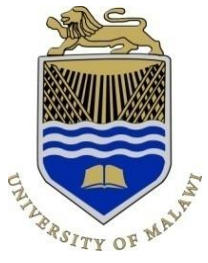
MODELLING IMPACT OF INDOOR RESIDUE SPRAY (IRS) MALARIA INTERVENTION WHILE ACCOUNTING FOR LAGGED AND NON-LINEAR EFFECTS OF CLIMATE IN MANGOCHI, MALAWI

MASTER OF SCIENCE (BIOSTATISTICS)

GODFREY SILUNGWE

UNIVERSITY OF MALAWI

JULY 2023



MODELLING IMPACT OF INDOOR RESIDUE SPRAY (IRS) MALARIA INTERVENTION WHILE ACCOUNTING FOR LAGGED AND NON-LINEAR EFFECTS OF CLIMATE IN MANGOCHI, MALAWI.

MSc. (BIOSTATISTICS) THESIS

BY

GODFREY SILUNGWE

BSc. (Mathematical Sciences Education – Statistics and Computing) – Malawi University of Business and Applied Sciences (MUBAS)

Submitted to the Department of Mathematical Sciences, Faculty of Science, in partial fulfillment of the requirement for the Degree of Master of Science in Biostatistics

University of Malawi

September 2023

DECLARATION

I the undersigned hereby declare that this thesis/dissertation is my own original work which has not been submitted to any other institution for similar purposes. Where other people‘s work has been used acknowledgements have been made.

GODFREY EMMANUEL SILUNGWE

Full Legal Name

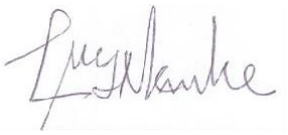
Signature

31st August 2023

Date

CERTIFICATE OF APPROVAL

The undersigned certify that this thesis represents the student‘s own work and effort and has been submitted with my approval in partial fulfillment of a Master of Science in Biostatistics at the University of Malawi.



Signature: _____ Date: 31/08/2023

Lawrence N Kazembe, PhD (Professor of Applied Statistics)

Main Supervisor

DEDICATION

I dedicate my dissertation work to my family and many friends. A special feeling of gratitude to my loving parents, Zakaliya and Monica whose words of encouragement and push for tenacity ring in my ears. My wife Beatrice and my daughters Monica and Sandra have never left my side. I also dedicate this dissertation to my many friends who have supported me throughout the process. You have been my best.

ACKNOWLEDGEMENTS

This paper and the research behind it would not have been possible without the exceptional support of my supervisor, Prof. Lawrence Kazembe. His enthusiasm, knowledge and exacting attention to detail have been an inspiration and kept my work on track. His generosity and expertise have improved this study in innumerable ways and saved me from many errors; those that inevitably remain are entirely my own responsibility.

ABSTRACT

The relationship between climate and malaria incidence is complex, with climate change influencing the epidemiological patterns of malaria in different regions. Understanding the nature of this relationship is essential for effective malaria programming, especially in the context of changing climate conditions. Modeling the relationship between climate factors and malaria incidence can help guide malaria control interventions based on current and projected climate data. It also serves as a basis for establishing malaria early warning systems to aid resource planning. This study aimed to investigate the relationship between climate factors, account for delayed climate effects, and evaluate the impact of indoor residue spray (IRS) intervention on malaria incidence. Monthly malaria incidence data were obtained from the national malaria control program, while climate data were collected from the Department of Meteorological Services and Climate Change in Mangochi district, Malawi. Two methods were employed: a distributed lag non-linear model to examine the nature of the relationship between climate variables and malaria incidence, and segmented regression to assess the impact of the IRS intervention while accounting for lagged climate effects and seasonal trends. The results revealed an immediate peak in malaria risk following extreme weather conditions, highlighting the importance of short-term effects of climate. The risk of malaria immediately doubles with extreme rains and humidity compared to average weather conditions. Notably, an immediate peak in malaria incidence was observed following exposure to all climatic factors, and the effects continued to manifest for up to three months (0-to-2-month lag). This suggests that previous climate conditions play a critical role in predicting current and future malaria incidence. In conclusion, the findings highlight the importance of short lags and the potential for immediate outbreaks following exposure to climatic factors. Incorporating these findings into malaria programming and control efforts can enhance the effectiveness of interventions and contribute to the development of proactive strategies to reduce the burden of malaria in the context of a changing climate.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC - Akaike Information Criterion

ASAQ - Artesunate-Amodiaquine

BIC - Bayesian Information Criterion

C - Celsius

DLNM – Distributed Lag Non-Linear Model

Gwanda - A region in Zimbabwe

HMIS - Health Management Information System

IPTp - Intermittent Prevention and Treatment in Pregnancy

IRS - Indoor Residual Spraying

LA - Lumefantrine-Artemether

MEWS - Malaria Early Warning System

P. - Plasmodium

spp. - Species

CHAPTER 1

INTRODUCTION

1.1. Introduction

This chapter provides a concise overview of the malaria burden in Malawi, including an examination of the historical trend. It also summarizes the national strategic plan and highlights the key interventions implemented to mitigate malaria incidence in the country. The chapter delves into a comprehensive explanation of the malaria transmission mechanism and explores the role of climate in shaping the dynamics of malaria transmission. Lastly, it articulates the problem statement, outlines the research objectives, and underscores the significance of this study in the context of malaria control efforts.

1.2. Background

In Malawi, malaria is endemic and all population is at risk and therefore affects large number of people. It continues to be a major public health problem, accounting for 20% of all outpatient visits in all age groups (HMIS, 2020). Malawi accounts for 2% of malaria cases worldwide and is among the top 15 countries with a high malaria burden (Chilanga, Delphine, Heather, & Claudia, 2020). In 2020 alone the country registered 6.9 million cases both confirmed and presumed (99.6% and 0.3% respectively) reported from health facilities and community case management program. Malawi has seen no major changes in malaria trend between 2014 and 2020 from 397 per 1000 population in 2014 to 385 per 1000 population in 2020.

Malaria burden is high among under-5 populations who are prone to severe malaria infection because they lack acquired immunity (Malawi Government, 2017). According to malaria indicator survey 2017, prevalence of malaria among under-5 populations has slightly declined from 33% in 2014 to 27% in 2017, the last malaria indicator survey was conducted in 2017. Vulnerable populations to malaria also include pregnant women and those living in areas that are prone to natural disasters, including floods and earth tremors. Populations living in hard-to-reach areas -

defined as more than 5km from the nearest health facility or limited to health services by geographical barriers - are also vulnerable (Malawi Government, 2017).

Malaria continues to be public health burden and has caused 15% of all deaths in public health facilities in the year 2020 (HMIS, 2020). According to health management information system (HMIS), malaria mortality rate has declined from 28 per 100,000 populations in 2014 to 13 per 100,000 populations in 2020.

Transmission is perennial in most areas and peaks after the start of the rainy season which begins in November/December, lasting through March/April. Malaria transmission intensity and risk of infection varies across the country and is highest in areas with high temperatures, rainfall and humidity, particularly along the low-lying lakeshore and Shire river valley areas and is lower, along the highland areas (Government of Malawi, 2020).

The Ministry of Health in Malawi, through national malaria control programme, aims at eliminating malaria by 2030. The overall goal of malaria strategic plan 2017 – 2022 was to reduce malaria incidence from 386 per 1000 population in 2015 to 193 per 1000 population by 2022 and malaria deaths by at least 50% of 2015 levels by 2022, (Government of Malawi, 2020). Key interventions include: Indoor Residual Spraying (IRS), Long Lasting Insecticide Treated Nets mass distribution campaign and malaria vaccination targeting specific population periodically. Other interventions implemented routinely are net distribution targeting pregnant women and newborn babies and intermittent prevention and treatment in pregnancy (IPTp) administration to pregnant women. Home remedies to control malaria are also widely practiced at individual level such as use of mosquito repellants and small-scale indoor residue spray (IRS). Lumefantrine-Artemether (LA) is the first-line treatment for uncomplicated malaria while Artesunate-Amodiaquine (ASAQ) as the second-line treatment for uncomplicated malaria (Government of Malawi, 2020).

1.3. Malaria Transmission Mechanism and Climate

Malaria burden is a result of interaction among three determinants namely host (age, sex and immunity), environment (climate and altitude) and parasite/agent (antigenicity, strain, resistance and behaviour) (Government of Malawi, 2020). The host component is shaped by several factors including genetic and acquired immunity, behavior, demographics, culture, socioeconomic

characteristics, and politics (Marcia & Castro, 2017). In highly malaria endemic areas like Malawi, children are mostly non-immune compared to adults which influence transmission intensity of malaria in children (Kazembe, Kleinschmidt, Holtz, & Sharp, 2006). The environmental component depends both on the natural environment—temperature, humidity, rainfall, soil quality, elevation/slope, land cover, hydrography, presence of natural enemies of mosquitoes and larvae, and natural disasters— and the human-made environment—land use, land change, deforestation, housing type, infrastructure (water, sanitation, and waste collection), urbanization, development projects (e.g., roads, railways, dams, irrigation, mining, resettlement projects, and oil pipelines), and disasters facilitated by human-made changes. The vector is shaped by the type of Anopheles species and associated feeding, resting, biting, and breeding behavior, flight range, vectorial capacity, mortality and reproduction rate, mosquito resistance to insecticides, and larval resistance to larvicides (Marcia & Castro, 2017).

Malaria is a parasitic disease caused by *Plasmodium species (spp.),* unicellular protozoan organisms in the phylum of *Apicomplexa* (Xin, Cui, & Deirdre, 2020). The species that infect humans include *Plasmodium falciparum, P. vivax, P. malariae, P. knowlesi,* and *P. ovale,* with *P. ovale* recently recognized as two subspecies called *Plasmodium ovale curtisi* (classic type) and *Plasmodium ovale wallikeri* (Sutherland, 2010). Whereas, *P. vivax* is the most widespread species, *P. falciparum* is the deadliest to humans.

It is often referred to as a climate-dependent disease, which is primarily because certain climatic conditions are needed for the complete maturation of sporozoites in mosquitoes (Varvara, Natalia, Mikhail, & Mikhail, 2019). Varvara et al (2019) reported that *Plasmodium vivax* requires lower temperatures for its development in the vector than other human malaria species. Climate conditions affects the bionomics of Anopheles mosquitoes, such as the speed of development of the aquatic stages (which depends on the temperature of the place of breeding), the speed of blood digestion (which depends on the temperature of the resting place), and their survival in general.

Climatic changes in the past have greatly affected the distribution of malaria and likely modified malaria geography (Phillippe & Myriam, 1995). Phillippe et al (1995) highlighted that making predictions regarding the geographical extent and intensity of malaria is difficult and the relationship between malaria and climate is complex. According to Phillippe et al (1995), temperature affects the survival of the parasite only during its life-cycle in the Anopheles vector

and modifies the vectorial capacity of the Anopheles. Optimal values of temperature, ranging between 22 and 300C, lengthen the life-span of the mosquitos and increase the frequency of blood meals taken by the females, to up to one meal every 48 hours. Higher temperatures also shorten the aquatic life cycle of the mosquitos from 20 to 7 days and reduce the time between emergence and oviposition, as well as the time between successive ovipositions.

Phillippe et al (1995) also indicated that rainfall generally means new opportunistic breeding places. Nonetheless, rainfall can also destroy existing breeding places; heavy rains can change breeding pools into streams, impede the development of mosquito eggs or larvae, or simply flush the eggs or larvae out of the pools. Conversely, exceptional drought conditions can turn streams into pools. The appearance of such opportunistic mosquito breeding sites sometimes precedes epidemics. The interaction between rainfall, evaporation, runoff, and temperature modulates the ambient-air humidity, which in turn affects the survival and activity of Anopheles mosquitos. To survive, they need at least 50% or 60% relative humidity. Higher levels lengthen the life-span of the mosquitos and enable them to infect more people.

Although Phillippe et al (1995) described a range of favorable conditions, the actual relation is not the same due to mediating/moderating factors, as such the relation of climate and malaria may vary from place to place (Phillippe & Myriam, 1995). As also observed by Chuang et al (2017), different administrative regions with varying vulnerability to climate show varying effects prompting the need to localize investigation to specific places with varying climate vulnerability (Chuang & Ting, 2017). Lisbeth et al (2017), in a study conducted in Guna Yala also called for further studies about weather impacts on malaria vector ecology, as well as the association of malaria vectors while paying attention to different socio-economic conditions such as poverty and cultural differences (Lisbeth, Jose, Chystrie, & Milagros, 2017).

Malawi's climate is subtropical with the three distinct seasons: rainy season extending from November to April, and the dry season from May to mid-August with temperatures at night reaching as low as 10-14o C and the hot season from between mid-August and November (The INFORM Project, 2014). Generally, the highlands are cooler and wetter while the low-lying regions are hotter and more humid.

1.4. Problem Statement

Malaria is indirectly associated with natural risk factors such as rainfall pattern, temperature and humidity which influence spread and transmission of malaria parasite (Ayansina, Isioma, Consolato, & Oluwatoyin, 2020). High rainfall increases mosquito breeding sites and therefore increasing transmission of malaria parasites among individuals. In the same way other risk factors; temperature and humidity also affect malaria transmission by regulating the rate of development of the mosquito larvae which influences mosquito survival rates (Gunda, Chimbar, Shamu, Sartorius, & Mukaratin, 2017). The link between climate variability and vector-borne diseases has also been established in a study by Gunda et al (2017) investigating association between malaria incidence and climate variables in rural Gwanda in Zimbabwe. Climate variability has the potential to either work for or against efforts to control the disease. (Gunda, Chimbar, Shamu, Sartorius, & Mukaratin, 2017).

An understanding of how malaria incidences vary as a result of climate variability (present and recent past) is important for planning for future malaria control programmes (Gunda, Chimbar, Shamu, Sartorius, & Mukaratin, 2017). It allows for the identification of the most suitable timing for implementing malaria interventions, considering the impact of past climate conditions. This understanding can serve as a guiding tool for program implementers, enabling them to incorporate past climate experiences into their decision-making process and enhance the effectiveness of malaria control initiatives. Also, exploring the connection between malaria incidence and lagged climatic conditions has the potential to contribute significantly to the development of malaria early warning systems.

Investigating the impact of past climate experiences on malaria transmission is also crucial for understanding local epidemiological shifts, some of which can be attributed to climate change. Future climate projections indicate a general warming trend, particularly in southern Malawi and over the lake (Vincent & Katharine, 2020). However, different climate models simulate diverse patterns of rainfall, leading to varying predictions. By 2030, the anticipated changes in annual mean rainfall range from a modest -8% decrease to a +20% increase. These changes become more pronounced by the 2070s, with projected ranges of -17% decrease to +27% increase. Conversely, there is a strong consensus among climate models regarding rising temperatures in Malawi.

Projections indicate a temperature increase of 0.5 to 1.5°C by the 2040s and a more substantial warming of 4 to 4.3°C by the year 2090 (Vincent & Katharine, 2020).

The changing climatic conditions highlight the importance of considering the complex interactions between climate factors and malaria transmission dynamics. Understanding the local climate patterns and their potential impact on malaria is also crucial for developing effective strategies to mitigate the disease and adapt to future climate challenges.

In addition to the aforementioned points, it is worth emphasizing that the statistical methods commonly employed in previous studies often assume a linear relationship between climate variables and malaria incidence. However, it has been reported in numerous studies that non-linear relationships exist in this context. Therefore, this study addressed this limitation by applying methods which captures non-linear relationship more accurately.

Furthermore, this research incorporated methods to model the delayed effects of climate on malaria transmission. Understanding the time lag between climate factors and their impact on malaria incidence is essential to comprehensively assess the relationship between the two variables. By applying appropriate methodologies to account for delayed climate effects, the study provides a more comprehensive and accurate analysis of the influence of climate on malaria transmission in Mangochi district.

Lastly, the study assessed the impact of indoor residue spray (IRS) intervention while accounting for the lagged and non-linear effects of climate in Mangochi, Malawi. Mangochi district is one of the areas burdened by high malaria cases, as highlighted in the Malawi Malaria Strategic Plan 2017-2022. Therefore, it serves as an ideal location to examine the interrelationship between climate change and malaria.

1.5. Research Objectives

The primary objective of this study was to model the impact of IRS malaria intervention while accounting for lagged and non-linear effects of climate in Mangochi, Malawi.

1.5.1. Specific Objectives

1.5.1.1. To model lagged and nonlinear effects of climate factors on malaria in Mangochi, Malawi.

1.5.1.2. To evaluate impact of indoor residue spray (IRS) malaria intervention while accounting for lagged effects of climate factors in Mangochi, Malawi.

1.6. Significance of the study

According to the Malawi Malaria Strategic Plan 2017-2022, the country aims to significantly reduce malaria incidence and deaths. The goal is to reduce malaria incidence by 50% from a baseline of 386 cases per 1000 population in 2016 to 193 cases per 1000 population, and reduce malaria deaths by 50% from 23 deaths per 100,000 population to 12 deaths per 100,000 population by 2022. To achieve this, Malawi aims to have at least 90% of the population utilizing one or more malaria preventative interventions. The national malaria control program will prioritize the implementation of quality indoor residual spraying (IRS) in selected epidemiological areas, guided by international/WHO standards and local climate trends. The success of IRS depends on the proper timing of implementation, aligned with peak malaria periods. This study will help forecast malaria peak periods based on past climate experiences, providing guidance for selecting the optimal implementation period for IRS to maximize its impact in reducing malaria transmission

The study is also justified by the need to develop action plans for malaria epidemic prevention and response in Malawi, particularly considering the increasing malaria incidence due to climate change (Government of Malawi, 2020). A malaria early warning system (MEWS) is seen as a promising tool to reduce the burden of malaria by accounting for the complex malaria-climate dynamics (Yoonhee et al., 2019). This study aims to contribute to the field by modeling the lagged effects of climate on malaria, which can be used to forecast future malaria epidemics based on past climate experiences. Enhancing the MEWS by incorporating climate-related epidemics is crucial for preventing malaria-related deaths during epidemics

This study aligns with Sustainable Development Goal (SDG) number 3, which aims to achieve good health and well-being for all. Target 3.B under this goal specifically focuses on improving early warning systems for global health risks. By developing an improved early warning system for malaria epidemics in Malawi, based on past climate experiences, this study contributes to the achievement of SDG 3.0. The enhanced early warning system will support the program's response to future malaria epidemics by utilizing climate data, ultimately helping to prevent avoidable deaths during disease outbreaks.

1.7. Organization of Thesis

The study is structured in a systematic manner to address the research objectives. Chapter 1 serves as an introduction, providing an overview of the study objectives. The subsequent chapters, namely Chapter 2 and Chapter 3, focus on specific aspects/objectives of the investigation. Each chapter includes a literature review, theoretical framework, materials and methods, results and discussion, and a conclusion specific to each objective. Theoretical framework outlined in chapter 2 also applies to chapter 3.

In Chapter 2, the primary focus is to model the lagged and nonlinear effects of climate factors on malaria in Mangochi, Malawi. This chapter delves into the intricate relationship between climate variables and malaria transmission dynamics, considering both the time delay and non-linear associations. By incorporating these factors into the modeling process, a more comprehensive understanding of the influence of climate on malaria incidence is achieved.

Chapter 3 takes the analysis further by incorporating lagged climate conditions when evaluating the impact of interventions on malaria incidence. By accounting for the time lag between climate factors and their effect on malaria transmission, this chapter provides a more accurate assessment of the benefits derived from interventions. The inclusion of lagged climate effects in the evaluation process enhances our understanding of the true impact of interventions and their effectiveness in reducing malaria cases in the specific context of Mangochi, Malawi.

Finally, Chapter 4 serves as the concluding chapter, providing an overall summary and conclusion that encompasses all the objectives examined throughout the thesis. This chapter synthesizes the findings from the preceding chapters, highlighting the key insights gained from the investigation and their implications for malaria control strategies.

CHAPTER 2

MODELLING LAGGED AND NONLINEAR EFFECTS OF CLIMATE FACTORS ON MALARIA IN MANGOCHI, MALAWI

2.1. Introduction

There is well documented evidence that malaria is influenced by climate factors which impact on vector dynamics hence influencing spread and transmission of malaria parasite among individuals (Lisbeth, Jose, Chystrie, & Milagros, 2017). In Malawi, malaria transmission is highest during rainy season (November to April) when there is also an increase in malaria vector breeding sites (Government of Malawi, 2020). Low lying areas have hot temperatures which is more favorable for mosquito breeding hence transmission is also highest in these areas. However due to climate change there is an observation that even highlands are experiencing rise of malaria cases attributed to increasing temperatures in highland areas that also favors malaria vector breeding.

Phillippe et al (1995) indicated that potential transmission of malaria is controlled by climatic factors such as temperature, humidity, and rainfall, which regulate the biology of development of both mosquito and parasite (Phillippe & Myriam, 1995). Ogden et al (2017) also emphasized the global concern of impacts of climate change on the 'big two diseases': malaria and dengue, which have now perhaps become the 'big five' of malaria, dengue, yellow fever, chikungunya and Zika (Ogden, 2017). According to the paper, the big five diseases are intrinsically sensitive to weather and climate (Ogden, 2017).

Many studies have established presence of non-linear relationship between climate variables and malaria risk in other countries (Gunda, Chimbar, Shamu, Sartorius, & Mukaratin, 2017). Gunda et al (2017) identified that administrative regions with varying vulnerability to climate show varying effects prompting the need to localize investigation of specific places with varying climate vulnerability. Lisbeth et al (2017) called for further studies about weather impacts on malaria vector ecology, as well as the association of malaria vectors with Gunas paying attention to their socio-economic conditions of poverty and cultural differences as an ethnic minority (Lisbeth, Jose,

Chystrie, & Milagros, 2017). The primary focus in this chapter is to model the lagged and nonlinear effects of climate factors on malaria in Mangochi, Malawi.

2.2. Literature review

Malaria, according to the World Health Organization, is one of the most serious and complex health problems facing humanity in the 20th century. In the past, climatic changes have greatly affected its geography (Phillippe & Myriam, 1995). Its seriousness and complexity are therefore likely to be compounded by an anthropogenic greenhouse effect. Phillippe et al (1995) indicated that the intensity and the extent of malaria potential transmission significantly change under the climate change scenarios generated by five atmospheric general circulation models. All five simulations revealed an increase in seasonal malaria at the expense of perennial malaria which is cause for great concern. Indeed, seasonal malaria is most likely to lead to epidemics among unprepared or nonimmune populations. Moreover, climate change may trigger massive migrations of environmental refugees. Such population movements would likely put national and international health infrastructures under severe stress. Today, malaria is a developing country issue but could spread to higher latitudes. The results obtained with Malaria Potential Occurrence Zone (MOZ) model suggest that malaria could become a public health problem for developed countries within decades.

There has been much debate as to whether or not climate change will have, or has had, any significant effect on risk from vector-borne diseases (Ogden, 2017). The debate on the former has focused on the degree to which occurrence and levels of risk of vector-borne diseases are determined by climate-dependent or independent factors, while the debate on the latter has focused on whether changes in disease incidence are due to climate at all, and/or are attributable to recent climate change. Ogden et al (2017) reviewed possible effects of climate change on vector-borne diseases, methods used to predict these effects and the evidence to date of changes in vector-borne disease risks that can be attributed to recent climate change. The findings are that predictions have both over- and underestimated the effects of climate change. Mostly under-estimations of effects are due to a focus only on direct effects of climate on disease ecology while more distal effects on society's capacity to control and prevent vector-borne disease are ignored. It also established increasing evidence for possible impacts of recent climate change on some vector-borne diseases

but for the most part, observed data series are too short (or non-existent), and impacts of climate-independent factors too great, to confidently attribute changing risk to climate change.

Risk assessment regarding the distribution of malaria vectors and environmental variables underpinning their distribution under changing climates is crucial towards malaria control and eradication (Godwin, Kayode, & Olakunle, 2019). Godwin et al (2019) estimated the potential future distribution of major transmitters of malaria in Nigeria—Anopheles gambiae sensu lato and its siblings: Anopheles gambie sensu stricto, and Anopheles arabiensis under low and high emissions scenarios. The study established higher magnitude of change in species prevalence predicted for the later part of the 21st century under high emission scenario, driven mainly by increasing and fluctuating temperature, alongside longer seasonal tropical rainfall accompanied by drier phases and inherent influence of rapid land use change, may lead to more significant increase in malaria burden when compared with other periods and scenarios during the century; especially in Humid forest, Derived savanna, Sahel and Sudan savannas.

Cyril et al (2018) also indicated that climate change is one of the greatest threats to human health in the 21st century (Cyril, Marie, & Annie, 2018). This is because it directly impacts on health through climatic extremes, air quality, sea-level rise, and multifaceted influences on food production systems and water resources. It also affects infectious diseases, which have played a significant role in human history, impacting the rise and fall of civilizations and facilitating the conquest of new territories. The paper by Cyril et al (2018) highlighted significant regional changes in vector and pathogen distribution reported in temperate, peri-Arctic, Arctic, and tropical highland regions during recent decades, changes that have been anticipated by scientists worldwide. The review established that further future changes are likely if we fail to mitigate and adapt to climate change. Many key factors affect the spread and severity of human diseases, including mobility of people, animals, and goods; control measures in place; availability of effective drugs; quality of public health services; human behavior; and political stability and conflicts. With drug and insecticide resistance on the rise, significant funding and research efforts must be maintained to continue the battle against existing and emerging diseases, particularly those that are vector borne

Life cycles of malaria mosquitoes and parasites are strongly affected by climate factors such as temperature and precipitation (Jung Eun Kim, 2018). The paper by Jung et al (2018) indicates that

optimal temperature for malaria transmission is around 25°C, which suggests that malaria transmission may occur predominantly between summer and early fall in Korea (Jung Eun Kim, 2018). In areas where climate has been gradually shifting may increase the risk of massive malaria outbreaks. Jung et al (2018) further emphasized the importance of investigating potential effect of climate change on P. vivax malaria transmission through a modeling study.

Rainfall and temperature are considered the main weather factors that highly determine malaria epidemics (Yoonhee, Ratnam, Takeshi, Yushi, & Swadhin, 2019). Yoonhee et al (2019) noted that high rainfall increases the number of breeding sites for mosquitoes and leads to increases in malaria transmission. He also acknowledged that some studies, however, have reported that intense rainfall could flush early-stage larvae and shrink mosquito populations in the short term. The study further reported that high temperatures increase the chance of transmission by shortening the duration of parasite growth in mosquitoes. Temperature changes also influence the development, reproduction, survival, and biting rate of mosquitoes (Yoonhee, Ratnam, Takeshi, Yushi, & Swadhin, 2019).

Soma et al (2019) reported major reasons for the persistence of malaria is the extensive geographic and climatic diversity of the country, which supports ideal ecological conditions for sustaining the parasites and their vectors (Soma, Vinay, Poonam, & Ramesh, 2019). The major climatic determinants of malaria are temperature, rainfall and humidity. The paper further noted that impact of climate change is not uniform around the globe: Some places may become warmer and drier, while others warmer and wetter. Hence, the threat of climate change is expected to have a profound effect on the mosquito's longevity, development of malaria parasites in the vectors, and consequently opening the windows of malaria transmission particularly in areas which are free due to temperature constrains. In other words, global climate change is likely to alter the spatial and temporal distribution of malaria. It further states that climate change will increase the opportunities for malaria transmission in traditionally non-malarious areas, and make it difficult to control in traditionally malarious areas due to an alteration in their growth cycle and transmission seasons. The study reported that although future repercussions of climate change on malaria transmission at the global level have already been explored; however, such evidences are limited in other areas.

A paper by Ayansina et al (2020) also mentioned the role of climatic variability and seasonality as significant in the spatiotemporal distribution of diseases (Ayansina, Isioma, Consolato, &

Oluwatoyin, 2020). The paper showed that the occurrence and spatial distribution of malaria are sensitive to the seasonality of climatic factors in most African countries, and other parts of the world with significant perinatal morbidity and mortality. This is because disease vectors also depend on suitable habitats to breed, which in turn depends heavily on climatic conditions and for understanding the nature of some illness. The malaria and meningitis (MM) transmission is highly seasonal due to climatic conditions; these occurrences are much more frequent in recent times due to climate change. The paper further states that climate and health are indistinctly interconnected, and this is the same for infectious diseases. Climate change is likely to increase malaria and meningitis incidence as the future environment might become more suitable for malaria transmission in many tropical highlands. It also reported a corresponding 0.90% increase in the number of malaria cases to each 1 °C temperature increase. Hertig et al (2019) also found that the occurrence of vector competent Anopheles species and favorable climatic conditions autochthonous malaria cases may re-emerge in countries where malaria was previously eradicated (Hertig, 2019).

Although malaria is one of the greatest historical killers of mankind, its range is limited by climate to the warmer regions of the globe (Steffen & Abba, 2018). The paper by Steffen et al (2018) further mentioned that anthropogenic global warming (and climate change more broadly) now threatens to alter the geographic area for potential malaria transmission, as both the Plasmodium malaria parasite and Anopheles mosquito vector have highly temperature-dependent lifecycles, while the aquatic immature Anopheles habitats are also strongly dependent upon rainfall and local hydrodynamics. A wide variety of process-based (or mechanistic) mathematical models have thus been proposed for the complex, highly nonlinear weather-driven Anopheles lifecycle and malaria transmission dynamics, but have reached somewhat disparate conclusions as to optimum temperatures for transmission, and the possible effect of increasing temperatures upon (potential) malaria distribution, with some projecting a large increase in the area at risk for malaria, but others predicting primarily a shift in the disease's geographic range. The paper also indicated that both global and local environmental changes drove the initial emergence of P. falciparum as a major human pathogen in tropical Africa some 10,000 years ago, and the disease has a long and deep history through the present.

Florence et al (2019) also highlighted that climate variables that directly influence vector-borne diseases' ecosystems are mainly temperature and rainfall (Florence & John, 2019). This is not only because the vectors bionomics are strongly dependent upon these variables, but also because most of the elements of the systems are impacted, such as the host behavior and development and the pathogen amplification. The paper also established that impact of the climate change on the transmission patterns of these diseases is not easily understood, since many confounding factors are acting together. Consequently, knowledge of these impacts is often based on hypothesis derived from mathematical models. Nevertheless, some direct evidences can be found for several vector-borne diseases. Evidences of the impact of climate change are available for malaria, arbovirus diseases such as dengue, and many other parasitic and viral diseases such as Rift Valley Fever, Japanese encephalitis, human African trypanosomiasis and leishmaniasis. The effect of temperature and rainfall change as well as extreme events, were found to be the main cause for outbreaks and are alarming the global community. Among the main driving factors, climate strongly influences the geographical distribution of insect vectors, which is rapidly changing due to climate change. Further, in both models and direct evidences, climate change is seen to be affecting vector-borne diseases more strikingly in fringe of different climatic areas often in the border of transmission zones, which were once free of these diseases with human populations less immune and more receptive. The impact of climate change is also more devastating because of the unpreparedness of Public Health systems to provide adequate response to the events, even when climatic warning is available. Although evidences are strong at the regional and local levels, the studies on impact of climate change on vector-borne diseases and health are producing contradictory results at the global level.

Yen et al (2020) explained that changes in the Earth's climate and weather continue to impact the planet's ecosystems, including the interface of infectious disease agents with their hosts and vectors (Yeh, Fair, Smith, & Torres, 2020). Environmental disasters, natural and human-made activities raise risk factors that indirectly facilitate infectious disease outbreaks. Subsequently, changes in habitat, displaced populations, and environmental stresses that affect the survival of species are amplified over time. The recurrence and spread of vector-borne (e.g., mosquito, tick, aphid) human, animal, and plant pathogens to new geographic locations are also influenced by climate change. The distribution and range of humans, agricultural animals and plants, wildlife and native plants, as well as vectors, parasites, and microbes that cause neglected diseases of the

tropics as well as other global regions are also impacted. In addition, genomic sequencing can now be applied to detect signatures of infectious pathogens as they move into new regions. Molecular detection assays complement metagenomic sequencing to help us understand the microbial community found within the microbiomes of hosts and vectors, and help us uncover mechanistic relationships between climate variability and pathogen transmission. The understanding of, and responses to, such complex dynamics and their impacts can be enhanced through effective, multi-sectoral One Health engagement coupled with applications of both traditional and novel technologies. Concerted efforts are needed to further harness and leverage technology that can identify and track these impacts of climate changes in order to mitigate and adapt to their effects.

In a study conducted by Lisbeth et al (2017) showed that EL Niño Southern Oscillation (ENSO), rainfall and Normalized Difference Vegetation Index (NDVI) were associated with the number of malaria cases in Guna Yala (Lisbeth, Jose, Chystrie, & Milagros, 2017). The study established high vulnerability of Guna populations to malaria and also that malaria infection is sensitive to climate change. They further called for further studies about weather impacts on malaria vector ecology, as well as the association of malaria vectors with Gunas paying attention to their socio-economic conditions of poverty and cultural differences as an ethnic minority.

Chuang et al (2017) in Swaziland investigated effects of climate to malaria in four administrative regions Lubombo, Hhohho, Manzini and Shiselweni (Chuang & Ting, 2017). This study indicated that climate conditions were more important in the Hhohho and Lubombo administrative regions, implying that residents in these areas are at higher risk of infection when temperatures and precipitation are suitable for malaria transmission. This clearly shows that places with different vulnerability to climate respond differently hence the need to localize and investigate area specific estimates of climate factors.

Non-linear relationship between climate factors and malaria incidence has been found in some studies. In Zimbabwe, Gunda et al (2017) investigated Malaria incidence trends and their association with climatic variables in rural Gwanda (Gunda, Chimbar, Shamu, Sartorius, & Mukaratin, 2017). The results showed significant association between malaria incidence and the climatic variables in Buvuma and Selonga wards at specific lag periods. In Ntalale ward, only precipitation (1- and 3-month lag) and mean temperature (1- and 2-month lag) were significantly

associated with incidence at specific lag periods. DLNM results suggest a key risk period in current month, based on key climatic conditions in the 1–4-month period prior.

In a paper conducted in republic of Korea, Jae et al (2012) suggested that malaria transmission in temperate areas is highly dependent on climate factors (Jae, Hae, & Young, 2012). In addition, lagged estimates of the effect of rainfall on malaria are consistent with the time necessary for mosquito development and P. vivax incubation.

Yoonhee et al (2019) proposed methods for malaria forecasting based on different modeling approaches such as statistical modeling (e.g., generalized linear model (GLM) and autoregressive integrated moving average (ARIMA) time series model), mathematical modeling (e.g., susceptible-exposed-infected-recovered (SEIR) model), and machine learning methods (e.g., neural network) (Yoonhee, Ratnam, Takeshi, Yushi, & Swadhin, 2019). However, the study acknowledged that no one method has been a gold standard because each method has different modeling assumptions and the optimal choice of the method depends upon the characteristics of a study population. Yoonhee et al (2019) applied a flexible statistical modeling approach, a GLM with a distributed lag nonlinear structure, to understand the complexity of nonlinear and delayed malaria-weather associations and develop a weather-based malaria prediction model accordingly. In this chapter, GLM with a distributed lag nonlinear structure is applied to model the nature of relationship between climate variables and malaria incidence in Mangochi, Malawi.

2.3. Conceptual framework

2.3.1. Generalized Linear Model

Malaria has been modeled using various statistical methods in the literature and significant malaria predictors (particularly climatic factors) have been identified in a variety of settings (Mukhopadhyay, Tiwari, Shetty, Gogtay, & Thatte, 2019). The statistical methods include linear regression and generalized linear regression with trend, seasonal parameters and weather covariates. The conditional distribution of malaria counts given the past, is assumed to follow a distribution from the exponential family (usually taken to be a Poisson/negative binomial distribution). Due to the presence of high dispersion in the data, a negative binomial distribution gives a better fit. This section reviews generalized linear modeling for time series data.

Generalized linear models (GLM) extend the concept of the well understood linear regression model. The linear model assumes that the conditional expectation of Y (the dependent or response variable) is equal to a linear combination $X^T\beta$ which could also be written as Y=$X^T\beta$+ε. Unfortunately, the restriction to linearity cannot take into account a variety of practical situations. For example, a continuous distribution of the error term implies that the response Y must have a continuous distribution as well. Hence, the linear regression model may fail when dealing with binary Y or with counts data.

GLM methodology is a specific class of nonlinear models for a general approach to nonlinear regression which assumes that the distribution of Y is a member of the exponential family. The exponential family covers a large number of distributions for example discrete distributions as the Bernoulli, binomial and Poisson which can handle binary and count data or continuous distributions as the normal, Gamma or Inverse Gaussian distribution.

A distribution is a member of the exponential family if its probability mass function (if Y discrete) or its density function (if Y continuous) has the following form.

$$f(y,\theta,\emptyset) = \exp\left(\frac{\mathrm{y}\theta - \mathrm{b}(\theta)}{\mathrm{a}(\emptyset)}\right) + \mathrm{c}(\mathrm{y},\emptyset)$$

Equation 1

The functions a(•), b(•) and c(•) varies for different Y distributions and the parameter of interest is θ, which is also called the canonical parameter (McCullagh and Nelder,1989). The additional parameter ∅, is only relevant for some of the distributions, and is considered as nuisance parameter.

2.3.2. Structure of GLM

A generalized linear model (or GLM) consists of three components:

A random component, specifying the conditional distribution of the response variable, Yi (for the ith of n independently sampled observations), given the values of the explanatory variables in the model $f(y,\theta,\emptyset) = \exp\left(\frac{\mathrm{y}\theta-\mathrm{b}(\theta)}{\mathrm{a}(\emptyset)}\right) + \mathrm{c}(\mathrm{y},\emptyset)$, where ∅ is a dispersion parameter and functions b(), a() and c() are known.

A linear predictor, that is a linear function of regressors $n_i = \alpha + B_1X_{i1} + B_2X_{i2} + B_kX_{ik}$ which may include quantitative explanatory variables, transformations of quantitative explanatory variables, polynomial regressors, dummy regressors, interactions, and so on.

Link Function, the link function which links the response variable, u_i = E(Yi), to the linear predictor: $g(\mu i) = n_i = \alpha + B_1X_{i1} + B_2X_{i2} + B_kX_{ik}$

2.3.3. GLM for Time Series

The ideas from generalized linear models are used in modeling time series data which is extended to handle time series where the data are dependent and the covariates are time dependent. Partial likelihood function transports main inferential features appropriate for independent data to time series which is not necessarily stationary. An essential component of partial likelihood is that it allows for temporal or sequential conditional inference with respect to a filtration generated by all that is known to the observer at the time of observation (Benjamin & Konstantinos, 2002). This enables very flexible conditional inference that can easily accommodate autoregressive components, functions of past covariates, and all sorts of interactions among covariates.

2.3.4. Parameter estimation

The likelihood is defined as the joint distribution of the data as a function of the unknown parameters. When the data are independent or when the dependence in the data is limited, the likelihood is readily available under appropriate assumptions on the factors in terms of which the joint distribution is expressed. In practice, however, things tend to be more complicated as the nature of dependence is not always known or even understood and consequently the likelihood is not within an easy reach (Benjamin & Konstantinos, 2002). This gives the impetus for seeking suitable modifications usually by means of conditioning. Partial likelihood is an example of such a modification.

If Y be a time series $\{y_t\}$, t = 1,. . . , N, with a joint density $y_\emptyset = (y_1 \ldots y_m)$ parametrized by a vector parameter ∅. In addition, if some auxiliary information (AI) is known throughout the period of observation. Then the likelihood is a function of ∅ defined by the equation

$$f_{\emptyset}(y_1 \dots y_1|AI) = f_{\emptyset}(y_1|AI) \prod_{t=2}^{N} f_{\emptyset}(y_t/y_1, y_2 \dots, y_{t-1}, AI)$$

Equation 2

The main difficulty the above likelihood function is that quite generally, if no additional assumptions are made, as the series size N increases so does the size of ∅. Hence, instead of getting more and more information about a fixed set of parameters, we obtain information but about an increasing number of parameters, a fact which raises consistency as well as modeling problems (Benjamin & Konstantinos, 2002). This is rectified when the conditional dependence in the data is limited and the increased amount of information obtained by a growing time series size concerns a fixed set of parameters. The appropriate assumptions and modifications of the general likelihood above are called for to accommodate dependent time series data such as the notion of partial likelihood (Benjamin & Konstantinos, 2002).

2.3.5. Partial likelihood function

If y_i is a response time series with the corresponding p–dimensional covariate process, $z_{t-1} = (z_{t-1,1} \dots z_{t-1,P})$ then define Ft−1 = σ{Y_{t-1}, Y_{t-2},..., Z_{t-1}, Z_{t-2},...}. The conditional expectation of the response given the past is defined as U_t = E[Y_{t-1}, | F_{t-1},].

2.3.6. GLM for count data

GLM for counts have as it's random component the Poisson Distribution. Observations of dependent counts can in many cases be modeled successfully through the Poisson distribution. The conditional density of the Poisson distribution with mean u_t can be written as.

$f(y_t, \theta_t, \emptyset/F_{t-1}) = \exp(y_t \log u_t - u_t) - \log_{y_t}!)$, t = 1,…,N,

Equation 3

where $E(y_t/F_{t-1}) = u_t$, $b(\theta) = u_t = \exp(\theta_t)$, $v(u_t)=u_t$, $\emptyset = 1$ and $w_t=1$,

the conical link is given by $g(u_t) = \theta(u_t) = \boldsymbol{n}_t = Z_t'B$.

if $Z_{t-1}' = (1, X_t, X_{t-1})'$ then the link function becomes

$\log(u_t) = B_0 + B_1 x_1 + B_2 Y_{t-2}$

with X_t standing for some covariate process, or a possible trend, or a possible seasonal component.

2.3.7. Modeling Rates: Including an Offset in the Model

Often the expected value of a response count Y_i is proportional to an index t_i. For instance, t_i might be an amount of time and/or a population size, such as in modeling crime counts. Or, it might be a spatial area, such as in modeling counts of a particular animal or plant species. Then the sample rate is Y_i/t_i, with expected value μ_i. With explanatory variables, a loglinear model for the expected rate has the form

$$\log(y_i|t_i) = \sum_{j=1}^{p} B_j x_{i,j}$$

Equation 4

Because $\log(y_i|t_i) = \log(y_i) - \log(t_i)$, the model makes the adjustment $-\log(t_i)$, to the log link of the mean. This adjustment term is called an offset. The fit corresponds to using $\log(t_i)$, as an explanatory variable in the linear predictor for $\log(u_t)$ and forcing its coefficient to equal 1. For this model, the expected response count satisfies

$$u_i = t_i exp(\sum_{j=1}^{p} B_j x_{i,j})$$

Equation 5

The mean has a proportionality constant for t_i that depends on the values of the explanatory variables.

2.3.8. Negative Binomial GLMS

For the Poisson distribution, the variance equals the mean. In practice, count observations often exhibit variability exceeding that predicted by the Poisson. This phenomenon is called overdispersion.

2.3.9. Over dispersion for a Poisson GLM

Common reason for overdispersion is heterogeneity: at fixed levels of the explanatory variables, the mean varies according to values of unobserved variables. Overdispersion is not an issue in ordinary linear models that assume normally distributed y, because that distribution has a separate variance parameter to describe variability. For Poisson and binomial distributions, however, the variance is a function of the mean.

Overdispersion is common in the modeling of counts. Suppose the model for the mean has the correct link function and linear predictor, but the true response distribution has more variability than the Poisson. Then the ML estimators of model parameters assuming a Poisson response are still consistent, converging in probability to the parameter values, but standard errors are too small. Extensions of the Poisson GLM that have an extra parameter account better for overdispersion.

2.3.10. Negative Binomial as a Gamma Mixture of Poissons

A mixture model is a flexible way to account for overdispersion. At a fixed setting of the explanatory variables actually observed, given the mean λ, suppose the distribution of y is Poisson(λ), but λ itself varies because of unmeasured covariates. Let μ = E(λ). Then unconditionally,

$$E(y) = E[E(y \mid \lambda)] = E(\lambda) = \mu,$$

Equation 6

$$var(y) = E[var(y \mid \lambda)] + var[E(y \mid \lambda)] = E(\lambda) + var(\lambda) = \mu + var(\lambda) > \mu.$$

This setup is called a mixture model for count data: suppose that given λ, y has a Poisson(λ) distribution, and λ has the gamma distribution. Recall that the gamma distribution has E(λ) = μ and var(λ) = μ^2/k for a shape parameter k > 0, so the standard deviation is proportional to the mean. Marginally, the gamma mixture of the Poisson distributions yields the negative binomial distribution for y. Its probability mass function is

$$p(y; u, k) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+k)} \left(\frac{u}{u+k}\right)^y \left(\frac{k}{u+k}\right)^k, \qquad y=1,2,3,\ldots\ldots$$

Equation 7

With k fixed, this is a member of an exponential dispersion family appropriate for discrete variables with natural parameter $\boldsymbol{log(\frac{u}{u+k})}$.

if $\gamma = 1/k$ then $E(y) = \mu$, $var(y) = \mu + \gamma u^2$

The index $\gamma > 0$ is a type of dispersion parameter. The greater the value of γ, the greater the overdispersion relative to the Poisson. As $\gamma \rightarrow 0$, $var(y) \rightarrow \mu$ and the negative binomial distribution converges to the Poisson.

The negative binomial distribution has much greater scope than the Poisson. For example, the Poisson mode is the integer part of the mean and equals 0 only when $\mu < 1$. The negative binomial is also unimodal, but the mode is 0 when $\gamma \geq 1$ and otherwise it is the integer part of $\mu(1 - \gamma)$. The mode can be 0 for any μ.

2.3.11. Negative Binomial GLMs

Negative binomial GLMs commonly use the log link, as in Poisson loglinear models, rather than the canonical link. For simplicity, we let the dispersion parameter γ be the same constant for all n observations but treat it as unknown, much like the variance in normal models. This corresponds to a constant coefficient of variation in the gamma mixing distribution, $\sqrt{\text{var}(\lambda)} / E(\lambda) = \sqrt{\lambda}$

The loglikelihood function for a negative binomial GLM with $\boldsymbol{n}$ independent observations is

$$L(B, \lambda, y) = \sum_{i=1}^{1} [\log\Gamma\left(y_i + \frac{1}{\lambda}\right) - \log\Gamma\left(\frac{1}{\lambda}\right) - \log\Gamma(y_i + 1)] + \sum_{i=1}^{1} [y_i \log\left(\frac{\lambda u_i}{1+\lambda u_i}\right) - \frac{1}{\lambda} \log(1 + +\lambda u_i)]$$

Equation 8

where u_i is a function of $\boldsymbol{\beta}$ through $n^i = g(u_i) = \sum_{j=1}^{p} B_j x_{i,j}$ with the link function g.

2.3.12. Negative Binomial Model with Variance Proportional to Mean

An alternative negative binomial parameterization results from writing the gamma density formula with $k\mu$ as the shape parameter,

$$f(\lambda, \mathrm{k}, \mathrm{u}) = \frac{k^{ku}}{\Gamma(ku)} exp(-k\lambda)\lambda^{ku-1}, \ \ \lambda \geq 0$$

Equation 9

So E(λ) = μ and var(λ) = μ/k. For this parameterization, the gamma mixture of Poisson distributions yields a negative binomial distribution with

E(y) = μ , var(y) = μ (1 + k)/k.

The variance is now linear rather than quadratic in μ . It corresponds to an inflation of the Poisson variance, converging to it as k → ∞.

The two parameterizations of the negative binomial are sometimes denoted by **NB1** (linear) and **NB2** (quadratic). Only the NB2 falls within the traditional GLM framework, being expressible as an exponential dispersion family distribution, and it is much more commonly used. Unlike the NB2 model, for an NB1 model $\boldsymbol{\beta}$ and k are not orthogonal parameters, and $\boldsymbol{\beta}\hat{}$ is not a consistent estimator when the model for the mean holds but the true distribution is not negative binomial (Cameron and Trivedi 2013).

2.3.13. Model diagnostics

Diagnostics in regression analysis consists of procedures for exploring and testing the adequacy and goodness of fit of fitted models. In the context of generalized linear models this refers in particular to the examination of several types of residuals and deviance analysis. Deviance analysis is carried out routinely through a statistic called the scaled deviance and the closely related information criteria AIC and BIC.

2.3.14. Model selection

Evaluation and selection among several competing models is based on Akaike's information criterion (AIC). The AIC criterion is defined as a function of the number of independent model parameters,

$$AIC = -2logPL\left(\widehat{\mathrm{B}}\right) + 2p,$$

Equation 10

where ,$\widehat{B}$ is the maximum partial likelihood estimator of B and p is the "model order", p = dim(B). We choose the model corresponding to p that minimizes AIC.

2.3.15. Residuals

Residual means a certain deviation of a fitted from an observed value. Residual analysis is important in assessing the goodness of fit-how well the fitted model explains the observed data-of a regression model, and in judging the impact and significance of covariates on the response. There are several ways to define residuals in the context of time series following generalized linear models. The most obvious definition is that of the so called raw or response residuals

$\hat{e}_t = Y_t - \hat{\mu}_t$, *t=1,...,N*

Three popular additional types of residuals, Pearson, working, and deviance, are defined in terms of the raw residuals as follows. The Pearson residuals are the standardized version of raw or response residues obtained by dividing each raw residual by the square root of the estimated variance to obtain the standardized Pearson residuals as follows;

$$\hat{r}_t = \frac{Y_t - \hat{\mu}_t}{\sqrt{V(\hat{\mu}_t)}},\ t = 1,...,N.$$

Equation 11

The working residuals are a different standardized version obtained after fitting a working model, which is an initial approximation to the true model,

$$\widehat{w}r_t = \frac{Y_t - \hat{\mu}_t}{d\mu_t/dn_t},\ t = 1,...,N.$$

Equation 12

where $d\mu_t/dn_t$is evaluated at $\widehat{B}$. The deviance residuals are given by

$$\hat{d}_t = sign(Y_t - \hat{\mu}_t)\sqrt{2[l_t(y_t) - l_t(\hat{\mu}_t)]},\ t=1,...,N$$

Equation 13

where the sum of squares of deviance residuals is equal to the deviance statistic.

2.3.16. Distributed lag non-linear model

Oftentimes, the effect of exposure to environmental stressors or other events is not limited to the time period when it occurs, but is delayed in time. That is the case of climatic variables effects (as adverse temperatures) that may last in time, spreading over several days. That being said, the impact of the environmental stressors at a given time may be explained by a combination of past exposures over several time lags, once it depends simultaneously on the intensity and timing of the exposures. A common way to model a non-linear effect with this additional time dimension is through distributed lag non-linear models (DLNMs)

The family of distributed lag non-linear models was developed to simultaneously estimate the non-linear dose-response dependencies and the delayed effects of temperature on mortality. It is based on a bi-dimensional space of functions, called "cross-basis", that describes the shape of the relationship simultaneously along the space of the predictor - temperature, in this case - and along its lag dimension, i.e., the time structure of the exposure–response relationship.

Initially, distributed lag models (DLMs) were developed for time series analysis and extensively used in econometric and social sciences before being adapted to epidemiology research. This family of models used to account for linear dependencies only, so Armstrong extended this methodology to distributed lag non-linear models (DLNMs), and it has since been used to simultaneously estimate the non-linear and delayed effects of temperature and air pollution on mortality or morbidity Hence, to understand a DLNM well one must first understand DLMs. A DLM is a dynamic model that estimates the effect of a regressor x on a response y over different time moments t. It can generally be represented as follows:

$$y_t = \alpha + B_0 x_1 + B_1 x_{t-1} + .. + B_L x_{t-L} + u$$

Equation 14

where α is the intercept, u is a stationary error term and L is the maximum lag allowed ($L \geq 1$). Each of the coefficients βl stands for the weight of the respective lag l ($l = 0, 1, ..., L$).

B_l coefficients may be interpreted either from a backward standpoint - the effect of the past exposition x_{t-1} on the present moment response y_t, i.e., the effect felt today due to the exposition

L days before, lag = l -, or from a forward standpoint - the effect of the current exposition x_t on the future response L moments later, y_{t+1}, i.e., the effect that today's exposure will cause L days from now . Thus, the coefficients B_l represent the lag weights and all together define the lag distribution.

Assuming there is a temporary change on the regressor variable x, which increases one unit only in moment t, x_t, then its immediate effect, y_t , will have an increase equal to the value of B_0. On the next moment (t+1) the effect y_{t+1}, will increase B_1 units. After that the effect y_{t+1} will increase B_2 units and so on, until the maximum lag, L, when the effect y_{t+L} increases B_L units. This is called the marginal effect of x on y.

Another hypothesis to consider would be a permanent change in the regressor variable. Assuming it increases one unit in moment t and remains that high in all future moments, then its immediate effect y_t will also have an increase equal to the value of B_0. However, in the future moment (t+1) the effect y_{t+1} will increase $B_0 + B_1$ values, after which the effect y_{t+2} will increase β0 +β1 +β2 values and so on, until the maximum lag L, when the effect y_{t+L} increases β0 +β1 +β2 +...+βL values. This is called the cumulative effect of x on y.

According to Gasparrini (2014) and assuming there is a linear exposure–response relationship, a general notation to describe the dependency in terms of exposure history to x evaluated at time t as:

$$S(x,t) = \int_{L_0}^{L} X_{t-l} w(L)\, dL$$

Equation 15

where $w(L)$ represents the weighting basis-function applied to constrain the coefficients B_l. $w(L)$ is directly defined in the lag dimension and determines the lag-response function that models the lag–response curve associated with exposure x.

$S(x,t)$ will then be included in a generalised linear model as a sum of linear terms, with related parameters η. The function $S(x,t)$ may be rewritten following a matrix notation, by applying the basis transformation over the lags - w(l) - and then combining it with the vector of expositions q_r.

$$S(x_t, n) = q_r n = w_t^T n$$

Equation 16

where

- $q_t = [x_t, \dots, x_{t-1}, \dots, x_{t-L}]^T$ is an original vector of ordered exposure histories, corresponding to a column of the n×(L+1) matrix Q. As such q_t changes along time, depending on the moment t when it is defined.
- $L = [0, \dots, l, \dots, L]^T$ is a vector of lags corresponding to the L+1 columns of Q
- C is a (L+1)× v_t matrix of lag-basis variables originated from the application of the basis-function to the lag vector l (where the basis-function is defined as w(l) with dimension v_t);
- η is a vector of unknown parameters.

Hereupon, w^T is a vector from the matrix $\boldsymbol{W} = \boldsymbol{QC}$, which is the matrix of the v_l transformed variables (obtained from the application of a basis lag function to the original lag vectors **l** - ***w(l)*** - combined with the original exposure histories, q_t). This v_t basis variables from the matrix $\boldsymbol{W}$ will be included in the design matrix to allow the estimation of the unknown parameters $\boldsymbol{\eta}$. The estimated parameters $\hat{\boldsymbol{n}}$ define the previously mentioned coefficients $\boldsymbol{B_l}$

$$\hat{\boldsymbol{B}} = C\hat{n}$$

Equation 17

The extension from DLM to distributed lag non-linear models (DLNM) was achieved by adding an exposure–response non-linear function along the dimension of the predictor x.

$$S(x_t, B) = z_t^T B$$

Equation 18

On this function, z_t^T. is the t^{tn} line of the matrix $\boldsymbol{Z - a\eta \times v_x}$ basis matrix resulting from the application of the basis-function - called f(x), of dimension $\boldsymbol{v_x}$, to the original vector of exposures x

Therefore, a generalization to DLNM will be adding a basis-function along the dimension of the predictor x to the already mentioned basis-function along the dimension of the ***lag l*** in;

$$S(x,t) = \int_{L_0}^{L} X_{t-l} w(L)\, dL$$

Equation 19

where ***f(x***) is the exposure-response function with dimension $\boldsymbol{v_x}$ and ***w(l)*** is the previously mentioned lag-response function. These are the two basis-functions, which may be chosen independently of each other. The basis-functions impose a set of completely known transformations of x, generating new variables, called basis variables.

The representation above assumes that the functions ***f(x)*** and ***w(l)*** are independent, i.e., that the exposure-response function is the same along all the lag space and that the lag structure is equal for all values of x. If we relax this assumption, admitting an interaction between the value of the predictor and its timing, then it may be more flexibly represented as:

$$s(x,t) = \int_{t0}^{L} f \cdot w(x_{-2-1}, L)\, dL$$

Equation 20

Following this notation, ***f·w(x,l)*** is a bivariate function, that models simultaneously the exposure response structure along x and the lag-response structure along l, defining the exposure–lag–response function . In other words, ***s(x,t)*** is a linear combination of the basis-functions ***f(x)*** and ***w(l***), integrated over the lag dimension, which defines a bi-dimensional space of functions that Armstrong (2006) called cross-basis function and which represent the core of DLNMs.

Each basis-function (***f(x)***= $\sum_{b-1}^{B} B_b X_b$ and ***w(l)*** = $\sum_{p-1}^{P} B_p l_p$) is fitted following the chosen function distribution, ***f()*** and ***w()*** respectively, originating **B**/**P** basis variables from the original ones. Then the new transformed variables will enter the regression model, instead of the original variables. The number of basis variables, **B** and **P**, determines the degrees of freedom (df) of the respective curves. **B** df for the dimension of the predictor and **P** df for the dimension of the lags. So the degrees of freedom of the cross basis-function is given by the product of the number of basis variables from ***f(x)*** and from ***w(l), B×P***.

The cross-basis function is better represented using matrix notation as Gasparrini et al (2021). Let $\boldsymbol{R}^{\cdot}$ be a $\boldsymbol{\eta} \times \boldsymbol{v_x} \times \boldsymbol{(L+1)}$ array of the lagged occurrences of each of the basis variables of $\boldsymbol{x}$, keeping $\boldsymbol{C}$ as the matrix of basis variables for the lag dimension,

$$s(x_t, B) = \sum_{j=1}^{v_x} * \sum_{l=l'}^{v_L} r_x^T c_k n_{jk} = w_t^T n$$

Equation 21

where r_{tj} is the vector of lagged exposures for the time $\boldsymbol{t}$ transformed through the basis-function j and c_k is the vector of lags transformed trough the basis-function $\boldsymbol{k}$. Now, $\mathbf{w_t^T}$ will be defined as the vector obtained by applying the $v_x \cdot v_L$ cross-basis functions to x_t . As such, both basis-functions are then simultaneously used to create the $v_x \cdot v_L$ basis variables, stored in the $\boldsymbol{W}$ matrix.

The cross-basis flexibly describes the relation along $\boldsymbol{x}$, allowing for linear and non-linear exposure-responses, combining it with the distributed lag-effects (an additional time-dimension). So, the $v_x \cdot v_L$ basis variables originated from the cross-basis function will enter the regression model instead of the original variables

We may now generally represent a basic DLNM, using a cross-basis function to express the non-linear relation between the predictor variable $\boldsymbol{x}$ and the response variable $\boldsymbol{y}$ along time. Following the notation of Gasparrini et al (2021), we obtain:

$$g(u_t) = \alpha + \sum_{j=1}^{J} s_j(x_{xj}; n_j) + \sum_{k=1}^{k} \gamma_k u_{tk}$$

Equation 22

where $\boldsymbol{u_t = E(y_t)}$ and $\boldsymbol{g}$ is a monotonic link function; $\boldsymbol{\alpha}$ is the intercept; s_j is the cross-basis function that denotes smoothed relationships between the variables x_{xj} and the linear predictor - defined by the parameter vectors $\mathbf{n_j}$, and u_{tk} are other predictors variables with linear effects over $\boldsymbol{Y}$, measured by the coefficients γ_k.

In this family of models, $\boldsymbol{Y}$ is assumed to follow a distribution from the exponential family. The two basis-functions used in the cross-basis $\boldsymbol{s_j}$ may be chosen from a wide variety of modelling options. Those options are:

Exposure space:

- a polynomial function, whose order must be determined;
- a stratified model, with chosen strata intervals;
- a spline function, for which the number and placing of knots must be chosen;
- linear thresholds;

Lag space:

- a polynomial function, whose order must be determined;
- a stratified model, with chosen strata intervals;
- a spline function, for which the number and placing of knots must be chosen;
- the coefficients may be unconstrained;

The options selected will determine the flexibility of the model. Simpler models (as linear thresholds) are usually less flexible but easier to interpret than the more complex ones (as natural cubic splines functions), while the latter may better adjust to the data and capture most of the relationship details and are less likely to leave residual confounding.

Other predictor variables may be included in the model. For instance, as happens in order to control for confounding, a smooth function of time to capture long-time trends and/or seasonality and some categorical variables, as day of the week are applied.

Also, the analysts must determine the maximum lag, L, which will depend on how long they believe an effect may be sustained in time. For instance, the maximum lag allowed should be higher if harvesting effects are expected, which may reflect on negative coefficients for longer lags.

All options have advantages and disadvantages, so the choice will depend on the purpose of the analysis, on a priori assumptions and/or the fitting of the model. The model fitting criteria mostly used in this are the Akaike's information criteria (**AIC**) and the Bayesian information criteria (**BIC).**

2.4. Methods and Materials

2.4.1. Study Context

Mangochi district is located in the southern region of Malawi and had a total population of 1,148,611, 6.5% of national population in 2018 (Malawi Housing and Population Census, 2018). It is a lakeshore district which is among 11 high malaria burden districts (Government of Malawi, 2020). It comprises of 1 district hospital and 41 health centers and several village clinics providing malaria services. All suspected uncomplicated malaria cases are tested using malaria rapid diagnostic test (mRDTs) at all levels (central, district, health center, clinic, community) (Government of Malawi, 2020). Light microscopy is used to test suspected complicated malaria cases (where capacity allows), diagnose severe malaria cases and confirm malaria treatment failure (Government of Malawi, 2020). Community case management aims to address three main childhood killers namely; malaria, pneumonia and diarrhoea. It promotes early recognition, prompt diagnostic testing, and appropriate treatment of malaria among children under five years in the home or community. It is an equity-focused strategy that aims to improve access for under-five children in hard-to-reach areas thereby improving timely and effective treatment of malaria.

2.4.2. Response Variable: Malaria incidence

Malaria is routinely collected at health facility level and uploaded into District Health Information System on monthly basis aggregated by health facility, district and national level. Reported malaria cases are cases confirmed through mRDT and Microscopy, unconfirmed malaria cases (clinical cases without confirmation) were excluded in this analysis. Malaria monthly data between 2015 and 2020 for Mangochi district was provided by national malaria control program in Malawi.

2.4.3. Primary Covariates: Climate data

In this study, data for monthly meteorological variables, including the daily maximum and minimum temperature; relative humidity; and the amount of rainfall were obtained from department of metrological services and climate change.

2.4.4. Distributed Lag Non-Linear Model

The outcome variable in this study was the number of monthly confirmed malaria cases spanning for a period of 6 years between January 2015 to December 2020. The explanatory variables were

climate factors; temperature (monthly minimum temperature and monthly maximum temperature), monthly rainfall (mm) and monthly relative humidity from January 2015 to December 2020. All the climate variables are continuous variables.

A time series regression was applied to model mean monthly malaria incidence as outcome variable and climate factors as explanatory variables expressed as follows. Let Y_t be monthly malaria cases, then

$$Y_t \sim NegBinomial(\mu_t)$$

$$log(u_t) = \alpha + \sum_{j=1}^{J} S_j(x_{t,j}, B_j) + \sum_{k=1}^{K} \psi_k(z_{tk}) + s(time, p) + log(population)$$

Equation 23

where y_t is a series of monthly malaria cases spanning for a period of 72 months, t=1,…72, u_t is expected monthly malaria cases, $\log(u_t)$ is a log link function, α is an intercept term and $s(time, p)$ is a natural cubic splines to control for seasonality in malaria time series. Non-linear effects of climate are modelled using distributed lag non-linear model (DLNM) specified by function $\sum_{j=1}^{J} S_j(x_{t,j}, B_j)$. The function S_j specify the relationships between the meteorological variables X_j at lag month j and the linear predictor defined by the parameter vectors B_j (Gasparrini, 2011). The variables z_{tk} are other predictors with linear effects specified by the related coefficients ψ_k, such as indoor residue spray (IRS) intervention in this study. A negative binomial family account for over dispersion of monthly malaria incidence while population offset to control for changes of population over time.

Data was analyzed using R version 3.2.4 (Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017).

2.5. Results

2.5.1. Descriptive statistics

A total of 1,935,056 malaria cases were reported between 2015 and 2020, with 860,745 cases occurring in children under the age of five and 1,074,311 cases occurring in individuals aged five and above. The highest number of malaria cases in a single year was 418,394, reported in 2018,

while the lowest number of cases was 210,279, reported in 2020. On a monthly basis, the average number of malaria cases was 26,876. The highest number of cases was observed in May 2018, with 53,449 cases, while the lowest number of cases was reported in August 2016, with 11,124 cases. The yearly averages and standard deviations for climate variables are provided in **Table 1**.

Table 1 Yearly means and standard deviations for climate variables

	2015		2016		2017		2018		2019		2020	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD
Malaria	20317	5699	24345	11155	30872	11776	34866	13636	33332	12125	17523	3503
Max Temp	31	3	32	3	31	2	31	3	30	2	31	3
Rainfall	86	156	64	90	97	121	55	62	87	115	54	83
Humidity	64	16	62	14	66	15	68	13	67	14	64	13

Monthly patterns of climate variables between 2015 and 2020 is as shown in Figure 1 below.

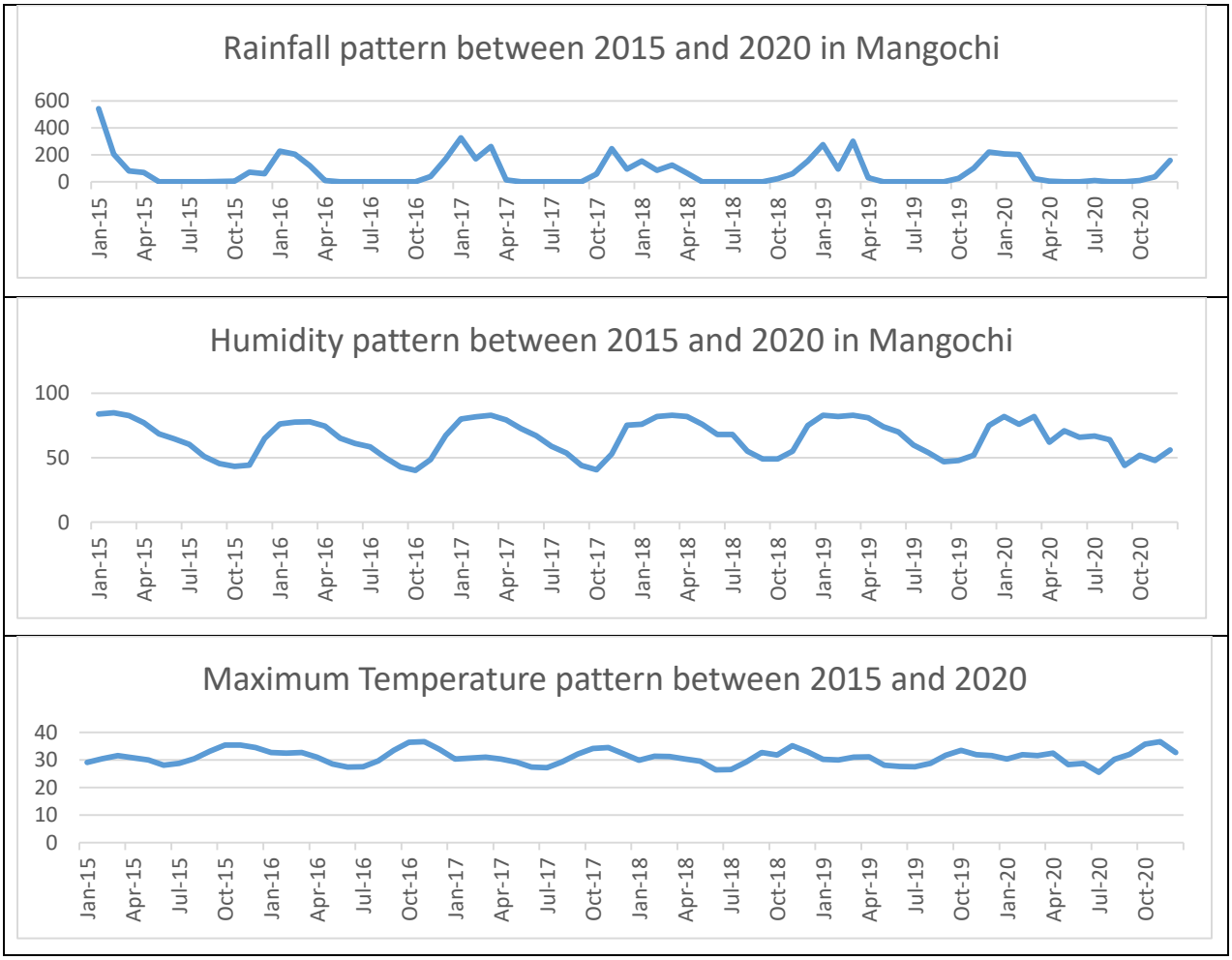


Figure 1 : Monthly pattern of climate variables between 2015 and 2020

Exploration of malaria seasonal trend visually coincide with seasonal trends of climate variables as shown in Figure 2*a-b-c-d*. This pictorial coincidence of climate variables and malaria cases suggests possible seasonal correlation.

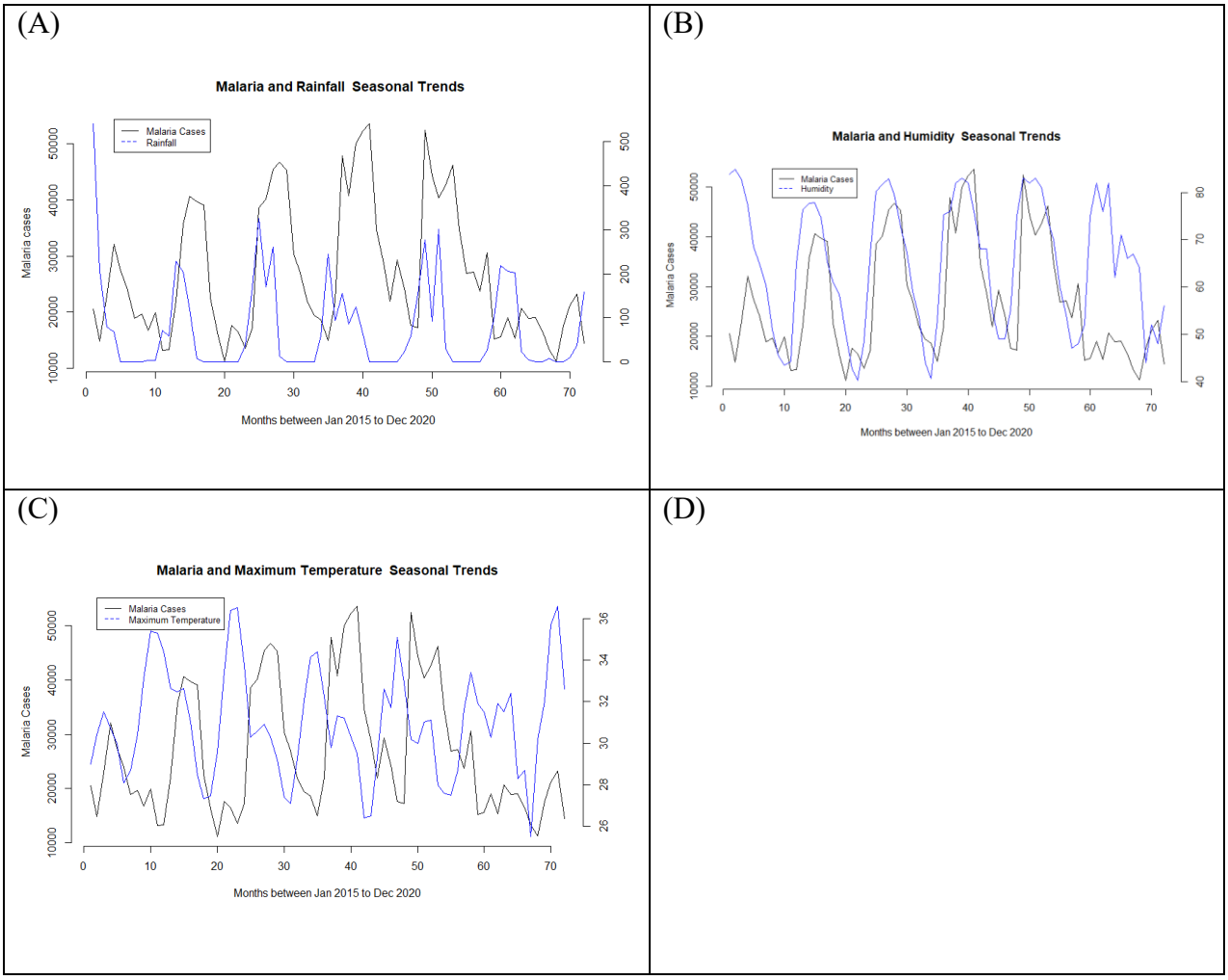


Figure 2: Seasonality of malaria cases and climate variables

Furthermore, scatterplots between malaria cases and climate variables depicts a linear pattern between humidity and malaria cases as shown in Figure 3*b* but does not show obvious patterns with rainfall and temperature Figure 3*a-c-d*.

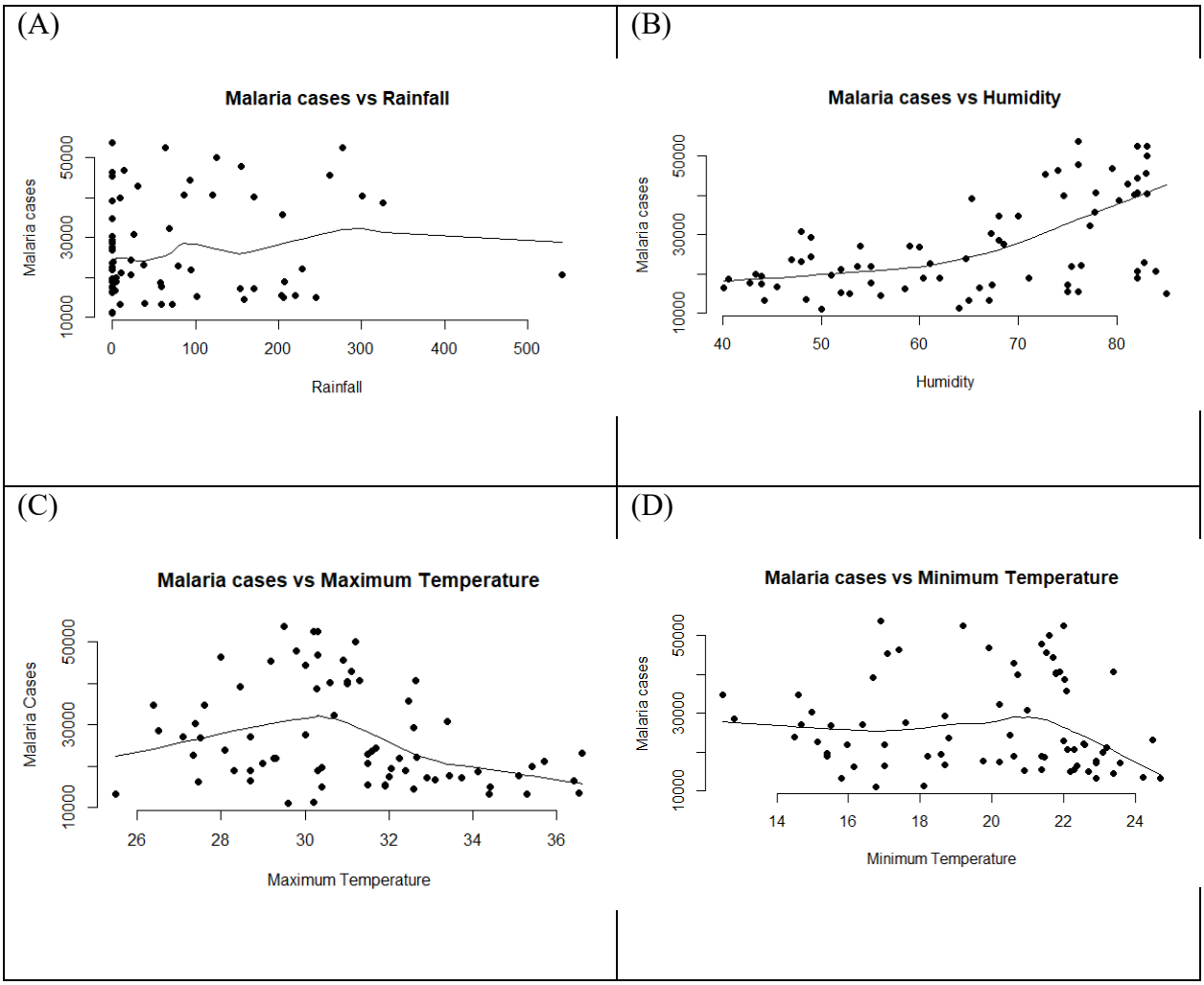


Figure 3: Association between malaria cases against climate variables and lowess line

2.5.2. Distributed Lag Non-Linear Model building

Table 2 shows variable correlation matrix which shows that malaria cases correlated positively with rainfall and humidity at lag-0, (r 0.112, 0.567), while a negative correlation was observed with minimum and maximum temperature (r -0.284 and -0.078). As reported by Chuang et al (2017), collinearity, or excessive correlation among explanatory variables, can complicate or prevent the identification of an optimal set of explanatory variables for a statistical model (Chuang & Ting, 2017). In this study, correlation between climate variables was assessed using pearson correlation to identify climate variables that have excessive collinearity. The correlation matrix in Table 2 shows positive correlation between variables: maximum and minimum temperature (correlation coefficient 0.813), minimum temperature and rainfall (correction coefficients 0.536) and humidity and rainfall (correlation coefficients 0.512).

Table 2: Cross correlation matrix of climate variables

Variables	Malaria	Rain	Humidity	Max Tem	Min Tem
Total malaria cases	1.00	0.112	0.567	-0.284	-0.078
Rain	0.112	1.00	0.512	0.133	0.536
Humidity	0.567	0.512	1.00	-0.377	0.122
Maximum Temperature	-0.284	0.132	-0.377	1.00	0.813
Minimum Temperature	-0.078	0.537	0.122	0.811	1.00

Variance inflation factor (VIF) was further applied to assess the impact of collinearities in the final model. Qinqin et al (2018) recommended dropping high correlated variables with VIF above 5 to minimize impact on model sensitivity (Qinqin, Runzi, Shannon, Cheng, & Yafei, 2018). The variance inflation factor is given by the formula below.

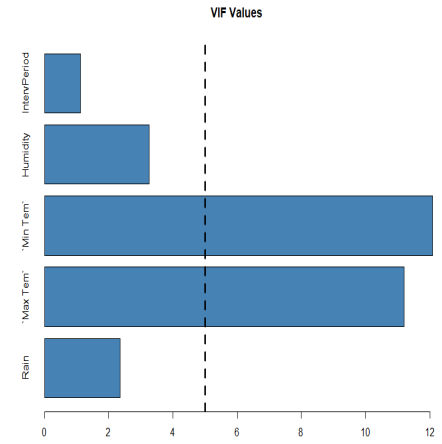
$$VIF_j = \frac{1}{1 - R_j^2}$$

Equation 24

where the VIF for variable j is the inverse of R^2 from the regression. A VIF is calculated for each explanatory variable and those with high values are removed.

The VIF results showed high VIF above 5 for minimum and maximum temperature as shown in Figure 4*a*. Following Qinqin recommendation, the two highly correlated climate variables cannot be combined as regressors in a model and therefore minimum temperature which has highest VIF was dropped and the resultant model has reduced variance inflation factor among all climate variables as shown in Figure 4b.

(a)VIF with all variables



(b)VIF with min Temp removed

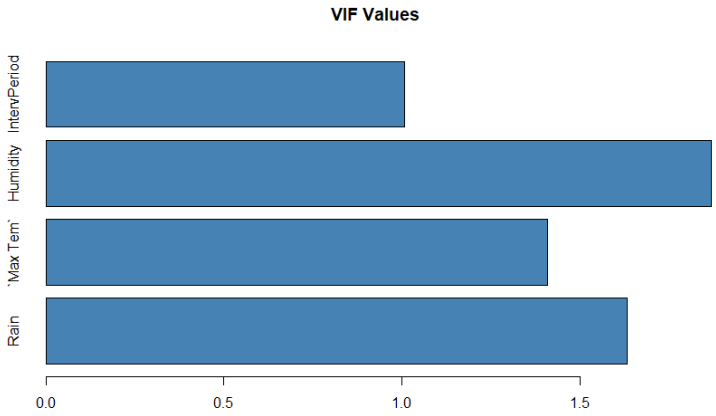


Figure 4: VIF values for climate variables

2.5.3. Exposure Lag Response relationship

Model specifications in the exposure-lag dimensions for climate variables were selected among a wide range of linear and no-linear functions by examining Akaike Information Criterion (AIC). A total of 18 model candidates with different specifications in exposure-lag dimensions were assessed as shown in Table 3.

Table 3 Exposure-lag response specifications for climate variables

Exposure-Lag response combination	Exposure-response f(x)	Lag response w(l)	Exposure (Degree and Knots)	Response(Degree and Knots)	AIC	BIC
Rainfall/Precipitation						
Linear-Linear	Linear	Linear			27370	27409.69
Linear-Polynomial	Linear	Polynomial			27330	27372.91
Polynomial-Linear	Polynomial	Linear			26650	26698.35
Polynomial-Polynomial	Polynomial	Polynomial	3D	2D	26410	26466.76
Natural Cubic Splines function-Linear	NS	Linear	Quantiles		26610	26663.01
Natural Cubic Spline function-Polynomial	NS	Polynomial	Quartiles	2D	26310	26369.45
Humidity						
Linear-Linear	Linear	Linear			27220	27261.94
Linear-Polynomial	Linear	Polynomial			27150	27198.23
Polynomial-Linear	Polynomial	Linear			27140	27190.62
Polynomial-Polynomial	Polynomial	Polynomial	3D	2D	26780	26839.57
Natural Cubic Splines function-Linear	NS	Linear	Quartiles		26920	26972.65
Natural Cubic Spline function-Polynomial	NS	Polynomial	Quartiles	2D	26610	26677.69
Maximum Temperature						
Linear-Linear	Linear	Linear			27390	27430.31
Linear-Polynomial	Linear	Polynomial			27360	27402.27
Polynomial-Linear	Polynomial	Linear			27180	27230.17
Polynomial-Polynomial	Polynomial	Polynomial	3D	2D	26830	26884.8
Natural Cubic Splines function-Linear	NS	Linear	Quartiles		27160	27215.84
Natural Cubic Spline function-Polynomial	NS	Polynomial	Quartiles	2D	26830	26899.46

The examination of AIC from different functions in the exposure–lag dimension for climate variables showed that cubic splines in exposure dimension and polynomial in lag response dimension have lower AIC across all the three climate variables. The natural cubic splines with knots placed at equal intervals in exposure space and polynomial in lag response dimension have lowest AIC. This confirms the existence of non-linear relationship in the predictor space. Finally, basis variables were generated and added in the final model using backward selection method. The backward variable selection showed model improvement associated with incorporation of climate basis variables.

The final specification in the exposure dimension for climate variables was a natural cubic spline with knots placed at equal intervals and polynomial of degree 2 in the lag response dimension. Seasonality is controlled by natural cubic splines.

Indoor Residual Spraying (IRS) intervention period was also included as a predictor in the final model. To control for changing population, an offset of log transformed population was included in the final model. GLM negative binomial family was applied to account for over dispersion of monthly malaria cases.

2.5.4. Lagged effects of rainfall

The exposure lag response relationship between rainfall and malaria incidence is illustrated in Figure 5*a* which shows nonlinear relationship between precipitation and malaria incidence. Increasing precipitation is associated with increased malaria risk which peaked at lag 0 when monthly rainfall reached maximum of 541mm, RR 2.4314162, CI 95% (2.0554548, 2.8761443) compared to risk when there is an average rainfall of 74.5mm. Predicted effects are sustained across all the 3 lags and are more certain (narrow confidence interval) as shown in Figure 5*d*. Figure 5*b* is a heat map showing virtual representation of relative risk which also indicates high risk associated with increasing rainfall and visually peaked at lag-0.

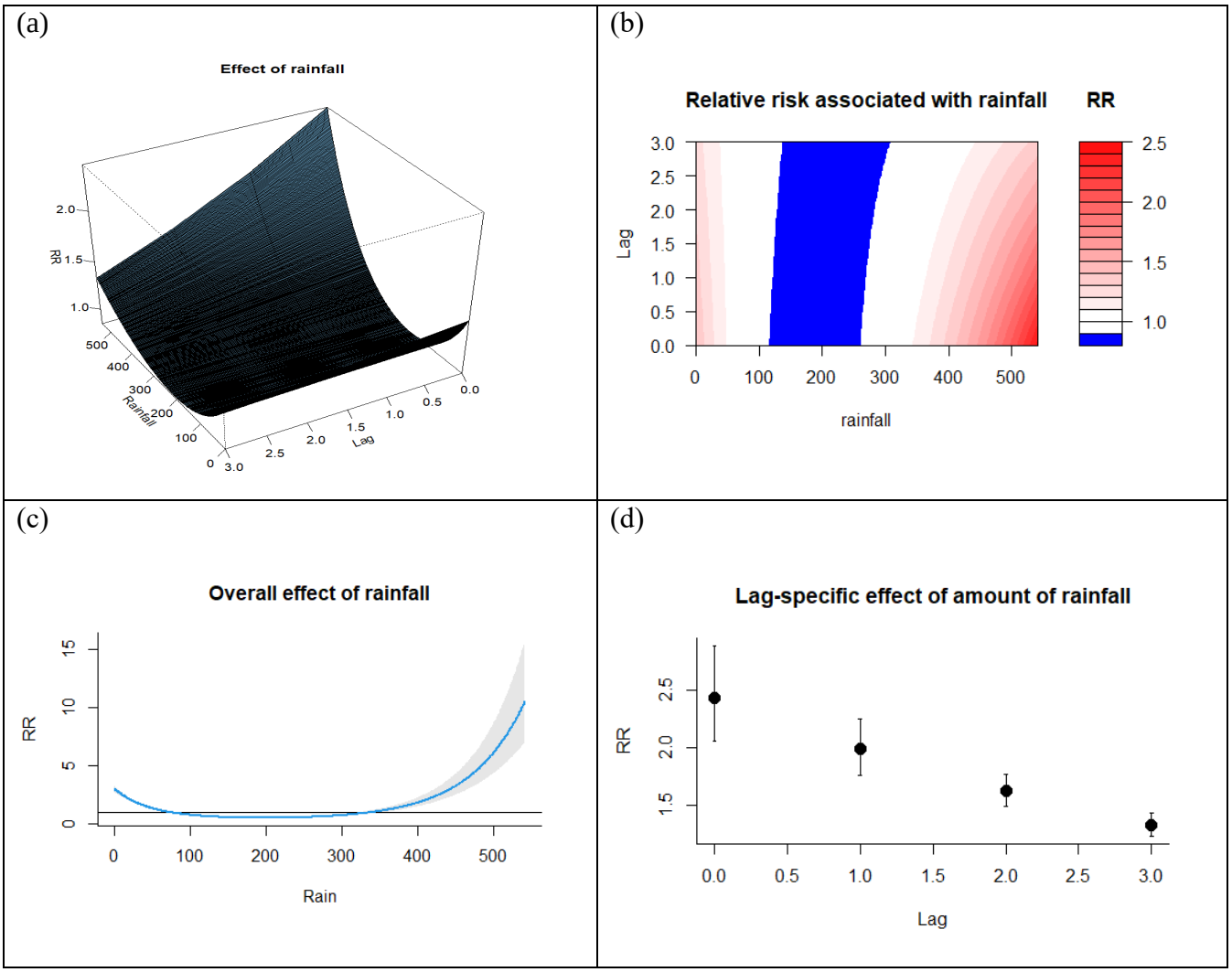


Figure 5: 3D relationship, relative risk and lag specific effects between malaria cases and rainfall

2.5.5. Lagged effects of humidity

The effect of humidity is illustrated in the Figure 6*a* which shows nonlinear relationship between humidity and malaria incidence. Increasing humidity is associated with increased relative risk which peaked at lag 0 (same month) when relative humidity reaches 84, RR 2.0537649 CI (1.9435101, 2.1702744) compared to relative risk at mean humidity. The overall effects of humidity are more certain (narrow confidence intervals) and sustained up to lag-2 as shown in Figure 6*d*. Figure 6*b* is a heat map indicating virtual representation of relative risk which also indicates high risk associated with increasing humidity and visually peaked at lag-0.

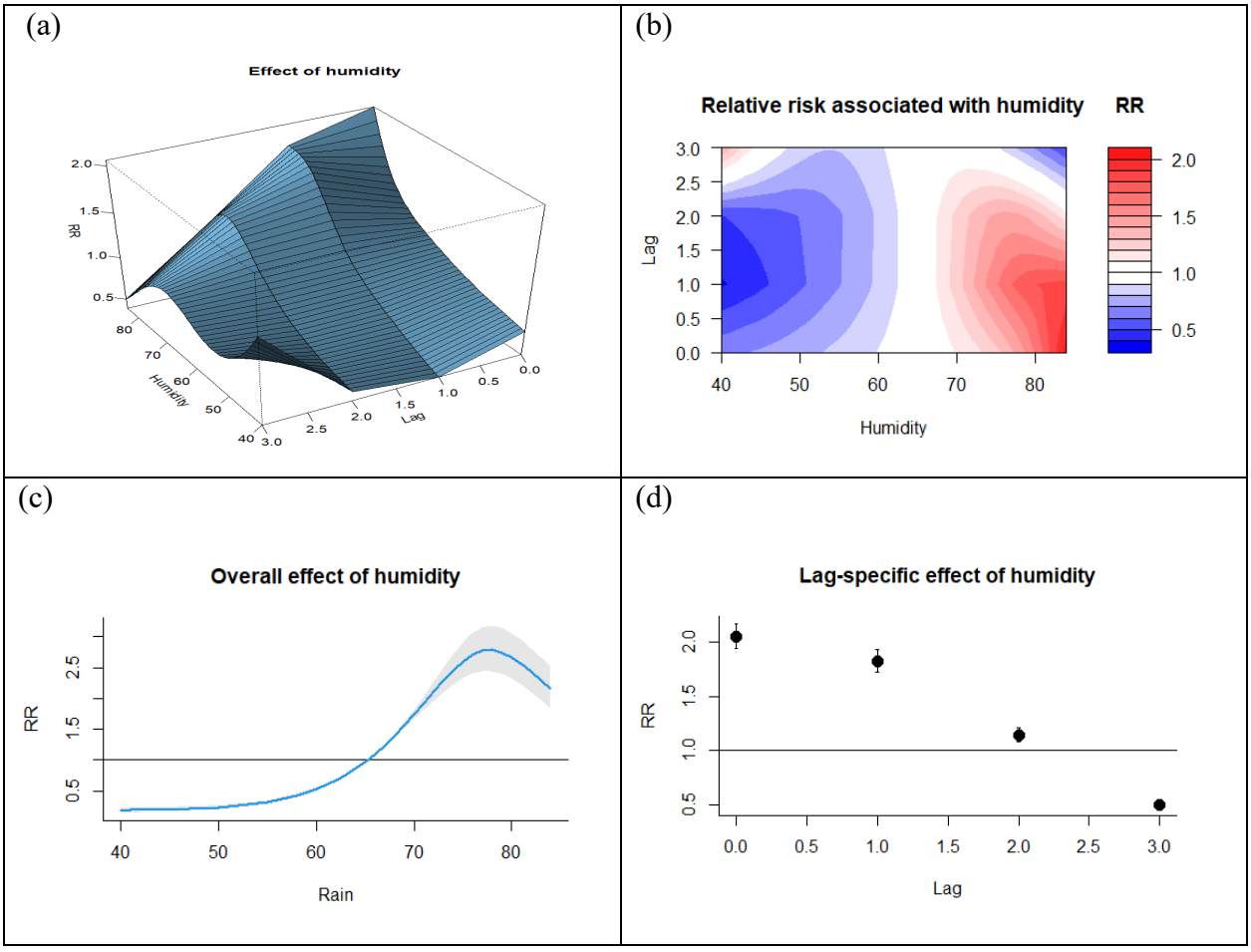


Figure 6: 3D relationship, relative risk and lag specific effects between malaria cases and humidity

2.5.6. Lagged effects of maximum temperature

The relationship between maximum temperature and malaria cases is shown in Figure 7*a* which depicts nonlinear exposure lag response relationship. Increasing maximum temperature is associated with increased risk which peaked at lag 0 (same month) when temperature reach 34 degrees Celsius, RR 1.3907299 CI (1.3523290, 1.4302212) and sustained up to lag 2 as shown in Figure 7*d*. Predicted effects are more certain, narrow confidence intervals, across all lags. Figure 7*b* is a heat map showing virtual representation of relative risk which also indicates high risk associated with increasing humidity and visually peaked at lag-0.

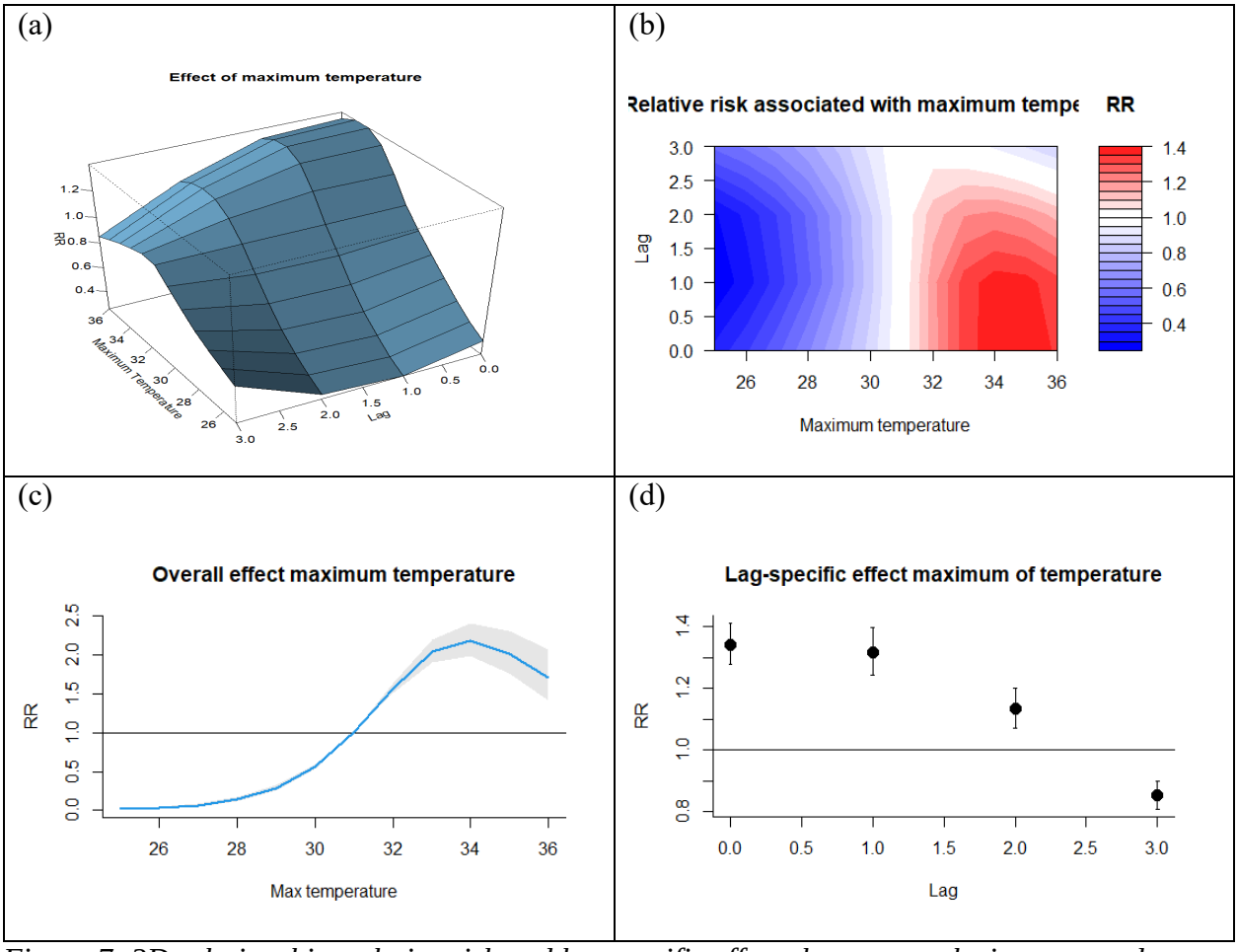


Figure 7: 3D relationship, relative risk and lag specific effects between malaria cases and maximum temperature

2.6. Discussions

The observed nonlinear relationship between rainfall and malaria incidence highlights the influence of precipitation in the breeding and survival of mosquito vectors consistent with many studies in this area. The positive association between increasing precipitation and malaria risk confirms that areas with higher rainfall are more conducive to mosquito breeding, leading to a higher prevalence of malaria cases. As reported by Yoonhee et al. (2019), rainfall is considered one of the main weather factors determining malaria epidemics. Also, the observed 0-2 months delayed effect of rainfall is consistent with results reported by some studies such as a paper by Jae et al (2012) who suggested that lagged estimates of the effect of rainfall on malaria are consistent with the time necessary for mosquito development and *P. vivax* incubation. More importantly, the

observed peak(double) of malaria risk immediately following periods of extreme rainfall conditions, highlights the significance of short-term effects of rainfall and possible outbreaks following extreme rainfall events.

The impact of humidity on malaria incidence has been extensively studied, and the study findings reaffirm the positive association between humidity and malaria risk. The observed immediate peak in malaria risk (same month), coinciding with extreme conditions of humidity, supports the notion that humidity acts as a crucial factor in influencing mosquito populations and subsequent malaria transmission. This result is also consistent with Philippe et al.'s (1995) report, which states that higher levels of humidity prolong the lifespan of mosquitoes and enable them to infect more people.

Similarly, the nonlinear relationship between temperature and malaria cases is consistent with previous research linking the disease and temperature conditions. The observed peak in malaria risk following high temperatures emphasizes the importance of temperature in shaping the dynamics of malaria transmission. As reported by Gunda et al. (2017), temperature affects malaria transmission by regulating the rate of development of mosquito larvae, which in turn influences mosquito survival rates.

In general, the results revealed an immediate peak (same month) in malaria risk following extreme weather conditions, highlighting the importance of short-term effects of climate. The risk of malaria immediately doubles with extreme rains and humidity compared to average weather conditions. This is consistent with Florence et al.'s (2019) report linking climate extreme events with outbreaks. The study also found delayed but diminishing effects of climate conditions from 0-2 months lag, indicating that the impact of climatic variables persists over multiple time intervals. This result is consistent with a report by Gunda et al (2017) which showed significant association between malaria incidence and the climatic variables at specific lag periods. (Gunda, Chimbar, Shamu, Sartorius, & Mukaratin, 2017). Similar to this study, the paper found that precipitation (1- and 3-month lag) and mean temperature (1- and 2-month lag) were significantly associated with incidence at specific lag periods in Ntalale ward. DLNM results suggest a key risk period in current month, based on key past climatic conditions. Mostly, the findings in this study are consistent with the existing literature, supporting the notion that climate variables play a crucial role in the transmission dynamics of malaria. Yoonhee et al.'s (2019) similarly reported that global

and local climate change can alter the spatial and temporal distribution of malaria, increasing opportunities for malaria transmission in traditionally non-malarious areas. Therefore, climate change has helped to create conditions that are conducive to vector breeding in certain areas of Malawi, which can help explain the observed shifts in malaria epidemiology.

Overall, the observed delayed effects of climate factors highlight the significance of climate conditions in the distribution and transmission of malaria. For instance, a study by Odgan et al. (2017) reported similar findings, indicating that diseases such as malaria are intrinsically sensitive to weather and climate. Moreover, our study provides additional evidence by incorporating lag effects, illustrating that the impact of these climatic variables persists over multiple time intervals. The results of the study suggest that the climate conditions experienced in the preceding months play a crucial role in predicting the incidence of malaria in the upcoming months.

The narrow confidence intervals observed in our predicted effects indicate a greater certainty in the association between climatic factors and malaria incidence. This suggests that the relationships identified are robust and consistent across different lag times, lending further credibility to the findings.

2.7. Limitations and recommendations

While the study contributes to the existing literature, it is important to acknowledge some limitations. Firstly, this study focused on a specific geographical region (Mangochi district), and therefore, caution should be exercised when generalizing the results to other locations. Secondly, the use of long monthly lags may affect the accuracy of the analysis, especially when climatic conditions exhibit significant variations over time. It would be beneficial to further investigate the effect of shorter lags, such as weekly data, to capture more immediate associations between climate and malaria incidence. Thirdly, the analysis did not account for important socio-economic factors and other potential confounders that can influence malaria transmission, such as local mosquito control programs. Unfortunately, the data on these factors were not available for inclusion in this study. Future research should strive to incorporate these variables to gain a more comprehensive understanding of the complex interplay between climate, socio-economic factors, and malaria transmission dynamics.

2.8. Conclusion

In conclusion, the study provides further evidence supporting the nonlinear relationship between climate conditions and the risk of malaria. These findings have implications for malaria control and prevention strategies, highlighting the importance of climate monitoring and forecasting in targeted interventions. By considering the impact of climatic factors, public health authorities can develop proactive measures to mitigate the spread of malaria, particularly in regions prone to extreme weather events. Continued research in this field will contribute to a better understanding of the complex interactions between climate and malaria transmission dynamics, ultimately aiding efforts to reduce the burden of this disease.

CHAPTER 3

EVALUATING IMPACT OF INDOOR RESIDUE SPRAY (IRS) MALARIA INTERVENTION WHILE ACCOUNTING FOR LAGGED EFFECTS OF CLIMATE FACTORS IN MANGOCHI, MALAWI

3.1. Introduction

In Malawi, indoor residue spray (IRS) remains key in malaria prevention and control although its implementation is generally low. It is a population level intervention applied at specific period of time which is expected to interrupt long-time trend of malaria in the post-intervention period. Due to ethical and practical barriers, large interventions such as indoor residue spray (IRS) do not have randomized control groups which limit the use of other statistical models such as randomized control trials (RCT) to model effects of an intervention. An attempt to randomize and establish control groups within indoor residue spray (IRS) implementation area may result in partial indoor residue spray (IRS) implementation which is prohibited as it promotes vector resistance (Vector Control Strategy-Malawi, 2015-2019). In the absence of randomization, interrupted time series (ITS) is principally appropriate tool for analyzing observational data where full randomization, or a case-control design, is not affordable or possible (Evangelos, Tim, David, & Iain, 2015).

ITS models such as segmented time series regression can be used to evaluate effectiveness of population-level health interventions such as indoor residue spray (IRS) that have been implemented at a clearly defined point in time (Lopez, Soumerai, & Gasparrini, 2018). The design takes advantage of natural experiments whereby an intervention is introduced at a known point in time and a series of observations on the outcome of interest exist both before and after the intervention (Lopez, Soumerai, & Gasparrini, 2018). ITS accounts for potential risk factors such as long time trends to hypothesize expected scenario under which an intervention had not taken place and the trend continues unchanged ('expected' trend, in the absence of the intervention, given the pre-existing trend) which is referred to as the 'counterfactual'. Interrupted time series designs are immune to many of the threats to validity compared to other observational designs (Wagner, Soumerai, Zhang, & Degnan, 2002).

In this study, a segmented time series regression was applied to evaluate impact of indoor residue spray (IRS) malaria intervention while controlling for delayed effects of climate conditions. Nested models with and without lagged climate variables were compared to select the best model that predict malaria cases. In addition, this study examined the influence of lagged climate variables on parameter estimates when assessing the effectiveness of indoor residue spray (IRS). The model was further used to project the expected outcomes if the intervention had not been implemented. Furthermore, the model can be utilized to predict potential future epidemics based on past climate experiences in Mangochi district.

3.2. Literature review

Although randomized controlled trials (RCTs) are considered the ideal approach for assessing the effectiveness of interventions, many interventions trials can be prohibitively expensive (Evangelos, Tim, David, & Iain, 2015). Evangelos et al (2015) emphasised that even well designed RCTs can be susceptible to systematic errors leading to biased estimates, particularly when generalizing results to "real world" settings. Observational studies address some of these shortcomings, but the lack of researcher control over confounding variables and the difficulty in establishing causation mean that conclusions from studies using observational approaches are generally considered to be weaker. Evangelos further pointed out to the eligibility criteria of RCT which is generally limiting between ranges from 3.5% to 50.7% in some studies due to presence of other constraining conditions. Evangelos et al (2015) however mentioned the strength of quasi-experimental study designs such as ITS which are able to estimate causal effects using observational approaches to evaluate the longitudinal effects of interventions, through regression modelling. ITS can be applied in the absence of randomization and is principally a tool for analyzing observational data where full randomization, or a case-control design, is not affordable or possible. Although other assumptions limit ITS application such as presence of linear trends, the intervention is introduced gradually or at more than onetime point, external time varying effects or autocorrelation (for example, seasonality), or the characteristics of the population change over time, such limitations can be potentially dealt with through modelling if the relevant information is known.

In another study to assess pre-ambulance care program, Monica estimated change in intercept and slope from pre- to post-intervention using segmented regression. The paper mentioned major

strengths of segmented time series regression which included its ability to distinguish the effect of the intervention from secular change, that is, change that would have happened even in the absence of the intervention (Monica, Joanne, Craig, & Jeremy, 2014). The paper further mentioned that with a few simple changes to the data set-up and model specification, segmented regression analysis can easily be implemented in standard statistical software packages. The design is flexible to estimate the effects of different intervention components by adding multiple 'interruptions' to the time series although this requires a sufficient number of time points between interventions for independent effects to be estimated. It also allows for phased and lag implementation by fitting a model with three segments, corresponding to the pre-implementation, implementation, and post-implementation periods. Although it recommended segmented time series analysis for analysis of data from an interrupted time series study, several modifications were proposed to the basic segmented regression analysis approach to deal with challenges arising in the evaluation of complex time series data. Gebski et al (2012) in paper to evaluate impact of prevention and control of infection program in health care, a modified step wedge design was used to model effects that might take weeks or months to become effective and might be implemented in different units at different times (Gebski, Ellingson, Jern, & Kle, 2012).

Segmented time series regression model has also been used in malaria studies to evaluate impact of interventions in malaria control. McLean et al (2018) applied the model to assess effect of integrated community health worker (CHW) programmes in reducing Plasmodium falciparum and Plasmodium vivax malaria incidence and malaria rapid diagnostic test (mRDT) positivity with each year of community health worker (CHW) operation (McLean, Alistair, Aung, Zay, & Hla, 2018). Through the model it was established that communities with CHWs providing malaria diagnosis and treatment experienced declines in *P. falciparum* and *P. vivax* malaria incidence of 70% (95% CI 66–73%) and 64% (59–68%) respectively each year of operation (McLean, Alistair, Aung, Zay, & Hla, 2018).

Faranak et al..,2003 also applied interrupted time series regression to evaluate an intervention to reduce inappropriate use of key antibiotics in the UK. The intervention was a policy for appropriate use of Alert Antibiotics (carbapenems, glycopeptides, amphotericin, ciprofloxacin, linezolid, piperacillin–tazobactam and third-generation cephalosporins) implemented through concurrent, patient-specific feedback by clinical pharmacists. The model captured increased use

of Alert Antibiotics before the intervention started but decreased steadily for 2 years thereafter (Faranak, Kirsteen, Dilip, Gabby, & Simon, 2003).

The main focus of this chapter was to assess the impact of indoor residue spray (IRS) malaria intervention, taking into consideration the influence of lagged and non-linear effects of climate factors. Multiple nested models were constructed and compared using AIC and residue deviance as evaluation metrics. The study specifically examined the delayed effects of climate and its implications on parameter estimation regarding the impact of the intervention. Moreover, the selected model was applied to estimate the effectiveness of indoor residue spray (IRS) and to forecast potential future epidemics, considering past climate experiences and the implementation of the intervention.

3.3. Malaria control in Mangochi

Indoor residual spraying (IRS) is one of the primary vector control interventions for reducing and interrupting malaria transmission (Government of Malawi, 2020). The WHO Global Strategy for Malaria (2016-2030) also recommends that all people living in high Malaria burden areas be protected through the provision, use and timely replacement of long-lasting insecticide treated nets (LLINs), and where appropriate application of indoor residue spray (IRS) (WHO, 2016). In line with this and the 2017-2022 revised Malawi Malaria Strategic Plan, IRS has been prioritized as high impact vector control intervention in high malaria burden districts of Mangochi, Balaka and Nkhata bay (Government of Malawi, 2020). Furthermore, there is entomological evidence that the vectors predominant in these districts are *A. funestus* which exhibit endophagic and endophilic behaviors making indoor residue spray (IRS) intervention suitable (Government of Malawi, 2020).

The national malaria control program conducted indoor residue spraying (IRS) in Mangochi district starting in November 2019. Other malaria control interventions implemented in the district recently include mass net distribution conducted in year 2016, routine net distribution targeting new born babies and pregnant women and larva source management promoted at community level.

3.4. Methods and Materials

3.4.1. Segmented Time Series Regression model

This study utilized a segmented time series regression analysis to assess the effectiveness of the indoor residue spray (IRS) malaria intervention, which was initiated in November 2019 in

Mangochi district. The segmented time series approach allowed for the evaluation of the intervention's impact on malaria incidence and the detection of any notable changes in the malaria trends after the intervention implementation.

The model frame is expressed as follows. Let Y_t be monthly malaria cases, then

$$Y_t \sim NegBinomial(\mu_t)$$

$$log(\mu_t) = \alpha + B_1 * Time\ in\ months + B_2 * Intervention + B_3 * sin\left(\frac{2\pi t}{T}\right) + B_4 * cos\left(\frac{2\pi t}{T}\right) + \sum_t B * Climate\ variables$$

Equation 25

In the model above, Y_t is number of malaria cases, μ_t is expected monthly malaria cases, $log(\mu_t)$ is log-link function, α is baseline intercept, B_1 is a coefficient representing a baseline trend, B_2 is coefficient representing the effect of intervention, $B_3 and\ B_4$ are coefficients for sine and cosine functions to control for malaria seasonality. As evident from the findings of this study, it was observed that climate factors, including rainfall, temperature, and humidity, exhibited delayed effects on malaria incidence. Therefore, delayed climate variables were included in the model represented by $\sum_t B$.

In time series data, seasonality is a common issue where outcomes in one month tend to be more similar to those in neighboring months within the same time of year, leading to autocorrelation (Bernal, Cummins, & Gasparrini, 2016). In this study seasonality was controlled using sine-cosine Fourier functions. To address the issue of over-dispersion in the data, the negative binomial regression model was employed in this study.

3.5. Results

3.5.1. Model building

The study examined nested models that included various components such as the baseline trend, seasonal control by applying sine and cosine functions, intervention effect (indoor residue spray or IRS), and climate variables (lagged and non-lagged).

The nested models used in this study are as follows:

Model-0: This model includes only the intercept term, serving as a baseline reference for comparison, equation 26.

Model-1: In addition to the intercept, this model includes a baseline trend over time, allowing for the analysis of temporal patterns in malaria incidence, equation 27.

Model-2: Along with the intercept and baseline trend, this model incorporates seasonality control by applying sine and cosine functions, capturing the cyclic nature of malaria incidence, equation 28.

Model-3: Building upon the previous models, this model introduces an intervention variable to evaluate the impact of indoor residue spray (IRS) while controlling for the baseline trend and seasonality, equation 29.

Model-4: In addition to the intercept, baseline trend, seasonality, and intervention variable, this model incorporates climate variables that are known to influence malaria. It assesses the direct association between these climate factors and malaria incidence, equation 30.

Model-5: Extending Model-4, this model includes lagged effects of climate variables. The selection of lagged terms is based on the findings from Chapter 2, indicating that the lagged effects peak at lag-0 and remain significant up to lag-3, diminishing thereafter. Thus, all climate variables in Model-5 have lags up to lag-2, representing the influence of previous climate experiences over the past two months, equation 31.

By comparing these nested models, the study aimed to determine the most influential factors and their effects on malaria incidence, providing valuable insights into the dynamics of malaria transmission and the impact of interventions and climate variables.

Mode-0: $log(\mu_t) = \alpha$

Equation 26

Mode-1: $log(\mu_t) = \alpha + B_1 * Time\ in\ months$

Equation 27

Mode-2: $lo\,g(\mu_t) = \alpha + B_1 * Time\ in\ months + B_2 * \sin\left(\frac{2\pi t}{T}\right) + B_3 * \cos\left(\frac{2\pi t}{T}\right)$

Equation 28

Model-3: $lo\,g(\mu_t) = \alpha + B_1 * Time\ in\ month + B_2 * \sin\left(\frac{2\pi t}{T}\right) + B_3 * \cos\left(\frac{2\pi t}{T}\right) + B_4 * Intervention$

Equation 29

Model-4: $lo\,g(\mu_t) = \alpha + B_1 * Time\ in\ month + B_2 * \sin\left(\frac{2\pi t}{T}\right) + B_3 * \cos\left(\frac{2\pi t}{T}\right) + B_4 * Intervention + \sum_t \dot{B_5} * climate\ variables$

Equation 30

Model-5: $lo\,g(\mu_t) = \alpha + B_1 * Time\ in\ month + B_2 * \sin\left(\frac{2\pi t}{T}\right) + B_3 * \cos\left(\frac{2\pi t}{T}\right) + B_4 * Intervention + \sum_t \dot{B_5} * lagged\ climate\ variables$

Equation 31

3.5.2. Model selection

Model selection was based on AIC from the all six models. As indicated in Table 4, model-5 has smallest AIC among all candidate models. This model has baseline trend, seasonality control (sine-cosine functions), intervention (IRS) and lagged climate variables.

Table 4 Comparison of AIC , BIC and QIC for candidate models

Model-ID	Model description	AIC	BIC	QIC
Model 0	Intercept only	1417.165	1423.824	177540.5
Model 1	Intercept and trend	1419.314	1428.192	183929.6
Model 2	Intercept, trend and seasonality	1399.986	1413.303	139525.5
Model 3	Intercept, trend, seasonality and intervention	1379.029	1394.565	98395.56
Model 4	Intercept, trend, seasonality, intervention and non-lagged climate variables	1385.953	1408.148	110041.4
Model 5	Intercept, trend, seasonality, intervention and lagged climate variables	1373.802	1409.314	92346.27

Examination of model residuals through ACF plots in Figure 8 also indicates pictorial diminishing of autocorrelation associated with incorporation of lagged climate variables in model-5. These finding indicate the importance of incorporating delayed effects of climate when modelling malaria incidence.

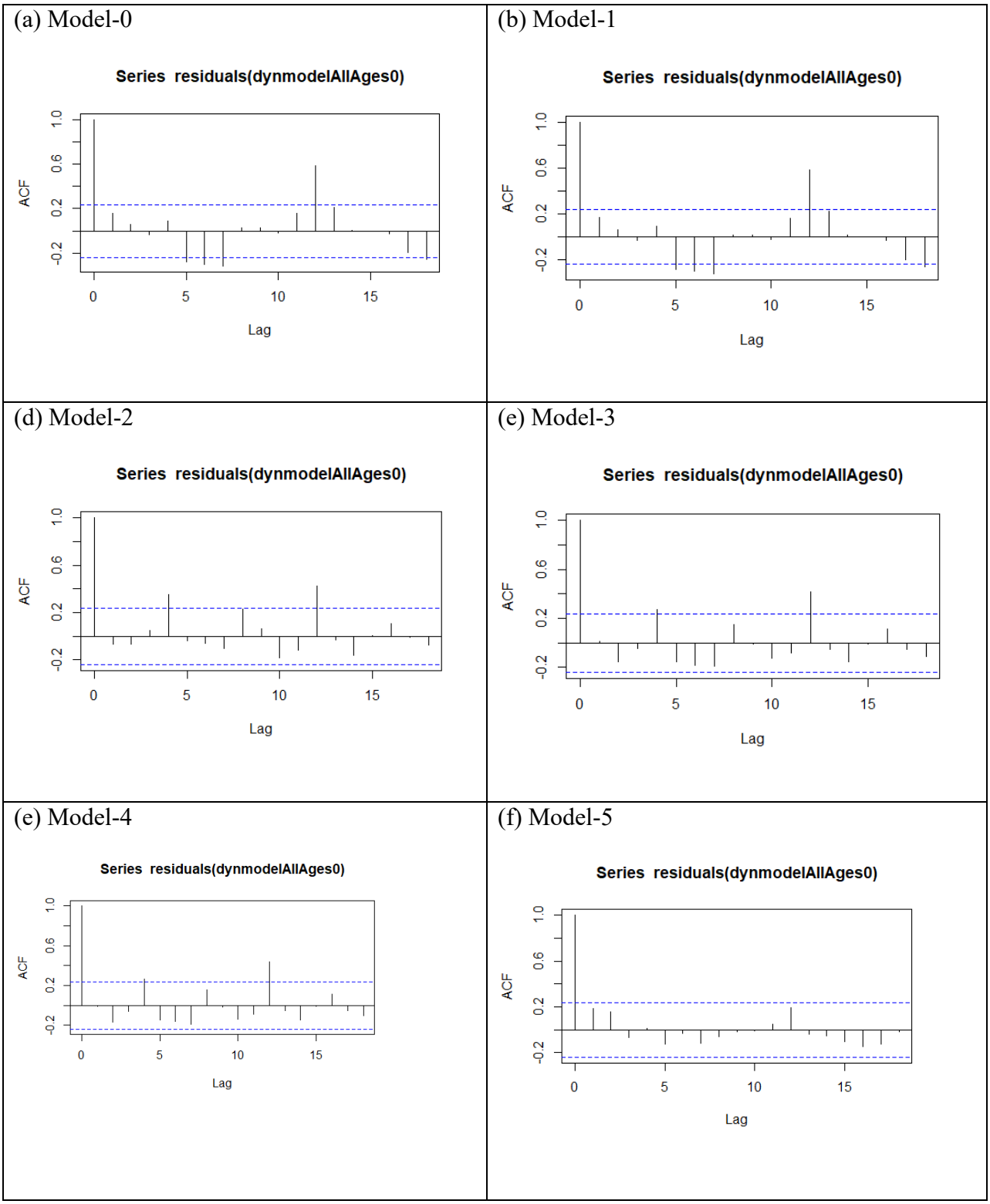


Figure 8 ACF plots of model residues

3.5.3. Model coefficients

The model coefficients for Model 5 are displayed in Table 5, and confidence intervals were examined to assess the significance of these coefficients by checking if they contain a zero effect. While the coefficients for rainfall did not show statistical significance, the study considered its known impact on vector breeding and disease transmission. The coefficients for humidity at lag-1 and lag-2 exhibited statistically significant positive associations with malaria incidence, suggesting that higher humidity levels in previous time periods may increase the likelihood of malaria. Similarly, the coefficients for maximum temperature at lag-1 was statistically significant, indicating that higher temperature levels in the previous month may also contribute to an increased likelihood of malaria. These findings align with results observed when investigating the relationships between these climate factors and disease transmission dynamics in the previous chapter.

Table 5 Model coefficients for lagged climate variable model

Coefficients:	Estimate	Std.Error	CI(lower)	CI(upper)
(Intercept)	2.58	2.093205	-1.527316	6.67789
beta_1	0.0377	0.028519	-0.018176	0.09361
sine	-0.442	0.219964	-0.872681	-0.01044
cosine	0.196	0.156435	-0.11074	0.50247
Trend	0.00455	0.002494	-0.000339	0.00944
Intervention	-0.647	0.106034	-0.854488	-0.43884
Rainfall lag-0	-0.0000358	0.000589	-0.001191	0.00112
Rainfall lag-1	-0.000177	0.000606	-0.001365	0.00101
Rainfall lag-2	0.000123	0.000582	-0.001018	0.00126
Humidity lag-0	0.00432	0.007054	-0.009506	0.01815
Humidity lag-1	0.0322	0.007399	0.017663	0.04667
Humidity lag-2	0.0178	0.00716	0.003771	0.03184
Maximum temp lag-0	0.0069	0.026026	-0.04411	0.05791
Maximum temp lag-1	0.0718	0.027937	0.017085	0.1266
Maximum temp lag-2	0.0377	0.027803	-0.016798	0.09219

3.5.4. Impact of lagged climate effect in modeling

The impact of indoor residue spray (IRS) intervention on malaria incidence was assessed using different candidate models, and the results are presented in Table 6. All models demonstrated a reduction in malaria incidence associated with IRS intervention. However, the magnitude of the impact varied across the different models. Notably, when climate variables were incorporated in models 4 and 5, the estimated impact changed from 48.7% to 46.93% and 47.62%, respectively. These findings highlight the importance of considering climate factors when modeling malaria incidence, as omitting them can potentially influence estimates of the true impact of interventions.

Table 6 Model parameters (exponentiated)

Model	Intercept	Trend	Intervention	AIC	Estimated % reduction of cases
Model 3	2078.18604	1.007421	0.512739	1379.029	48.73%
Model 4	7.383166	0.006607004	0.5306841	1385.953	46.93%
Model 5	13.135119	1.0045596	0.5237897	1373.802	47.62%

3.5.5. Impact of indoor residue spray (IRS) malaria intervention in Mangochi

The study utilized model-5 to evaluate the impact of IRS on malaria cases, considering the delayed effects of climate factors. The findings revealed a significant reduction of malaria cases by 48% (CI: 46%-49%) in the general population due to IRS intervention. However, when analyzing the data based on different age groups and accounting for lagged effects of climate factors, varying impacts were observed, as presented in Table 7. Notably, the under-5 age category exhibited a substantial reduction in malaria cases, with IRS resulting in a 51% decrease (CI: 49%-54%), while the over-5 category experienced a lower reduction of 44% (CI: 43%-47%).

Table 7 Model estimates for impact by age category

		95% CI	
Age group	Estimated % reduction of cases	Lower	Upper
All population	48%	46%	49%
Under - 5	51%	49%	54%
Over-5	44%	43%	47%

The graphs in Figure 9 below shows long time trends that hypothesize expected scenario under which an intervention had not taken place and the trend continues unchanged ('expected' trend, in the absence of the intervention, given the pre-existing trend) which is referred to as the 'counterfactual'. The counter factual is obtained by removing intervention effect so that the historical trend is allowed to continue without interruption beyond in the post intervention period as shown below.

$$Counter\ factual = \frac{Model\ Predicted}{exp(B_2 * Intervention)}$$

Equation 31

In all age groups the counter factual trend depicts high malaria cases compared to what is observed and predicted by the model as shown in Figure 9*a-c*.

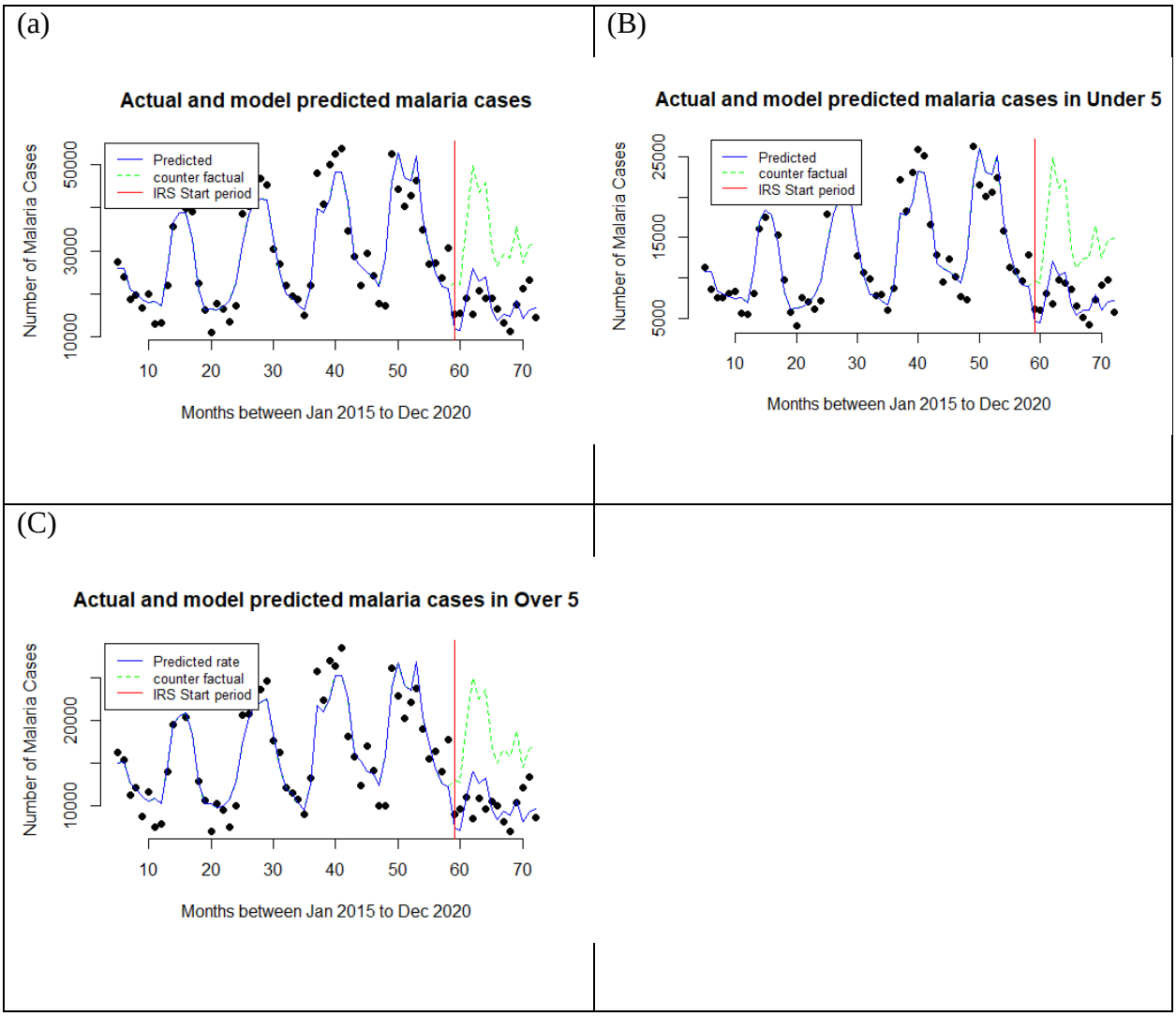


Figure 9 Graph showing predicted monthly incidence against counterfactual climate adjusted for climate factors

3.6. Discussion

The findings of this study align with the existing literature on the effectiveness of interrupted time series (ITS) designs to evaluate the effectiveness of interventions in real-world settings where randomized controlled trials (RCTs) may not be feasible or practical (Evangelos et al., 2015). The study further showed that incorporating lagged climate effects in ITS models improved the accuracy of modelling impact of IRS interventions. The study also demonstrated that lagged climate conditions can significantly impact disease transmission dynamics and, therefore, should be accounted for when evaluating impact of an intervention such as IRS.

Generally, the study indicates a significant reduction in malaria cases associated with IRS intervention. Furthermore, it has been demonstrated that the delayed effects of climate conditions significantly influenced malaria incidence during specific periods. Therefore, when evaluating the impact of interventions, it is crucial to thoroughly consider and account for the influence of lagged climate conditions. After accounting for lagged climate conditions, the study revealed a notable reduction in malaria incidence by 48% (CI: 46%-49%) associated with the implementation of Indoor Residue Spray (IRS) in the general population. However, when analyzing the data based on different age groups, the study found varying impacts. The under-5 age category exhibited a significant reduction in malaria cases (51%, CI: 49%-54%), whereas the over-5 category experienced a comparatively lower reduction (44%, CI: 43%-47%). The finding is important considering previous research that has highlighted the vulnerability of young children to severe malaria and the potential benefits of targeted interventions for this age group such as a study by Kazembe et al. (2015). The study by McLean et al. (2018) also reported differential impacts of community health worker (CHW) programs on malaria incidence among different population groups.

The analysis also revealed a distinct counterfactual trend, demonstrating a higher incidence of malaria cases compared to what was observed and predicted by the model. The substantial reduction of 48% (CI: 46%-49%) associated with the implementation of Indoor Residue Spray (IRS) supports the efficacy of this intervention strategy in combating malaria.

3.7. Limitations and recommendations

The analysis in this study did not account for socio-economic factors and other potential confounders, such as local mosquito control programs, which can also influence malaria transmission. Unfortunately, these data were not available for inclusion in the analysis. Furthermore, it should be noted that a time-varying population offset was not used because TScount R package did not support the implementation of a variable offset at the time of this study. However, it is important to mention that the influence of population changes over the monthly time period is expected to be minimal in this study, as there were negligible population variations observed between the months.

3.8. Conclusion

In conclusion, the study highlights the significant impact of IRS intervention on reducing malaria incidence in Malawi. The inclusion of climate variables and accounting for lagged effects enhanced the accuracy of the predictions and provided a comprehensive understanding of the intervention's effects. The findings support the continued implementation and scale-up of IRS as a population-level intervention for malaria prevention and control, particularly in areas with high malaria burden. Furthermore, the differential impact observed among age groups emphasizes the importance of targeting interventions to specific populations, such as young children who are at higher risk of malaria. These findings contribute to the designing effective malaria control programs that consider both intervention strategies and environmental factors.

CHAPTER 4

CONCLUSION

Overall, the findings of this study contribute to the growing body of literature on the relationship between climatic factors and malaria incidence. The study confirms the presence of a nonlinear relationship between rainfall, humidity, maximum temperature, and the risk of malaria, consistent with previous research. These findings align with studies that emphasize the role of climate variables in the transmission dynamics of malaria.

The study highlights the significance of rainfall in creating breeding grounds for mosquito vectors and increasing the prevalence of malaria cases. This finding is supported by previous research, indicating that higher rainfall is associated with a higher risk of malaria. The study also underscores the impact of humidity on mosquito populations and subsequent malaria transmission, in line with previous studies highlighting the positive association between humidity and malaria risk. Similarly, the study demonstrates the importance of temperature in shaping the dynamics of malaria transmission, particularly in relation to mosquito larval development and survival.

The study incorporated lagged effects, considering that the impact of climatic variables persists over multiple time intervals. The findings align with previous research emphasizing the sensitivity of malaria and other diseases to weather and climate conditions. The study's narrow confidence intervals provide a higher level of certainty in the association between climatic factors and malaria incidence, further supporting the robustness of the findings.

However, the study has limitations. It focused on a specific geographical region, limiting the generalizability of the results. The use of long monthly lags may affect the accuracy of the analysis, and future research could explore the effect of shorter lags to capture more immediate associations. The study did not account for important socio-economic factors and potential confounders, such as local mosquito control programs, which can influence malaria transmission. Incorporating these variables in future research would provide a more comprehensive understanding of the complex interplay between climate, socio-economic factors, and malaria transmission dynamics.

In conclusion, the study contributes to the understanding of the nonlinear relationship between climatic factors and malaria incidence. The findings have implications for malaria control and prevention strategies, emphasizing the importance of climate monitoring and forecasting in targeted interventions. By considering the impact of climatic factors, public health authorities can develop proactive measures to mitigate the spread of malaria, particularly in regions prone to extreme weather events. Continued research in this field will further enhance efforts to reduce the burden of malaria by unraveling the complex interactions between climate and malaria transmission dynamics

REFERENCES

Ayansina, A., Isioma, N., Consolato, S., & Oluwatoyin, A. (2020). Early warning climate indices for malaria and meningitis in tropical ecological zones. *Scientific Reports, 10*(1), 1-4.

Benjamin, K., & Konstantinos, F. (2002). *Regression Models for Time Series Analysis* (1 ed., Vol. 45). John Wiley & Sons, Inc.

Bernal, L. J., Cummins, S., & Gasparrini, A. (2016). Interrupted time series regression for the evaluation of public health interventions. *Education Corner, 1*(1), 1-3.

Chilanga, E., Delphine, C.-V., Heather, M., & Claudia, M. (2020). Prevalence and determinants of malaria infection among children of local farmers in Central Malaw. *Malaria Journal, 1*, 1-6.

Chuang, & Ting, W. (2017). Assessment of climate-driven variations in malaria incidence in Swaziland. *Malaria Journal, 16*(1), 1-3.

Cyril, C., Marie, M., & Annie, J. (2018). Impact of recent and future climate change on vector-borne diseases. *Climate Sciences*.

Evangelos, K., Tim, D., David, S., & Iain, B. (2015). Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *Research Methods & Reporting, 350*, 1-4.

Faranak, A., Kirsteen, G., Dilip, N., Gabby, P., & Simon, O. (2003). Outcomes of an intervention to improve hospital antibiotic prescribing: Interrupted time series with segmented regression analysis. *Journal of Antimicrobial Chemotherapy, 52*(5), 1-3.

Florence, F., & John, R. (2019). Impact of past and on-going changes on climate and weather on vector-borne diseases transmission: a look at the evidence. *Infectious Diseases of Poverty, 8*(1), 1-6.

Gasparrini, A. (2011). Distributed Lag Linear and Non-Linear Models in R. *Journal of Statistical Software, 13*(1), 1-7.

Gebski, Ellingson, Jern, I., & Kle, I. (2012). Modelling interrupted time series to evaluate prevention and control of infection in healthcare. *Epidemiol. Infection, 140*(12), 1-3.

Godwin, A., Kayode, A., & Olakunle, O. (2019). Potential distribution of dominant malaria vector species in tropical region under climate change scenarios. *PLOS/ONE, 14*(6), 1-3.

Government of Malawi. (2020). *Revised Malaria Strategic Plan.* Lilongwe: NMCP.

Gunda, R., Chimbar, M., Shamu, S., Sartorius, B., & Mukaratin. (2017). Malaria incidence trends and their association with climate variables in rural Gwanda. *Malaria Journa, 16*(1), 1-3.

Hertig, E. (2019). Distribution of Anopheles vectors and potential malaria transmission stability in Europe and the Mediterranean area under future climate change. *Parasites & Vectors, 12*(1), 1-2.

HMIS. (2020). *Health Mangemnet Information System* . Lilongwe: MoH.

Jae, P., Hae, C., & Young, K. (2012). Estimated Effect of Climatic Variables on the Transmission of Plasmodium vivax Malaria in the Republic of Korea. *Environmental Health Perspectives, 120*(9), 1-3.

Jung Eun Kim, Y. C. (2018). Effects of climate change on Plasmodium vivax malaria transmission dynamics: A mathematical modeling approach. *Applied Mathematics and Computation, 347*(0), 616-618.

Kazembe, L., Kleinschmidt, I., Holtz, T., & Sharp, B. (2006). Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. *International journal of health geographics, 5*(1), 2-3.

Lisbeth, A. H., Jose, C., Chystrie, R., & Milagros, C. (2017). Climatic fuctuations and malaria transmission dynamics, prior to elimination, in Guna Yala, República de Panamá. *Malaria Journal, 17*(1), 1-3.

Lopez, B., Soumerai, & Gasparrini. (2018). A methodological framework for model selection in interrupted time series studies. *Journal of clinical epidemiology, 103*(1), 1-5.

Malawi Government. (2017). *Malawi Malaria Indicator Survey.* Lilongwe: Malawi Ministry of Health.

Malawi Housing and Population Census. (2018). Zomba, Malawi: NSO.

Marcia, & Castro. (2017). Malaria Transmission and Prospects for Malaria Eradication: The Role of the Environment. *Cold spring harbor perspective in medicine, 1*(1), 1-3.

McLean, Alistair, Aung, T. M., Zay, S. K., & Hla, P. W. (2018). Malaria elimination in remote communities requires integration of malaria control activities into general health care: an observational study and interrupted time series analysis in Myanmar. *BMC Medicine, 16*(1), 1-7.

Ministry of health. (2019). *2017-2022 Malaria Strategic Plan Mid-Term Review Report.* Lilongwe: NMCP.

Monica, T., Joanne, M., Craig, R., & Jeremy, G. (2014). The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care. *Implementaion science, 9*(1), 1-4.

Mukhopadhyay, Tiwari, Shetty, Gogtay, & Thatte. (2019). Modeling and Forecasting Indian Malaria Incidence Using Generalized Time Series Models. *Communications in Statistics: Case Studies, Data Analysis and Applications, 1*(1), 1-3.

Ogden, N. (2017). Climate change and vector-borne diseases of public health significance. *Environmental Microbiology, 364*(19), 1-6.

Phillippe, M., & Myriam, L. (1995). Malaria and Climate: Sensitivity of Malaria Potential Transmission to Climate. *Royal Swedish Academy of Sciences, 24*(4), 200-207.

Qinqin, X., Runzi, L., Shannon, R., Cheng, L., & Yafei, L. (2018). Using a distributed lag non-linear model to identify impact of temperature variables on haemorrhagic fever with renal syndrome in Shandong Province. *Epidemiology and Infection, 1*(1), 2-4.

Soma, S., Vinay, G., Poonam, S., & Ramesh, D. (2019). Shift in Potential Malaria Transmission Areas in India, Using the Fuzzy-Based Climate Suitability Malaria Transmission (FCSMT) Model under Changing Climatic Conditions. *Environmental Research and Public health, 16*(18), 2-3.

Steffen, E., & Abba, G. (2018). Mathematical modeling of climate change and malaria transmission dynamics: a historical review. *Mathematical biology, 1*(1), 1-4.

The INFORM Project. (2014). *An Epidemiological Profile of Malaria and its Control in Malawi.* Blantyre: Malaria Alert Centre.

Trocaire. (n.d.). *Malawi climate change case study.* Filling the heat.

Varvara, M., Natalia, S., Mikhail, V., & Mikhail, G. (2019). Effects of Climate Change and Heterogeneity of Local Climates on the Development of Malaria Parasite (Plasmodium vivax) in Moscow Megacity Region. *Environmental Research and Public health, 16*(5), 1-3.

(2015-2019). *Vector Control Strategy-Malawi.* Lilongwe: NMCP.

Vincent, & Katharine. (2020). *Future Climate Projections for Malawi.*

Wagner, Soumerai, Zhang, & Degnan, R. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics, 27*(4), 1-3.

WHO. (2016). *GLOBAL TECHNICAL STRATEGY FOR MALARIA 2016–2030.* United Kingdom: WHO Library Cataloguing-in-Publication Data.

Xin, Z., Cui, Z., & Deirdre, J. (2020). Host-Malaria Parasite Interactions and Impacts on Mutual Evolution. *Frontiers in cellurar and infection microbiology, 10,* 1-5.

Yeh, K., Fair, J., Smith, W., & Torres, T. (2020). Assessing Climate Change Impact on Ecosystems and Infectious Disease: Important Roles for Genomic Sequencing and a One Health Perspective. *Tropical Medicine and Infectious Diseases, 5*(2), 1-5.

Ying, Z., Peng, B., & Janet, H. (2010). Meteorological variables and malaria in a Chinese temperate city: A twenty-year time-series data analysis. *Environment International, 36*(5), 1-2.

Yoonhee, K., Ratnam, Takeshi, D., Yushi, M., & Swadhin, B. (2019). Malaria predictions based on seasonal climate forecasts in South Africa: A time series distributed lag nonlinear model. *NatureResearch, 9*(1), 2-8.

APPENDIX

Original Research Article

Modelling lagged and nonlinear effects of climate factors on malaria in Mangochi, Malawi

Godfrey Silungwe[1], Lawrence N. Kazembe[2]

Department of Mathematical Sciences, University of Malawi

Article Info

Publishable academic manuscript

Key words

DLNM

Abstract

This manuscript presents a study that aimed to analyze the relationship between climate factors and malaria incidence. The research utilized monthly malaria incidence data from the national malaria control program and climate data from the Department of Meteorological Services and Climate Change in Mangochi district, Malawi. The analysis employed a distributed lag non-linear model to examine the nature of the relationship between climate variables and malaria incidence. The results revealed an immediate peak in malaria risk in the same month following extreme weather conditions, highlighting the importance of short-term effects. In this study, the risk of malaria immediately doubles with extreme rains and humidity compared to risk at average climate conditions. The effects continued for up to two months but gradually subsided thereafter. These findings underscore the significance of considering previous climate conditions in predicting current and future malaria incidence. The study emphasizes the importance of understanding the relationship between climate and malaria incidence for informing targeted interventions, establishing malaria early warning systems, and mitigating the effects of climate change on malaria transmission. Incorporating the knowledge gained from this research into malaria programming and control efforts can enhance the effectiveness of interventions and contribute to proactive strategies for reducing the burden of malaria in the context of a changing climate.

1. Introduction

Malaria transmission is influenced by natural risk factors such as rainfall patterns, temperature, and humidity, which affect the spread of the malaria parasite (Ayansina, Isioma, Consolato, & Oluwatoyin, 2020). Increased rainfall leads to the proliferation of mosquito breeding sites, thereby increasing the transmission of malaria parasites among individuals. Similarly, temperature and humidity impact malaria transmission by regulating the rate of mosquito larvae development and mosquito survival rates (Gunda, Chimbar, Shamu, Sartorius, & Mukaratin, 2017). Studies, such as the one conducted by Gunda et al. (2017) in rural Gwanda, Zimbabwe, have established the link between climate variability and vector-borne diseases like malaria. Climate variability has the potential to either facilitate or hinder efforts to control the disease (Gunda, Chimbar, Shamu, Sartorius, & Mukaratin, 2017).

Understanding the variation in malaria incidence due to climate variability, both present and recent past, is crucial for planning future malaria control programs (Gunda, Chimbar, Shamu, Sartorius, & Mukaratin, 2017). It enables the identification of optimal timing for implementing malaria interventions, taking into account the impact of previous climate conditions. This understanding serves as a valuable tool for program implementers, allowing them to incorporate past climate experiences into decision-making processes and enhance the effectiveness of malaria control initiatives. Furthermore, understanding the relationship between malaria incidence and lagged climatic conditions contributes to the development of robust malaria early warning systems.

Examining the influence of past climate experiences on malaria transmission is also vital for understanding local epidemiological shifts, some of which can be attributed to climate change. Future climate projections indicate a general warming trend, particularly in southern Malawi and over the lake (Vincent & Katharine, 2020). However, different climate models provide varying predictions regarding rainfall patterns.

These changing climatic conditions highlight the importance of considering the complex interactions between climate factors and malaria transmission dynamics. The understanding of local climate patterns and its potential impact to malaria is crucial for developing effective strategies to mitigate the disease and adapt to future climate challenges.

Previous studies however commonly overlook the existence of lagged and non-linear relationship between climate conditions and malaria incidence. This study utilized methods that capture lagged and non-linear relationships more accurately. By applying appropriate methodologies to account for delayed climate effects, this study provides a more comprehensive and accurate analysis of the influence of climate on malaria transmission in the Mangochi district.

2. Distributed lag non-linear model

The distributed lag non-linear model (DLNM) is a methodology used to model the non-linear and delayed effects of environmental stressors or events. It incorporates a combination of past exposures over several time lags to explain the impact of the stressors at a given time. The DLNM is based on a bi-dimensional space of functions called "cross-basis" that describes the shape of the relationship along the predictor (e.g., temperature) and its lag dimension.

Equation 1: Distributed Lag Model (DLM)

$$y_t = \alpha + B_0 x_1 + B_1 x_{t-1} + .. + B_L x_{t-L} + u$$

In the DLM, y_t represents the response variable at time t, x is the regressor, and $B_L$represents the weight of the respective lag L. The coefficients B_L can be interpreted as the effect of the past exposure on the present moment response or the effect of the current exposure on the future response.

Equation 2: Dependency on Exposure History

$$S(x,t) = \int_{L_0}^{L} X_{t-l} w(L)\, dL$$

Equation 2 describes the dependency of the response on exposure history. S(x,t) represents the exposure-response

relationship at time t, X_{t-l} is the lagged exposure, and w(L) is the weighting basis function applied to constrain the lag coefficients. The integral sums up the effect of exposures over different time lags.

Equation 3: Cross-Basis Function in DLNM

$$S(x_t, n) = q_r n = w_t^T n$$

In the DLNM, the cross-basis function represents the non-linear relationship between the predictor variable x and the response variable y over time. The function $S(x_t, n)$ is a linear combination of the basis functions f(x) and w(l), which describe the exposure-response structure along x and the lag-response structure along l, respectively.

Equation 4: DLNM Equation

$$g(u_t) = \alpha + \sum_{j=1}^{J} s_j(x_{xj}; n_j) + \sum_{k=1}^{k} \gamma_k Z_{tk}$$

The DLNM equation incorporates the cross-basis function s_j, which denotes smoothed relationships between the predictor variables x_{xj} and the linear predictor. The model assumes that the response variable Y follows a distribution from the exponential family. Other predictor variables Z_{tk} with linear effects are also included, and the parameters α, n_j, and γ_k are estimated.

These equations capture the key concepts and relationships within the distributed lag non-linear model methodology. They provide a basis for understanding the modeling approach and how it incorporates the non-linear and delayed effects of environmental stressors.

3. Application: analysis of results

A distributed lag non-linear model (DLNM) was applied to analyze the relationship between climate factors and malaria incidence in Mangochi district. The model used monthly malaria cases as the dependent variable and climate variables, such as temperature, rainfall, and relative humidity, as the independent variables.

The equation used in the DLNM was:

$$Y_t \sim NegBinomial(\mu_t)$$

$$log(u_t) = \alpha + \sum_{j=1}^{J} S_j(x_{t,j}, B_j) + \sum_{k=1}^{K} \psi_k(Z_{tk}) + s(time, p) + log(population)$$

- The dependent variable, u_t, represents the expected malaria cases on monthly basis.
- The independent variables are the climate factors, including monthly temperature, monthly rainfall, and monthly relative humidity.
- NegBinomial is the negative binomial distribution used to account for overdispersion in the data.
- α is the intercept term.
- $S_j(x_{t,j}, B_j)$ captures the non-linear effects of climate factors at different lag months.
- $\psi_k(z_{tk})$ represents the linear effects of other predictors, such as the IRS intervention.
- $S_j(x_{t,j}, B_j)$ is a natural cubic spline used to control for seasonality in the malaria time series.
- log(population) is the offset term that adjusts for changes in population size over time.

By applying this model and estimating the parameters, the researchers aimed to understand the relationship between climate factors and malaria incidence in Mangochi district, considering the non-linear effects, delayed effects, and potential confounding factors.

3.1.Distributed Lag Non-Linear Model building

The correlation between climate variables was assed using Pearson correlation to identify climate variables that exhibit excessive collinearity. The correlation matrix, as shown in Table 2, revealed positive correlations among the climate variables. Specifically, the maximum and minimum temperatures exhibited a correlation coefficient of 0.81, indicating a strong positive relationship. Similarly, there was a positive correlation between the minimum temperature and rainfall, with a correlation coefficient of 0.54. Furthermore, the humidity and rainfall variables showed a positive correlation, with a correlation coefficient of 0.51.

These findings suggested some degree of collinearity among climate variables which pose challenges in the statistical analysis as it can lead to multicollinearity.

Table 2 Correlation matrix of climate variables

Climate variables	Malaria Total	Rain	Humidity	Max Tem	Min Tem
Malaria Total	1.00	0.112	0.567	-0.284	-0.078
Rain	0.112	1.00	0.512	0.133	0.536
Humidity	0.567	0.512	1.00	-0.377	0.122
Maximum Temp	-0.284	0.132	-0.377	1.00	0.813
Minimum Temp	-0.078	0.537	0.122	0.811	1.00

In order to assess the impact of collinearity in the final model, the Variance Inflation Factor (VIF) was applied. The VIF is a statistical index that quantifies the extent to which the variance of the estimated regression coefficients is increased due to collinearity among the independent variables. The researcher followed the recommendation by Qinqin et al. (2018), who suggested removing variables with a VIF above 5 to minimize the impact of collinearity on model sensitivity. Following VIF analysis, variables with high VIF were dropped and the resultant model had reduced variance inflation factor among all climate variables.

3.2. **Exposure Lag Response relationship**

In the model selection process, a variety of linear and non-linear functions in the exposure-lag dimensions for climate variables were considered. A total of 18 different model candidates were evaluated, each with different specifications in the exposure-lag dimensions.

To determine the most appropriate model specifications, the Akaike Information Criterion (AIC) was used. The AIC provides a measure of the model's goodness of fit while considering its complexity. By comparing the AIC values across different functions in the exposure-lag dimensions for climate variables, the researchers identified the model specifications that yielded lower AIC values.

The examination of AIC revealed that using natural cubic splines with knots placed at equal intervals in the exposure dimension and a polynomial in the lag response dimension resulted in the lowest AIC values.

Backward variable selection was then performed to further refine the model. This process involved generating basis variables based on the selected model specifications and incorporating them into the final model.

The backward variable selection method demonstrated improvement in the model when including these climate basis variables.

To control for seasonality, natural cubic splines were employed. Additionally, the model included the predictor of Indoor Residual Spraying (IRS) intervention period.

3.3.Lagged effects of rainfall

Figure 5a illustrates a nonlinear relationship between precipitation and malaria incidence. The study found that increasing precipitation compared to its average is associated with a higher risk of malaria, reaching its peak at lag 0 when monthly rainfall reaches a maximum of 541mm. The relative risk (RR) at this point is 2.4314162, with a 95% confidence interval (CI) of (2.0554548, 2.8761443).

The predicted effects of rainfall on malaria incidence remain consistent across all three lag periods and exhibit a narrow confidence interval, as depicted in Figure 5d. Figure 5b which shows a heat map, representing relative risk, visually shows a high risk associated with increasing rainfall, peaking at lag-0 (bright red). Additionally, the study identified delayed but diminishing effects of climate conditions from 0-3 months lag, indicating that the impact of climatic variables persists over multiple time intervals.

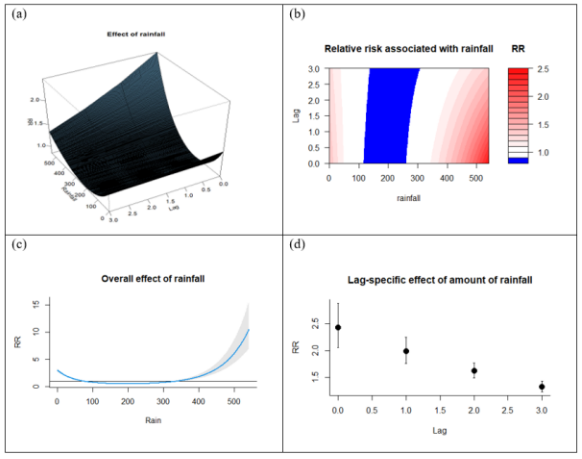


Figure 5 3D relationship, relative risk and lag specific effects between malaria cases and rainfall

3.4.Lagged effects of humidity

Figure 6a depicts a nonlinear relationship between humidity and malaria incidence. The study revealed that increasing humidity compared to its average is associated with a higher relative risk, peaking at lag 0 when relative humidity reaches a maximum of 84. At this point, the relative risk (RR) is 2.0537649, with a 95% confidence interval (CI) of (1.9435101, 2.1702744).

The effects of humidity on malaria incidence exhibit a narrow confidence interval and are sustained up to lag-2, as shown in Figure 6d. Figure 6b which shows risk heat map, representing relative risk, visually illustrates a high risk associated with increasing humidity, similarly reaching its peak at lag-0.

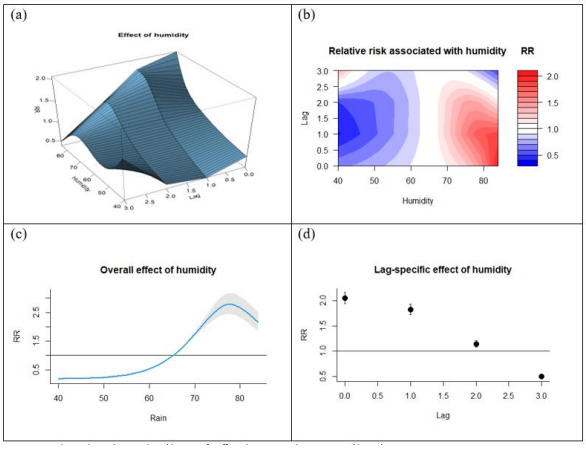


Figure 6 3D relationship, relative risk and lag specific effects between malaria cases and humidity

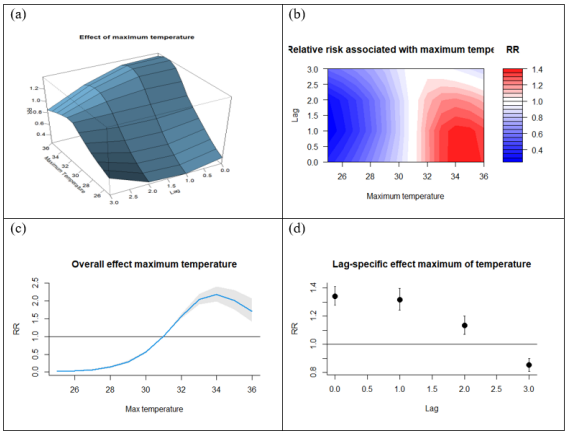


Figure 7 3D relationship, relative risk and lag specific effects between malaria cases and maximum temperature

3.5.Lagged effects of temperature

Figure 7a depicts a nonlinear exposure lag response relationship between maximum temperature and malaria cases. The analysis revealed that increasing temperature compared to its average is associated with an elevated risk of malaria, with the risk peaking at lag 0 when the temperature reaches a maximum of 34 degrees Celsius. At this point, the relative risk (RR) is 1.3907299, with a 95% confidence interval (CI) of (1.3523290, 1.4302212).

The effects of temperature on malaria cases are sustained up to lag 2, as shown in Figure 7d, and the predicted effects exhibit narrow confidence intervals across all lags. Figure 7b which shows risk heat map, representing relative risk, visually indicates a high risk associated with increasing maximum temperature, reaching its peak at lag-0.

4. Discussion and Conclusion

The findings in this study are consistent with previous reports such as Gunda et al. (2017), which demonstrated a significant association between malaria incidence and precipitation at specific lag periods. Similar to the current study, Gunda et al. found that precipitation (at 1- and 3-month lags) and mean temperature (at 1- and 2-month lags) were significantly associated with malaria incidence at specific lag periods. The distributed lag non-linear model (DLNM) analysis in this study also suggests present high-risk period based on past 3-months rainfall conditions. Specifically, the results in this study highlight a peak in malaria risk which doubled immediately following extreme rainfall conditions, aligning with Florence et al.'s (2019) report linking climate extreme events with malaria outbreaks. The effect of rainfall continued to be sustained for a period of 3 months and diminishes beyond 3 months.

This study, in conjunction with previous research, provide evidence that changes in rainfall pattern could contribute to the creation of conditions conducive to vector breeding, which could explain the observed shifts in malaria epidemiology in certain

areas of Malawi. A study by Yoonhee et al. (2019) reported that global and local climate change can alter the spatial and temporal distribution of malaria, increasing opportunities for transmission in traditionally non-malarious areas.

The impact of humidity on malaria incidence has also been extensively studied, and the findings of this study reaffirm the positive association between humidity and malaria risk. The observed peak in malaria risk which doubles immediately following highest recorded relative humidity, supports the understanding that humidity plays a crucial role in influencing mosquito populations and subsequent malaria transmission. This result aligns with Philippe et al.'s (1995) report, which states that higher humidity levels prolong the lifespan of mosquitoes, enabling them to infect more individuals.

Similarly, the nonlinear relationship between temperature and malaria cases is consistent with previous research that has established a connection between the disease and temperature conditions. The observed peak in malaria risk following high temperatures emphasizes the importance of temperature in shaping the dynamics of malaria transmission. As reported by Gunda et al. (2017), temperature affects malaria transmission by regulating the rate of development of mosquito larvae, which in turn influences mosquito survival rates. Thus, considering temperature as a crucial factor in understanding and addressing malaria transmission dynamics is essential.

In conclusion, the findings of this study contribute to the growing body of literature on the relationship between climatic factors and malaria incidence. The results demonstrate a delayed and nonlinear relationship between climate conditions and the risk of malaria, consistent with previous research. The study found that precipitation and mean temperature at specific lag periods were significantly associated with malaria incidence. There was a peak in malaria risk following extreme weather conditions, highlighting the importance of short-term effects. The positive association between increasing rainfall and malaria risk confirms the influence of precipitation on mosquito breeding and the prevalence of malaria cases. Humidity was also found to be positively associated with malaria risk, influencing mosquito populations. Similarly, high temperatures were associated with an increased risk of malaria. The observed delayed effects of climate factors underscore the significance of climate conditions in malaria distribution and transmission. Overall, these findings enhance our understanding of the impact of climatic variables on malaria epidemiology.

Original Research Article

Evaluating impact of indoor residue spray (IRS) malaria intervention while accounting for lagged effects of climate factors in Mangochi, Malawi.

Godfrey Silungwe[1], Lawrence N. Kazembe[2]

Department of Mathematical Sciences, University of Malawi

Article Info

Publishable academic manuscript

Key words

IRS
Segmented time series regression

Abstract

This objective focused on the evaluation of the impact of Indoor Residue Spray (IRS) intervention on malaria incidence in Mangochi district, Malawi, using a segmented time series regression. The study incorporated lagged climate conditions to improve the modeling and evaluation of IRS intervention. The analysis revealed a significant reduction in malaria cases associated with the implementation of IRS, with a notable overall decrease of 48% (CI: 46%-49%) in the general population. Different age groups exhibited varying impacts, with the under-5 age category experiencing a significant reduction of 51% (CI: 49%-54%). These findings supported the efficacy of IRS as a malaria prevention strategy and highlighted the importance of considering lagged climate conditions in the evaluation of interventions. The study contributed to the literature on interrupted time series designs and their application in evaluating population-level health interventions.

4. Introduction

This study aimed to evaluate the impact of Indoor Residue Spray (IRS) intervention on malaria incidence in Mangochi district, Malawi, using a segmented time series regression analysis. An interrupted time series (ITS) was employed as an alternative to randomized controlled trials (RCTs) in assessing the effectiveness of interventions in real-world settings. This approach was particularly useful for population-level interventions such as IRS, where RCTs were impractical and costly in resource-constrained settings. These designs allowed for the analysis of observational data in the absence of full randomization or a case-control design.

Segmented time series regression analysis was particularly well-suited for evaluating interventions with clearly defined implementation periods. By distinguishing the effects of interventions from secular trends, this approach provided a robust framework for assessing the impact of IRS interventions on malaria incidence. Previous studies, such as Monica et al. (2014) and Gebski et al. (2012), had demonstrated the effectiveness of segmented regression analysis in various healthcare settings.

In the context of malaria control, segmented time series regression had been used to evaluate the effects of interventions, such as community health worker programs (McLean et al., 2018). These studies had shown promising results in reducing malaria incidence and had highlighted the potential of this analytical approach.

The study aimed at assessing the impact of IRS intervention on malaria incidence in

Mangochi district, Malawi. Furthermore, the study incorporated lagged and non-linear climate effects in the analysis to account for the influence of climate conditions on disease transmission dynamics. By applying segmented time series regression, the study estimated the causal effects of the IRS intervention and provide an opportunity to forecast potential future epidemics based on past climate experiences and the implementation of the intervention.

5. Application: analysis of results

This study utilized a segmented time series regression analysis to assess the effectiveness of the indoor residue spray (IRS) malaria intervention, which was initiated in November 2019 in Mangochi district. The segmented time series approach allowed for the evaluation of the intervention's impact on malaria incidence and the detection of any notable changes in the malaria trends after the intervention implementation.

The model frame is expressed as follows. Let Y_t be monthly malaria cases, then

$$Y_t \sim NegBinomial(\mu_t)$$

$$log(u_t) = \alpha + B_1 * Time\ in\ months + B_2 * Intervention + B_3 * sin\left(\frac{2\pi t}{T}\right) + B_4 * cos\left(\frac{2\pi t}{T}\right) + \sum_t B * Climate\ variables$$

In the model above, Y_t is number of malaria cases, $log(u_t)$ is log-link function, α is baseline intercept, B_1 is a coefficient representing a baseline trend, B_2 is coefficient representing the effect of intervention, $B_3 and\ B_4$ are coefficients for sine and cosine functions to control for malaria seasonality. As evident from the findings of this study, it was observed that climate factors, including rainfall, temperature, and humidity, exhibited delayed effects on malaria incidence. Therefore, climate variables were included in the model represented by $\sum_t B$. To address the issue of over-dispersion in the data, the negative binomial regression model was employed in this study.

4.1.Model building

The study examined nested models that included various components such as the baseline trend, seasonal control using sine and cosine functions, intervention (indoor residue spray or IRS), and climate variables (lagged and non-lagged).

The nested models used in this study are as follows:

Model-0: This model includes only the intercept term, serving as a baseline reference for comparison, equation 3.2.

Model-1: In addition to the intercept, this model includes a baseline trend over time, allowing for the analysis of temporal patterns in malaria incidence

Model-2: Along with the intercept and baseline trend, this model incorporates seasonality control using sine and cosine waves, capturing the cyclic nature of malaria incidence

Model-3: Building upon the previous models, this model introduces an intervention variable to evaluate the impact of indoor residue spray (IRS) while controlling for the baseline trend and seasonality

Model-4: In addition to the intercept, baseline trend, seasonality, and intervention variable, this model incorporates climate variables that are known to influence malaria. It assesses the direct association between these climate factors and malaria incidence

Model-5: Extending Model-4, this model includes lagged effects of climate variables. The selection of lagged terms is based on the findings from Chapter 3, indicating that the lagged effects peak at lag-0 and remain significant up to lag-2, diminishing thereafter. Thus, all climate variables in Model-5 have lags up to lag-2, representing the influence of previous climate experiences over the past two months.

By comparing these nested models, the study aimed to determine the most influential factors and their effects on malaria incidence, providing valuable insights into the dynamics of malaria transmission and the impact of interventions and climate variables.

Model selection was based on comparing the AIC from the all six models. The results showed that model-5 has smallest AIC among all candidate models. This model has baseline trend, seasonality control (sine-cosine functions), intervention (IRS) and lagged climate variables.

4.2.Impact of lagged climate effect in modeling

The impact of indoor residue spray (IRS) intervention on malaria incidence was assessed using different candidate models, and the results are presented in Table 6. All models demonstrated a reduction in malaria incidence associated with IRS intervention. However, the magnitude of the impact varied across the different models. Notably, when climate variables were incorporated in models 4 and 5, the estimated impact changed from 48.7% to 46.93% and 47.62%, respectively.

Table 6 Model parameters (exponentiated)

Model	Intercept	Trend	Intervention	AIC	Estimated % reduction of cases
Model 3	2078.18604	1.007421	0.512739	1379.029	48.73%
Model 4	7.383166	0.006607004	0.5306841	1385.953	46.93%
Model 5	13.135119	1.0045596	0.5237897	1373.802	47.62%

5. Impact of indoor residue spray (IRS) malaria intervention in Mangochi

The study finally utilized model-5 to evaluate the impact of IRS on malaria cases, considering the delayed effects of climate factors. The findings revealed a significant reduction of malaria cases by 48% (CI: 46%-49%) in the general population due to IRS intervention. However, when analyzing the data based on different age groups and accounting for lagged effects of climate factors, varying impacts were observed, as presented in Table 7. Notably, the under-5 age category exhibited a substantial reduction in malaria cases, with IRS resulting in a 51% decrease (CI: 49%-54%), while the over-5 category experienced a lower reduction of 44% (CI: 47%-43%).

Table 7 Model estimates for impact by age category

		95% CI	
Age group	Estimated % reduction of cases	Lower	Upper
All population	48%	46%	49%
Under - 5	51%	49%	54%
Over-5	44%	43%	47%

The graphs in *figure 9* below shows long time trends that hypothesize expected scenario under which an intervention had not taken place and the trend continues unchanged ('expected' trend, in the absence of the intervention, given the pre-existing trend) which is referred to as the 'counterfactual'. In all age groups the counter factual trend depicts high malaria cases compared to what is observed and predicted by the model as shown in *figure 9a-c*.

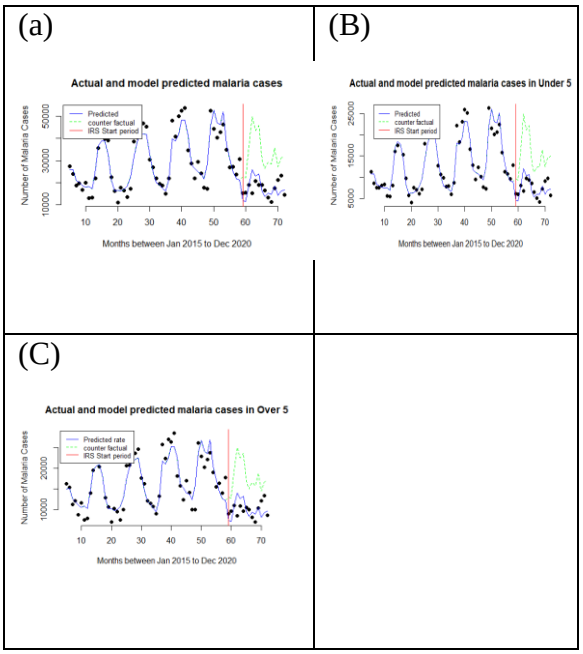


Figure 9 Graph showing predicted monthly incidence against counterfactual climate un adjusted for climate factors

6. Discussion and Conclusion

The findings of this study align with the existing literature on the effectiveness of interrupted time series (ITS) designs to evaluate the effectiveness of interventions in real-world settings where randomized controlled trials (RCTs) may not be feasible or practical (Evangelos et al., 2015). The study further showed that incorporating lagged climate effects in ITS models improved the accuracy of modelling impact of IRS interventions. The study also demonstrated that lagged climate conditions can significantly impact disease transmission dynamics and, therefore, should be accounted for when evaluating impact of an intervention such as IRS.

Generally, the study indicates a significant reduction in malaria cases associated with IRS intervention. Furthermore, it has been demonstrated that the delayed effects of climate conditions significantly influence malaria incidence during specific periods. Therefore, when evaluating the impact of interventions, it is crucial to thoroughly consider and account for the influence of lagged climate conditions. After accounting for lagged climate conditions, the study revealed a notable reduction in malaria incidence by 48% (CI: 46%-49%) associated with the implementation of Indoor Residue Spray (IRS) in the general population. However, when analyzing the data based on different age groups, the study found varying impacts. The under-5 age category exhibited a significant reduction in malaria cases (51%, CI: 49%-54%), whereas the over-5 category experienced a comparatively lower reduction (44%, CI: 47%-43%). The finding is important considering previous research that has highlighted the vulnerability of young children to severe malaria and the potential benefits of targeted interventions for this age group such as a study by Kazembe et al. (2015). The study by McLean et al. (2018) also reported differential impacts of community health worker (CHW) programs on malaria incidence among different population groups.

The analysis also revealed a distinct counterfactual trend, demonstrating a higher incidence of malaria cases compared to what was observed and predicted by the model. The substantial reduction of 48% (CI: 46%-49%) associated with the implementation of Indoor Residue Spray (IRS) supports the efficacy of this intervention strategy in combating malaria.

In conclusion, the study highlights the significant impact of IRS intervention on reducing malaria incidence in Malawi. The inclusion of climate variables and accounting for lagged effects enhance the accuracy of the predictions and provide a comprehensive understanding of the intervention's effects.

The findings support the continued implementation and scale-up of IRS as a population-level intervention for malaria prevention and control, particularly in areas with high malaria burden. Furthermore, the differential impact observed among age groups emphasizes the importance of targeting interventions to specific populations, such as young children who are at higher risk of malaria. These findings contribute to the designing effective malaria control programs that consider both intervention strategies and environmental factors

RCODE

##LOADING REQUIRED PACKAGES IN R

```
library(dyn)

library(dlnm)

library(splines)

library(MASS)

library(nlme)

library(mgcv)

library(tscount)

library(car)
```

Transforming variables to zoo class

```
RainFall<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$Rain)

Humidity<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$Humidity)

MinTemperature<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$`Min Tem`)

MaxTemperature<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$`Max Tem`)

MinTem<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$`Min Tem`)

MaxTem<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$`Max Tem`)

TotalMalariaCases<-Mangoch_Dataset_2015_2020_U_and_O5$MalariaTotal


TotalMalariaU5<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$MalariaU5)

TotalMalariaO5<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$MalariaO5)

TotalPopulation<-Mangoch_Dataset_2015_2020_U_and_O5$Population

Time <-zoo(Mangoch_Dataset_2015_2020_U_and_O5$Time)

Intervention <- zoo(Mangoch_Dataset_2015_2020_U_and_O5$Population)
```

##CHAPTER 2

##FIGURE 2

##SEASONAL PATTERN OF CLIMATE VARIABLES AND MALARIA

##Malaria cases and rainfall seasonal pattern##

```
seasonalAllCases<-plot(Time, TotalMalariaCases,main = " Malaria and Rainfall  Seasonal Trends", type = "l", axes = TRUE, bty = "n", xlab = "Months between Jan 2015 to Dec 2020", ylab = "Malaria cases")

par(new=TRUE)

plot(Time,RainFall,type = "l", axes = FALSE, bty = "n", xlab = "", ylab = "",col="blue")

axis(side=4)

legend(4, 550, legend=c("Malaria Cases", "Rainfall"),col=c("black", "blue"), lty=1:2, cex=0.8)
```

##Malaria cases and humidity seasonal pattern##

```
seasonalAllCases<-plot(Time, TotalMalariaCases,main = "Malaria and Humidity  Seasonal Trends", type = "l", axes = TRUE, bty = "n", xlab = "Months between Jan 2015 to Dec 2020", ylab = "Malaria Cases")

par(new=TRUE)

plot(Time,Humidity,type = "l", axes = FALSE, bty = "n", xlab = "", ylab = "",col="blue")

axis(side=4)

legend(8, 85, legend=c("Malaria Cases", "Humidity"),col=c("black", "blue"), lty=1:2, cex=0.8)
```

##Malaria cases and maximum temperature seasonal pattern##

```
seasonalAllCases<-plot(Time, TotalMalariaCases,main = "Malaria and Maximum Temperature  Seasonal Trends", type = "l", axes = TRUE, bty = "n", xlab = "Months between Jan 2015 to Dec 2020", ylab = "Malaria Cases")

par(new=TRUE)

plot(Time,MaxTemperature,type = "l", axes = FALSE, bty = "n", xlab = "", ylab = "",col="blue")

axis(side=4)
```

```
legend(2, 37, legend=c("Malaria Cases", "Maximum Temperature"),col=c("black", "blue"), lty=1:2, cex=0.8)
```

##Malaria cases and minimum temperature seasonal pattern##

```
seasonalAllCases<-plot(Time, TotalMalariaCases,main = "Malaria and Minimum Temperature  Seasonal Trends", type = "l", axes = TRUE, bty = "n", xlab = "Months between Jan 2015 to Dec 2020", ylab = "Malaria Cases")

par(new=TRUE)

plot(Time,MinTemperature,type = "l", axes = FALSE, bty = "n", xlab = "", ylab = "",col="blue")

axis(side=4)

legend(50, 25, legend=c("Malaria Cases", "Minimum temperature"),col=c("black", "blue"), lty=1:2, cex=0.8)
```

##FIGURE 3

SCATTTER PLOTS OF CLIMATE VARIABLES AND MALARIA CASES##

```
scatterRain<-plot(RainFall, TotalMalariaCases, main = "Malaria cases and rainfall",   xlab = "Rainfall", ylab = "Malaria cases", pch = 19, frame = FALSE)

lines(lowess(RainFall, TotalMalariaCases))

scatterHum<-plot(Humidity, TotalMalariaCases, main = "Malaria cases and humidity",   xlab = "Humidity", ylab = "Malaria cases", pch = 19, frame = FALSE)

lines(lowess(Humidity, TotalMalariaCases))

scatterMaxT<-plot(MaxTemperature, TotalMalariaCases, main = "Malaria cases and Maximum temperature",   xlab = "Maximum Temperature", ylab = "Malaria Cases", pch = 19, frame = FALSE)

lines(lowess(MaxTemperature, TotalMalariaCases))

scatterMinT<-plot(MinTemperature, TotalMalariaCases, main = "Malaria cases and Minimum temparature",   xlab = "Minimum Temperature", ylab = "Malaria cases", pch = 19, frame = FALSE)

lines(lowess(MinTemperature, TotalMalariaCases))
```

##VARIABLE COLLINEARITY TEST##

##TABLE 2

##Pearson correlation table

```
cor(Mangoch_Dataset_2015_2020_U_and_O5[, c("MalariaTotal","Rain","Humidity","Max Tem","Min Tem")])
```

##FIGURE 4

Generating VIF values for model with collinear climate variables

```
modelAllModelVIF <- glm.nb(TotalMalariaCases ~  RainFall +  Humidity + MinTem + MaxTem + Intervention ,offset(log2(TotalPopulation)), data=Mangoch_Dataset_2015_2020_U_and_O5)

vif_values<-vif(modelAllModelVIF)

vif_values
```

Plotting VIF Bar Graph for with collinear climate variables

```
barplot(vif_values,main="VIF Values", horiz=TRUE,col="steelblue")

abline(v=5,lwd=3,lty=2)
```

Generating VIF values for model with collinear climate variables

```
modelAllModelVIFReducedVIF <- glm.nb(TotalMalariaCases ~  RainFall + Humidity + MaxTem  + Intervention, offset(log2(TotalPopulation)), data=Mangoch_Dataset_2015_2020_U_and_O5)

vif_valuesReduced<-vif(modelAllModelVIFReducedVIF)

vif_valuesReduced

## Plotting VIF Bar Graph for without collinear climate variables

barplot(vif_valuesReduced,main="VIF Values", horiz=TRUE,col="steelblue")
```

##TABLE 4

##EXPONSURE LAG RESPONSE SPECIFICATION IN DLNM AND GENERATING BASIS VARIABLES#

##Model Specification for rainfall

##Linear in exposure space and linear in lag space

```
basis.Rain <- crossbasis(RainFall, lag=3, argvar = list(fun="lin"),arglag=list(fun="lin"))

summary(basis.Rain)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Rain + ns(Time, 15) ,offset(log2(TotalPopulation)),
data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Linear in exposure space and Polynomial in lag space

```
basis.Rain <- crossbasis(RainFall, lag=3, argvar = list(fun="lin"),arglag=list(fun="poly",2))

summary(basis.Rain)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Rain + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Polynomial in exposure space and Linear in lag space

```
basis.Rain <- crossbasis(RainFall, lag=3, argvar = list(fun="poly",degree=3),arglag=list(fun="lin"))

summary(basis.Rain)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Rain + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Polynomial in exposure space and Polynomial lag space

```
basis.Rain <- crossbasis(RainFall, lag=3, argvar =
list(fun="poly",degree=3),arglag=list(fun="poly",degree=2))

summary(basis.Rain)
```

```
modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Rain + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Natural Cubic Splines function with knots in Quantiles in exposure soace and Linear in lag space

```
basis.Rain <- crossbasis(RainFall, lag=3, argvar = list(fun="ns",4),arglag=list(fun="lin"))

summary(basis.Rain)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Rain + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Natural Cubic Spline function in exposure space and Polynomial in lag space

```
basis.Rain <- crossbasis(RainFall, lag=3, argvar = list(fun="ns",4),arglag=list(fun="poly",degree=2))

summary(basis.Rain)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Rain + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

Model specification for Humidity

##Linear in exposure space and Linear in lag space

```
basis.Hum <- crossbasis(Humidity, lag=3, argvar = list(fun="lin"),arglag=list(fun="lin"))

summary(basis.Hum)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Hum + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel
```

```
BIC(modelAllModel)
```

##Linear in exposure space and Poly in lag space

```
basis.Hum <- crossbasis(Humidity, lag=3, argvar = list(fun="lin"),arglag=list(fun="poly",2))

summary(basis.Hum)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Hum + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Poly in exposure space and Linear in lag space

```
basis.Hum <- crossbasis(Humidity, lag=3, argvar = list(fun="poly",degree=3),arglag=list(fun="lin"))

summary(basis.Hum)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Hum + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Poly in exposure space and Polynomial in lag space

```
basis.Hum <- crossbasis(Humidity, lag=3, argvar =
list(fun="poly",degree=3),arglag=list(fun="poly",degree=2))

summary(basis.Hum)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Hum + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##NS in exposure space and Linear in lag space

```
basis.Hum <- crossbasis(Humidity, lag=3, argvar = list(fun="ns",4),arglag=list(fun="lin"))
```

```
summary(basis.Hum)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Hum + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Natural Cubic Spline function in exposure space and Polynomial in lag space

```
basis.Hum <- crossbasis(Humidity, lag=3, argvar = list(fun="ns",4),arglag=list(fun="poly",degree=2))

summary(basis.Hum)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.Hum + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Model Specification for Maximum Temperature

##Linear in exposure space and Linear lag space

```
basis.MaxTem <- crossbasis(MaxTemperature, lag=3, argvar = list(fun="lin"),arglag=list(fun="lin"))

summary(basis.MaxTem)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.MaxTem + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Linear in exposure space and Poly in lag space

```
basis.MaxTem <- crossbasis(MaxTemperature, lag=3, argvar = list(fun="lin"),arglag=list(fun="poly",2))

summary(basis.MaxTem)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.MaxTem + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)
```

```
modelAllModel

BIC(modelAllModel)
```

##Poly in exposure space and Linear in lag space

```
basis.MaxTem <- crossbasis(MaxTemperature, lag=3, argvar = list(fun="poly", 3),arglag=list(fun="lin"))

summary(basis.MaxTem)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.MaxTem + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##Polynomial in exposure space and Polynomial in lag space

```
basis.MaxTem <- crossbasis(MaxTemperature, lag=3, argvar =
list(fun="poly",3),arglag=list(fun="poly",2))

summary(basis.MaxTem)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.MaxTem + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)

##NS in exposure space and  Linear in lag space

basis.MaxTem <- crossbasis(MaxTemperature, lag=3, argvar = list(fun="ns",4),arglag=list(fun="lin"))

summary(basis.MaxTem)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.MaxTem + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##NS in exposure space and Poly in lag space

```
basis.MaxTem <- crossbasis(MaxTemperature, lag=3, argvar = list(fun="ns",4),arglag=list(fun="poly",2))

summary(basis.MaxTem)

modelAllModel <- glm.nb(TotalMalariaCases ~ basis.MaxTem + ns(Time, 15)
,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel

BIC(modelAllModel)
```

##FINAL MODEL WITH BASIS VARIABLES FOR CLIMATE VARIABLES

```
basis.Rain <- crossbasis(RainFall, lag=3, argvar = list(fun="ns",2),arglag=list(fun="ns",2))

basis.Hum <- crossbasis(Humidity, lag=3, argvar = list(fun="ns",4),arglag=list(fun="poly",degree=2))

basis.MaxTem <- crossbasis(MaxTemperature, lag=3, argvar = list(fun="ns",4),arglag=list(fun="poly",2))

modelAllModel <- glm.nb(TotalMalariaCases ~  basis.Rain  + basis.MaxTem + basis.Hum  + Intervention
+ ns(Time, 15) ,offset(log2(TotalPopulation)),data=Mangoch_Dataset_2015_2020_U_and_O5)

modelAllModel
```

##NON LINEAR EFFECT PLOTING AND INTERPRETATION

##FIGURE 5

##Lagged effect of Rainfall

```
predRain <- crosspred(basis.Rain, modelAllModel,coef=NULL, vcov=NULL, at=c(0:541.4),cen = 74.5)

predRain

plot(predRain, "overall",lwd=2,col = 4, xlab = "Rain", ylab = "RR", main = "Overall effect of rainfall")

plot(predRain, ptype="slices", type = "p", pch = 19, cex = 1.5, var = 541,ci = "bars", ylab = "RR", main =
"Lag-specific effect of amount of rainfall")

plot(predRain, xlab = "Rainfall", theta = 240, phi = 40,ltheta = -185, zlab = "RR", main = "Effect of
rainfall")

plot(predRain, "contour", plot.title = title(xlab = "rainfall",ylab = "Lag", main = "Relative risk associated
with rainfall"), key.title = title("RR"))
```

##FIGURE 6

Lagged effect of Humidity

predHum <- crosspred(basis.Hum, modelAllModel,coef=NULL, vcov=NULL, at=c(40:84),cen = 65.2)

predHum

plot(predHum, "overall",lwd=2,col = 4, xlab = "Rain", ylab = "RR", main = "Overall effect of humidity")

plot(predHum, ptype="slices", type = "p", pch = 19, cex = 1.5, var =84,ci = "bars", ylab = "RR", main = "Lag-specific effect of humidity")

plot(predHum, xlab = "Humidity", theta = 240, phi = 40,ltheta = -185, zlab = "RR", main = "Effect of humidity")

plot(predHum, "contour", plot.title = title(xlab = "Humidity",ylab = "Lag", main = "Relative risk associated with humidity"), key.title = title("RR"))

##FIGURE 7

Lagged effect of Maximum temperature

predMaxTem <- crosspred(basis.MaxTem, modelAllModel,coef=NULL, vcov=NULL, at=c(25:36),cen = 30.99)

predMaxTem

plot(predMaxTem, "overall",lwd=2,col = 4, xlab = "Max temperature", ylab = "RR", main = "Overall effect maximum temperature")

plot(predMaxTem, ptype="slices", type = "p", pch = 19, cex = 1.5, var = 36,ci = "bars", ylab = "RR", main = "Lag-specific effect maximum of temperature")

plot(predMaxTem, xlab = "Maximum Temperature", theta = 240, phi = 40,ltheta = -185, zlab = "RR", main = "Effect of maximum temperature")

plot(predMaxTem, "contour", plot.title = title(xlab = "Maximum temperature",ylab = "Lag", main = "Relative risk associated with maximum temperature"), key.title = title("RR"))

##CHAPTER 3

Transforming variables to zoo class

MalariaTotal1<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$MalariaTotal)

```
Rain1<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$Rain)

Humidity1<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$Humidity)

MinTemp1<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$`Min Tem`)

MaxTemp1<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$`Max Tem`)

MalariaU51<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$MalariaU5)

MalariaO51<-zoo(Mangoch_Dataset_2015_2020_U_and_O5$MalariaO5)

Time1 <-zoo(Mangoch_Dataset_2015_2020_U_and_O5$Time)
```

extracting other variables in the data set

```
interventions <- interv_covariate(n = 72, tau = c(59), delta = c(1))

sinwave <- sin(2*pi/12*Mangoch_Dataset_2015_2020_U_and_O5$Time)

cosinwave <- cos(2*pi/12*Mangoch_Dataset_2015_2020_U_and_O5$Time)

PopAll<- Mangoch_Dataset_2015_2020_U_and_O5$Population

PopU5<-Mangoch_Dataset_2015_2020_U_and_O5$U5Pop

PopO5<-Mangoch_Dataset_2015_2020_U_and_O5$O5Pop
```

##Table 5 and Figure 8 showing AIC , BIC and QIC for candidate models and residual plots

Model-0 with intercept only no regressors

```
regressorsAll <- cbind()

response<-window(TotalMalariaCases, start=5)##trancating first 4 months to be able to use 4 month delayed effects

dynmodelAllAges0<-tsglm(response,
                        model = list(past_obs = c(1)), link = "log", distr = "nbinom",
                        xreg=NULL)

summary(dynmodelAllAges0)

coefficients(dynmodelAllAges0)
```

```
exp(coef(dynmodelAllAges0))

confint(dynmodelAllAges0)

acf(residuals(dynmodelAllAges0))
```

Model-1 intercept and baseline trend only

```
regressorsAll <- cbind(Time1)

regressors <- window(regressorsAll, start=5)

response<-window(TotalMalariaCases, start=5)

dynmodelAllAges0<-tsglm(response,

                model = list(past_obs = c(1)), link = "log", distr = "nbinom",

                xreg=regressors)

summary(dynmodelAllAges0)

coefficients(dynmodelAllAges0)

exp(coef(dynmodelAllAges0))

confint(dynmodelAllAges0)

acf(residuals(dynmodelAllAges0))
```

seasonal control and baseline trend

```
regressorsAll <- cbind(sinwave, cosinwave, Time1)

regressors <- window(regressorsAll, start=5)##trancating first 4 months to be able to use 4 month delayed effects

response<-window(TotalMalariaCases, start=5)

dynmodelAllAges0<-tsglm(response,

                model = list(past_obs = c(1)), link = "log", distr = "nbinom",

                xreg=regressors)
```

```
summary(dynmodelAllAges0)

coefficients(dynmodelAllAges0)

exp(coef(dynmodelAllAges0))

confint(dynmodelAllAges0)

acf(residuals(dynmodelAllAges0))
```

Model seasonal control, baseline trend and intervention

```
regressorsAll <- cbind(sinwave, cosinwave, Time1, interventions)

regressors <- window(regressorsAll, start=5)

response<-window(TotalMalariaCases, start=5)

dynmodelAllAges0<-tsglm(response,

                model = list(past_obs = c(1)), link = "log", distr = "nbinom",

                xreg=regressors)

summary(dynmodelAllAges0)

coefficients(dynmodelAllAges0)

exp(coef(dynmodelAllAges0))

confint(dynmodelAllAges0)

acf(residuals(dynmodelAllAges0))
```

Model seasonal control, baseline trend, intervention and non-lagged climate variables

```
regressorsAll <- cbind(sinwave, cosinwave, Time1, interventions,Rain1,Humidity1,MaxTemp1)

regressors <- window(regressorsAll, start=5)

response<-window(TotalMalariaCases, start=5)

dynmodelAllAges0<-tsglm(response,

                model = list(past_obs = c(1)), link = "log", distr = "nbinom",

                xreg=regressors)
```

```
summary(dynmodelAllAges0)

coefficients(dynmodelAllAges0)

exp(coef(dynmodelAllAges0))

confint(dynmodelAllAges0)

acf(residuals(dynmodelAllAges0))
```

Model with baseline, intervention, seasonal control and lagged climate variables

```
regressorsAll <- cbind(sinwave, cosinwave, Time1, interventions,lag(Rain1,0:-2),lag(Humidity1,0:-2),lag(MaxTemp1,0:-2))

regressors <- window(regressorsAll, start=5)

response<-window(TotalMalariaCases, start=5)

dynmodelAllAges0<-tsglm(response,
                        model = list(past_obs = c(1)), link = "log", distr = "nbinom",
                        xreg=regressors)

dynmodelAllAges0

summary(dynmodelAllAges0)

coefficients(dynmodelAllAges0)

exp(coef(dynmodelAllAges0))

confint(dynmodelAllAges0)

acf(residuals(dynmodelAllAges0))
```

PREDICTING USING MODEL-5 WITH LAGGED CLIMATE VARIABLES

truncating variables to suit truncated anaylsis

```
truncatedPop <- window(Mangoch_Dataset_2015_2020_U_and_O5$Population, start=5)

truncatedU5Pop <- window(Mangoch_Dataset_2015_2020_U_and_O5$U5Pop, start=5)

truncatedO5Pop <- window(Mangoch_Dataset_2015_2020_U_and_O5$O5Pop, start=5)
```

```
truncatedInterv <- window(Mangoch_Dataset_2015_2020_U_and_O5$IntervPeriod, start=5)

truncatedtime <- window(Mangoch_Dataset_2015_2020_U_and_O5$Time, start=5)

trResponseU5 <- window(Mangoch_Dataset_2015_2020_U_and_O5$MalariaU5, start=5)

trResponseO5 <- window(Mangoch_Dataset_2015_2020_U_and_O5$MalariaO5, start=5)
```

predicted rate in all ages

```
regressorsAll <- cbind(sinwave, cosinwave, Time1, interventions,lag(Rain1,0:-2),lag(Humidity1,0:-2),lag(MaxTemp1,0:-2))

regressors <- window(regressorsAll, start=5)

response<-window(TotalMalariaCases, start=5)

dynmodelAllAges0<-tsglm(response,

                        model = list(past_obs = c(1)), link = "log", distr = "nbinom",

                        xreg=regressors)
```

##Table 6 Model coefficients for lagged climate variable model-5

```
summary(dynmodelAllAges0)
```

##UNDER ALL AGES ANALYSIS

##Figure 9 Graphs showing predicted monthly incidence against counterfactual climate un adjusted for climate factors

##Model predicted malaria cases in all ages

```
predAllAges<-predict.glm(dynmodelAllAges0,type = "response")

predAllAges
```

##predicted rate

```
predRateAllAges<-predAllAges/truncatedPop
```

##Observation Rate

```
ActualRateAllAges <- response/truncatedPop
```

##Obtaining counter factual estimates from the model by removing intervention effect

predAllAgesCounter<-predAllAges/exp(-6.47e-01*truncatedInterv) ##-6.47e-01 is coefficient for intervention variable in the model

predRateAllAgesCounter<-predAllAgesCounter/truncatedPop

##plotting Predicted rate and counter factual

seasonalAllCases<-plot(truncatedtime, response,main = "Actual and model predicted malaria cases", pch = 19, axes = TRUE, bty = "n", xlab = "Months between Jan 2015 to Dec 2020", ylab = "Number of Malaria Cases")

lines(truncatedtime,predAllAgesCounter,lty="dashed", bty = "n", xlab = "", ylab = "",col="green")

lines(truncatedtime,predAllAges,type = "l", bty = "n", xlab = "", ylab = "",col="blue")

abline(v=59,col="Red")

legend(3, 55000, legend=c("Predicted", "counter factual","IRS Start period"),col=c("blue","green","Red"), lty=1:2, cex=0.8)

##UNDER FIVE ANALYSIS

transforming population into log for an offset use

logPopulation <- log(truncatedU5Pop)

regressorsAll <- cbind(sinwave, cosinwave, Time1, interventions,lag(Rain1,0:-2),lag(Humidity1,0:-2),lag(MaxTemp1,0:-2))

regressors <- window(regressorsAll, start=5)

dynmodelU5<-tsglm(trResponseU5,

model = list(past_obs = c(1)), link = "log", distr = "nbinom",

xreg=regressors)

predicted rate

predUnder5<-predict.glm(dynmodelU5,type = "response")

predRateUnder5<-predUnder5/truncatedU5Pop

##Model residues

##residuals.glm(dynmodelU5,type = "response")

##Observation Rate

```
ActualRateUnder5 <- trResponseU5/truncatedU5Pop
```

##Obtaining counter factual estimates from the model by removing intervention effect

```
predUnder5Counter<-predUnder5/exp(-0.7230173*truncatedInterv)

predRateUnder5Counter<-predUnder5Counter/truncatedU5Pop
```

##plotting Predicted rate and counter factual

```
seasonalAllCases<-plot(truncatedtime, trResponseU5,main = "Actual and model predicted malaria cases in Under 5", pch = 19, axes = TRUE, bty = "n", xlab = "Months between Jan 2015 to Dec 2020", ylab = "Number of Malaria Cases")

lines(truncatedtime,predUnder5Counter,lty="dashed", bty = "n", xlab = "", ylab = "",col="green")

lines(truncatedtime,predUnder5,type = "l", bty = "n", xlab = "", ylab = "",col="blue")

abline(v=59,col="Red")

legend(6, 27000,legend=c("Predicted", "counter factual","IRS Start period"),col=c( "blue","green","Red"), lty=1:2, cex=0.8)
```

##OVER FIVE ANALYSIS

transforming population into log for an offset use

```
logPopulation <- log(truncatedO5Pop)

regressorsAll <- cbind(sinwave, cosinwave, Time1, interventions,lag(Rain1,0:-2),lag(Humidity1,0:-2),lag(MaxTemp1,0:-2))

regressors <- window(regressorsAll, start=5)

dynmodelO5<-tsglm(trResponseO5,
                  model = list(past_obs = c(1)), link = "log", distr = "nbinom",
                  xreg=regressors)
```

predicted rate

```
predOver5<-predict.glm(dynmodelO5,type = "response")
```

predRateOver5<-predOver5/truncatedO5Pop

##Observation Rate

ActualRateOver5 <- trResponseO5/truncatedO5Pop

##Obtaining counter factual estimates from the model by removing intervention effect

predOver5Counter<-predOver5/exp(-5.85e-01 *truncatedInterv)

predRateOver5Counter<-predOver5Counter/truncatedO5Pop

##plotting Predicted rate and counter factual

seasonalAllCases<-plot(truncatedtime, trResponseO5,main = "Actual and model predicted malaria cases in Over 5", pch = 19, axes = TRUE, bty = "n", xlab = "Months between Jan 2015 to Dec 2020", ylab = "Number of Malaria Cases")

lines(truncatedtime,predOver5Counter,lty="dashed", bty = "n", xlab = "", ylab = "",col="green")

lines(truncatedtime,predOver5,type = "l", bty = "n", xlab = "", ylab = "",col="blue")

abline(v=59,col="Red")

legend(3, 29000, legend=c("Predicted rate", "counter factual","IRS Start period"),col=c("blue","green","Red"), lty=1:2, cex=0.8)