a result, this is a global space-time heterogeneity effect that is typically modelled as white noise.

Again, interaction (I) can refer to any non-permanent factor that can induce a minor rise or decrease in malaria rates in a given constituency-year, allowing for random-independent-oscillations around the expected rates (López-Quilez and Munoz, 2009).

II. Temporally structured interaction

This interaction assumes that the unstructured spatial effect (v_i) is interacting with the structured temporal effect (γ_t) and the structure matrix is denoted as follows:

$$R_\delta = R_v \otimes R_\gamma ,$$

where $R_v = I$ and R_γ is the neighbouring structure defined through second – order random walk (RW2). This results in the assumption that for the i^{th} constituency, the parameter vector $\{\delta_{1i}, \dots, \delta_{iT}\}$ has a time-dependent autoregressive structure component which does not depend from the ones of the other constituencies (Orford, 2001).

In this interaction, each zone (constituency) has its own structure that is distinct/independent from nearby constituencies, and the evolution structure for each constituency can take on as many forms as the temporal main impact itself. However, this does not imply that each constituency evolves independently of the others, as they may have a similar temporal main effect. Independence has little effect on deviations from the global trend (López-Quilez and Munoz, 2009).

From literature, this interaction is best when it comes to fitting diseases models to identify risk factors in specific areas e.g., constituencies as well as bringing deviations from the global trend to provide a good balance between fit and complexity (López-Quilez and Munoz, 2009).

III. Spatially structured interaction (last fitted interaction)

This Interaction assumes that the unstructured temporal effects (ϕ_t) is interacting with the structured spatial effect (u_i) and the structure matrix is denoted as follow:

$$R_\delta = R_\phi \otimes R_u ,$$

where $R_\phi = I$ and R_u is the neighbouring structure defined by the second – order

random walk (*RW2*) algorithm. This leads to the presumption that the i^{th} (year) time point $t = \delta_t, \dots, \delta_{nt}$ has a spatial structure distinct from the previous time points (Orford, 2001).

Similarly, each period of the interaction can be regarded to have its own spatial structure, irrespective of previous periods (its neighbours in time). For example, spatial clustering effects is classically modelled with a CAR distribution for each year (Orford, 2001). This interaction can also yield to interaction I (non-spatially or temporally interaction) because of the insertion of a spatial term for heterogeneity (such as in the BYM specification) (Orford, 2001).

As a result, it is expected that each unique region may deviate somewhat from the global trend, but that this divergence will be similar to that of neighbouring regions while remaining independent of the prior or subsequent period of time.

This is also one of the important interactions to address in spatial temporal modelling because it depicts instances in which an unobserved regional factor affects an area including two or more nearby zones but is not persistent in time (Orford, 2001).

For other parameters e.g., β_0 , we allocated previous distributions to scaled precision matrix parameters based on their diagonal marginal standard deviations. The prior distribution follows a weakly informative prior with zero-mean Gaussian distribution with a large variance which is usually an appropriate prior in most of the cases since there is always not enough information available to fully specify a precise prior. A variety of models were run, and the model with the lowest DIC value was deemed the most effective.

3.7 Parameters estimation, spatial autocorrelation, and best model selection criteria

Parameters estimation

We defined a queen nb neighbourhood as adjacent counties with adjacency weights $W_{ij} = 1$ if constituency i and j share a common boundary and $W_{ij} = 0$ otherwise. The parameters were estimated using the Integrated Nested Laplace approximation (INLA), which takes considerably less time than the Markov Chain Monte Carlo Methods (MCMC) (Gómez-Rubio, Bivand, and Rue, 2021). The priors for all parameters were set to follow log gamma distribution because the study used negative binomial model which is a generalisation of the Poisson model with mean = 0 and very low precision = 0.001 that corresponds to a large variance. Sensitivity analysis was performed on all the models due to the inherent challenges that arise with each formulation.

The same priors were considered to fit another model BYM2. This model improves parameter control by allowing the parameters to be seen independently of one another, as opposed to BYM, which does not allow the spatially organised component to be seen independently of the unstructured component (Orford, 2001).

Spatial autocorrelation testing

Several R packages were used in the analysis. Moran's I was used to test the data for global spatial autocorrelation. Malaria occurrence clusters were then identified using local Moran and Getis and Ord spatial statistics. The mean population of each constituency from 2018 to 2020, as well as the constituency mean for all climatic variables included in the model as possible factors/covariates (fixed effects), were also computed using yearly data from 2018, 2019, and 2020. According to the literature, disease mapping considers the sum of cases as the dependent variable in spatial and spatial temporal analysis (Orford, 2001). In this study, the sum and mean of all confirmed malaria positive cases for three years (2018, 2019, and 2020) was

computed and used as the dependent variable for disaggregated and aggregated data respectively. There was a unique identification code (object ID) for all records in the dataset for spatial analysis purposes, where total malaria cases per constituency were cross-linked to constituency names by a unique identification code, allowing constituency analysis to be performed.

The Moran I Index was calculated to determine whether there is a global spatial autocorrelation in the data/malaria case clusters in Namibia. Moran's I statistical test's null hypothesis was that the incidence of malaria observed in constituency i is independent of that observed in the neighbouring constituency (j), which was defined by linking each constituency to its immediate neighbour. The local moran statistic and local G maps were then used to determine where there is high/low clustering.

Best model selection

Using the malaria data, all of the Poisson Regression assumptions were tested, and the equidispersion assumption was violated, indicating that the data is dispersed. As a result, in the non-spatial, spatial, and spatial temporal final models, a Negative binomial (Poisson-gamma mixture) distribution was considered to allow the modelling of Poisson heterogeneity through gamma distribution and log link function to assess the significance of variability in the incidence ratio.

Preliminary analysis of univariate and multivariate associations of covariates considered, and incidence rate were obtained from both Poisson and Negative Binomial non-spatial models, and the result from the Negative Binomial non-spatial model conspired to explain the effects better (low AIC value) than the result from the Poisson non-spatial mode.

TABLE 3.5: List of Analysis performed in the study

ANALYSIS DESCRIPTION	
Analysis 1	Descriptive analysis
Analysis 2	Poisson non-spatial modelling (using 2018-2020 aggregated dataset)
Analysis 3	Negative Binomial non-spatial modelling (using both 2018-2020 aggregated dataset and 2017-2019 individual dataset)
Analysis 4	Negative Binomial spatial modeling (using both 2018-2020 aggregated dataset and 2017-2019 individual dataset)
Analysis 5	Negative Binomial spatial-temporal model (using 2018-2020 disaggregated dataset)

Table 3.5 have summarized the analysis performed in this study.

3.8 Model selection criterion

3.8.1 Akaike Information Criterion (AIC)

The AIC is a statistical approach to assess how well a non-spatial model fits the data it was generated from. In statistics, AIC is used to compare different possible models and determine which one is the best fit for the data. It is generated using the number of independent variables examined in the model and the model's highest likelihood estimate (how well the model reproduces the data).

$$AIC = 2k - 2\ln(L), \quad (3.18)$$

where k represents the number of parameters in the statistical model and L represents the maximum likelihood function value for the estimated model.

The best-fit model according to AIC is the one that explains the greatest amount of variation using the fewest possible independent variables. Several models are computed from the same dataset, and the model with the lower AIC is deemed the optimal model, explaining the largest amount of variation with the fewest potential independent variables (Moffatt, 2017). The phrase $2k$ refers to a penalty for overfitting, and it discourages the use of too many variables in models, which always leads

to a decreased likelihood.

3.8.2 Bayesian Information Criterion (BIC)

Schwarz (1998) proposed the BIC as a model selection criterion which is similar to AIC in that it is based on empirical log-likelihood functions. When fitting a model, adding parameters might enhance the likelihood, but this can lead to overfitting. As a result, the BIC addresses this issue by incorporating a penalty term for the amount of parameters in the model. The penalty term in BIC is higher than in AIC, and the model with the lowest BIC value is the best fit.

The BIC is given by:

$$BIC = k \ln(n) - 2 \ln(\hat{L}), \quad (3.19)$$

(Coban and Sayil, 2019), where

\hat{L} = is the model's likelihood function's maximum value, and
 k = the number of free parameters to be approximated n = the number of observations, or, more precisely, the sample size

3.8.3 Deviance Information Criterion (DIC)

This is a hierarchical generalisation of the AIC and BIC, which are commonly used in spatial and spatio-temporal model selection, such as Bayesian model selection using MCMC simulation. Although this is not always true, the derivation of DIC assumes that the genuine model is included by the stated parametric family of probability distributions that generate future data. In DIC, the observed data are used to evaluate the estimated models. As a result, DIC prefers models that are over-fitted. Nonetheless, the best model is the one with the lowest DIC value. The definition of DIC is as follows:

$$D(\theta) = 2 \log(f(y)) - 2 \log(p(y|\theta)), \quad (3.20)$$

(Wikipedia contributors, 2021; Celeux et al., 2006 ; Bürkner, 2017), where y are the data, θ are the unknown parameters of the model, and $p(y|\theta)$ is the likelihood function.

The model's effective number of parameters is calculated as follows:

$$pD = \overline{D(\theta)} - D(\bar{\theta}), \quad (3.21)$$

(Wikipedia contributors, 2021; Bürkner, 2017), where $\bar{\theta}$ is the expectation of θ and pD is the effective number of parameters.

Then, DIC is then calculated as:

$$DIC = pD - \bar{D} \quad (3.22)$$

Models are also penalised by the value of \bar{D} , which promotes a strong fit.

CHAPTER 4

RESULTS

The obtained results are displayed and discussed in this chapter. Sections have been created from the received results. The initial portion offers more extensive information about the data (descriptive analysis), which is followed by additional sections, such as results from non-spatial, spatial, and spatial-temporal models.

4.1 Description and distribution of 2018 to 2020 malaria incidence rate in Kavango E and W, Zambezi and Ohangwena region

The description and distribution of 2018 to 2020 malaria incidence rate in Kavango E and W, Zambezi and Ohangwena region is displayed in Table 3.1. Results obtained from 2018 to 2020 aggregated malaria data show that a total of 31 619 people in the 4 regions (Kavango East and West, Ohangwena, and Zambezi region) tested malaria positive by RTD in 2018, while in 2019 and 2020, the total number of people who tested malaria positive was 2 990 and 10 035 respectively, for a total of 44 644 malaria positive cases reported over a three-year period from 2018 to 2020 (Table 4.1). Furthermore, for all the years, most cases were reported during Autumn (March, April, and May) and Summer (December, January, and February) (Figure 4.1 and Table 4.1). Namibia's peak malaria season typically occurs between January and May following the rainy season, as most cases were reported during this period in Autumn (March, April, and May) and also Summer (December, January, and February), which is also the season with the highest rainfall in the three years stretching from 2018 to 2020

(Figure 4.1 and Table 4.1).

Malaria cases transmission were found to be highly seasonal (Figures 4.1 and 4.3). Figure 4.1 shows that the number of malaria cases observed is directly proportional to the amount of rainfall received. The more rain there was, the more malaria cases there were.

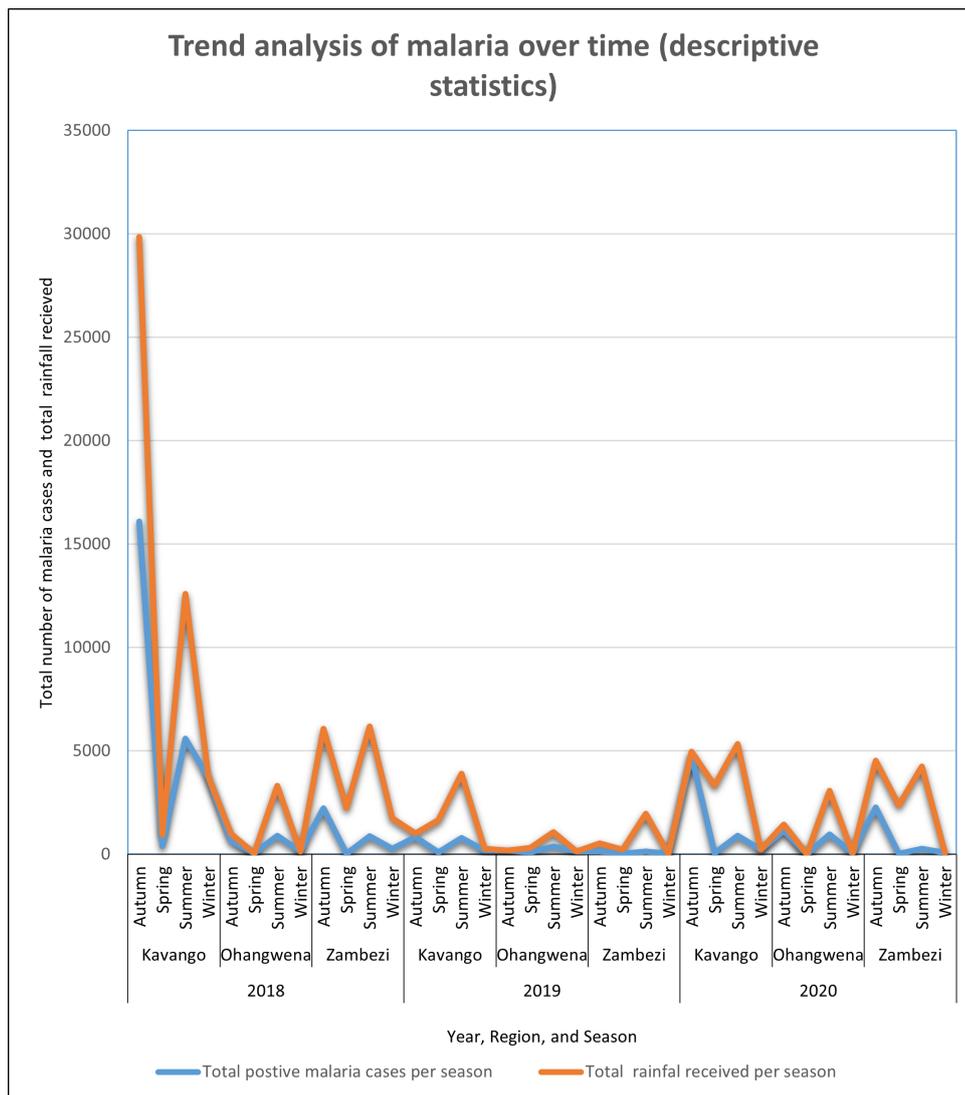


FIGURE 4.1: Trend analysis of malaria (2018-2020)

TABLE 4.1: Total number of people suspected to have malaria and tested positive per region per season

Year	Region	Season	Suspected malaria cases	Positive malaria cases
2018	Kavango E & W	Autumn	63834	17103
		Spring	34079	380
		Summer	30854	5103
		Winter	52504	3712
		Total	181271	26298
	Ohangwena	Autumn	20556	664
		Spring	15001	44
		Summer	13413	869
		Winter	18085	132
		Total	67055	1709
	Zambezi	Autumn	8991	2214
		Spring	4439	31
Summer		5147	880	
Winter		7333	277	
Total		25910	3402	
2018 Total			274236	31619
2019	Kavango E & W	Autumn	22438	828
		Spring	26805	98
		Summer	22732	793
		Winter	30010	172
		Total	101985	1891
	Ohangwena	Autumn	10805	172
		Spring	16911	92
		Summer	12341	384
		Winter	21102	135
		Total	61159	783
	Zambezi	Autumn	4667	126
		Spring	6179	20
Summer		4227	143	
Winter		6382	27	
Total		21455	316	
2019 Total			184599	2990
2020	Kavango E & W	Autumn	32841	4757
		Spring	13713	45
		Summer	19220	897
		Winter	16410	214
		Total	82184	5913
	Ohangwena	Autumn	16304	1131
		Spring	4401	8
		Summer	11432	955
		Winter	5240	39
		Total	37377	2133
	Zambezi	Autumn	7454	2266
		Spring	2700	30
Summer		3567	270	
Winter		2335	66	
Total		16056	2632	
2020 Total			135617	10035
Grand Total			594452	44644

had the lowest average incidence rate (Figure 4.3).

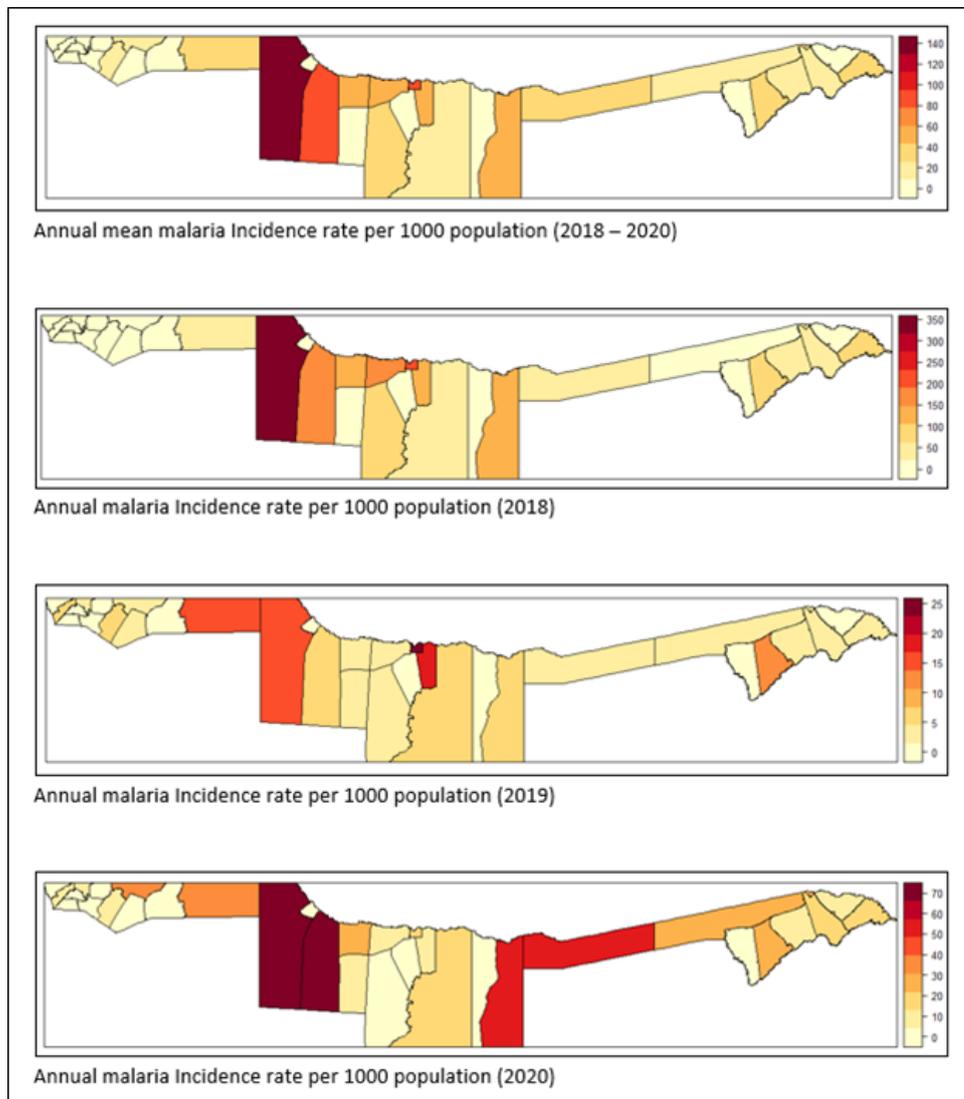


FIGURE 4.3: Malaria Incidence rate per 1000 population at constituency level

TABLE 4.2: Summary of the 11 constituency that revealed the highest malaria incidence rate per 1000 population in 2018, 2019, and 2020

2018		2019		2020	
Constituency	IR	Constituency	IR	Constituency	IR
Mpungu	335	Rundu Urban	24	Tondoro	70
Rundu Urban	195	Rundu Rural	20	Mpungu	67
Kapako	168	Okongo	15	Mukwe	51
Tondoro	164	Mpungu	15	Ndiyona	49
Musese	135	Linyanti	11	Okongo	37
Rundu Rural	116	Engela	6	Omundaungilo	33
Ndiyona	110	Ndiyona	6	Musese	27
Ncamagoro	80	Tondoro	6	Katima Mulilo Urban	26
Linyanti	79	Mashare	5	Kongola	25
Kabbe South	64	Eenhana	4	Linyanti	23
Mukwe	50	Musese	3	Katima Mulilo Rural	21

IR: Incidence rate per year

Table 4.2 shows the 11 constituencies with the highest malaria incidence per 1000 inhabitants during the three years (2018 – 2020). Mpungu, Rundu Urban and Rural, Linyanti, Tondoro, Kapako, Mukwe, Engela, Mukwe, Okongo, Ndiyona, Mashare, Kongola, Kabbe south, Ncamagoro, and Eenhana constituencies had the highest malaria incidence rate in the study's 34 constituencies (Table 4.2).

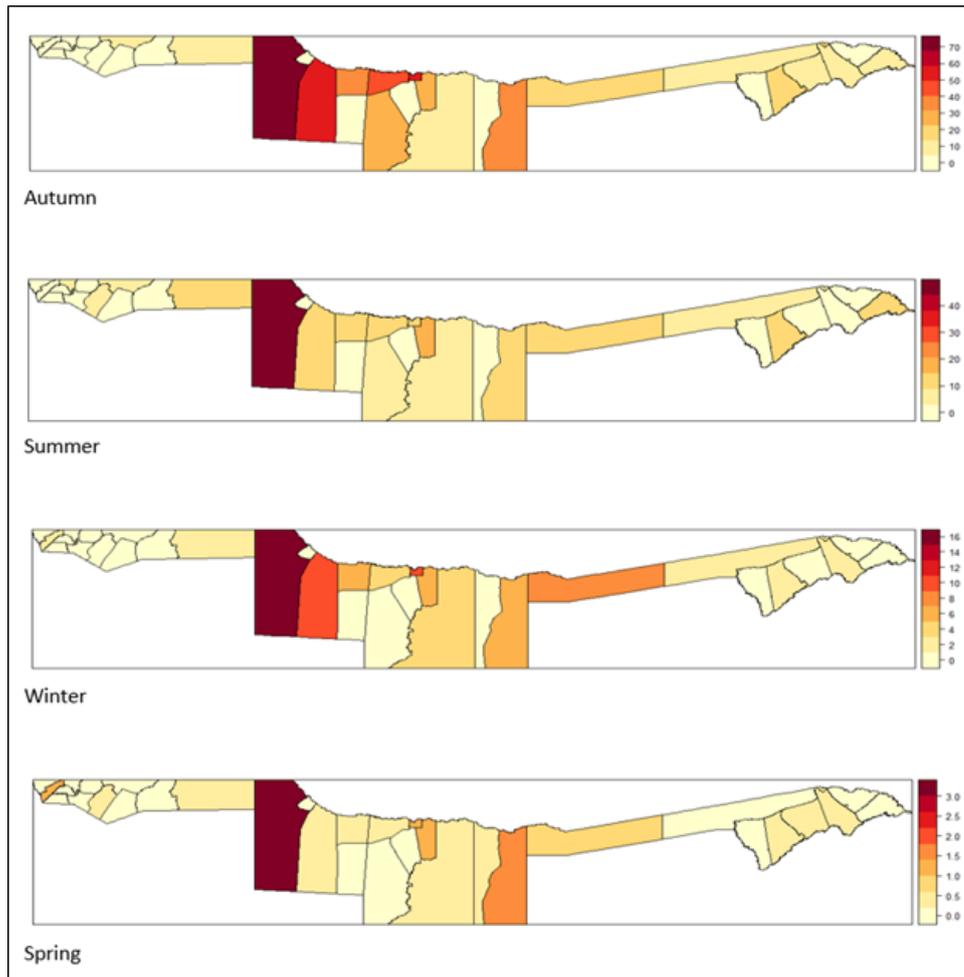


FIGURE 4.4: Annual average malaria incidence rate per 1000 population per season

Furthermore, the Autumn and Summer seasons had a higher average malaria incidence rate when compared to other seasons (Figure 4.4). During the spring months (September, October, and November), only Mpungu had an average of 3 cases per 1000 people, while the rest (33 constituencies) had an average of 1 case per 1000 people, indicating that few malaria cases are reported during those months (Figure 4.4).

TABLE 4.3: Summary of computed malaria incidence per year

Year	Estimated population	Number of cases	Mean rate	IR/1000 population (Namibia)	IR/1000 population in 4 regions
2018 -2020	2494579	14881	0.00597	6	23
2018	2448301	31619	0.01291	12	49
2019	2494530	2990	0.00120	1	5
2020	2540905	10035	0.00395	4	15

IR: Incidence rate per 1000 population

The mean incidence rate was calculated at the country level using Namibia's current population of 2,594,488 as of Thursday, September 2, 2021, as compiled by Worldometer from the most recent United Nations data, excluding the less than 5% of cases recorded in the other ten regions of the country.

(<https://worldpopulationreview.com/countries/namibia-population>).

The three-year average incidence rate (2018-2020) was found to be 6 cases per 1000 population, with 12 cases per 1000 population in 2018. In 2019, the incidence rate was 1 case per 1000 people, then increased to 4 cases per 1000 people in 2020 (See Table 4.3).

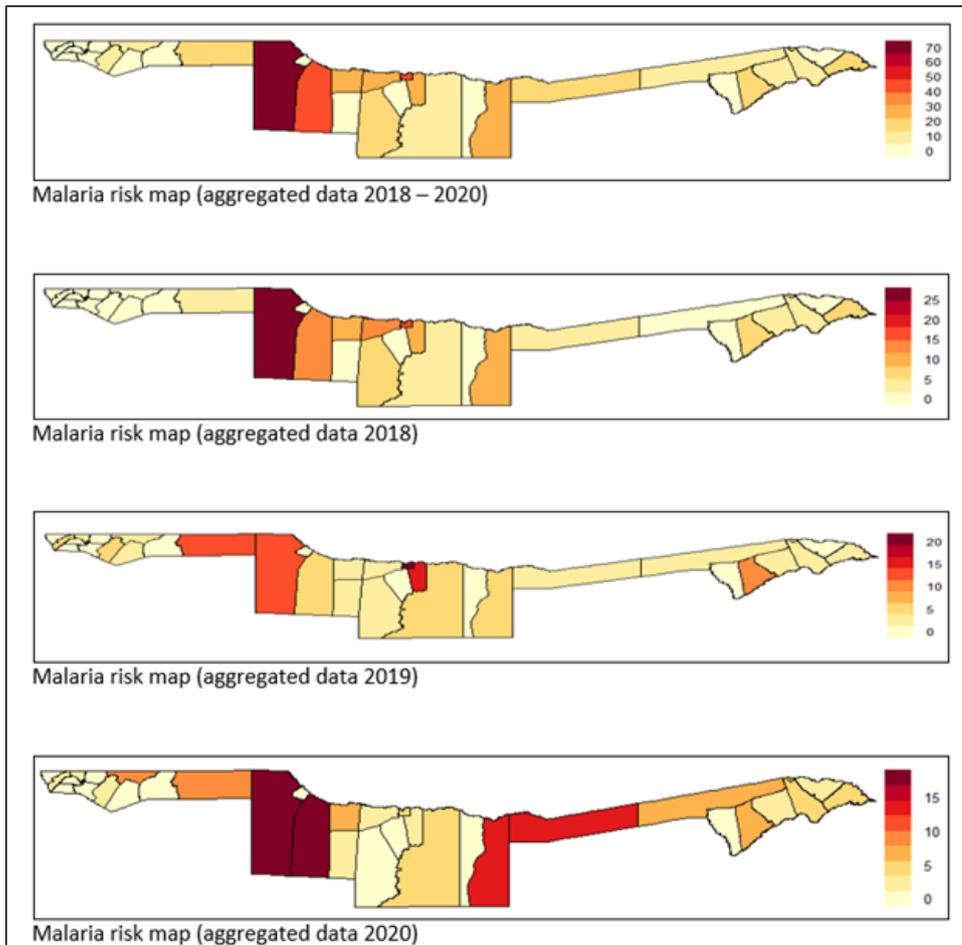


FIGURE 4.5: Malaria risk map (unpredicted)

Individuals in Mpungu were found to be at higher risk, followed by those in Tondoro and Rundu Urban constituencies for both years (Figure 4.5). The number of constituencies at high risk of malaria increased in 2020, with Ndiyona and Mukwe in the East of Kavango becoming one of them (Figure 4.5).

4.2 Non-spatial examination of malaria risk variables in Namibia (Non-spatial modelling)

One of the study's objectives was to identify risk factors/covariates associated with malaria. As a result, for both datasets (aggregated dataset 2018-2020 and individual dataset 2017-2019), a variety of models were fitted using R statistical software by adding and removing variables, and the model was chosen using the AIC value, with the model with the lowest AIC value considered the best model (See R-code presented in Appendix A).

Variables added to the non-spatial model from the malaria aggregate dataset are listed in Table 4.4, whereas variables added to the non-spatial model from the malaria individual dataset are listed in Table 4.5.

TABLE 4.4: Variable considered for non-spatial Negative Binomial regression model (aggregated data 2018-2020)

Variable code	Variable name
Region	Region of the individual tested malaria positive by RTD
Constituency	Constituency of the individual tested malaria positive
ConstID	constituency ID
Population	Total number of people living in that constituency divided by 1000 (offset)
Season	Season (Autumn, Summer, Winter and Spring)
PosRDT	Annual mean of people tested malaria positive by RTD (Depended variable)
HPD	human population density of the constituency
T (avg)	Annual average temperature (mean)
T (min.)	Annual minimum temperature (mean)
T (max.)	Annual maximum temperature (mean)
RF	Annual total rainfall (mean)
WS (avg)	Annual average wind speed (mean)
ST (avg)	Annual average soil temperature (mean)
LW (avg)	Annual average leaf wetness (mean)
H	Annual Humidity (mean)

TABLE 4.5: Variable considered for non-spatial Negative Binomial regression model (Individual data 2017-2019)

Variable code	Variable name
District	Region of the person tested malaria positive
Constituency	Constituency
ConstID	constituency ID
PosRDT	Total number of people tested malaria positive by RTD
p3	proportion of household made with sand, dung, wood, planks, palm, polished wood and others as floor material in the constituency
P4	proportion of household made with vinyl ceramic tiles, cement and carpet as floor material in the constituency
p5	proportion of household made with cane, truck reed, stone with mud as wall material in the constituency
p6	proportion of household made with cement stone with lime cement, wood, planks and bricks as wall material in the constituency
p7	proportion of female in the constituency
p8	proportion of male in the constituency
p9	proportion of people less than 5 years in the constituency
p10	proportion of people aged 5 to 19 years in the constituency
p11	proportion of people aged 20 to 39 years in the constituency
p12	proportion of people aged 40 to 59 years in the constituency
P13	proportion of people aged 60 and above years in the constituency
p15	proportion of children Learners and Students in the constituency
p20	proportion of people that are employed in the constituency
p22	proportion of people who does not own mosquito net not owning a mosquito bed net in the constituency
p32	proportion of houses sprayed in the past 12 months in the constituency
Population	Total number of people living in that constituency divided by 1000 (offset)

Note: For each constituency, all environmental variables represent the annual average

TABLE 4.6: Comparisons of the non-spatial implemented models (2018 – 2020 aggregated data)

	(Model 1)	(Model 2)
	Poisson model	NBM
Null deviance	155190	786.11
Residual deviance	53857	448.92
degrees of freedom	385	385
AIC	55350	3148.8

NBM: Negative Binomial model

Models fitted using aggregated data only looked at environmental factors, whereas models fitted with individual elements looked at several other probable malaria factors thought to have an effect on malaria (Tables 4.4 and 4.5). The resulting non-spatial model is illustrated in Table 4.6 and it found the non-spatial Negative Binomial model to be the best model since it had a lower AIC value (3148.8) than the non-spatial Poisson model (55350).

The residual deviance indicates how well the model predicts the response (total number of malaria cases) when the predictors are included. The lower the residual deviance (Model 2), the better the goodness of fit. Over-dispersion can also be determined by dividing the residual (scaled) deviance by the degrees of freedom; if the result is not close to one, the data is dispersed. This implies that the negative binomial distribution produces significantly better (more accurate) results, although only a few variables were found to be significant at $\alpha = 0.05$ in Model 2 versus Model 1 (Table 4.6).

TABLE 4.7: Poisson and Negative Binomial regression non-spatial model (aggregated data 2018 -2020)

Parameters	Poisson non - spatial model			Negative Binomial non - spatial model			Negative Binomial model(estimates)				
	β	SE	z value	Pr(> z)	β	SE	z value	Pr(> z)	β	P-value	95% CI
Intercept	-0.951	0.03	-24.09	<0.001	2.307	0.238	9.677	<0.001	10.044	0.000	(1.8397, 2.8034)
Spring	3.7260	0.04	95.25	<0.001	-3.359	0.261	-12.89	<0.001	0.0348	0.0000	(-3.8170, -2.8439)
Summer	3.1270	0.04	79.82	<0.001	-4.518	0.249	-1.82	0.06913	0.6365	0.0691	(-0.0945, 0.0419)
Winter	1.8070	0.04	46.29	<0.001	-1.873	0.249	-7.52	<0.001	0.1568	0.0000	(-2.3798, -1.3663)
Autumn (Ref)	1.000								1.000		
Ohangwena	-2.345	0.02	-125.3	<0.001	-1.853	0.219	-8.45	<0.001	0.3866	0.0000	(-2.3045, -1.4004)
Zambezi	-0.805	0.025	-47.67	<0.001	-0.951	0.236	-4.03	<0.001	0.9997	0.0001	(-1.4152, -0.4676)
Kavango's (Ref)	1.000								1.000		
HPD	-6.594e-04	0.00	-28.54	<0.001	-0.000	0.000	-0.77	0.43731	1.0000	0.4373	(-0.0011, 0.0006)
T (avg)	8.105e-07	0.00	14.44	<0.001	0.000	0.000	1.33	0.18268	0.9952	0.1827	(0.0000, 0.0000)
T (min.)	-0.005	-0.00	-36.05	<0.001	-0.005	0.003	-1.83	0.06669	1.0085	0.0667	(-0.0115, 0.0012)
T (max.)	0.009	0.00	58.23	<0.001	0.009	0.003	3.29	<0.001	1.0018	0.0010	(0.0033, 0.0140)
RF	-2.120e-04	0.00	-7.09	<0.001	0.002	0.01	1.78	0.07425	0.9994	0.0742	(-0.0005, 0.0050)
WS (avg)	-1.873e-04	0.00	-2.811	0.00494	-0.000	0.001	-0.50	0.6155	1.0011	0.6155	(-0.0021, 0.0021)
ST (avg)	1.369e-03	0.00	36.83	<0.001	0.001	0.000	1.17	0.24051	0.9997	0.2405	(-0.0004, 0.0028)
H	-1.066e-03	0.00	-14.03	<0.001	-0.0003	0.002	-0.18	0.85729	1.0130	0.8573	(-0.0038, 0.0310)
LW (avg)	1.640e-04	0.00	-11.04	<0.001	-0.0001	0.000	-0.41	0.68578	1.0001	0.8573	(-0.0220, 0.0007)

Sign if $Pr(> |z|) < 0.05$

Sign if P-value < 0.05

Note: each climatic variable in the table represent annual average of each constituency.

HPD(human population density, per km),T (avg)(average temperature, °C)

T (min.)(minimum temperature, °C),T (max.)(maximum temperature, °C), RF(rainfal,mm)

WS (avg)(average wind speed,m/s),ST (avg) (average soil temperature, °C)

H(humidity,p),LW (avg)(average leaf wethness,-)

NB: All climatic variables were aggregated as annual means and annual total malaria cases was considered as the depended variable.

As a result, the Negative Binomial model was chosen as the final best model for this study based on the accuracy of the results, an offset was added to the model (population/1000). Ignoring the structured and unstructured spatial effects, the results revealed that when all predators in the model were excluded, the estimated mean rate of malaria was 10 cases per 1000 population (Table 4.7). At the 95 percent confidence level, season, region, and maximum temperature were found to be significantly associated with the effect of malaria. While holding all other variables in the model constant, every single °C increase in maximum temperature in a specific constituency increases the estimated rate ratio by 0.0018 (0.18 percent), and there was enough evidence to conclude this at $\alpha = 0.05$ (Table 4.7). This means that constituencies with high maximum temperatures have a slightly higher incidence rate of malaria than others (Table 4.7)

Keeping other variables constant, for every unit increase in the total amount of rainfall received it will decrease the estimated incidence rate ratio of malaria by 0.0006 (0.6%) and one unit increase in humidity will increase the estimated rate by 0.13 (13%). This implies that a constituency with high rainfall and high humidity is likely to have a low malaria incidence rate although there was not enough evidence to conclude this ($\text{Exp}(\beta) = 0.9994$, p-value = 0.2405 and ($\text{Exp}(\beta) = 1.0130$, p-value = 0.8573 respectively) (Table 4.7).

Controlling for other variables, the estimated incidence rate per 1000 population in Ohangwena region is 61% less as compared to Kavango region. This implies that people living in Kavango region were 41% more likely at risk than the people living in Ohangwena region ($\text{Exp}(\beta) = 0.3866$, p-value < 0.001). Moreover, the incidence rate in Zambezi region were found to be 0.3% less as compared to Kavango region. This implies that the estimated incidence rate in the two regions (Zambezi and Ohangwena) was more less and there was enough evidence to conclude this ($\text{Exp}(\beta) = 0.9997$, p-value < 0.001) (Table 4.7). Furthermore, when all other variables were held constant, the incidence rate was lower in the spring and winter than in the autumn (($\text{Exp}(\beta) = 0.0348$, p-value = 0.0000 and ($\text{Exp}(\beta) = 0.1568$, p-value = 0.0000), respectively (see Table 4.7).

TABLE 4.8: Negative Binomial regression non-spatial model (Individual data 2017–2019)

Results of the individual dataset			
Parameters	$\hat{\beta}$	P-value	95% CI
Intercept	0.032	0.000	-3.577, -3.298
Gender			
Male	1.089	0.001	0.035, 0.136
Female (Ref)	1.000		
Age Group			
5 to 19	1.236	0.000	0.136, 0.289
20 to 39	1.112	0.018	0.018, 0.195
40 to 59	1.027	0.636	-0.083, 0.135
60 and above	0.881	0.056	-0.258, 0.002
0 < 5 years (REF)	1.000		
Place of Residence			
Village	1.030	0.307	-0.027, 0.088
Town (Ref)	1.000		
Type of Health Facility			
Health Centre	1.088	0.029	0.008, 0.160
Hospital	1.130	0.000	0.063, 0.182
Clinic (Ref)	1.000		
Occupation			
mosquito-infested employees	0.705	0.000	-0.476, -0.225
Professionals	0.689	0.000	-0.571, -0.182
Small Business	0.898	0.004	-0.181, -0.035
Unemployed	0.894	0.001	-0.181, -0.043
Youth (REF)	1.000		
Slept under mosquito bed net in the last 3 night			
yes	1.404	0.000	0.266, 0.414
no (Ref)	1.000		
Home sprayed in past 12 months			
Yes	0.990	0.718	-0.062, 0.043
no (Ref)	1.000		
District			
Eenhana	1.132	0.115	-0.033, 0.276
Engela	0.847	0.087	-0.361, 0.020
Katima – Mulilo	1.010	0.931	-0.215, 0.223
Nankudu	2.083	0.000	0.646, 0.821
Nyangana	2.055	0.000	0.629, 0.811
Okongo	1.106	0.356	-0.120, 0.309
Rundu	1.383	0.000	0.246, 0.403
Andara (Ref)	1.000		
Season			
Spring	0.808	0.046	-0.428, -0.010
Summer	0.998	0.946	-0.058, 0.054
Winter	0.829	0.000	-0.269, -0.107
Autumn	1.000		

From Table 4.8

mosquito-infested employees:	Farmers, Fishers, Cleaners, Security guards, Police officers, Truck drivers, Cattle header, and Road construction workers
Youth:	Children, leaners and students
Small Business:	Small market sales, traders, and other manual labourers

Using the non-spatial approach indicated earlier, we fitted a non spatial model using individual data from 2017 to 2019. As can be observed from Table 4.8, the variables gender, age group, place of residence, type of health facility, occupation, employment status, sleeping under mosquito bed nets, and district were found to be significantly associated with malaria incidence. More specifically, males were found to be 1.089 times more likely than females to have malaria after controlling other variables (95% CI, and p- value < 0.001 individuals aged 5 to 19 years were found to be 1.236 times more likely than individuals aged less than 5 years, and this was significant at 5% level of significance. The rate of testing positive for malaria in a villager was found to be 3% higher than in a town dweller, but there was insufficient evidence to conclude this at 5% level of significance (Table 4.8).

4.3 Modeling spatial patterns of malaria

Using the spatial approach indicated earlier to identify spatial clusters, several measures of spatial correlation were performed, including global and local measures of spatial autocorrection as well as Getis-Ord approach measures of spatial autocorrection. A neighbourhood structure at the constituency level was created in R, along with an adjacency matrix and a weight matrix through queens contiguity but one could also consider comparing different neighbourhood matrix structure e.g., queen with rook neighbourhood structure. One constituency was discovered to have one neighbour (least connected constituency) and 5 constituencies were discovered to have 5 neighbourhoods (most connected constituency), 10 constituencies with 3

neighbourhoods, another 10 with 4 neighbourhoods, and 8 with 2 neighbourhoods (Figure 4.6).

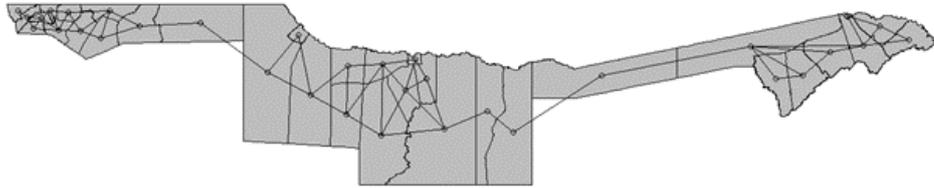


FIGURE 4.6: Neighbourhoods at constituency level

According to the literature, the Global Moran's is the most powerful and widely used spatial autocorrelation tool for determining the overall strength of spatial dependence in data (Rezaeian et al., 2007; Lai, So, and Chan, 2008; Holowaty et al., 2010; Gruebner et al., 2011). At $\alpha = 0.05$, the Global Moran's I test was used to test the null hypothesis H_0 of no significance clustering of malaria incidence within constituencies. Moran's I statistics value for aggregated data 2018 – 2020 was 0.1863 (p – value = 0.0429) with a variance of 0.0159, whereas Moran's I for 2017 – 2019 individual dataset was 0.3003 (p – value = 0.005) with a variance of 0.0159.

The mean value of Moran's I was positive and the p value for the two Morans' I was statistically significant at $\alpha = 0.05$. This indicates that there was spatial autocorrelation in the data at the constituency level, indicating that malaria is spatially clustered in North Namibia. Values in neighbouring constituencies tend to cluster, with high values clustering next to other high values and low values clustering next to other low values. As a result, the findings encourage one to continue using the local Moran and Local G Statistics to determine where high/low clustering occurs (hot spot and cold spot of malaria in the 4 regions).

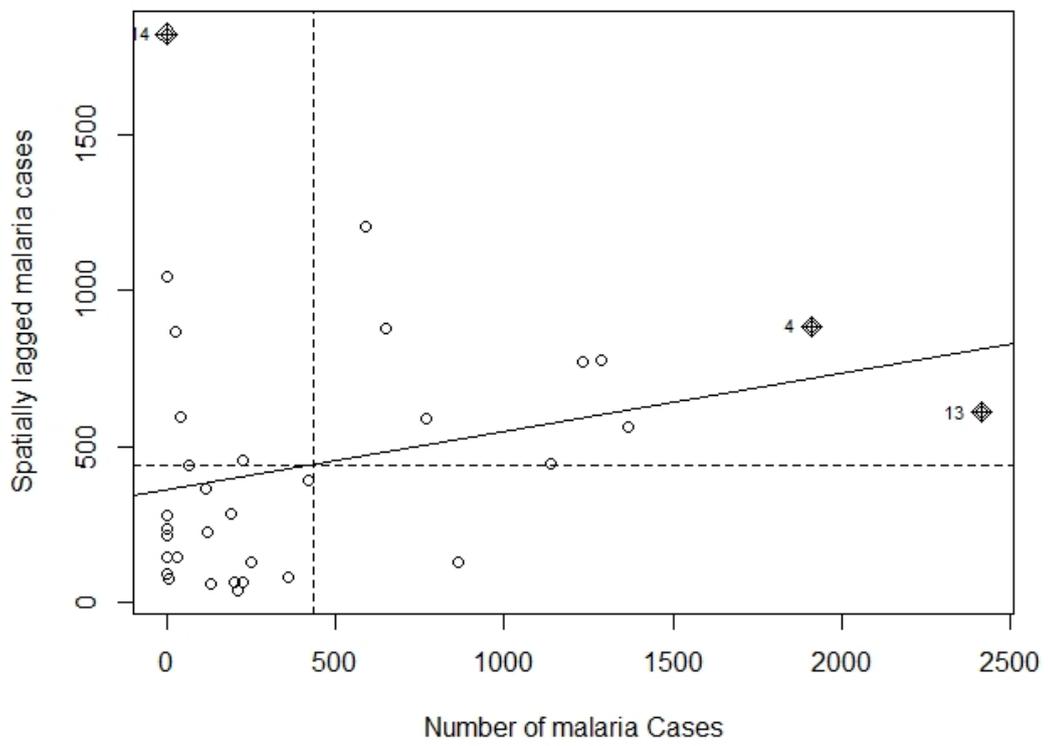


FIGURE 4.7: Moran's I scatter plot

Figure 4.6 shows a standardised malaria incidence scatter plot of a constituency in the x-axis versus the spatial lag of the constituency's standardised malaria incidence rate, which disaggregates the spatial autocorrelation into four types of association. The first association is known as (HH): these are points in Quadrant I that show a location (constituency) with a high malaria incidence surrounded by another location (constituency) with a high malaria incidence. The second association is (LL): these are points in Quadrant II that show a low malaria incidence location (constituency) surrounded by a low malaria incidence location (constituency). The third association is (HL): these are points in Quadrant III that show a location (constituency) with a high malaria incidence surrounded by a location with a low malaria incidence. The final association is (LH): these are Quadrant IV points that show a location (constituency) with low malaria incidence surrounded by a location (constituency) with high malaria incidence.

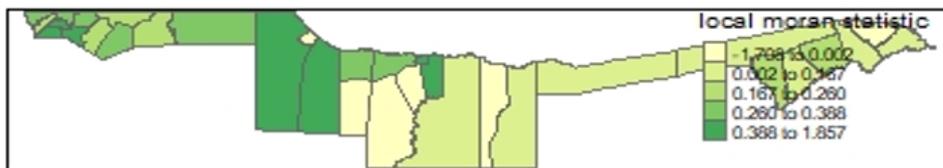


FIGURE 4.8: Local Moran's I statistics map (2018 -2020).

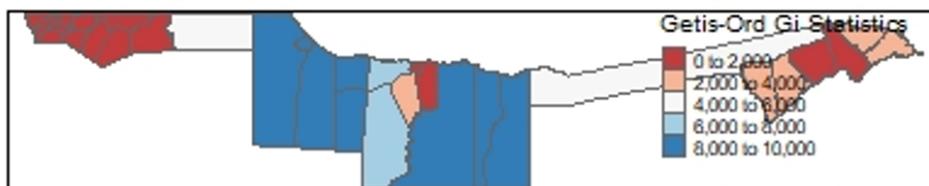


FIGURE 4.9: Getis - Ord Gi statistic cluster map I significant cluster map (2018 - 2020)

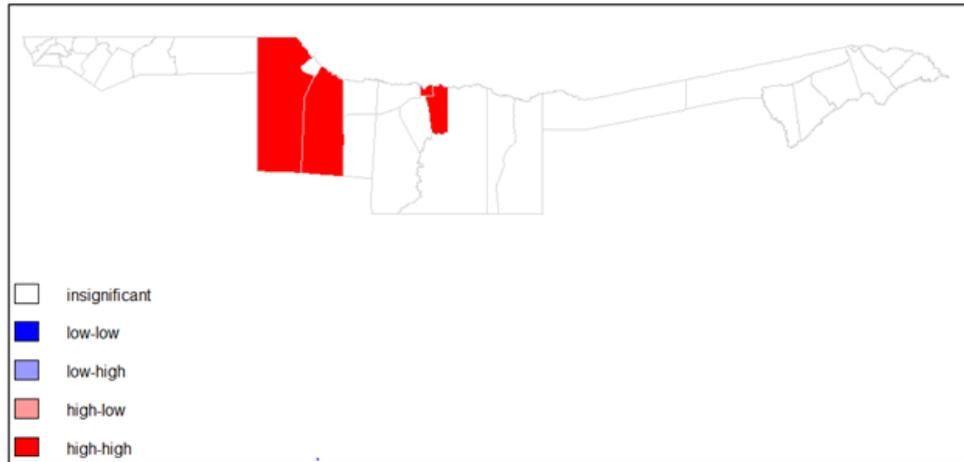


FIGURE 4.10: Local Moran's I significant cluster map (2018 - 2020)

A Moran's statistics map and Gi statistics revealed a spatial clustering across constituencies. Because most of the constituencies had a positive value for local Moran's I, it follows that most of the constituencies were surrounded by constituencies with high values, and constituencies in Kavango West and Ohangwena revealed a high spatial clustering (highlighted in dark green) when compared to others (Figure 4.8). Local Gi statistics (Figure 4.9), represented as a Z-score, also revealed a high clustering of malaria cases across all constituencies, with most cases concentrated in Kavango West and East.

All of the constituencies highlighted in dark green and blue (Figures 4.8 and 4.9) were malaria hotspot areas, with Mpungu, Tondoro, Rundu urban, and Rundu rural in Kavango West being the significant malaria hotspots as highlighted in red (Figure 4.10), while the rest of the constituencies revealed a non-significant clustering as highlighted in white (Figure 4.10).

4.3.1 Results of the spatial model

A sensitivity analysis is one of the most critical components for completely understanding Bayesian results in an applied research situation. Priors can have a significant impact on the posterior distribution, according to the simulation study. It is impossible to separate the impact of the prior from the role of the data in the model estimate phase without a sensitivity analysis. A sensitivity analysis can aid

the researcher in determining the impact of the prior vs the impact of the data. To put it another way, this approach can assist determine how much theory [via informed theory or lack of theory (e.g., diffuse priors)] influences the final model findings and how much the results are driven by patterns in the sample data (Depaoli, 2020). In this study, sensitivity analysis was not performed as Priors for the spatial random effects were set to follow log gamma distribution with mean = 0, precision = 0.001 (since it was a negative binomial model), while the default prior assigned to the associated coefficients (and the intercept) was a Gaussian distribution and this is assumed to have a smaller variance as per literature.

The findings of a spatial model using malaria aggregated data (2018-2020) are provided in the tables below. According to Table 4.9, adding covariates to the Negative Binomial geographic model increased its fit to the smaller DIC value (493.17), making it fit the data better than the Negative Binomial spatial model without covariates (516.85).

TABLE 4.9: Comparisons of the spatial implemented models (2018 – 2020 aggregated data)

	(Model 3)	(Model 4)
	NBSM (without covariates)	NBSM (with covariates)
DIC (NoEP)	516.85 (1.77)	493.17 (10.51)
(1/overdispersion)	0.42	0.782

NBSM: Negative Binomial spatial model

TABLE 4.10: Pearson's correlation between sum of positive tested malaria cases with human population density as well as climatic variables included in the spatial model (aggregated data 2018 – 2020)

	PosRDT	HPD	T (avg)	T (min.)	T (max.)	RF	WS (avg)	ST (avg)	H
HPD	-.048	1							
T (avg)	.791**	.152	1						
T (min.)	.457**	.279	.693**	1					
T (max.)	.342*	.260	.591**	.924**	1				
RF	.751**	.148	.968**	.764**	.712**	1			
WS (avg)	.518**	-.05	.621**	.683**	.498**	.594**	1		
ST (avg)	.443**	.177	.657**	.501**	.597**	.710**	.279	1	
H	.637**	.257	.901**	.734**	.743**	.925**	.502**	.751**	1
LW (avg)	.343*	.118	.646**	.525**	.663**	.745**	.188	.875**	.761**

** . Correlation is significant at the 0.01 level (2-tailed)

* . Correlation is significant at the 0.05 level (2-tailed)

TABLE 4.11: log scale Parameter for the Negative Binomial BYM model (aggregated data 2018 - 2020)

Parameter	M3 (Spatial model without covariates)					M4 (Spatial model with covariates)				
	P	SD	Quantiles			P	SD	Quantiles		
	mean		25	50	75	mean		25	50	75
Intercept	1.376	0.268	0.884	1.363	1.99	-53.3	11.8	-78.47	-52.61	-31.98
HPD	-	-	-	-	-	0.000	0.001	-0.002	0.000	0.003
ST (avg)	-	-	-	-	-	0.011	0.014	-0.017	0.012	0.038
T (avg)	-	-	-	-	-	0.982	0.491	0.040	0.972	1.978
T (min.)	-	-	-	-	-	0.053	0.187	-0.299	0.047	0.438
T (max.)	-	-	-	-	-	0.840	0.198	0.524	0.812	1.296
RF	-	-	-	-	-	0.058	0.026	0.011	0.057	0.113
WS (avg)	-	-	-	-	-	1.068	0.253	0.658	1.037	1.645
H	-	-	-	-	-	-0.089	0.036	-0.162	-0.088	-0.020
LW (avg)	-	-	-	-	-	-0.015	0.006	-0.028	-0.015	-0.004
PoNB	0.42	0.08	0.30	0.41	0.58	0.782	0.166	0.498	0.768	1.15
u_i	1190	1204	111	838	4366	828.0	753.8	61.3	614.6	2857.2
v_i	1219	1232	112	858	4457	931.5	859.6	76.0	688.2	3182.4

P mean (Posterior mean)

SD (standard deviation)

PoNB (Parameter of Negative binomial)

u_i (structured random effect)

v_i (unstructured random effect)

Using the spatial approach indicated earlier, all of the fixed effects variables in the spatial Negative Binomial BYM model revealed a relationship with the malaria constituency annual mean (Table 4.10). Annual monthly average temperature (mean), annual monthly maximum temperature (mean), annual monthly total rainfall (mean), and annual monthly average wind speed (mean) all had a significant positive effect on annual mean malaria incidence, whereas annual monthly average humidity (mean) and annual average leaf wetness (mean) had a significant negative effect (Table 4.11). However, annual monthly soil temperature (mean) and annual monthly minimum temperature (mean) were both found to be positively related to malaria annual mean incidence rate, but this was not significant, and human population density was found to have no effect on malaria incidence rate in Namibia (Table 4.11).

If the positive and negative posterior mean effects were exponentiated, they would be interpreted as decreases and increases in relative risks, respectively. For example, $e^{0.058} = 1.06(0.06)$ and this means that for every one mm increase in annual monthly

total rainfall (mean) it will increase the mean incidence rate of malaria by 6%. Also, for every one unit increase in annual monthly average temperature (mean), annual monthly maximum temperature (mean), and annual monthly average wind speed (mean), it will increase the log mean of malaria by 0.011, 1.015, 0.763, and 0.988, respectively (Table 4.11).

Malaria incidence mean was estimated to decrease by 8.52%, $e^{-0.089} = 0.9148$ (0.0852) for every one % increase in annual monthly humidity (mean). Also, annual monthly leaf wetness was found to decrease the mean incidence of malaria by 1.39%, $e^{-0.014} = 0.986$ (0.0139) (Table 4.11).

The proportion of males who tested positive for malaria, the proportion of individuals aged less than 5 years, 5 to 19 years, 20 to 39, 40 to 59, and 60 and above, the proportion of individuals who live in villages, the proportion of individuals who were employed, and many other variables (Table 4.5) were examined using the malaria individual dataset (2017-2019) to run a spatial model. The results obtained are as follows:

TABLE 4.12: Comparison of the two models (Individual data 2017-2019)

	Model 5	Model 6
	NBSM (without covariates)	NBSM (with covariates)
WAIC ((NoEP)	468.72	465.05
DIC ((NoEP)	468.7 (1.57)	463.06 (5.18)
(1/overdispersion)	0.275	0.38

NBM: Negative Binomial modely, NBSM: Negative Binomial spatial model
NoEP: Number of Effective Parameters

Using the malaria individual data, it was still discovered that adding some of the other possible factors such as age group proportion, gender, and so on, as well as climatic factors, improved model estimates much better than just considering random effects, as the DIC value of the model with both spatial effects and some other possible factors revealing a lower DIC: 463.06 with a greater number of effective parameters (5.18), as compared to a model with only spatial effects (Table 4.12).

A spatial model was fitted using malaria individual data. All variables included such as human population density, proportion of people aged 5 to 19 years in the constituency, proportion of people employed in the constituency, proportion of houses sprayed in the past 12 months in the constituency, proportion of people whose households are made with cement, stone with lime cement, wood, planks, and bricks as wall material in the constituency. Both structured and unstructured random effects revealed a significant positive effect on malaria cases' posterior means of 1295 and 1252, respectively (Table 4.13).

4.4 Modelling spatial and time patterns of malaria

Creating a smoothed space-time malaria incidence graph, maps are critical for decision making when it comes to the malaria endemic in Namibia. This study examined malaria cases data from a spatiotemporal perspective using Bayesian models with temporal random effects and space-time interaction terms to identify significant predictors associated with malaria incidence risk and to generate contemporary smoothed maps of disease risk in Namibia's northern risk constituencies.

TABLE 4.13: log scale Parameter for the Bayesian Negative Binomial
BYM model (Individual data 2017 - 2019)

Results of the individual dataset (2017 -2019)										
	M5 (Spatial model without covariates)					M6 (Spatial model with covariates)				
Parameter	P mean	SD	Quantiles			P mean	SD	Quantiles		
			25	50	75			25	50	75
Intercept	1.33	0.30	0.70	1.28	2.00	-0.02	1.24	-2.45	-0.02	2.41
P3	-	-	-	-	-	-0.005	0.01	-0.03	-0.01	0.02
p5	-	-	-	-	-	0.00	0.01	-0.02	0.00	0.03
p8	-	-	-	-	-	-0.05	0.05	-0.14	-0.05	0.04
p9	-	-	-	-	-	-4.02	4.56	-12.29	-4.26	5.61
p10	-	-	-	-	-	1.55	2.16	-2.65	1.57	5.70
p11	-	-	-	-	-	-5.38	3.74	-12.43	-4.29	2.24
p12	-	-	-	-	-	2.71	9.66	-15.11	2.48	22.87
p13	-	-	-	-	-	12.97	15.14	-16.71	12.95	42.74
p20	-	-	-	-	-	4.55	3.61	-2.77	4.64	11.41
p32	-	-	-	-	-	1.50	1.50	-1.40	1.48	4.50
PoNB	0.28	0.05	0.19	0.27	0.38	0.32	0.06	0.22	0.64	0.99
u_i	502	451	24	370	3640	1252	1414	115	829	4961
v_i	1231	1007	182	964	3255	1295	1445	116	863	5076

$p3$ = household made with sand or dung as floor material,

$p4$ = household made stone with mud as wall material,

$p8$ = proportion of male, $p9$ = proportion of people less than 5 years, $p10$ = of people aged 5 to 19 years,

$p11$ = people aged 20 to 39 years, $p12$ = people aged 40 to 59 years, $p13$ = people aged 60 and above ,

$p13$ = people aged 60 and above years, $p20$ = people that are employed, and

$p32$ = houses sprayed in the past 12 months

NB: all variables in this table were proportion

4.4.1 Results of the spatio-temporal model

Variables that were found to have a significant relationship with malaria in the previous spatial analysis (Table 4.11) were added to the spatio-temporal model Eq. (3.14 and 3.15) extension of the spatial Besag-York-Mollie (BYM) model Eq. (3.9) with Negative binomial family since the data was dispersed to disclose both spatial and temporal trend pattern/trend of malaria. Added variables are monthly average and maximum mean temperature, mean of monthly total rainfall received, mean of monthly average wind speed, humidity mean, and monthly average leaf wetness.

The model in Eq.(3.9) was a Conditional Autoregressive (CAR) convolution model with two random impacts, one spatially organised area-specific random impact and one unstructured area-specific random impact, and the term reflected the temporally structured effect γ_t in Eq. (3.15), was dynamically simulated using a random walk of order 2 to account for extra heterogeneity in the counts due to unseen (spatially unstructured effects) risk factors. The outcomes obtained are as follows:

***i.* Results of the parametric spatio-temporal model**

Results obtained from the spatio-temporal model are displayed as follows:

TABLE 4.14: Comparison of the spatio-temporal models (aggregated data 2018-2020)

	Model 7	Model 8	Model 9	Model 10
	Parametric trend model	Unstructured interactive model	Temporal structured interactive model	Spatial structured interactive model
WAIC (NoEP)	1218.73	1101.78	1107.32	1109.65
DIC (NoEPNoEP)	1219.83	1099.71(33.27)	1103.55	1107.77
			(8.65)	(65.86)
(1/overdispersion)	0.398	1.51	1.95	1.427

NoEP: Number of effective parameters

TABLE 4.15: Log scale parameter for the Bayesian Negative Binomial (BYM) unstructured interactive spatio-temporal model with added significant climatic variables from spatial model using 2018-2020 aggregated dataset

Model 8					
Parameters	P mean	SD	Quantiles		
			25	50	75
Intercept	0.436	0.661	-0.826	0.423	1.77
T (avg)	0.000	0.000	0.000	0.000	0.000
T (max.)	0.006	0.006	0.005	0.005	0.017
RF	0.006	0.003	0.000	0.005	0.011
WS (avg)	-0.003	0.003	-0.009	-0.003	0.004
H	0.002	0.004	-0.006	0.002	0.009
LW (avg)	-0.001	0.001	-0.002	-0.001	0.001

T (avg), T (max), RF, WS (avg), H, and LW (avg) were the same variable as in Table 4.11

The spatio-temporal model with unstructured interaction was considered as the best model since it had the smallest DIC value (Table 14). Results show that annual average rainfall was found to have a significant effect on malaria through space and time. For every one mm increase in annual rainfall in a certain constituency it will increase the annual mean cases of malaria by 0.6%. Annual average maximum temperature was also found to be significantly associated with malaria. For every one °C increase in annual maximum temperature in a certain constituency it will increase the annual mean cases of malaria by 0.6%. The rest of the variables revealed a non-significant relationship with malaria from spatial and temporal perspectives at 95% confidence intervals.

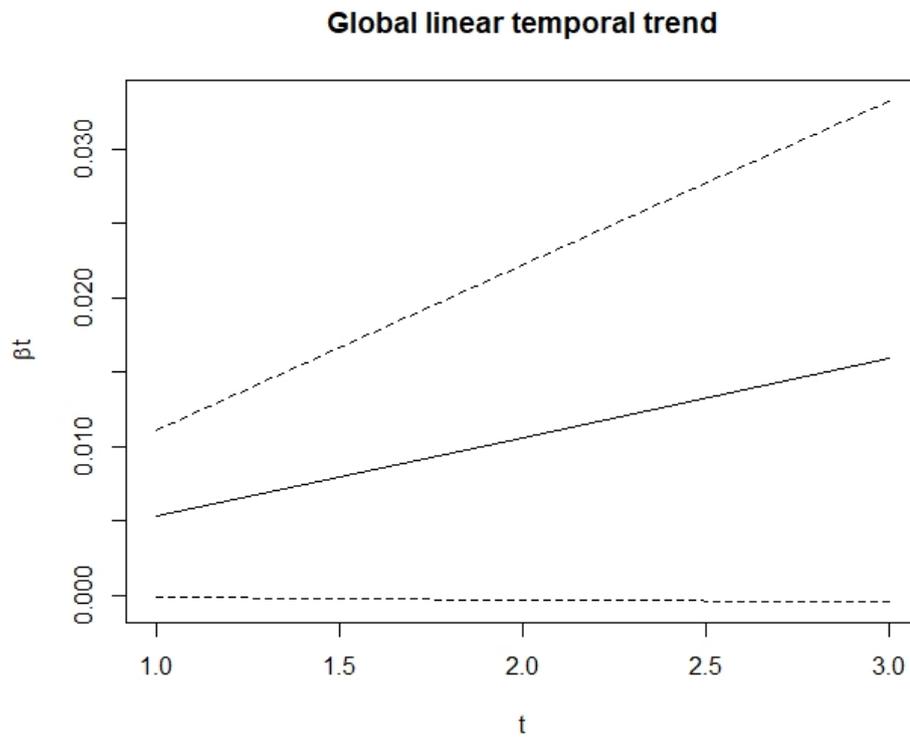


FIGURE 4.11: Global linear temporal trend for malaria (spatio temporal model with added covariates)

The solid line in the middle represents the posterior mean for the main linear trend β_t while the dotted line represents the 95 percent credibility interval, and the model discovered a global time effect. The plot of the posterior mean of the main time effect (years) clearly shows a slight increase in global trend as time passes, for example, a high estimate of malaria was observed from 2018 to 2020 (Figure 4.11).

Moreover, $\zeta = u_i + v_i$ the posterior mean of the spatial (structured and unstructured) effects was obtained. Figure 4.12 shows the obtained results.

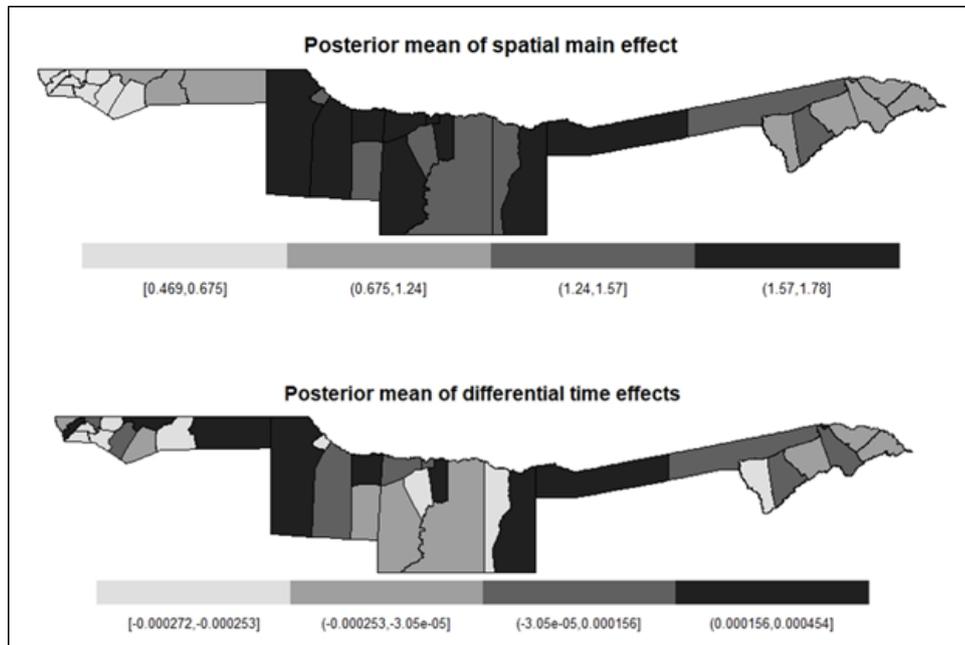


FIGURE 4.12: Spatial main effect and differential temporal maps of the spatiotemporal model 3 (posterior mean obtained using a random walk of order 2)

The results of the model with added annual average rainfall received monthly as a fixed factor revealed a greater spatial effect in constituencies located on the outskirts of Kavango East and Kavango West. Regrettably, the two regions on the east and west sides of the Kavango, two in the middle of the Kavango, and two in the Ohangwena region showed a higher differential trend than the average (Figure 4.12).

ii. Results of the spatio–temporal interaction model

INLA allows for the definition of linear combinations on various latent effects, as well as the estimation of their posterior marginals. It is worth noting that these linear combinations have no effect on the model’s fit (as the use of a predictor matrix did). The option `lincomb` is supplied to the `inla()` function to represent linear combinations. Controlling the argument is also important. `Lincomb` can be used to specify the parameters that govern how the posterior marginals of linear combinations are calculated.

The two effects space and time was then combined through linear combination using `inla.make.lincombs` before fitting the final model of the study “spatio–temporal interaction model (Eq. 3.17)”. For each year, a linear combination: $I \times \gamma_t + I \times \phi_t$ was obtained by combining the r^{th} element of the two-diagonal matrix one for structured and one for unstructured temporal parameters (Elliott, Wakefield, Best, & Briggs, 2001). We then added `lincomb = lcs` to the model Eq. (3.17) to allow for interaction between space and time effects, which explained the differences in the time trend of malaria cases across different constituencies.

The study also looked at three interactions. The first interaction involved unstructured spatial effects interacting with unstructured temporal effects, the second was the unstructured spatial effects interacting with structured temporal effects and the last interaction was between unstructured temporal effects interacting with structured special effects. Figure 4.13 shoows the results of Eq. (3.17).

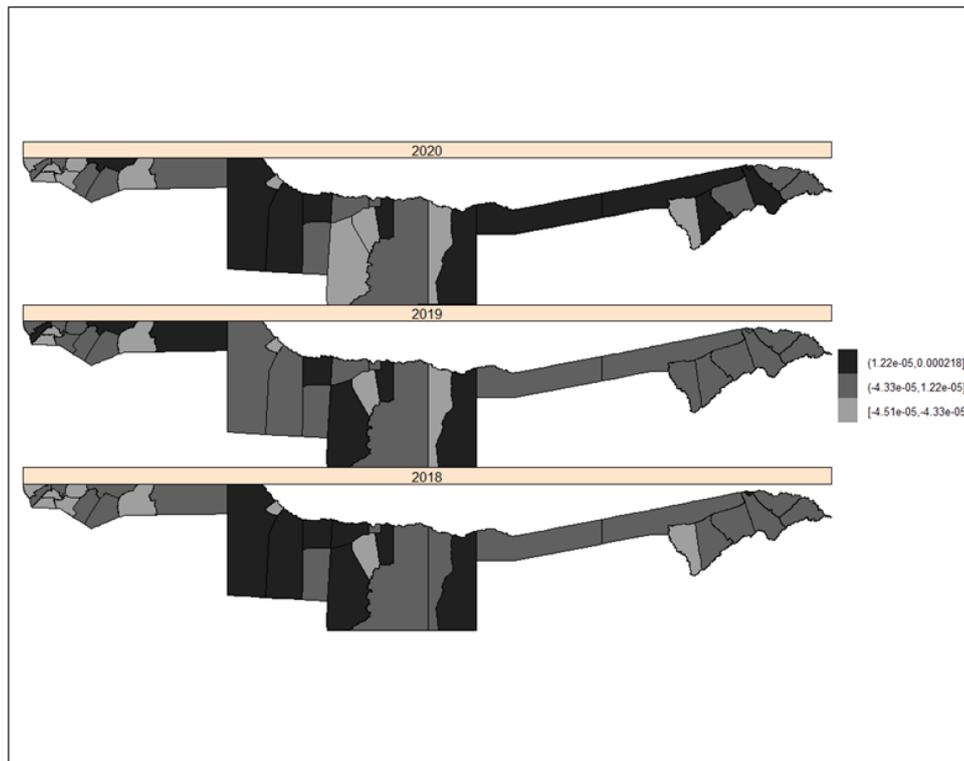


FIGURE 4.13: Posterior mean for malaria incidence (Nonspatially or Temporally interaction) of the non-parametric spatio-temporal model using RW2 with added covariate effect

In 2018 and 2020, a pattern of positive significant unstructured random effects (spatial and temporal) was observed, primarily in the west of Kavango and constituencies bordering Kavango East with Zambezi region. Furthermore, an increase in the number of constituencies having turned black (positively significant) was observed between 2018 and 2020. However, in some of the Kavango and Ohangwena constituencies, a negative significant unstructured random effect was detected (negative posterior mean in light grey) (Figure 4.13).

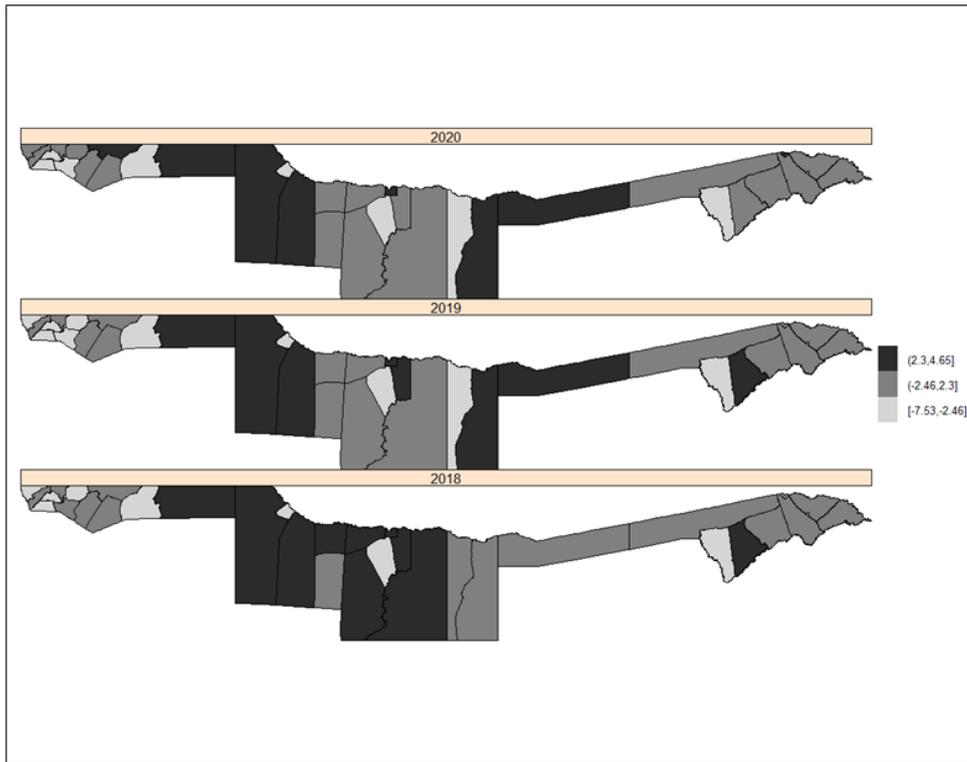


FIGURE 4.14: Posterior mean for malaria incidence (Temporally Structured interaction) of the non-parametric spatio-temporal model using RW2 with added covariate effect

In addition, a pattern of temporal structured random effects was observed in both years. Even though most malaria hotspots were detected in the East and Middle of Kavango constituencies in 2018, the East and Middle of Kavango constituencies still had the highest posterior estimates (malaria hotspots) in 2019 and 2020. Furthermore, there were no changes in the constituencies at risk of malaria in 2019 and 2020 (Figure 4.14).

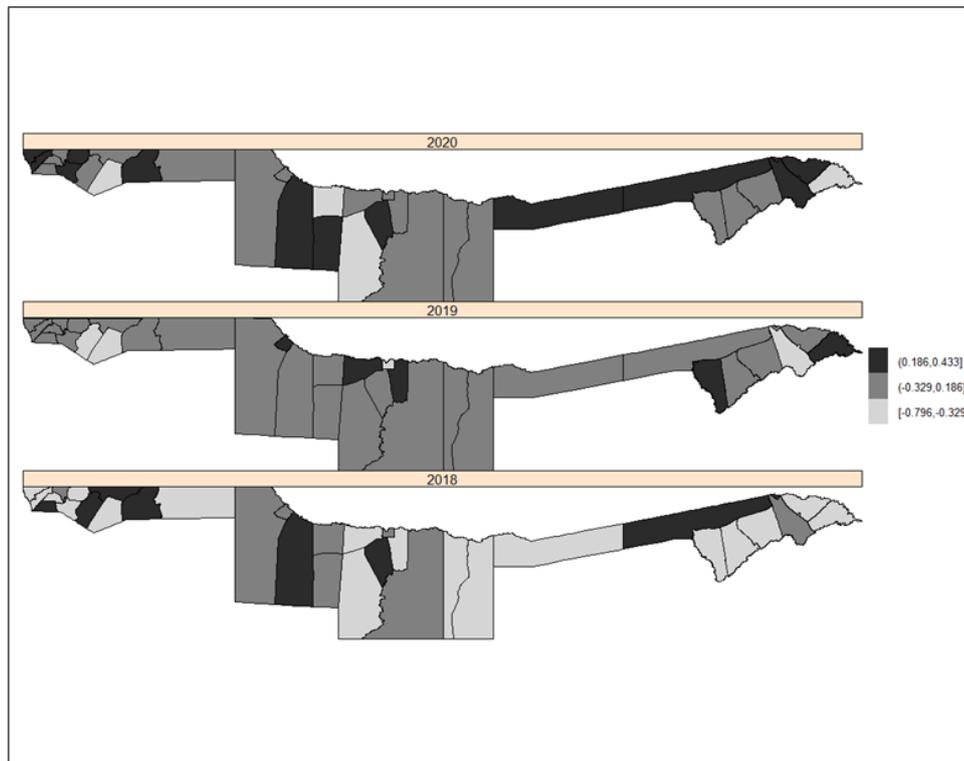


FIGURE 4.15: Posterior mean for malaria incidence (Spatially Structured interaction) of the non-parametric spatio-temporal model using RW2 with added covariate effect

Figure 4.15 revealed that both negative and positive spatial structured variations were observed throughout the four regions where constituencies with negative significant spatial trends to cluster. However, a positive significant spatial structured random effect (highlighted in black) in both regions was visible in 2020 .

CHAPTER 5

DISCUSSIONS, CONCLUSIONS AND RECOMMENDATIONS

5.1 Discussions

The four regions (Kavango East and West, Ohangwena, and Zambezi) reported a total of 44644 malaria cases over a three-year period, with 31 619 instances reported in 2018, 2 990 in 2019, and 10 035 in 2020, and this accounted for an average of approximately 90% of all reported malaria cases in Namibia throughout the last three years (2018-2020). The malaria epidemic has been a major challenge in Kavango West and East, Zambezi and Ohangwena regions for many years now compared to other regions (Table 3.1 and Figure 4.1). Namibia was indeed expected to have 0 cases per 1000 population in 2020. However, ignoring a few percentages of cases recorded in other regions, a 92% incidence rate decrease was achieved in 2019 as the incidence decreased from 12 cases per 1000 population in 2018 to 1 case in 2019, but unfortunately, the incidence rate increased again from 1 case in 2019 to 4 cases per 1000 population in 2020 (Table 4.3) with high malaria transmission still in Kavango West and East, Ohangwena and Zambezi constituencies.

Due to complicated data, where malaria cases were recorded as weekly clinic cases, to avoid many errors during data cleaning since the dataset needed to be re-aggregated per constituencies for easy mapping due to unavailability of some clinic shapefiles, the study was limited to constituencies located in the four high reported malaria transmission for the past 3 years which is Kavango West and East, Ohangwena and Zambezi region. Constituencies in the two Kavangos (East and

West) produced 77% of the 44644 malaria cases reported in the four regions in 2018-2020, followed by constituencies in the Zambezi region with 13% and constituencies in Ohangwena region with 10%. Most of the cases were reported during Autumn and Summer as represented by 61% and 26% respectively (December to May) and this is the period the country reported high average rainfall mostly in constituencies that reported a high number of malaria cases (Table 3.1).

After computing the Standardized Incidence Ratio (SIR), which is a simple measure of disease risk in a specific population each year, Mpungu constituency recorded the highest malaria incidence rate with an annual average of 137 cases per 1000 population, followed by Tondoro, Rundu urban, and Rundu rural constituencies (descriptive statistics). This means that these are the constituencies that observed an average of more malaria cases than expected as per the descriptive statistics while the majority of the Ohangwena constituencies revealed less malaria risk (Figure 4.5).

When comparing the two malaria risk maps (malaria risk computed using the SIR method and presented in Figures 4.5) and malaria risk predicted from spatio-temporal models (Figure 4.13 & 4.15), they both observed an increase in the number of constituencies having turned black (high malaria risk) in the east of Kavango in 2020 compared to 2018 and 2019.

In addition, the spatio-temporal unstructured interactive model predicted high malaria risk in some of the Zambezi constituencies eg., Kongola, Katima Mulio urban, Katima Mulio rural, and Sibbinda in 2020 that could not be detected by the SIR methods because of high rainfall and temperature together with some unstructured spatial and random interactions in those constituencies.

The spatio-temporal spatial structured interactive model also predicted high malaria risk in some of the Kavango and Zambezi constituencies including Ongenga, Engela, Ondobe, Omulonga, and Epembe constituency in Ohangwena region in 2020 (Figures 4.13 and 4.15), although the unstructured interactive risk map (Figure 4.13) explains the malaria risk better than the temporal structured interactive risk map

(Figure 4.15).

This suggests that there is a greater need to be done in recognised malaria risk constituencies (hotspots) from both analyses SIR techniques and risk maps anticipated by the model (Figure 4.2, 4.13,4.14, and 4.15). For example, offering full-service malaria control and implementing programmes in schools and communities to teach people about battling this endemic disease (malaria) as a way to achieve the malaria-targeted plan to eliminate malaria in Namibia by 2030.

Figures C.1, C.2, and C.3 (Appendix C) depict the results of the spatio-temporal models without the addition of covariates. Ignoring rainfall as malaria fixed effects will have an effect on mapping malaria risk through space and time. Comparing the two figures (Figures 4.12 and C.1), Figure C.1 could not detect all the malaria hotspot as compared to Figure 4.12 that considered the amount of rainfall received in a specific constituency. This evidenced that rainfall has an effect on the distribution of malaria through space and time. The map of non-spatial or temporally interaction map before adding fixed effects (climatic factors) did not uncover strong spatial effects as compared to the non-spatial or temporally interaction map with added fixed effects (Figures C.2 and 4.13). Moreover, there were not many changes in posterior mean for malaria incidence temporally structured interaction map before and after adding fixed effects (see Figures C.3 and 4.14). In addition, adding fixed effect on the spatially structured interaction model has aided in detecting spatial effects in the North constituencies of Zambezi region in 2020 unlike the spatially structured interaction model without added fixed effects (Figures C.4 and 4.15).

Similar studies on malaria spatial modelling conducted earlier, not only in African but also in Asian countries (e.g. Sipe, 2003; Hay et al., 2006; Kazembe, 2007; Victor, 2009; Tuyishimire, 2016; Yunxia ,2015; Joao, 2018) have found the distribution of malaria cases to have a seasonality characteristic setting. The non-spatial model (Table 4.8) also presented a significant malaria seasonality with high mean malaria incidence being reported during Autumn (March to May) as compared to other seasons. The mean malaria incidence rate was found to decrease by 97% and 84%

during Spring (September to November) and Summer (December to February) respectively as compared to Autumn (March to May).

The study found the spatial and temporal variation of malaria risk to be due to a combination of climatic factors both observed and unobserved, where average annual total rainfall and annual average maximum temperature were found to explain the spatial and temporal variation of malaria infection in Namibia from both temporal and spatial perspectives and this was similar to the results obtained by Kazembe (2007) in Malawi.

The population living in the far East and West of Kavango and Zambezi region constituencies was predicted to be more at malaria risk as compared to others (Figures 4.13, and 4.15), primarily, people aged 5 to 19 , 20 to 39 years and unemployed people that are living in villages (Table 4.9). This confirmed that the occurrence of malaria cases in constituencies might be high and closely related to the two ecological factors, namely, maximum temperature and amount of rainfall received (Table 4.15) as those are the constituencies that receive high annual rainfall and an annual record high temperature (Figure 4.1).

Most countries have eradicated malaria by mapping malaria anopheles mosquito breeding sites using other spatial approaches such as geographical information system (GIS) and spatial statistics, which allow for the assessment of the degree of infection clustering to explore the spatial clustering of dengue (Ali et al., 2003; Zhang et al., 2008; Vazquez-Prokopec et al., 2010). Tuyishimire (2016), for example, used extensive malaria occurrence data to model malaria risk factors in Rwanda and was able to locate malaria breeding locations in rural areas in Rwanda's south-eastern region.

People who tested positive for malaria were followed up on within one week after being diagnosed with malaria at their homes so that the geo-coordinates of breeding site locations could be documented. However, the geo-coordinates of breeding were not available on the DHIS2 system during the period of data collection. As a result,

the study was unable to further detect and map malaria anopheles mosquito breeding sites.

5.2 Conclusions

The goal of this study was to use an ecological spatial-temporal regression model that is known to be a completely Bayesian model by Neema and Bohning (2012) based on the Besag, York and Mollie framework (Besag, York & Mollie; 1991) to characterise geographical variation in malaria risk and evaluate possible connections between disease risk, and environmental factors at the constituency level in Namibia's malaria-endemic areas.

Through the whole posterior inference approach, a detailed examination of the uncertainty in the unobserved random factors that also contribute to the volatility of malaria mean rate was done. The procedure was accomplished by adding the observed and unobservable random effects into the whole hierarchical Bayesian model.

The random effects that were evaluated included structured space-time heterogeneity, which measured the effect of constituencies clustering, unstructured heterogeneity, time trend effect, which represented the three respective periods 2018 - 2020, interaction between space and time, and the covariates effect of climatic variable, as well as some other possible variables that were found to have a significant effect on malaria in other countries according to literature. R-INLA was used to perform the analysis, which included prior and hyper-prior distribution specifications for the parameter and hyper parameters.

A variety of datasets were merged to achieve the study's goals, such as Malaria tested positive by rapid diagnosis datasets (RDT) surveillance agreed datasets that contain daily records of people who tested positive for malaria, individual malaria datasets that contain more information about people who tested positive for malaria, climatic weather data that contains environmental variables thought to be related to malaria from literature, and the Namibia shapefiles dataset provided from the Namibia Statistics Agency (NSA) for mapping purposes.

The study discovered that the spatial temporal model with both random and fixed effects best fit the model, which demonstrated a strong spatial and temporal heterogeneity distribution of malaria cases (spatial pattern) with high risk in most of the Kavango West and East outskirt constituencies where high malaria peak was discovered to occur during Autumn and Summer (January to May) and annual average rainfall, annual average maximum temperature together with some unobserved random effects were found to be significantly associated with malaria cases distribution through space and time.

Furthermore, the findings of the best BYM model's posterior mean estimations of the parameters revealed that unstructured random effects contributed to most of the malaria variations in Namibia which makes little difference in the identified malaria hotspots of the spatio-temporal models with and without covariates.

5.3 Recommendations

The use of a Bayesian approach to estimate the contributions of climatic/environmental indicators on the spatial-temporal pattern of pandemic diseases should be encouraged in statistics for sensitivity data errors and as a way to guide appropriate actions and better allocation of limited health care resources. This work is therefore dedicated to people working in the government and non-governmental organisations for policy development, specifically the MoHSS to help them to achieve malaria target 3.3 that aims eliminating malaria before 2030 by utilising the findings of this study. For example, giving priority to the individual who lives in constituencies that require closer interventions (identified malaria hotspot areas) at the appropriate time (January – May "rainy season") more specifically individuals aged 5 to 19, the youth e.g., children, learners, and students, and individuals that are not employed or working in Small Business (e.g., small market sales, traders, and other manual labourers, specifically males as these are the people that were found to be more exposed to the malaria risk.

Future studies should consider examining all possible putative sources of malaria transmission including travel histories and networks, and treatment seeking behavior and should mostly focus on finding and mapping potential anopheles mosquito habitat that was missed in this study due to a lack of information in the datasets on anopheles mosquito breeding locations (e.g., irrigated agriculture). In addition, malaria data to be used for future study should be collected via PCR test for high accuracy of the confirmed cases and some adjustments to standardize the covariates must be considered since all this was missed in this study. Furthermore, malaria death records should be captured on the DHIS2 system in such a way that spatial and temporal survival analysis in identified high malaria risk constituencies can also be performed.

BIBLIOGRAPHY

- Abel, Daniel (2021). "Modelling Malaria Vulnerability Hotspot by Using Geospatial Techniques: The case of Kindo Koyscha Woreda, Wolaita Zone, Ethiopia". PhD thesis. ASTU.
- Achu, AL and RS Rose (2016). "GIS Analysis of Crime Incidence and Spatial Variation in Thiruvananthapuram City". In: *International Journal of Remote Sensing Applications* 6, pp. 1–7.
- Alegana, Victor A et al. (2013). "Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial–temporal models". In: *Spatial and spatio-temporal epidemiology* 7, pp. 25–36.
- Alemu, Abebe et al. (2011). "Climatic variables and malaria transmission dynamics in Jimma town, South West Ethiopia". In: *Parasites & vectors* 4.1, pp. 1–11.
- Ali, Mohammad et al. (2003). "Use of a geographic information system for defining spatial risk for dengue transmission in Bangladesh: role for *Aedes albopictus* in an urban outbreak". In: *The American journal of tropical medicine and hygiene* 69.6, pp. 634–640.
- Allenby, Greg M and Peter E Rossi (2006). "Hierarchical bayes models". In: *The handbook of marketing research: Uses, misuses, and future advances*, pp. 418–440.
- Anselin, Luc (1998). "Exploratory spatial data analysis in a geocomputational environment". In: *Geocomputation: A Primer*. Chichester New York: Wiley.
- Assuncao, Renato M and Edna A Reis (1999). "A new proposal to adjust Moran's I for population density". In: *Statistics in medicine* 18.16, pp. 2147–2162.
- Belitz, Christine et al. (2009). "BayesX". In: *Software for Bayesian Inference in Structured Additive Regressions Models. Methodology Manual, München*.
- Besag, J (1991). "York J. Mollie A. Bayesian image-restoration, with 2 applications in spatial statistics". In: *Ann. Inst. Stat. Math* 43, pp. 1–20.

- Besag, Julian (1974). "Spatial interaction and the statistical analysis of lattice systems". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.2, pp. 192–225.
- Blangiardo, Marta and Michela Cameletti (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.
- Bürkner, Paul-Christian (2017). "Advanced Bayesian multilevel modeling with the R package brms". In: *arXiv preprint arXiv:1705.11123*.
- Cameron, A Colin and Pravin K Trivedi (2001). "Essentials of count data regression". In: *A companion to theoretical econometrics* 331.
- (2013). *Regression analysis of count data*. Vol. 53. Cambridge university press.
- Celeux, Gilles et al. (2006). "Deviance information criteria for missing data models". In: *Bayesian analysis* 1.4, pp. 651–673.
- Coban, Kaan Hakan and Nilgun Sayil (2019). "Evaluation of earthquake recurrences with different distribution models in western Anatolia". In: *Journal of Seismology* 23.6, pp. 1405–1422.
- De Jong, P, C Sprenger, and F Van Veen (1984). "On extreme values of Moran's I and Geary's c". In: *Geographical Analysis* 16.1, pp. 17–24.
- Dehnavieh, Reza et al. (2019). "The District Health Information System (DHIS2): A literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries". In: *Health Information Management Journal* 48.2, pp. 62–75.
- Dey, Dipak K, Sujit K Ghosh, and Bani K Mallick (2000). *Generalized linear models: A Bayesian perspective*. CRC Press.
- DiMaggio, Charles (2012). "Spatial Epidemiology Notes". In: *Applications and Vignettes in R*.
- Feachem, Richard GA et al. (2019). "Malaria eradication within a generation: ambitious, achievable, and necessary". In: *The Lancet* 394.10203, pp. 1056–1112.
- Ferrao, Joao L et al. (2018). "Mapping and modelling malaria risk areas using climate, socio-demographic and clinical variables in Chimoio, Mozambique". In: *International journal of environmental research and public health* 15.4, p. 795.
- Gangnon, Ronald E and Murray K Clayton (2003). "A hierarchical model for spatially clustered disease rates". In: *Statistics in Medicine* 22.20, pp. 3213–3228.

- Gardner, William, Edward P Mulvey, and Esther C Shaw (1995). "Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models." In: *Psychological bulletin* 118.3, p. 392.
- Gelfand, Alan E and Penelope Vounatsou (2003). "Proper multivariate conditional autoregressive models for spatial data analysis". In: *Biostatistics* 4.1, pp. 11–15.
- Gómez-Rubio, Virgilio (2020). *Bayesian inference with INLA*. CRC Press.
- Gómez-Rubio, Virgilio, Roger S Bivand, and Håvard Rue (2021). "Estimating spatial econometrics models with integrated nested Laplace approximation". In: *Mathematics* 9.17, p. 2044.
- Greene, William (2008). "Functional forms for the negative binomial model for count data". In: *Economics Letters* 99.3, pp. 585–590.
- Gruebner, Oliver et al. (2011). "A spatial epidemiological analysis of self-rated mental health in the slums of Dhaka". In: *International Journal of Health Geographics* 10.1, pp. 1–15.
- Guerra, Carlos A, Robert W Snow, and Simon I Hay (2006). "Mapping the global extent of malaria in 2005". In: *Trends in parasitology* 22.8, pp. 353–358.
- Guisan, Antoine, Thomas C Edwards Jr, and Trevor Hastie (2002). "Generalized linear and generalized additive models in studies of species distributions: setting the scene". In: *Ecological modelling* 157.2-3, pp. 89–100.
- Haiyambo, Daniel H et al. (2019). "Molecular detection of *P. vivax* and *P. ovale* foci of infection in asymptomatic and symptomatic children in Northern Namibia". In: *PLoS neglected tropical diseases* 13.5, e0007290.
- Hamza, Suha Elhaj Suliman et al. (2015). "Evaluation of Malaria Diagnostic Methods in Medical Military Hospital, Khartoum State". PhD thesis. Sudan University of Science & Technology.
- Hardin, James W et al. (2007). *Generalized linear models and extensions*. Stata press.
- Hardy, Andy et al. (2019). "Automatic detection of open and vegetated water bodies using Sentinel 1 to map African malaria vector mosquito breeding habitats". In: *Remote Sensing* 11.5, p. 593.
- Hedeker, Donald (2005). "Generalized linear mixed models". In: *Encyclopedia of statistics in behavioral science*.

- Hempelmann, Ernst and Kristine Krafts (2013). "Bad air, amulets and mosquitoes: 2,000 years of changing perspectives on malaria". In: *Malaria journal* 12.1, pp. 1–14.
- Hilbe, Joseph M (2011). *Negative binomial regression*. Cambridge University Press.
- Holowaty, Eric J et al. (2010). "Feasibility and utility of mapping disease risk at the neighbourhood level within a Canadian public health unit: an ecological study". In: *International journal of health geographics* 9.1, pp. 1–14.
- Hsiang, Michelle S et al. (2020). "Effectiveness of reactive focal mass drug administration and reactive focal vector control to reduce malaria transmission in the low malaria-endemic setting of Namibia: a cluster-randomised controlled, open-label, two-by-two factorial design trial". In: *The Lancet* 395.10233, pp. 1361–1373.
- Hussein, Mogahed Ismail Hassan et al. (2020). "Malaria and COVID-19: Unmasking their ties". In: *Malaria Journal* 19.1, pp. 1–10.
- Kazembe, Lawrence N (2007). "Spatial modelling and risk factors of malaria incidence in northern Malawi". In: *Acta Tropica* 102.2, pp. 126–137.
- Lai, Poh-Chin, Fun-Mun So, and Ka-Wing Chan (2008). *Spatial epidemiological approaches in disease mapping and analysis*. CRC press.
- Lamb, Tracey (2012). *Immunity to parasitic infection*. John Wiley & Sons.
- Lawson, Andrew B (2018). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. Chapman and Hall/CRC.
- Lee, Duncan and Richard Mitchell (2013). "Locally adaptive spatial smoothing using conditional auto-regressive models". In: *Journal of the Royal Statistical Society: Series C: Applied Statistics*, pp. 593–608.
- Lee, Kil Seong and Sang Ug Kim (2008). "Identification of uncertainty in low flow frequency analysis using Bayesian MCMC method". In: *Hydrological Processes: An International Journal* 22.12, pp. 1949–1964.
- Lesaffre, Emmanuel and Andrew B Lawson (2012). *Bayesian biostatistics*. John Wiley & Sons.
- Li, Hongfei, Catherine A Calder, and Noel Cressie (2007). "Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model". In: *Geographical Analysis* 39.4, pp. 357–375.

- López-Quilez, Antonio and Facundo Munoz (2009). "Review of spatio-temporal models for disease mapping". In: *Final Report for the EUROHEIS 2*.
- Lord, Catherine et al. (2006). "Autism from 2 to 9 years of age". In: *Archives of general psychiatry* 63.6, pp. 694–701.
- Luerken, Erick L (2009). *Aggregate excess-of-loss under extreme risk: A reinsurance model with Fréchet claims*. University of Nevada, Reno.
- Maina, Joseph et al. (2019). "A spatial database of health facilities managed by the public health sector in sub Saharan Africa". In: *Scientific data* 6.1, pp. 1–8.
- McCreesh, Patrick et al. (2018). "Subpatent malaria in a low transmission African setting: a cross-sectional study using rapid diagnostic testing (RDT) and loop-mediated isothermal amplification (LAMP) from Zambezi region, Namibia". In: *Malaria journal* 17.1, pp. 1–11.
- McKenna, Maryn (2008). *Beating Back the Devil: On the Front Lines with the Disease Detectives of*. Simon and Schuster.
- Moffatt, M (2017). "An Introduction to Akaike's Information Criterion (AIC)". In: *Retrieved from*.
- Moraga, Paula (2019). *Geospatial health data: Modeling and visualization with R-INLA and shiny*. Chapman and Hall/CRC.
- Morris, Mitzi et al. (2019). "Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan". In: *Spatial and spatio-temporal epidemiology* 31, p. 100301.
- Mwahi, Etuhole M (2014). "Shared-component model with application to mapping gender specific pattern in HIV testing and condom use in Namibia". PhD thesis. Citeseer.
- Newby, Gretchen et al. (2016). "The path to eradication: a progress report on the malaria-eliminating countries". In: *The Lancet* 387.10029, pp. 1775–1784.
- Orford, Scott (2001). *Spatial Epidemiology: Methods and Applications: P Elliott, J Wakefield, N Best, D Briggs (eds)*. Oxford: Oxford University Press, 2000, pp. 494, £ 65.00. ISBN: 0-19-262941-7.
- Organization, World Health (2016). *World malaria report 2015*. World Health Organization.

- Parham, Paul Edward and Edwin Michael (2010). "Modeling the effects of weather and climate change on malaria transmission". In: *Environmental health perspectives* 118.5, pp. 620–626.
- Reed, J Michael et al. (2002). "Emerging issues in population viability analysis". In: *Conservation biology* 16.1, pp. 7–19.
- Rezaeian, Mohsen et al. (2007). "Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary". In: *Journal of Epidemiology & Community Health* 61.2, pp. 98–102.
- Sellers, Kimberly F and Galit Shmueli (2010). "A flexible regression model for count data". In: *The Annals of Applied Statistics*, pp. 943–961.
- Sipe, Neil G and Pat Dale (2003). "Challenges in using geographic information systems (GIS) to understand and control malaria in Indonesia". In: *Malaria journal* 2.1, pp. 1–8.
- Smith, Jennifer L et al. (2017). "Spatial clustering of patent and sub-patent malaria infections in northern Namibia: Implications for surveillance and response strategies for elimination". In: *PLoS One* 12.8, e0180845.
- Stevenson, Mark Anthony (2003). "The spatio-temporal epidemiology of Bovine spongiform encephalopathy and Foot-and-mouth disease in Great Britain: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy at Massey University". PhD thesis. Massey University.
- Swartout, Kevin M et al. (2015). "What is the best way to analyze less frequent forms of violence? The case of sexual aggression." In: *Psychology of violence* 5.3, p. 305.
- Turner, Louise et al. (2013). "Severe malaria is associated with parasite binding to endothelial protein C receptor". In: *Nature* 498.7455, pp. 502–505.
- Tuyishimire, J et al. (2016). "Spatial modelling of malaria risk factors in Ruhuha sector in the east of Rwanda". In: *Rwanda Journal* 1.
- Ugarte, MD, AF Militino, and T Goicoa (2008). "Prediction error estimators in empirical Bayes disease mapping". In: *Environmetrics: The official journal of the International Environmetrics Society* 19.3, pp. 287–300.
- Umer, Muhammad Farooq et al. (2019). "Effects of socio-environmental factors on malaria infection in Pakistan: a Bayesian spatial analysis". In: *International journal of environmental research and public health* 16.8, p. 1365.

- Union, African (2006). "African Common Position on the Review of the Millennium Declaration and the Millennium Development Goals". In.
- Vazquez-Prokopec, Gonzalo M et al. (2010). "Quantifying the spatial dimension of dengue virus epidemic spread within a tropical urban environment". In: *PLoS neglected tropical diseases* 4.12, e920.
- WHO (2021). *WHO guideline for malaria, 13 July 2021*. Tech. rep. World Health Organization.
- Wikipedia contributors (2021). *Deviance information criterion — Wikipedia, The Free Encyclopedia*. [Online; accessed 17-December-2021]. URL: https://en.wikipedia.org/w/index.php?title=Deviance_information_criterion&oldid=1050625817.
- Wilesmith, JW et al. (2003). "Spatio-temporal epidemiology of foot-and-mouth disease in two counties of Great Britain in 2001". In: *Preventive veterinary medicine* 61.3, pp. 157–170.
- Worrall, Eve, Aafje Rietveld, and Charles Delacollette (2004). "The burden of malaria epidemics and cost-effectiveness of interventions in epidemic situations in Africa." In: *The American journal of tropical medicine and hygiene* 71.2 Supp, pp. 136–140.
- Wu, Xiaoxu et al. (2016). "Impact of climate change on human infectious diseases: Empirical evidence and human adaptation". In: *Environment international* 86, pp. 14–23.
- Zhang, Wenyi et al. (2008). "Spatial analysis of malaria in Anhui province, China". In: *Malaria journal* 7.1, pp. 1–10.

APPENDIX A

R code

A.1 non-spatial and spatial Analysis

```
setwd ("C:/Users/rkatale/Desktop/analysis")
```

```
install.packages(c("sp", "rgeos", "ggplot2", "ggmap", "dplyr", "raster", "tmap", "leaflet",  
"spatstat", "spdep", "rgdal", "spdep", "rgdal", "rgeos", "latticeExtra", "RColorBrewer",  
"gridExtra", "maps", "mapproj", "CARBayes", "Matrix", "tidyverse", "spgwr", "tm_shape",  
"lattice"))
```

```
install.packages("INLA", repos = c(getOption("repos"), INLA = "https://inla.r-inla-download.org/R/stable"),  
dep = T)
```

```
#Call these packages
```

```
library(rgeos) library(rgdal)  
library(sp)  
library(ggplot2)  
library(ggmap)  
library(dplyr)  
library(raster)  
library(leaflet)  
library(spatstat)  
library(spdep)  
library(latticeExtra)  
library(RColorBrewer)  
library(gridExtra)  
library(Matrix)  
library(lattice)  
library(maptools)  
library(foreign)  
library(ggmap)  
library(BayesX)  
library(splines)  
library(spdep)  
library(foreach)  
library(parallel)  
library(INLA)  
library(shapefiles)
```

```

library(maps)
library(mapproj)
library(RColorBrewer)
library(CARBayes)
library(sf)
library(reshape2)
library(rstanarm)
require(sandwich)
library(tidyverse)
library(rgeos)
library(tmap)
library(tmaptools)
library(spgwr)
library(grid)
library(tm_shape)

#Load in shapefile
NW <- readOGR(dsn = ".", layer= "SelectedReg")
plot(NW)
data1 <- read.csv(file="SelectedReg.csv", header=TRUE, sep=",")
library(spData)

queen.nb = poly2nb(NW)
summary(queen.nb)
queen.listw=nb2listw(queen.nb) #convert nb to listw type
listw=queen.listw

#===== plot neighbourhood
plot(NW, border=gray(.5))
plot(queen.nb, coordinates(NW), add=TRUE)

#===== Merge the two data files
data2 <- merge(NW, data1, by='OBJECTID')
# =====Computing moran's I in R for spatial Data
mi <- moran.test(data2$NoCasesA, listw = nb2listw(queen.nb))
mi

#===== Plot Moran I scatter plot, Moran I (local),and probability of most signif-
icant Moran I (Chapter 3, Figure 3.1)

moran.plot(data2$NoCasesA, listw = queen.listw, xlab="Number of malaria Cases",
ylab="Spatially lagged malaria cases")

locm <- localmoran(data2$NoCasesA, listw = nb2listw(queen.nb)) summary(locm)
knitr::opts_chunk$set(echo = TRUE)

# Change the presentation of decimal numbers to 4 and avoid scientific notation op-
tions(prompt="R> ", digits=4, scipen=999)

```

```

# =====Load in shapefile
NW <- readOGR(dsn = ".", layer= "SelectedReg") plot(NW)
data1 <- read.csv(file="SelectedReg.csv", header=TRUE, sep=",")
library(spData)
queen.nb = poly2nb(NW) summary(queen.nb)
queen.listw=nb2listw(queen.nb) #convert nb to listw type listw=queen.listw

#===== plot neighbourhood
plot(NW, border=gray(.5)) plot(queen.nb, coordinates(NW), add=TRUE)

#===== Merge the two data files
OA.Census<- merge(NW, data1, by='OBJECTID')
OA.Census_sf <- st_as_sf(OA.Census)
#8 Spatial distribution

tm_shape(shp, is.master = NA, projection = NULL, bbox = NULL, unit = "metric", simplify
= 1, line.center.type = c("segment", "midpoint"), ...)

tm_shape(OA.Census_sf) + tm_fill("NoCasessig", palette = "Reds", style = "quantile", title
= " Sum of malaria postive cases") + tm_borders(alpha=.8)

# Neighbour structure with colour red for visibility
#Find queen neighbours
neighbours <- poly2nb(OA.Census) neighbours
neighbours_sf <- poly2nb(OA.Census_sf) neighbours_sf

#Plot queen neighbours links plot(OA.Census, border = 'lightgrey')
(neighbours, coordinates(OA.Census), add=TRUE, col='red')

listw <- nb2listw(neighbours2) listw
# Global spatial autocorrelation
# Compute local Moran
# binds results to our polygon shapefile

moran.map <- cbind(OA.Census, local)

tm_shape(moran.map) + tm_fill(col = "Ii", style = "quantile", title = "local moran statis-
tic")+ tm_borders(alpha=.900)

#Plot LISA clusters
quadrant <- vector(mode="numeric",length=nrow(local))

# centers the variable of interest around its mean
m.NoCasessig <- OA.Census$NoCasessig - mean(OA.Census$NoCasessig)

# centers the local Moran's around the mean

m.local <- local[,1] - mean(local[,1])
# significance threshold

```

```

signif <- 0.1 # builds a data quadrant
quadrant[m.NoCasessig >0 & m.local>0] <- 4 quadrant[m.NoCasessig <0 & m.local<0] <-
1
quadrant[m.NoCasessig <0 & m.local>0] <- 2
quadrant[m.NoCasessig >0 & m.local<0] <- 3
[local[,5]>signif] <- 0

```

```
# plot in r
```

```

brks <- c(0,1,2,3,4)
colors <- c("white", "blue", rgb(0,0,1,alpha=0.4)
rgb(1,0,0,alpha=0.4), "red") plot(OA.Census, border="lightgray", col=colors [findInter-
val(quadrant, brks, all.inside=FALSE)]) box()

```

```

legend("bottomleft", legend = c("insignificant", "low-low", "low-high", "high-low", "high-
high"), fill=colors, bty="n")

```

```
#Getis-Ord approach
```

```

#creates centroid and joins neighbours within 0 and 800 units
nb <- dnearneigh(coordinates(OA.Census), 0,2)

```

```

#creates listw
nb_lw <- nb2listw(nb, style = 'B')

```

```

#Plot data and neighbours
plot(OA.Census, border = 'lightgrey')
plot(nb, coordinates(OA.Census), add=TRUE, col = 'red')

```

```

#Getis-Ord Gi statistic
local_g <- localG(OA.Census$NoCasessig, nb_lw)
local_g <- cbind(OA.Census, as.matrix(local_g))
names(local_g)[8] <- "gstat"

```

```
#Cluster map
```

```

tm_shape(local_g) + tm_fill("gstat", palette = "RdBu", style = "pretty", title = "Getis-Ord
Gi Statistics") + tm_borders(alpha=1)

```

```

#=====
Non spatial model (INDIVIDUAL DATASET)
#=====

```

```

mydata <- read.csv(file="Individualdata2021.csv", header=TRUE, sep=";")
summary(mydata)
str(mydata)
attach (mydata)

```

```
poisreg1offset2individual <- glm(PosRDT Gender+ Agegroup+ Placeofresidence+TypeofFacility+
Occupation+Sleptunderbednetlast3nights+Homesprayedpast12months+District+
Season +offset(log(Population/1000)),family = poisson, data = mydata)
```

```
summary (poisreg1offset2individual)
```

```
#===== Extracts the column vector of p values
coef(summary(poisreg1offset2individualnegbinomial))[,4]
```

```
#====finding confidence interval for non spatial negative binomial model
confint(poisreg1offset2individualnegbinomial) (est <- cbind(Estimate =
coef(poisreg1offset2individualnegbinomial), confint(poisreg1offset2individualnegbinomial)))
```

```
#=====finding the exponentiated coefficients=====
```

```
exp(coef(poisreg1offset2individualnegbinomial))
```

```
#=====
```

```
#END Non spatial model (INDIVIDUAL DATASET)
```

```
#=====
```

```
#=====
```

```
#Non spatial model ( Aggregated data)
```

```
#=====
```

```
mydata <- read.csv(file="Malalia26.csv", header=TRUE, sep=",")
```

```
summary(mydata)
```

```
str(mydata)
```

```
attach (mydata)
```

```
library(GLMsData)
```

```
mydata$Region <-relevel(factor(mydataRegion),ref = "Ohangwe")
```

```
mydata$Season <-relevel(factor(mydataSeason),ref = "Spring")
```

```
# =====Non Spatial Poison Model=====
```

```
poisreg1offset2 <- glm(PosRDT Season +Region + hpd +mavgtemp + mmintemp +
mmaxtemp + mtotalrainfal + mavgwindspeed + mavegoiltemp + humidity +avegleafwet-
ness+offset(log(Population/1000)),family = poisson, data = mydata)
```

```
summary (poisreg1offset2)
```

```
#=====Non Spatial Negative binomial Model=====
```

```
poisreg1offsetnegbinomial <- glm.nb(PosRDT Season +Region + hpd+mavgtemp +
mmintemp + mmaxtemp + mtotalrainfal + mavgwindspeed + mavegoiltemp + humidity
+avegleafwetness+offset(log(Population/1000) ))
```

```
summary (poisreg1offsetnegbinomial)
```

```
#===== Extracts the column vector of p values
```

```
coef(summary(poisreg1offset2))[,4]
```

```
coef(summary(poisreg1offsetnegbinomial))[,4]
```

```

#=====

#===finding confidence interval for non spatial negative binomial model===

confint(poisreg1offsetnegbinomial) (est <- cbind(Estimate = coef(poisreg1offsetnegbinomial),
confint(poisreg1offsetnegbinomial)))
#=====finding the exponentiated coefficients=====

exp(coef(poisreg1offsetnegbinomial))

poisreg1offsetDIC <- (PosRDT Season + hpd + mavgtemp + mmintemp + mmaxtemp
+ mtotalrainfal + mavgwindspeed + mavegoiltemp + humidity +avegleafwetness +off-
set(log(Population/1000))+ f(CONST, model = "iid", hyper = prior.prec) )
summary (poisreg1offsetDIC)

#=====

#=====END OF NON SPATIAL MODELS=====

```

```

#=====SPATIAL MODEL AGGREGATED DATA

require(sp) # package to work with spatial data
require(rgdal) # package to work with spatial data

# =====Load in shapefile NW <- readOGR(dsn = ".", layer= "SelectedReg") plot(NW)
data1 <- read.csv(file="aggregateddata2020.csv", header=TRUE, sep=",")
colnames(data1)[1]<-"OBJECTID"

#===== Merge the two data files

data2 <- merge(NW,data1,by='OBJECTID')
ddd2<-data2@data
require(RColorBrewer)
# Create a colour palette to use in graphs
my.palette <- brewer.pal(n = 9, name = "YlOrRd")

#===== Visualise the number of scats across space
splot(data2, zcol = "RR", col.regions = my.palette, cuts = 8)

#===== Specify the adjacency matrix
mal_Temp <- poly2nb(data2) # construct the neighbour list
nb2INLA("mal.graph", mal_Temp) # create the adjacency matrix in INLA format
mal.adj <- paste(getwd(),"/mal.graph",sep="")

# name the object

inla.setOption(scale.model.default = F)

H <- inla.read.graph(filename = "mal.graph") # and save it as a graph
# Plot adjacency matrix image(inla.graph2matrix(H), xlab = "", ylab = "")

data2$IDcode<-1:34

#=====

THE NEGATIVE BINOMIAL NULL MODEL
#=====

formula3 <- NoCases 1 + # fixed effect
f(IDcode, model = "bym", # spatial effect: IDcode is a numeric identifier for each area in the
lattice (does not work with factors)
graph = mal.adj) # this specifies the neighbouring of the lattice areas
mal_Data<-data2@data

```

```

# Finally, we can run the model using the inla() function
Mod_mal3 <- inla(formula3, family = "nbinomial", # since we are working with count data
data = mal_Data, control.compute = list(cpo = T, dic = T, waic = T), E=expe)
# CPO, DIC and WAIC metric values can all be computed by specifying that in the control.compute option
# These values can then be used for model selection purposes if you wanted to do that
# Check out the model summary
summary(Mod_mal3)

formula3_p <- NoCases ~ 1 +f( IDcode, model = "bym", graph = mal.adj, scale.model = TRUE,
hyper = list( prec.unstruct = list(prior = "loggamma", param = c(1,0.001)), # precision for the unstructured effect (residual noise)
prec.spatial = list(prior = "loggamma", param = c(1,0.001)) # precision for the spatial structured effect ) )

Mod_mal3_p <- inla(formula3_p, family = "nbinomial", data = mal_Data, control.compute = list(cpo = T), E=expe )

summary(Mod_mal3_p)

# We can extract the summary of the fixed effects (in this case only GS)

zeta.cutoff <- c(0, 1,2,3,4) # we make a categorisation to make visualisation easier
cat.zeta <- cut(unlist(zeta), breaks = zeta.cutoff, include.lowest = TRUE)

# Create a dataframe with all the information needed for the map

maps.cat.zeta <- data.frame(IDcode = mal_Data$IDcode, cat.zeta = cat.zeta)

# Create a new polygon from Fox_Lattice and add the value of the posterior mean

mal_Lattice_post <- data2 data.mal.post <- attr(mal_Lattice_post, "data")
attr(mal_Lattice_post, "data") <- merge(data.mal.post, maps.cat.zeta, by = "IDcode")

my.palette.post <- rev(brewer.pal(n = 9, name = "YlGnBu"))
splot(obj = mal_Lattice_post, zcol = "cat.zeta", col.regions = my.palette.post)

#=====
a <- 0
prob.zone <- lapply(zone.index, function(x) 1 - inla.pmarginal(a, x))
prob.zone.cutoff <- c(0, 0.1, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)

cat.prob.zone <- cut(unlist(prob.zone), breaks = prob.zone.cutoff, include.lowest = T)

# Create a new polygon from Fox_Lattice and add the value of the posterior sd

```

```
maps.cat.prob.zone <- data.frame(IDcode = mal_Data$IDcode, cat.prob.zone = cat.prob.zone)
```

```
mal_Lattice_var <- data2 data.fox.var <- attr(mal_Lattice_var, "data") attr(mal_Lattice_var,  
"data") <- merge(data.fox.var, maps.cat.prob.zone, by = "IDcode")
```

```
my.palette.var <- brewer.pal(n = 9, name = "BuPu") spplot(obj = mal_Lattice_var, zcol =  
"cat.prob.zone", col.regions = my.palette.var, add = T)
```

```
#=====
THE NEGATIVE BINOMIAL SPATIAL MODEL (WITH RANDOM
EFFECTS)
#=====
```

```
formula2 <- NoCases 1 +hpd_mean+mavegoiltemp_mean+ mavgtemp_mean + mmintemp_mean+  
mmaxtemp_mean+mtotalrainfal_mean+mavgwindspeed_mean  
+humidity_mean+avegleafwetness_mean+ # fixed effect f(IDcode, model = "bym", # spatial  
effect: IDcode is a numeric identifier for each area in the lattice (does not work with factors)  
graph = mal.adj) # this specifies the neighbouring of the lattice areas
```

```
mal_Data<-data2@data
```

```
# Finally, we can run the model using the inla() function
```

```
Mod_mal <- inla(formula2, family = "nbinomial", # since we are working with count data  
data = mal_Data, control.compute = list(cpo = T, dic = T, waic = T), E=expe)
```

```
# CPO, DIC and WAIC metric values can all be computed by specifying that in the con-  
trol.compute option
```

```
# These values can then be used for model selection purposes if you wanted to do that
```

```
# Check out the model summary
```

```
summary(Mod_mal)
```

```
formula2_p <- NoCases 1 +hpd_mean +mavegoiltemp_mean + mavgtemp_mean +  
mmintemp_mean+mmaxtemp_mean+mtotalrainfal_mean+mavgwindspeed_mean+  
humidity_mean+avegleafwetness_mean + # fixed effect +f( IDcode, model = "bym", graph =  
mal.adj, scale.model = TRUE, hyper = list( prec.unstruct = list(prior = "loggamma", param  
= c(1,0.001)), # precision for the unstructured effect (residual noise) prec.spatial = list(prior  
= "loggamma", param = c(1,0.001)) # precision for the spatial structured effect ) )
```

```
Mod_mal_p2 <- inla(formula2_p, family = "nbinomial", data = mal_Data, control.compute
= list(cpo = T), E=expe )
```

```
summary(Mod_mal_p2)
# We can extract the summary of the fixed effects (in this case only GS)
```

```
round(Mod_mal$summary.fixed, 3)
```

```
# Calculating the number of areas
```

```
Nareas <- length(mal_Data[,1])
```

```
“ zone.index <- Mod_mal$marginals.random$IDcode[1:Nareas]
```

```
# exponentiate each of the zone marginals to return it to its original values (remember that
this is a poisson model so all the components of the model are log-transformed)
```

```
zetaE <- lapply(zone.index,function(x) inla.emarginal(exp,x))
```

```
zetaE.cutoff <- c(0, 1,2,3,4)
```

```
zeta.cutoff <- c(0,0.1, 1,2,3,4)# we make a categorisation to make visualisation easier
```

```
cat.zetaE <- cut(unlist(zetaE), breaks = zetaE.cutoff, include.lowest = TRUE)
```

```
# Create a dataframe with all the information needed for the map
```

```
maps.cat.zetaE <- data.frame(IDcode = mal_Data$IDcode, cat.zetaE = cat.zetaE)
```

```
# Create a new polygon from Fox_Lattice and add the value of the posterior mean
```

```
mal_Lattice_postE <- data2
```

```
data.mal.postE <- attr(mal_Lattice_postE, "data") attr(mal_Lattice_postE, "data") <-
merge(data.mal.postE, maps.cat.zetaE, by = "IDcode")
```

```
my.palette.postE <- rev(brewer.pal(n = 9, name = "YlGnBu"))
```

```
sppplot(obj = mal_Lattice_postE, zcol = "cat.zetaE", col.regions = my.palette.postE)
```

```
#=
```

```
a <- 0
```

```
prob.zoneE <-
```

```
lapply(zone.index,function(x) 1 - inla.pmarginal(a, x))
```

```
prob.zone.cutoffE <- c(0, 0.1, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, cat.prob.zoneE <-
```

```
cut(unlist(prob.zoneE), breaks = prob.zone.cutoffE, include.lowest = T)
```

```

# Create a new polygon from Fox_Lattice and add the value of the posterior sd
maps.cat.prob.zoneE <- data.frame(IDcode = mal_Data$IDcode, cat.prob.zoneE = cat.prob.zoneE)

mal_Lattice_varE <- data2 data.fox.var <- attr(mal_Lattice_varE, "data")
attr(mal_Lattice_varE, "data") <- merge(data.fox.var, maps.cat.prob.zoneE, by = "IDcode")

my.palette.varE <- brewer.pal(n = 9, name = "BuPu") spplot(obj = mal_Lattice_varE, zcol
= "cat.prob.zoneE", col.regions = my.palette.varE, add = T)

#=====
#THE NEGATIVE BINOMIAL SPATIAL MODEL for INDIVIDUAL DATA (WITH
RANDOM EFFECTS)
#=====

formula5 <- NoCases 1 +p3+p5+p8+p9+p10+p11+p12+p13+p20+p32+ # fixed effect
f(IDcode, model = "bym", # spatial effect: IDcode is a numeric identifier for each area in
the lattice (does not work with factors) graph = mal.adj) # this specifies the neighbouring of
the lattice areas
mal_Data5<-data2@data

# Finally, we can run the model using the inla() function
Mod_mal5 <- inla(formula5, family = "nbinomial", # since we are working with count data
data = mal_Data5, control.compute = list(cpo = T, dic = T, waic = T),E=expe)

# CPO, DIC and WAIC metric values can all be computed by specifying that in the con-
trol.compute option
# These values can then be used for model selection purposes if you wanted to do that

# Check out the model summary
summary(Mod_mal5)

formula5_p <- NoCases 1+p3+p5+p8+p9+p10+p11+p12+p13+p20+p32+ # fixed effect +f(
IDcode, model = "bym", graph = mal.adj, scale.model = TRUE, hyper = list( prec.unstruct
= list(prior = "loggamma", param = c(1,0.001)), # precision for the unstructured effect
(residual noise) prec.spatial = list(prior = "loggamma", param = c(1,0.001)) # precision for
the spatial structured effect ) )

Mod_mal5_p <- inla(formula5_p, family = "nbinomial", data = mal_Data5, control.compute
= list(cpo = T), E=expe )

```

```

summary(Mod_mal5_p)

# We can extract the summary of the fixed effects (in this case only GS)
round(Mod_mal5$summary.fixed, 3)

# Calculating the number of areas
Nareas <- length(mal_Data5[,1])
zone.index <- Mod_mal5$marginals.random$IDcode[1:Nareas]

# exponentiate each of the zone marginals to return it to its original values (remember that
this is a poisson model so all the components of the model are log-transformed)
zeta <- lapply(zone.index,function(x) inla.emarginal(exp,x))

zeta.cutoff <- c(0, 1,2,3,4) # we make a categorisation to make visualisation easier cat.zeta <-
cut(unlist(zeta), breaks = zeta.cutoff, include.lowest = TRUE)

# Create a dataframe with all the information needed for the map
maps.cat.zeta <- data.frame(IDcode = mal_Data5$IDcode, cat.zeta = cat.zeta)

# Create a new polygon from Fox_Lattice and add the value of the posterior mean
mal_Lattice_post <- data2 data.mal.post <-
attr(mal_Lattice_post, "data") attr(mal_Lattice_post, "data") <- merge(data.mal.post,
maps.cat.zeta, by = "IDcode")

my.palette.post <- rev(brewer.pal(n = 9, name = "YlGnBu"))
spplot(obj =
mal_Lattice_post, zcol = "cat.zeta", col.regions = my.palette.post)
#=====

a <- 0
prob.zone <- lapply(zone.index, function(x) 1 - inla.pmarginal(a, x))

prob.zone.cutoff <- c(0, 0.1, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)
cat.prob.zone <- cut(unlist(prob.zone), breaks = prob.zone.cutoff, include.lowest = T)

# Create a new polygon from Fox_Lattice and add the value of the posterior sd
maps.cat.prob.zone <-
data.frame(IDcode = mal_Data$IDcode, cat.prob.zone = cat.prob.zone)

mal_Lattice_var <- data2

data.fox.var <- attr(mal_Lattice_var, "data") attr(l_Lattice_var, "data") <- merge(data.fox.var,

```

```
maps.cat.prob.zone, by = "IDcode")
```

```
my.palette.var <- brewer.pal(n = 9, name = "BuPu")
```

```
spplot(obj = mal_Lattice_var, zcol = "cat.prob.zone", col.regions = my.palette.var, add = T)
```

```
# END OF SPATIAL ANALYSIS
```

A.2 spatio - temporal Analysis

```
setwd("C:Users... Dataset") remove(list=ls())
my.dir <- paste(getwd(),"/",sep="")
require(INLA) inla.setOption(scale.model.default=FALSE)
```

```
require(splancs)
require(sp)
require(fields)
require(mapttools) require(lattice)
require(abind) library(spdep)
library(RColorBrewer)
```

```
datak <- read.csv(paste(my.dir,"katale/aggregateddata 2018to20.csv",sep="")) malariak <-
readShapePoly(paste(my.dir,"katale/SelectedReg",sep="")) ID.area1<- datak$ID.area
```

```
temp <- poly2nb(malariak) nb2INLA("malaria.graph", temp)
malaria.adj <- paste(getwd(),"/malaria.graph",sep="")
```

parametric trend model

```
formula.par <- PosRDT sum 1 +mtotalrainfalmean +mavgtemp mean +mmaxtemp
mean+mavgwindspeed mean +humidity mean+avegleafwetness mean+ f(ID.area,
model="bym",graph=malaria.adj, constr=TRUE) + f(ID.area1,ID.year,model="iid", con-
str=TRUE) + ID.year
```

```
model.par <- inla(formula.par,family="nbinomial",data=datak,E=expe, control.predictor=list(compute=TRUE,
control.compute=
list(dic=TRUE,cpo=TRUE))
```

```
round(model.par$summary.fixed[,1:5],3)
```

```
# ***
```

#Plotting the poaterior mean of the main time effects

```
x <- seq(1,3) # Years
```

```
y<-model.par$summary.fixed[2,1]x
```

```
plot(x,y, type="l", main="Global linear temporal trend",xlab="t",ylab=expression(betat),
ylim=c(min(model.par$summary.fixed[2,3]x),max(model.par$summary.fixed[2,5]x)))
lines(model.par$summary.fixed[2,3]x,lty=2)
```

```
lines(model.par$summary.fixed[2,5]x,lty=2)
```

* #Plotting the posterior mean of spatial effect**

```

m <- model.par$marginals.random[[1]][1:34]
zeta.ST1 <- unlist(lapply(m,function(x)inla.emarginal(exp,x)))

SMR.cutoff<-summary(zeta.ST1)[-4]
xi.factor <- cut(zeta.ST1,breaks=SMR.cutoff,include.lowest=TRUE)

m <- model.par$summary.random[[2]][1:34,2]

int.cut <-summary(m)[-4]

int.factor <- cut(m,breaks=int.cut,include.lowest=TRUE)

data.malariak <- attr(malariak, "data")

attr(malariak, "data")=data.frame(data.malariak, xi=xi.factor)

trellis.par.set(axis.line=list(col=NA))

spplot(obj=malariak, zcol= "xi", col.regions=gray(3.5:0.5/4),colorkey = list(space = "bottom", height = 0.9),main="Posterior mean of spatial main effect")

# ***
#Plotting the posterior mean of defferential time effect

datak <- read.csv(paste(my.dir,"katale/aggregateddata 2018to20.csv",sep=""))
SelectedReg <- readShapePoly(paste(my.dir,"katale/SelectedReg",sep=""))
data.SelectedReg <- attr(SelectedReg, "data")

# ***
#Non - parametric dynamic trend model

datak <- read.csv(paste(my.dir,"katale/aggregateddata 2018to20.csv",sep=""))
SelectedReg <- readShapePoly(paste(my.dir,"katale/SelectedReg",sep=""))

data.SelectedReg <- attr(SelectedReg, "data")

year <- numeric(0)

for(i in 1:3) year<- append(year,rep(i,34))

y = as.numeric(datak$PosRDT sum) ID.area<-as.numeric(datak$ID.area) ID.area1<-
as.numeric(datak$ID.area)
ID.year<-as.numeric(datak$ID.year)
ID.year1<-as.numeric(datak$ID.year)
ID.area.year<-as.numeric(datak$ID.area.year)
E<-datak$expe mtotalrainfalmean
<-as.numeric(datak$mtotalrainfalmean )

```

```

mavgtemp mean <-as.numeric(datak$mavgtemp mean ) mmaxtemp mean
<-as.numeric(datak$mmaxtemp mean )
mavgwindspeed mean <-as.numeric(datak$mavgwindspeed mean ) humidity mean <-
as.numeric(datak humidity mean )
avegleafwetness mean <-as.numeric(datak$avegleafwetness mean )

datakk <- data.frame(y= y, E= E, ID.area=as.numeric(ID.area), ID.area1=as.numeric(ID.area),
year=year, ID.year = ID.year, ID.year1=ID.year, ID.area.year =
as.numeric(ID.area.year),mtotalrainfalmean = as.numeric(mtotalrainfalmean),mavgtemp mean
= as.numeric(mavgtemp mean),mmaxtemp mean = as.numeric(mmaxtemp mean),mavgwindspeed
mean = as.numeric(mavgwindspeed mean),humidity mean = as.numeric(humidity mean),
avegleafwetness mean = as.numeric(avegleafwetness mean))

```

Spatial graph

```

temp <- poly2nb(malariak)
nb2INLA("malaria.graph", temp)
malaria.adj <- paste(getwd(),"/malaria.graph",sep="")

```

```

temp <- poly2nb(SelectedReg)
nb2INLA("malaria.graph", temp)
# Temporal graph

```

```

malaria.adj <- paste(getwd(),"/malaria.graph",sep="")

```

```

formula.ST1 <- y 1+mtotalrainfalmean +mavgtemp mean +mmaxtemp mean+mavgwindspeed
mean +humidity mean+avegleafwetness mean+f(ID.area,model="bym",graph=malaria.adj)
+f(ID.year,model="rw2") + f(ID.year1,model="iid") lcs <- inla.make.lincombs(ID.year =
diag(3), ID.year1 = diag(3))
model.ST1 <- inla(formula.ST1,family="nbinomial",data=datakk,E=E, control.predictor=
list(compute=TRUE), lincomb=lcs)

```

```

temporal.CAR <- lapply(model.ST1$marginals.random$ID.year, function(X) marg <-
inla.tmarginal(function(x) exp(x), X) inla.emarginal(mean, marg) )

```

```

temporal.IID <- lapply(model.ST1$marginals.random$ID.year1, function(X) marg <-
inla.tmarginal(function(x) exp(x), X) inla.emarginal(mean, marg) )

```

* # Posterior mean for temporal trend graph**

```

plot(seq(1,3),seq(0.999,1.001,length=3),type="n",xlab="t",ylab=expression(exp(phi[t])))
lines(unlist(temporal.IID)) lines(unlist(temporal.CAR),lty=2) abline (h=1,lty=1) # ***

```

#— Type I interaction —#

```

formula.intI<- y 1+mtotalrainfalmean +mavgtemp mean +mmaxtemp mean+mavgwindspeed
mean +humidity mean+avegleafwetness mean+f(ID.area,model="bym", graph=malaria.adj)
+ f(ID.year,model="rw2") + f(ID.year1,model="iid") + f(ID.area.year,model="iid")

```

```

mod.intI <- inla(formula.intI,family="nbinomial",data=dataakk,E=E, control.predictor=
list(compute=TRUE), control.compute=list(dic=TRUE,cpo=TRUE))

###
delta.intI <- data.frame(delta=mod.intI$summary.random$ID.area.year[,2],
year=dataakk$ID.year,ID.area=dataakk$ID.area) delta.intI.matrix <- matrix(delta.intI[,1], 34,3,by-
row=FALSE) rownames(delta.intI.matrix)<- delta.intI[1:34,3]

# Check the absence of spatial trend for (intI) sumarrydelta<-summary(c(delta.intI[,1]))

####
cutoff.interaction<-sumarrydelta[c(-3,-4)] delta.intI.factor <- data.frame
(CONST=data.SelectedReg$CONST) for(i in 1:3) delta.factor.temp <-
cut(delta.intI.matrix[,i],breaks=cutoff.interaction,include.lowest=TRUE) delta.intI.factor <-
cbind(delta.intI.factor,delta.factor.temp)

colnames(delta.intI.factor)<- c("NAME",seq(2018,2020))
# plot intI
attr(SelectedReg, "data") <- data.frame(data.SelectedReg, intI=delta.intI.factor)
trellis.par.set(axis.line=list(col=NA))
spplot(obj=SelectedReg, zcol=c("intI.2018","intI.2019","intI.2020"), col.regions=gray(2.5:0.5/4),
names.attr=seq(2018,2020),main="")

# ***
# END of Int I
#— Type II interaction —#

ID.area.int <- dataakk$ID.area
ID.year.int <- dataakk$ID.year
mtotalrainfalmean.int <- dataakk$mtotalrainfalmean
mavgtemp mean.int <- dataakk$mavgtemp mean
mmaxtemp mean.int <- dataakk$mmaxtemp mean
mavgwindspeed mean.int <- dataakk$mavgwindspeed mean
humidity mean.int <- dataakk$humidity mean avegleafwetness mean.int <-
dataakk$avegleafwetness mean

formula.intII<- y 1+ mtotalrainfalmean+mavgtemp mean +mmaxtemp mean+mavgwindspeed
mean +humidity mean+avegleafwetness mean+f(ID.area,model="bym",graph=malaria.adj)
+f(ID.year,model="rw2")+ f(ID.year1,model="iid")+ f(ID.area.int,model="iid", group=ID.year.int,
control.group=list(model="rw2"))

mod.intII <- inla(formula.intII,family="nbinomial",data=dataakk,E=E,
control.predictor=list(compute=TRUE), control.compute=list(dic=TRUE,cpo=TRUE))
###
delta.intII <- data.frame(delta=mod.intII$summary.random$ID.area.int [,2],
year=dataakk$ID.year,ID.area=dataakk$ID.area)

delta.intII.matrix <- matrix(delta.intII[,1], 34,3,byrow=FALSE)
rownames(delta.intII.matrix)<- delta.intII[1:34,3]

```

```

***
# Check the absence of spatial trend for (intII)

sumarrydelta<-summary(c(delta.intIII[,1]))

****

cutoff.interaction<-sumarrydelta[c(-3,-4)]

***

delta.intII.factor <- data.frame(CONST=data.SelectedReg$CONST) for(i in 1:3) delta.factor.temp
<- cut(delta.intII.matrix[,i],breaks=cutoff.interaction,include.lowest=TRUE) delta.intII.factor
<- cbind(delta.intII.factor,delta.factor.temp)

colnames(delta.intII.factor)<- c("NAME",seq(2018,2020))

attr(SelectedReg, "data") <- data.frame(data.SelectedReg,intII=delta.intII.factor) trel-
lis.par.set(axis.line=list(col=NA)) ***

# *** plot Int II

spplot(obj=SelectedReg, zcol=c("intII.2018", "intII.2019", "intII.2020"), col.regions=gray(2.5:0.5/3),
names.attr=seq(2018,2020),main="")

# ***
# END of Int II
#— Type III interaction —#

formula.intIII<- y 1+mtotalrainfalmean+mavgtemp mean +mmaxtemp mean+mavgwindspeed
mean +humidity mean+avegleafwetness mean+f(ID.area,model="bym",graph=malaria.adj)
+ f(ID.year,model="rw2") + f(ID.year1,model="iid") + f(ID.year.int,model="iid",
group=ID.area.int,control.group=list(model="besag", graph=malaria.adj))

mod.intIII <- inla(formula.intIII,family="nbinomial",data=dataakk,E=E, control.predictor=
list(compute=TRUE), control.compute=list(dic=TRUE,cpo=TRUE))
****
delta.intIII <- data.frame(delta=mod.intIII$summary.random$ID.year.int[,2],year=dataakk$ID.year,
ID.area=dataakk$ID.area) delta.intIII.matrix <- matrix(delta.intIII[,1], 34,3,byrow=FALSE)
rownames(delta.intIII.matrix)<- delta.intIII[1:34,3]
****
# Check the absence of spatial trend for (intII)

sumarrydelta<-summary(c(delta.intIII[,1]))
**** cutoff.interaction<-sumarrydelta[c(-3,-4)]

****

delta.intIII.factor <- data.frame(CONST=data.SelectedReg$CONST) for(i in 1:3) delta.factor.temp
<- cut(delta.intIII.matrix[,i],breaks=cutoff.interaction,include.lowest=TRUE) delta.intIII.factor

```

```

<- cbind(delta.intIII.factor,delta.factor.temp)

colnames(delta.intIII.factor)<- c("NAME",seq(2018,2020))

  *** attr(SelectedReg, "data") <- data.frame(data.SelectedReg, intIII=delta.intIII.factor)
trellis.par.set(axis.line=list(col=NA))

  ****
  # *** plot Int III

spplot(obj=SelectedReg, zcol=c("intIII.2018","intIII.2019","intIII.2020"), col.regions=gray(2.5:0.5/3),
names.attr=seq(2018,2020),main="")

# ***
# END OF SPATIO - TEMPORAL ANALYSIS

```

APPENDIX B

Malaria active case detection questionnaire



MALARIA RAPID CASE NOTIFICATION FORM
(MUST BE COMPLETED WITHIN 24 HOURS OF ALL CONFIRMED MALARIA DIAGNOSIS)
 National Vector-Borne Diseases Control Programme, MoHSS



Investigation Date _____ (DD/MM/YY)
 Region _____ District _____ Health Facility _____
 Reporting Health Worker Name _____ Health Worker Phone # _____

PATIENT DETAILS

Patient Name _____
 Patient phone number _____
 Alternative phone number _____
 Current village/town _____ Headman name _____
 Household Health Number _____ Head of Household name _____
 Age in completed years _____

Gender Male Female Pregnant Yes No

Current occupation Unemployed Student Guard Farming/Agriculture Other Manual Labour
 Small-market sales or trade Nurse/Teacher/Professional Fisherman
 Other: Specify _____

DIAGNOSIS

Detection setting: Passive (at health facility) Active (in the field)
 Diagnosis confirmed by: RDT Microscopy Clinical Symptoms
 Species: *P. falciparum* Other species Mixed Infection
 Diagnosis Type: Uncomplicated Severe

TREATMENT

Treatment Prescribed: Artemether-lumefantrine (AL) Oral Quinine IV/IM Artesunate
 IV Quinine IM Quinine Primaquine Other _____

TRAVEL HISTORY

Have you spent a night outside of Namibia in the past 4 weeks? Yes No
 If no, skip 'Trip Details'

Trip Details (Please fill out the country and village name if Imported/Non local)

Trip History (Must be filled if Imported/Non local)	First Night (DD/MM/YY)	Last Night (DD/MM/YY)

Case classification Local Imported/Non-local
 If local, please state in which region/district transmission most likely occurred: _____

PREVENTION MEASURES

Has your home been sprayed in the past 12 months? Yes No
 Do you own a bed net? Yes No
 Did you sleep under a bed net each of the last 3 nights? Yes No

Nationality of patient: Namibia Angola Zambia Botswana DRC Zimbabwe Other _____

FIGURE B.1: Malaria active case detection questionnaire

APPENDIX C

Additional material

C.1 Maps for the spatio-temporal model without added covariates effects

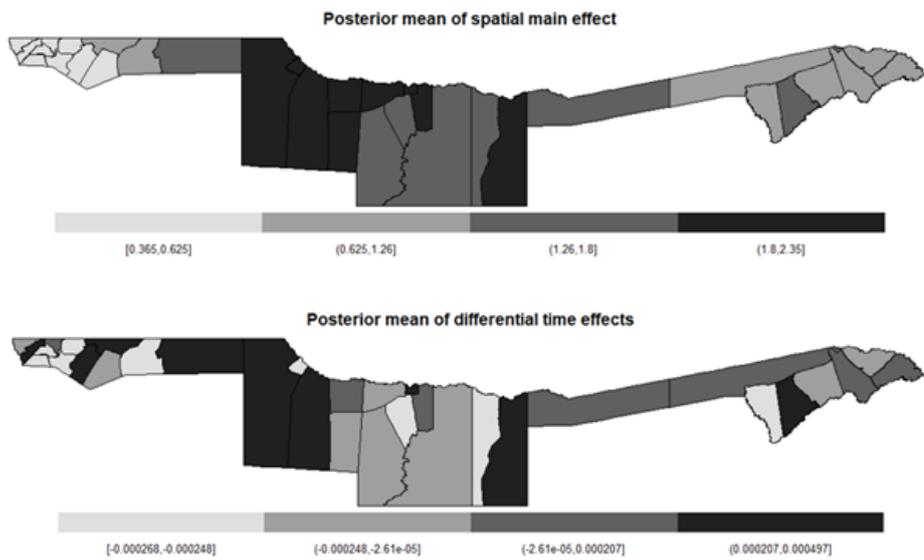


FIGURE C.1: Spatial main effect and differential temporal maps of the spatiotemporal model without added covariates (posterior mean obtained using a random walk of order 2)

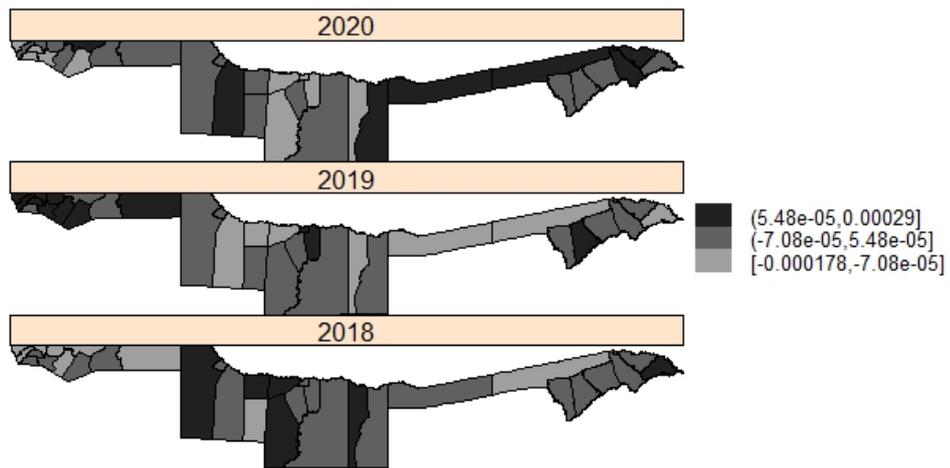


FIGURE C.2: Posterior mean for malaria of the nonspatially or temporal interaction model without added covariates using RW2

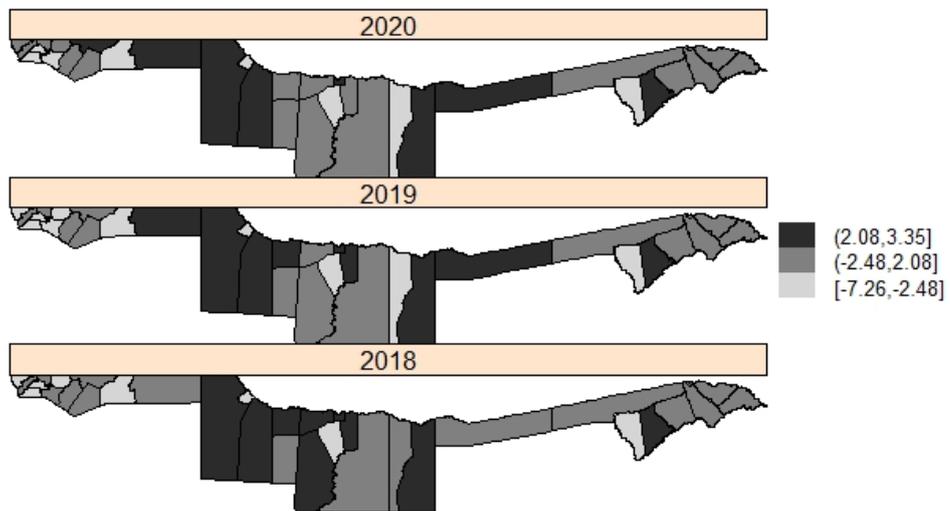


FIGURE C.3: Posterior mean for malaria of the temporal structured interaction model without added covariates using random walk of order 2

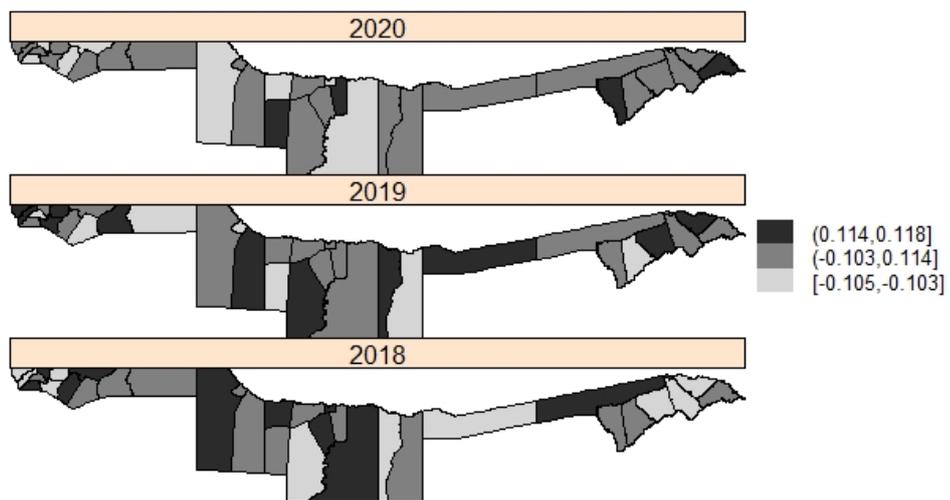


FIGURE C.4: Posterior mean for malaria of the spatial structured interaction model without added covariates using random walk of order 2