A HOLISTIC APPROACH TO DYNAMIC MODELLING OF MALARIA TRANSMISSION

B. MODU

Ph.D

2020

A Holistic Approach to Dynamic Modelling of Malaria Transmission

An Investigation of Climate-Based Models used for Predicting Malaria Transmission

Babagana MODU

Submitted for the degree of

Doctor of Philosophy

Faculty of Engineering and Informatics

University of Bradford

2020

Babagana Modu

Title: A Holistic Approach to Dynamic Modelling of Malaria Transmission

Sub-title: An Investigation of Climate-Based Models used for Predicting Malaria Transmission

Keywords: Agent-based modelling, Climate-factors, Mathematical modelling, Prediction, Machine learning, Malaria transmission, Prevention, Control, Intervention

Abstract

The uninterrupted spread of malaria, besides its seasonal uncertainty, is due to the lack of suitable planning and intervention mechanisms and tools. Several studies have been carried out to understand the factors that affect the development and transmission of malaria, but these efforts have been largely limited to piecemeal specific methods, hence they do not offer comprehensive solutions to predict disease outbreaks. This thesis introduces a 'holistic' approach to understand the relationship between climate parameters and the occurrence of malaria using both mathematical and computational methods. In this respect, we develop new climate-based models using mathematical, agent-based and data-driven modelling techniques. A malaria model is developed using mathematical modelling to investigate the impact of temperature-dependent delays. Although this method is widely applicable, but it is limited to the study of homogeneous populations. An agent-based technique is employed to address this lim-

itation, where the spatial and temporal variability of agents involved in the transmission of malaria are taken into account. Moreover, whilst the mathematical and agent-based approaches allow for temperature and precipitation in the modelling process, they do not capture other dynamics that might potentially affect malaria. Hence, to accommodate the climatic predictors of malaria, an intelligent predictive model is developed using machine-learning algorithms, which supports predictions of endemics in certain geographical areas by monitoring the risk factors, e.g., temperature and humidity. The thesis not only synthesises mathematical and computational methods to better understand the disease dynamics and its transmission, but also provides healthcare providers and policy makers with better planning and intervention tools.

Declaration

The thesis entitled **"A Holistic Approach to Dynamic Modelling of Malaria Transmission"** was supervised by **Dr. Savas Konur**. I declare that the work submitted is my own and that appropriate credit has been given where reference has been made to the work of others. Also, this work has not been previously submitted for any other degree or program elsewhere.

Babagana Modu

Dedication

I dedicated this work to the memory of my

beloved parents,

Mal. Modu Bulama & Falmata Modu Bulama

I pray Almighty Allah to grant them

Jannatul Firdaus, Aameen.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Almighty Allah for making this PhD journey successful, despite a change to my supervisors.

Special thanks goes to my supervisor, Dr. Savas Konur, for his unconditional support and encouragement that led to the accomplishment of this PhD. Due to Savas' humbleness, he has never treats me as his student but rather as a learning colleague. Academically, I have learned extensively from his expertise, especially in agent-based modelling, data science and programming. Furthermore, his guidance grounded me well and boosted my confidence in scientific writing, research methods and as an academic.

My wholehearted thanks go to my previous supervisors, Prof Yonghong Peng and Dr Taufiq A. Asyhari. They started supervising me well, but later they moved to the Universities of Sunderland and Cranfield, respectively. These acknowledgements would have been incomplete without mentioning them, due to their contributions towards my PhD. The limited time I spent with them can never be forgotten, because they also regarded me as learning colleague and made themselves available at the times of need. I will never forget the support they provided me both at the University of Bradford and in their new places of work, and they ensured a smooth tran-

sition to my new supervisor. No amount of words is enough to capture all they have given me.

I would like to acknowledge the contribution of Dr. Muhammed Jibrin Ngala and his team for initiating the Postgraduate Foundation Programme that led to my PhD study in this university.

Furthermore, I am thankful to all my fellow PhD colleagues, particularly those within the computing research lab and engineering. It is my pleasure to know you all, and indeed the time that we had together was fruitful for me; I hope the academic partnership established will be maintained. I wish to acknowledge the Nigerian community dwelling in Bradford and Leeds, particularly the Nigeria Muslim Forum (NMF) members - I thank you all.

Particular thanks go to my entire family for their unconditional love, parental support, encouragement, and unending prayers for my success. In fact, you have never ceased to support me within your little means. To my friends and well-wishers, I would like to use this opportunity to acknowledge you for the concern that you have shown for me. This has proven to me that distance can never be a barrier so long as people care for one another.

I would like to say very big thank you to my beloved wife, Fatima Hassan Gana, for enduring my absence, managing family stresses and serving humanity as a midwife. I pray Almighty Allah to reward you abundantly for the sacrifices you have made in maintaining the family's well being; may it be the source of your joy and comfort now and hereafter, Aameen. Last, but not the least, my thanks go to our children, Fatima Babagana (Ammi),

Hassan Babagana (Walid) and Muhammad Babagana (Al-Amin). I pray Almighty Allah to bless them, and make them reasons for our happiness now and hereafter, Aameen.

Publications and Presentations

Journal and Conference Papers

- Modu, B., Asyhari, A. T., & Peng, Y. (2016, December). Data Analytics of climatic factor influence on the impact of malaria incidence. In 2016 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1-8). IEEE. (**Thesis contribution**)

- Modu, B., Asyhari, A. T., Konur, S., & Peng, Y. (2017). An Assessment on the Hidden Ecological Factors of the Incidence of Malaria. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 1, No. 3, p. 131). (**Thesis contribution**)

- Modu, B., Polovina, N., Lan, Y., Konur, S., Asyhari, A., & Peng, Y. (2017). Towards a predictive analytics-based intelligent malaria outbreak warning system. Applied Sciences, 7(8), 836. (**Thesis contribution**)

- Modu, B., Konur, S., Polovina, N., & Lan, Y. (2018). Machine Learning Analysis and Agent-Based Modelling of Malaria Transmission. In Fuzzy Systems and Data Mining IV (pp. 465 - 472). Amsterdam: Frontiers in Artificial Intelligence and Applications. Retrieved from http://ebooks.iospress.nl/volumearticle/50703. (**Thesis contribution**)

Table 1: Glossary

Acronyms	Meaning
ABMs	Agent-Based Models
ASEI	Aquatic Susceptible Exposed Infected
API	Application Programming Interface
ACF	Autocorrelation Function
AR	Autoregression
CFSR	Climate Forecast System Reanalysis
CCF	Cross-Correlation Function
DDE	Delay Differential Equation
DFE	Disease Free Equilibrium
DEE	Disease Endemic Equilibrium
DT	Decision Tree
DMC	Disease Model Cradle
EFA	Exploratory Factor Analysis
EIP	Extrinsic Incubation Period
EW	Equal Width
EF	Equal Frequency
FA	Factor Analysis
HTML	Hyper Text Markup Language
IIP	Intrinsic Incubation Period
IPCC	Inter-Governmental Panel of Climate Change
JSON	Java Script Object Notation
KNN	K-Nearest Neighbours
K-M	K-Means
LVs	Latent Variables
LiR	Linear Regression
LoR	Logistic Regression
LMM	Liverpool Malaria Model
LIBSVM	Library for Support Vector Machines
LHS	Latin Hypercube Sampling
MA	Moving Average
MVs	Measurement Variables
NB	Naive Bayes
NCEP	National Centre for Environmental Prediction
ODE	Ordinary Differential Equation

Table 2: Glossary Continued

Acronyms	Meaning
PLS-PM	Partial Least Squares Path Modelling
PDF	Probability Distribution Function
PACF	Partial Autocorrelation Function
PRCC	Partial-Rank Correlation Coefficient
Q-Q	Quantile-Quantile
SEM	Structural Equation Modelling
SVM	Support Vector Machine
SEIR	Susceptible Exposed Infected Recovery
SIS	Susceptible Infected Susceptible
SEI	Susceptible Exposed Infected
SA	Sensitivity Analysis
SSB	Sum of Square Between
SST	Sum of Square Total
VIF	Variance Inflated Factor
WHO	World Health Organisation
XML	Extensible Markup Language

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents the motivation of writing this thesis, statement of the problem, research question, aim and objectives and justification of the study. In addition, the summary of contributions established out of this thesis and the connection between them is illustrated using a well-labelled diagram.

1.1 Introduction

The unceasing spread of malaria, besides its seasonal uncertainty, has long threatened people. The most vulnerable regions are Sub-Saharan Africa and India, and together they account for almost 80% of the global burden [6]. Although malaria is preventable and treatable, the lack of suitable mechanisms and tools for intervention planning prior to its season or outbreak remain a challenge. However, efforts to develop malaria intervention mechanisms have been undermined by spatial and temporal fluctuations due to factors influencing its transmission.

The factors influencing malaria are multifaceted, for example, human movement patterns [19], climate factors [113], human age-structures [157], environmental condi-

tions [87], and the re-emergence of anti-malaria parasites and mosquitoes [22] among other. However, the complex nature of these factors within the occurrences of malaria makes any understanding of its dynamics and spatio-temporal pattern of transmission difficult. The most compelling factor influencing malaria's spread is climate, which includes temperature, precipitation and humidity among other. Moreover, several studies ([81, 101, 112, 132]) also show that malaria's prevalence, seasonal incidences and outbreaks are driven by these climate factors, and each of these influence its transmission in different circumstances. The temperature influences the mosquito biting rate [172], survival [51], parasite development [37] and immature mosquito maturation [33], while precipitation and humidity provide mosquitos with breeding sites for reproduction [81] and increased biting competence [148] respectively. Furthermore, among the climate factors, temperature is the large-scale driver of malaria's transmission that influences all stages of a mosquito's life cycle and supports the development of *plasmodium species* (the parasite causing the malaria infection).

Malaria represents a serious health issue; more than 3.4 billion people worldwide [6], are exposed to its risks and it predominantly spreads in tropical and sub-tropical regions. This is because the climatic conditions experienced in the territory sufficiently support the mosquito's survival and enables the *plasmodium species* to develop and reproduce faster. The influence of climate factors, particularly temperature, on the distribution of malaria, will greatly increase risks in the future, due to the likely increase in the earth temperature as a result of global warming. A recent campaign on the consequences of global warming, and the likelihood of malaria's prominence in previously unexposed areas has prompted the World Health Organization (WHO) and other healthcare providers to further strengthen efforts towards intervention, prevention and control. The Inter-governmental Panel of Climate Change (IPCC) says that

gradual rises in temperature over time can alter the natural habitats of the mosquito by changing its prevalence and prolonging the season of malaria's spread [145]. A marginal 0.5°C increase in temperature could cause an approximate 30–100% increase in the mosquito population [162]. Similarly, a small shift in temperature, from 2°C to 3°C could increase the number of humans vulnerable to malaria by up to 5% [167]. This shows that an increase in temperature as a result of anticipated global warming is a risk factor for malaria's proliferation; hence, this needs careful consideration as part of a control strategy.

To address these challenges, numerous studies [18, 36, 120, 129, 131, 133] among others, were carried out to investigate the influence of climate factors on the dynamics of malaria's transmission as well as its spatial and temporal considerations. However, these efforts were largely limited to specific 'piecemeal' methods, and therefore do not offer comprehensive solutions to predict the outbreak of malaria. This thesis introduces an integrated approach to investigate the causal relationship that exists between climate factors and the occurrence of malaria by deploying mathematical and computational methods. Climate-based models to predict malaria's transmission will be utilized to carry out this investigation. In doing so, a new climate-based model will be develop using mathematical, agent-based and data-driven modelling techniques.

1.2 Statement of the Problem

The effects of malaria are far more profound in Third World countries due to very limited medical resources. When an intense outbreak occurs, most of these countries cannot cope with the high number of patients due to a lack of medicine, equipment and hospital facilities. Moreover, due to poverty, and economic insatiability, the pre-

vention or reduction of risk factors of this disease is very challenging, especially in these countries. Technology can offer alternative solutions by providing early detection mechanisms that help to control the spread of the disease and allow for the advance management of treatment facilities to ensure a more timely health service, which can save thousands of lives. The availability of an early detection system will not only prevent or decrease a large scale spread of malaria by creating quarantine zones, but will also help healthcare providers to deliver the necessary medical care on time by managing resources and calling for international aid and support, if needed.

1.2.1 Research Question

This thesis focuses on investigating the relationship between climate factors and dynamics in malaria transmission in order to develop a comprehensive model to predict the outbreak of malaria. Therefore, to address the problem, the following research question is formulated: How can a climate-based model be utilised for effective planning and the management of intervention. To effectively answer this research question, the question is further streamlined to form three sub-questions as follows:

i. How can the impact of temperature on mosquito influences human malaria?

ii. How can the climate factors influences spatial and temporal distribution of malaria?

iii How can the climate factors be used for aiding developing tool and strategising malaria intervention?

These research question could be achieved using the aim and objectives outlined in subsection 1.2.2.

1.2.2 Aim and Objectives

In order to bring a successful conclusion to the outlined questions under 1.2.1, the following specific objectives are considered:

i. To develop a mathematical model and assess human infectiousness from the impact of temperature in malaria dynamics.

ii. To develop and validate an agent-based model to investigate the dynamics of malaria on humans by incorporating population heterogeneity and spatio-temporal changes.

iii. To develop an intelligent system using embedded machine learning algorithms that are capable of predicting the likelihood of a high incidence malaria season or outbreak

These objectives are addressed in Chapter 3, 4 and 5 respectively.

1.2.3 Significance of the Study

The dynamic modelling of a climate-based malaria transmission model will enhance an understanding of the disease characteristics particularly at a micro and macro scale. This will harness a spatio-temporal understanding of the disease dynamics in different patches of a population so as to guide intervention. Furthermore, the model has an economic benefit in maximizing the intervention to the target population at the time of need.

1.3 Thesis Outline

The key contributions of this thesis are exhaustively discussed in Chapter 3, 4 and 5. These contributions are categorized under three components, as shown in Figure 1.1, as Mathematical modelling, Agent-based modelling and Data-driven modelling. In subsection 1.3.1, the logical connections between components, based on their limitations, are explained in detail.

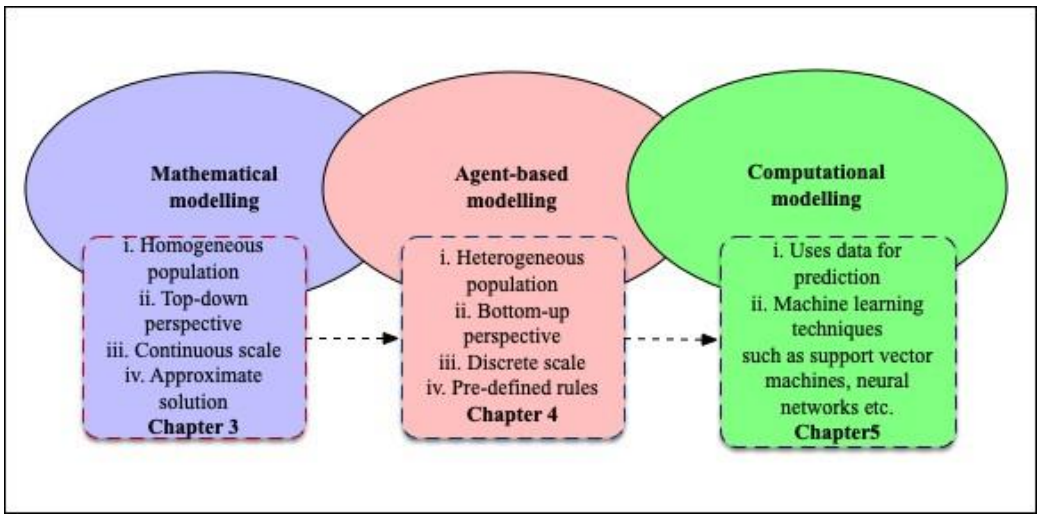

Figure 1.1: The components of this thesis contributions.

1.3.1 Main Contribution

This subsection presents the following summary of the key contributions of this thesis:

a. Among the factors influencing malaria, temperature is key as it drives the parasite's development, the mosquito's survival and increases the vector competence. The effect of temperature on malaria has been extensively dealt with in the context of incubation in humans. However, this has been under-explored in the

context of an incubation period in mosquitos. This limitation results in a partial understanding of the relationship between temperature and malaria; hence, any prevention mechanism to determine an appropriate time for intervention can be highly ineffective. This is addressed by investigating a climate dependent malaria model with a delay in mosquito dynamics.

b. Efforts have been taken to mitigate the transmission of malaria, but the complexity arising from population heterogeneity makes mathematical modelling a less effective strategy as mechanism for prevention. To effectively address the limitations of mathematical approach, an agent-based modelling was utilised. The model developed is not only capable of predicting the pattern of malaria dynamics in a population but also its peak season (see Chapter 4). Some part of this contribution has been previously presented in a conference on Fuzzy Systems and Data Mining IV. The paper appears in the proceedings published in book series entitled 'Frontiers in Artificial Intelligence and Application' (http://ebooks.iospress.nl/volumearticle/50703.

c. The lack of mechanism to predict malaria outbreaks or high seasons, particularly in developing countries is notable. The correlation between climate factors and the occurrence of malaria was utilised to develop a malaria predictive model; this was achieved by deploying machine learning algorithms. The model serves as an intelligent system for predicting the likelihood of a high incidence of malaria, and to thus inform healthcare providers to enable the better planning and management of interventions (see Chapter 5). This contribution containing two conference papers: (i) Data Analytics of Climatic Factor Influence on the Impact of Malaria Incidence, presented during Symposium Series on Computational In-

telligence and published in IEEE explore. (ii) An Assessment on the Hidden Ecological Factors of the Incidence of Malaria, presented during conference on Digitalisation for Sustainable Society and published in Applied Science, MDPI.

Full details of these contributions are discussed in Chapter 3, 4 and 5 respectively.

1.3.2 Organization of the Thesis

The remaining chapters of this thesis are organized as follows. Chapter 2 presents the complete cycle of malaria's transmission and its mathematical representation is drawn from the classical model of malaria. Chapter 3 outlines the mathematical modelling of temperature influences on delay due to extrinsic incubation and maturation to the human infectiousness. A detailed theoretical evaluation, numerical simulation and sensitivity analysis of the model inputs were presented. In Chapter 4, the agent-based modelling technique of the mathematical model presented in equation 3.9 was analyzed. Chapter 5 presents a data-driven predictive model using embedded machine learning algorithms. Finally, Chapter 6 provides the conclusion of the entire thesis and identifies areas for further research.

1.3.3 Summary

This chapter has provided an overview of this thesis, whilst chapter 2 presents the theoretical framework of malaria's transmission. This consists of the malaria parasite cycle in humans and mosquitos, the mosquito's development stages, and the mathematical formulation of malaria's transmission and its properties.

Chapter 2

Background

This chapter presents the complete cycle of malaria and mathematical formulations drawing from the classical model of malaria. Extensions of the classical model and its properties are also presented.

2.1 Malaria

Malaria is an infectious disease caused by the *protozoan* parasites of the *genus plasmodium* [49]. It is transmitted from human to human through effective bites by an adult female mosquito (see Figure 2.1). The female mosquitoes are responsible for the spread of malaria to humans [182]; in a biological context, they are called *anopheles*. These mosquitoes live in an environment where their conditions suitably support their survival and reproduction.

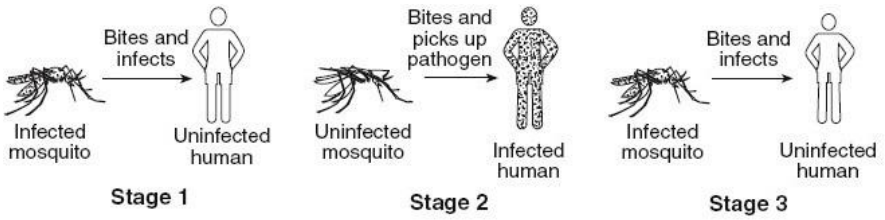

Figure 2.1: Describes how malaria is transmitted from humans to humans, initiated by an infectious mosquito.

2.1.1 Global distribution

Globally, malaria occurs in tropical and sub-tropical regions, as shown in Figure 2.2. This owes to the favourable weather conditions of the regions support the mosquito's survival and the parasites causing malaria to replicate faster. The tropical and sub-tropical regions are characterized by hot and humid summers with essential rainfall occurring in the hottest months [174].

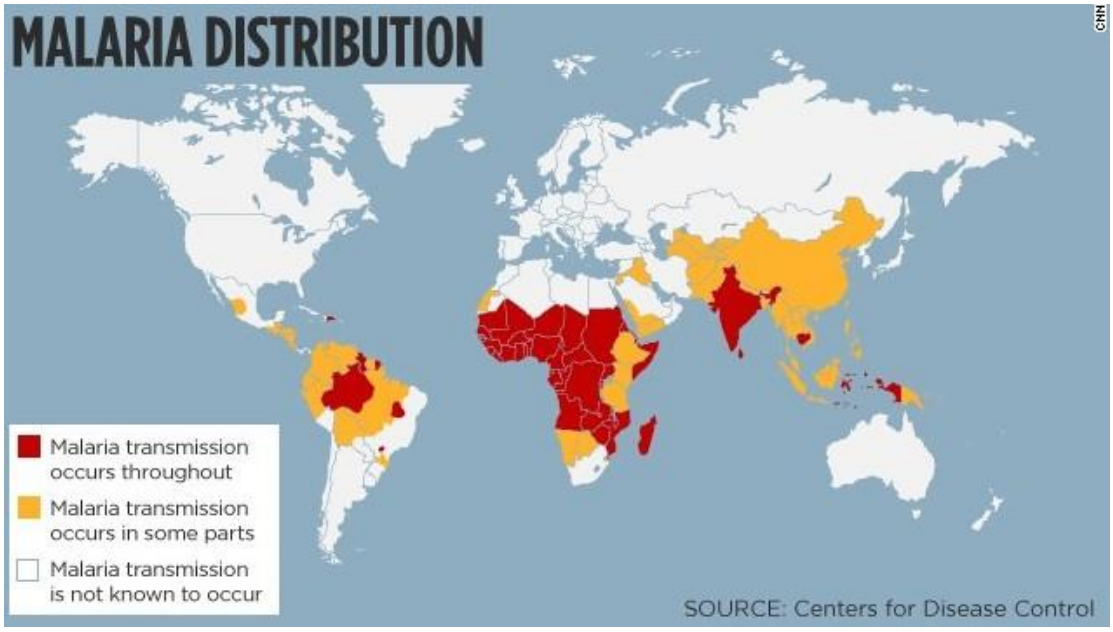

Figure 2.2: The world map shows the magnitude of malaria's occurrence within tropical and sub-tropical regions.

As shown in Figure 2.2, the transmission of malaria is high throughout sub-Saharan Africa. However, it shows that some areas in South-East Asia, the Eastern Mediterranean, Western Pacific, and the Americas are at risks of malaria. It was reported [6] that more than two-thirds of the world's population are at risk of being infected. From this, over 200 million cases and a significant number of deaths close to half a million have been annually reported since 2010.

To strengthen efforts to address the menace of malaria within endemic countries, the WHO and other healthcare providers have been appropriating substantial funds to fight malaria.

2.1.2 Malaria Funding

In 2017 alone, the governments of malaria endemic countries and international partners invested $3.1 billion towards its elimination worldwide. This amount was slightly higher than 2016, and has remained relatively stable since 2010 until 2017. Despite this spending, the progress made has been insignificant in achieving the 2016–2030 target. The target aims to reduce the incidence and mortality rates from malaria at least by 40% worldwide before the end of 2020. As the date has drawn nearer and the target has become unrealistic, the WHO Global Technical Strategy (GTS) has further pushed the target to 2030. This has attracted an increase in funds to combat malaria to at least $6.6 billion per year with effect from 2020 [6].

Besides continued awareness campaigns against mosquito bites and protective measures, research development and routine intervention are among the key areas that consumes enormous funds.

2.1.3 Malaria Intervention

Between 2015–2017, about 624 million Insecticide Treated-Nets (INTs) were distributed worldwide. Of these, about 85% of ITNs were distributed through free mass distribution campaigns, 8% in antenatal care facilities, and 4% as part of immunization programmes. Also, its was estimated that 2.74 billion treatment courses of Artemisinin-based Combination Therapy (ACT) were procured by countries over the period 2010–2017 [6].

The effective distribution of intervention materials requires time; it enables planning and optimization by reaching out to the target population at a time of need. Thus, the lag effects between climate factors and the prevalence of malaria will be utilized.

2.1.4 Factors responsible

Several studies [18, 21, 30, 35, 36, 101, 132] have shown the prevalence of malaria in places that are known to be associated with climate factors. These factors are rainfall, temperature and humidity; they support mosquitos by providing breeding sites for reproduction, an ecological niche for survival, and biting competence, respectively. In different circumstances, the human mobility pattern [105], insecticide-resistant mosquitoes [54], environmental conditions [87] and drug-resistant parasites [43] among others, also influence the prevalence of malaria.

2.2 Malaria transmission cycle

Succinct illustrations of malaria transmission are presented in section 2.1, whilst this subsection explains a complete cycle of malaria in detail. This enables an understand-

ing of the risk factors associated with malaria's spread, and proffers precautionary measures.

Three key elements are significant in malaria's transmission: humans (striking hosts), mosquitoes (vectors) and the physical environment (habitat). Figure 2.3 illustrates how these elements influence each other in the transmission of malaria.

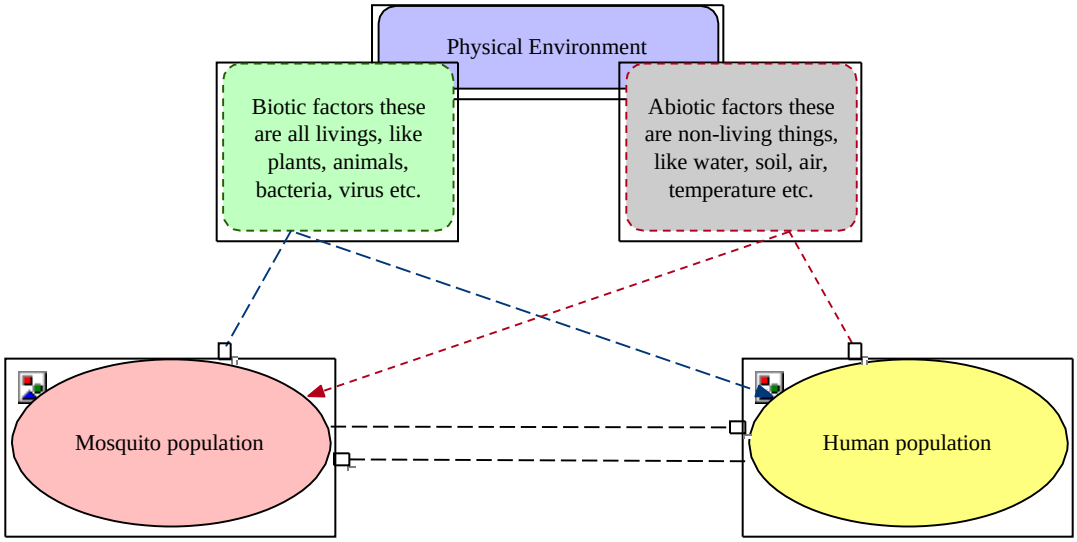

Figure 2.3: The influence of the physical environment on malaria's transmission

Malaria's transmission cycle is essentially divided into two sub-cycles which are called the parasite cycle and mosquito cycle, as shown in Figure 2.4.

2.2.1 Parasite cycle

Globally, over 100 species of malaria parasites exist, but only four are known to cause malaria [102], which are: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale* and *Plasmodium malariae*. The *Plasmodium falciparum* is the most virulent and life-threatening among the species [41], hence, it is responsible for about 75% of the cases reported worldwide [71]. The malaria parasite development cycle takes place in

both human and mosquito hosts, hence Figure 2.4 illustrates the processes involved.

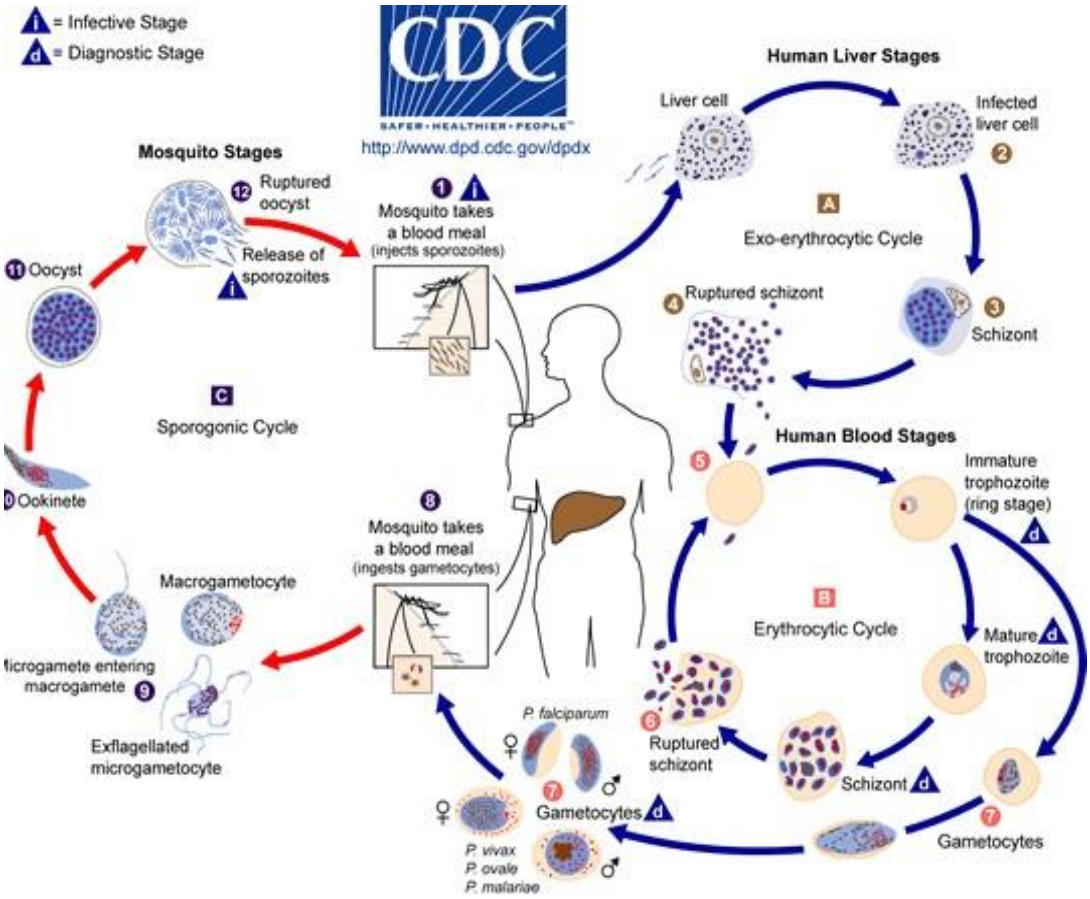

Figure 2.4: The life-cycle of malaria parasites

2.2.1.1 Human host perspective

As mentioned in section 2.1, *Anopheles* is the potential vector for transmitting malaria into the human population. It bites a human and injects saliva containing the *plasmodium sporozoites* into the human bloodstream. When the *sporozoites* enters the human's liver cells, it will grow to become a *schizont*, which later multiplies (*schizogony*) and produces 2,000–4,000 *merozoites*. Consequently, the *merozoites* will either infect other liver cells and repeat the process or proceed into the blood and enter the *erythrocyte*

(red blood cell) in the *schizogonous* cycle. The *sporozoites* in the liver last for about 6–16 days, and an individual *merozoite* in a red blood cell will produce 6–24 *merozoites*. This process is repeated many times so that the number of infected *erythrocytes* in the bloodstream increases greatly. The duration of the cycle is the duration between successive bouts of fever, which corresponds with rupture to the red blood cells [182]. A part of *merozoites* will stop multiplying and asexually transform into male and female *gametocytes*.

2.2.1.2 Mosquito host perspective

On the other hand, when a susceptible mosquito bites an infected human, it injects saliva and sips blood containing *merozoites* and *gametocytes*. The *merozoites* are digested in the mosquito's gut but the *gametocytes* survive. Consequently, the female type *gametocytes* will transform into *macrogametes* (female *gametes*) and male counterparts to *microgametes* (male *gametes*). The *macrogametes* will be fertilised by the exflagellation of *microgametes* to give a *zygote*. The *zygote* develops into an *ookinete* over 12–48 hours and penetrates the wall of the midgut to become an *oocyst*. Therefore, the *oocyst* grows its content and divides into about 10,000 elongated *sporozoites*. However, this does not occur unless ambient temperatures range from 16–33°C. The *sporozoites* burst out of the *oocyst* into the *haemocoele* of the mosquito and the majority migrate to the salivary glands [182].

2.2.2 Mosquito cycle

Among the climate factors, rainfall provides a breeding site for mosquito reproduction, and thus completes the mosquito's cycle. The *Anopheles* mosquito needs stagnant wa-

ter, a water body or a dump area to lay its eggs. A complete mosquito cycle comprises four stages: egg, larva, pupa and adult. The first three stages are called the aquatic or immature stages (see Figure 2.4). At these stages, the immature mosquito has to live inside water until they emerge as an adult.

Only the *Anopheles* bites human beings and sucks blood for nourishment and the development of eggs. Each *Anopheles* can develop several hundred eggs at each blood meal and lay them in or around a water body. The eggs are either attached to each other to form a raft or individually float on water. Usually, eggs are hatched within 24–48 hours after releasing larvae, which will later change to pupae before emerging as adult mosquito. This process may take one to several weeks, but depends on species and ambient temperatures. The newly emerged mosquito has to stand still around water to dry its wings before flying away. Thereafter, it will fly searching for humans or animals to feed on for several days. Adult mosquito can live up to four or eight weeks [182].

2.2.3 Symptoms of malaria

When an infectious mosquito bites a healthy person, it takes about 7–30 days for the first malaria signs to manifests. The duration of this period is called intrinsic incubation period (IIP). This period is usually shorter when infected by *P. falciparum* and longer for *P. malariae*. Sometimes, the manifestation of the first symptoms can be delayed by *prophylaxis* (preventive measures), whether the malaria is uncomplicated or uncomplicated [93].

2.2.3.1 Uncomplicated symptoms

Mild or uncomplicated malaria usually lasts between 6–10 hours after the attack and with a cold stage (sensation of cold, shivering). This stage is followed by a hot stage (fever, headaches, vomiting; seizures in young children) and a sweating stage (sweats, return to normal temperature, tiredness). Most often the attacks occur every second day with *P. falciparum*, *P. vivax* and *P. ovale* and every third day with *P. malariae*.

2.2.3.2 Complicated symptoms

Severe or complicated malaria normally occurs in people with low or no previous immunity. The *P. falciparum* is often the parasite causing severity, although this varies considerably from person to person. Particularly affected are individuals from unstable and non-malaria areas, children under five years of age, pregnant women and tourists. Severe malaria causes cerebral complications, such as the impairment of consciousness, seizures and coma.

2.2.4 Diagnosis of malaria

All the malaria symptoms mentioned in subsection 2.2.3 gives a sensitive diagnosis that can be confirmed through laboratory tests. However, a definite diagnosis of malaria depends on the presence of parasites in a blood smear examined under a microscope or via antigen identification through a Polymerase Chain Reaction (PCR).

2.3 Malaria Models

In epidemiological studies, mathematical models are used to investigate and understand the dynamics of infectious diseases and to suggest strategies and policies for preventive measures.

The first mathematical model of malaria transmission that appeared in the literature [140, 141] was later extended [100], and is now popularly known as Ross-Macdonald. Since then, several extensions have been made by incorporating new variables to the Ross-Macdonald model, which have addressed a wide range of problems [82, 135, 136, 139, 152, 159, 166].

Figure 2.5 represents malaria transmission cycle shown in Figure 2.4 within a compartmental model. The compartmental model is a technique used in mathematical modelling to describe the progression of a disease within a homogeneous population. This technique was first applied [140, 141] and then received greater attention, particularly in the modelling of infectious diseases. Figure 2.5 outlines the transition of the human and mosquito populations through the compartments:

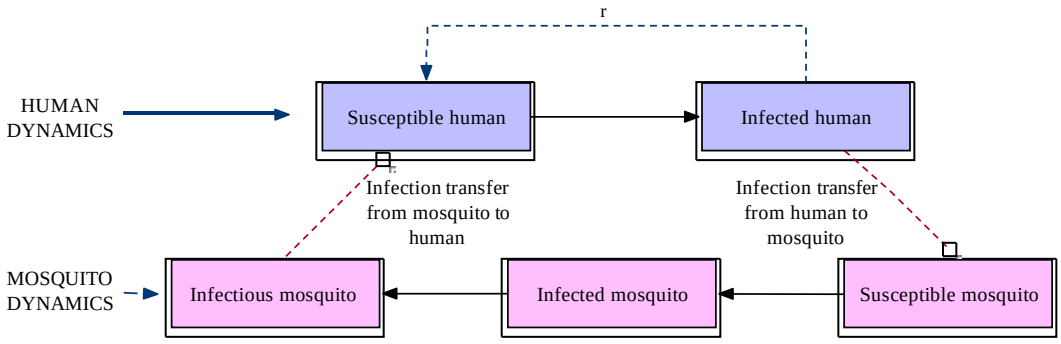

Figure 2.5: The Ross-Macdonald model structure describes the dynamics of malaria transmission.

Table 2.1: The definition of parameters used in the Ross-Macdonald model

Parameters	Description
m	Number of mosquitoes per human host.
a	Biting rate of the mosquitoes on their human host.
b_1	Infectiousness of humans to mosquitoes.
b_2	Susceptibility of human.
μ	Mortality of adult mosquitoes.
T	Incubation period of parasites within the host vector.
r	Rate of recovery of infected humans.

The transition of human and mosquito population in the Figure 2.5 through the compartments are outlined below

i. Flow of human from susceptible compartment to the infected compartment through recovery from infection.

ii. Flow of mosquitoes from a susceptible compartment to the infected compartment and a move to the infectious compartment.

iii. Human and mosquito populations are linked through the red dotted line showing the direction of transmission.

As shown in Figure 2.1, malaria transmits from mosquito to human as well as from human to mosquito. Hence, in the following subsections, a mathematical formulation describing the dynamics will be explained for each transmission scenario.

2.3.1 Transmission from mosquito to human

Suppose we consider the first human infected by a mosquito, and that each female mosquito bites a host on average a times per night. Assuming the density of female mosquito is denoted by m for every one human, each human is thus bitten ma times per

night. Of the density of female mosquitoes, m, only a fraction, w, of the mosquitoes have *sporozoites* in their salivary glands, and only a fraction, b_2, of these are actually infectious (capable of transmitting malaria virus) to the human. This reduces the number of infective bites per human per night to b_2maw. Hence, an infective bite will lead to a new infection provided the person bitten is not previously infected. If y is the infected proportion of the human population, then new infections emerge the rate $b_2maw(1 - y)$. Once the human is infected, the recovery will take place at rate r, i.e., the average time for an infectious humans to be cleared of malaria is $1/r$. Therefore, the differential equation governing the proportion of infected humans, y, can be written as:

$$\dot{y} = mab_2w(1 - y) - ry \tag{2.1}$$

where $\dot{y}$ denotes the change of the proportion of infections per unit of time. The definition of the parameters used in the formulations of equation 2.1 are presented in Table 2.1.

2.3.2 Transmission from human to mosquito

On the other hand, the transmission of malaria from human to mosquito is described in this subsection. Suppose the mosquito population is divided into three categories: a susceptible proportion $1 - v - w$ that is uninfected; an infectious proportion w with *sporozoites* in the salivary glands, and a latent proportion v that is infected, but not yet infectious (*sporozoites* needs to be developed). The susceptible mosquitoes become infected by biting infected humans, of which a fraction, b_1, contain *gametocytes* and are infectious. The proportion of latent mosquitoes increase at the rate $b_1ay(1-v-w)$. Hence, the newly infected mosquitoes would becomes infectious to humans if they

survive the incubation period, T , required for the development of the *gamatocytes* into *sporozoites*. Assuming a mortality rate of μ, i.e., an average life span of $1/\mu$, a proportion $e^{-\mu T}$ survive this period. Of the rate, $b_1 a y(1 - v - w)e^{-\mu T}$, move from the latent state to mosquitoes infected T days earlier. Note that proportions of the latent and infectious mosquitoes decrease exponentially through mortality.

The process of infectiousness in a mosquito's epidemiological state is governed by the following differential equations.

$$\begin{aligned} \dot{v} &= ab_1 y(1 - v - \hat{w}) - ab_1 \hat{y}(1 - \hat{v} - \hat{w})e^{-\mu T} - \mu v \\ \dot{w} &= ab_1 \hat{y}(1 - \hat{v} - \hat{w})e^{-\mu T} - \mu w \end{aligned} \tag{2.2}$$

where: $\hat{y} \equiv y(t - T)$, $\hat{v} \equiv v(t - T)$ and $\hat{w} \equiv w(t - T)$

2.3.3 Basic reproduction number

The basic reproduction number is a threshold quantity used to determine whether a disease will persist or become extinct in a population. This quantity is sometimes called the basic reproduction ratio, and is symbolically denoted by R_o. Thus, R_o quantifies the average number of people infected by a single infectious person in a susceptible population [24, 100].

The basic reproduction number, R_o, can be derived for equation (2.1) and system (2.2). By using equation (2.1) through system (2.2), R_o is defined by

$$R_0 = \frac{ma^2 b_1 b_2 e^{-\mu T}}{r\mu} \tag{2.3}$$

In the context of equation (2.1) to system (2.2), the quantity R_o is a threshold used

to determine the intensity of malaria's transmission in a human population. Hence, malaria can be spread into a human population, if $R_o > 1$ (i.e., each infection gives rise to at least one additional case).

By equating the equation (2.1) and system (2.2) to zero, and with the help of equation (2.3), the prevalence of infection in both humans and mosquitoes can be given as:

$$\begin{aligned} \hat{y} &= \frac{R_o - 1}{R_o - \frac{a}{\mu}} \\ \hat{w} &= \frac{R_o - 1}{R_o} \frac{\frac{a}{\mu}}{1 + \frac{a}{\mu}} e^{-\mu T} \end{aligned} \tag{2.4}$$

Therefore, provided $\hat{y}$ and $\hat{w}$ will be positive. This gives the basis for Ross's threshold of malaria transmission [140]. Based on this condition, the following conclusions are resolved:

i. The amount of malaria in a locality tends towards a fixed limit determined by the number of malaria-bearing mosquitoes and other factors.

ii. If the number of malaria-bearing mosquitoes is below a certain threshold, that limit will be zero.

The basic reproduction number, R_o, and herd protection threshold (HPT) for various infectious diseases are presented in Table 2.2 to better understand the pattern of spread and preventive measures.

This HPT is the essential level of protection given to a population through a mass vaccination programme in order to limit the transmission of a pathogen [144]. As it is practically impossible to effectively achieve 100% mass vaccination coverage, it is critical to determine the optimal threshold required for a population to be protected.

Table 2.2: Basic reproduction numbers and implied crude HPT for various diseases [23, 144, 168].

Infections	R_o	HPT(%)
Diphtheria	6–7	84–85
Influenza	2–4	50–75
Malaria	5–100	80–99
Measles	9–18	83–94
Mumps	4–14	75–93
Pertussis	5–35	90–94
Polio	2–4[a], 8–14[b]	80–86
Rubella	6–7	83–86
Smallpox	5–7	80–85

[a] Populations with good hygiene, [b] Populations with poor hygiene

For a population to be protected, the number of immune individuals must be higher than HPT [23, 168]. Thus, the essential level of coverage can be illustrated using the formula: $\geq (1 - s)$.

Table 2.2 presents the HPT values for various infectious diseases, and this calculation is as follows. In a situation where 30% of a population is susceptible: HPT$\geq (1 - 0.30)$, at least 70% of the population must be covered by the vaccination to obtain herd protection or immunity. The HPT value is different for each infectious disease, and this will provide a valuable target for immunization programmes, thus influencing the critical minimum level of vaccine coverage [23, 168]. Furthermore, the HPT varies between regions as well as by the characteristics of a given population and its mixing patterns.

2.3.4 Model with variability

The inclusion of variability in a malaria model aims to capture the bite frequency by mosquitoes on a human population. In this model, it is assumed that humans can be infected several times by various strains of malaria parasites. It describes the mean

number of infections, say X, harboured by any one human host. On the other hand, mosquitoes are assumed to be infected only once with no recovery. This description enables the mosquito population to be separated into susceptible, latent, and infective mosquitoes.

The generalization of this model follows from the model described in the literature [122]. The model describes the human population under N separate categories, each of which makes a proportion, φ_i, of the total population. This is based on the assumption that each category is homogeneous with respect to malaria's infection, and this differs from other categories in terms of susceptibility to infection, b_2, biting rate, a, and duration of disease, $\rho = 1/r$. Thus, each category of human is assumed to be infected randomly, so that the prevalence within a category would be $P_i = 1 - e^{-X_i}$ [122].

As shown in equation 2.1, humans are infected at a rate $ma_ib_{2,i}w$, where w denotes the proportion of infective mosquitoes. Susceptible mosquitoes are infected within the human category, i at a rate $a_ib_iP_i$. Hence, the rate at which mosquitoes become infected and the average of all categories is given by

$$h = \sum_i \varphi_i a_i b_i (1 - e^{-X_i}), \text{ for } i = 1, 2, \cdots, N \tag{2.5}$$

Then, equation 2.5 can be further expressed as:

$$\dot{X} = ma_ib_{2,i}w - r_iX_i \tag{2.6}$$

Using equation 2.5 and 2.6, the average human infectiousness corresponding to category i is given by

$$\dot{u} = \mu - \sum_i \varphi_i a_i b_1 (1 - e^{-X_i})u - \mu u \tag{2.7}$$

From equation 2.7, the frequency of susceptible mosquitoes is expressed as:

$$\dot{w} = e^{-\mu T} \sum_i \varphi_i a_i b_1 (1 - e^{-\hat{X}_i})\hat{u} - \mu w \tag{2.8}$$

where $\hat{X} = X_i(t - T)$ and $\hat{u} = u(t - T)$ for the frequency of infectious mosquitoes. Using equation 2.5 through 2.8, the basic reproduction number of the malaria model with variability can be given as:

$$R_0 = \bar{R}_0 \left[1 + \frac{var(a)}{\bar{a}^2} + 2\frac{cov(a, b_2)}{\bar{a}\bar{b}_2} + 2\frac{cov(a, \rho)}{\bar{a}\bar{\rho}} + \frac{cov(\rho, b_2)}{\bar{\rho}\bar{b}_2}\right]^1 \tag{2.9}$$

where $\bar{R}_0 = \frac{m\bar{a}b_1\bar{b}_2 e^{-\mu T}}{r\mu}$ denotes the basic reproduction number due to the mean parameters in the population, $\bar{x}$ denotes the mean of x, $var(x)$ denotes the variance of x, $cov(x, y)$ denotes the covariance of x and y.

2.3.5 Model with immunity

The inclusion of immunity is essential to realistically describe the dynamics in malaria transmission. Using a model without considering immunity might lead to an unrealistic understanding of the malaria transmission pattern. The general model of malaria with immunity consists of three differential equations denoting changes in the proportions of susceptible humans, denoted by x, infected humans, denoted by y, and immune humans, denoted by z (see Figure 2.6).

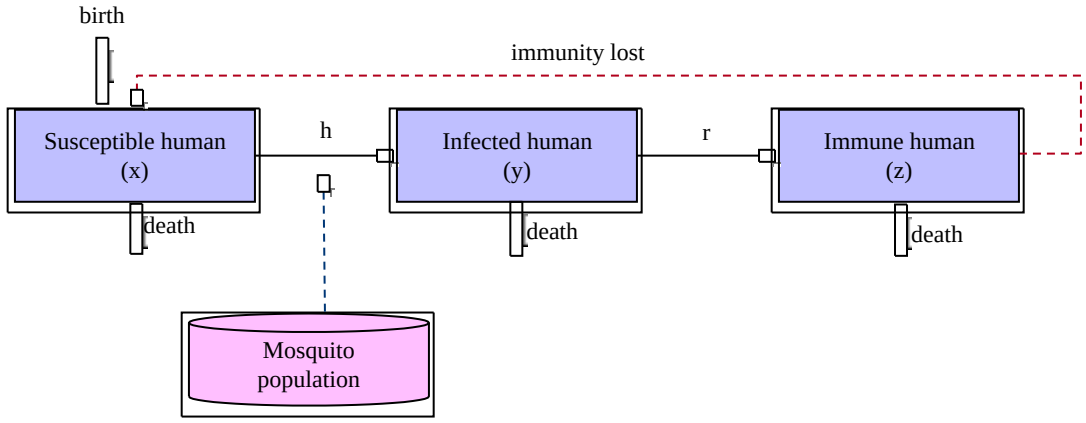

Figure 2.6: The malaria model with immunity in human population.

The transition in Figure 2.6:

i. Flow of susceptible humans' transit to infected class at rate h.

ii. Flow of infected humans' transit to immune class at rate r.

iii. Flow of immune humans after loss of immunity and transit to susceptible class in the absence of any form of intervention at rate γ.

The malaria model with immunity is an extension of the model proposed in [26, 27], and the differential equations governing Figure 2.6 are given by system 2.10, while the definition of the parameters are presented in Table 2.3.

$$\begin{aligned} \dot{x} &= \delta - \delta x - hx - \gamma z \\ \dot{y} &= hx - (r + \delta)y \\ \dot{z} &= ry - (\gamma + \delta)z \end{aligned} \tag{2.10}$$

The immunity for disease like malaria is boosted by new infections [175], and only lasts for τ years in the absence of new infections. If τ is equal to the mean residence

Table 2.3: The definition of parameters used in Ross-Macdonald model with immunity.

Parameters	Description
h	Rate at which susceptible individuals becomes infected.
r	Rate at which infected individuals are recovered and move into immune class.
γ	Rate at which immune individuals becomes susceptible again.
δ	Deaths rate occur and are unaffected by disease status.
$1/\delta$	Life-expectancy.

time in the immune class, that is $\tau = 1/(\gamma + \delta)$, where the parameter γ is given by equation 2.11 according to the definition in [26].

$$\gamma(h) = \frac{(h+\delta)e^{-(h+\delta)\tau}}{1 - e^{-(h+\delta)\tau}} \tag{2.11}$$

Thus, the infection rate, h, can be given as follows

$$h = ma^2 b_1 b_2 e^{-\mu T} \frac{y}{\mu + ay} \tag{2.12}$$

The analysis of the malaria model with immunity in system 2.10 at the equilibrium point leads to a basic reproduction number, shown below

$$R_o = \frac{ma^2 b_1 b_2 e^{-\mu T}}{(r+\delta)\mu} \tag{2.13}$$

2.3.6 Model with asexual stage vaccine

The asexual stage vaccines act within the parasites to protect individuals from infection (anti-*sporozoite*) or from developing parasitaemia and the disease (anti-blood stage vaccines). In the mass vaccination of a proportion p (of all newborns for example) an asexual stage vaccine is stimulated. This is achieved by letting a proportion p be

born as immune and a proportion $1 - p$ as susceptible. In using the malaria model with immunity described in system 2.10, and the resultant model with an asexual stage vaccination built upon it with no other changes to the model, this is given by

$$\begin{aligned} \dot{x} &= \delta(1-p) - \delta x - hx - \gamma x \\ \dot{y} &= hx - (r+\delta)y \\ \dot{z} &= \delta p + ry - (\gamma + \delta)z \end{aligned} \tag{2.14}$$

By solving equation 2.14 along the equilibrium point, the basic reproduction number is given by

$$R_0^t = R_0\left(1 - \frac{\delta}{\delta + \gamma_0}p\right) \tag{2.15}$$

where R_o denotes the basic reproduction number of the model with no vaccination shown in equation 2.13 and γ_0 denotes the rate of loss of immunity in the absence of any infection ($h = 0$). Hence, malaria cannot invade a population if the basic reproduction number $R_0^t < 1$. This condition leads to equation 2.16, and thus indicates malaria's eradication threshold.

$$p > \left(1 - \frac{1}{R_0}\right)\left(1 + \frac{\gamma_0}{\delta}\right) \tag{2.16}$$

2.3.7 Model of transmission-blocking vaccine

A malaria transmission model with a blocking vaccine acts within the asexual stages of the parasites but do not protect individuals from infection. In this model, the vaccine blocks the transmission of malarial infection from human to human through a mosquito bite.

However, a mass vaccination scenario with a transmission-blocking vaccine can

be stimulated by splitting the described model with immunity into two categories (see system 2.10). The first category represents $(1 - p)$ proportion of the population that is not vaccinated. This is almost identical to the model with immunity except that immunity against the asexual stage of the parasite is lifelong. Then, in the second category, the proportion, p of the population receives the vaccine. This category is infected and later becomes immune at the same rate as the unvaccinated category. However, it does not make any contribution to transmission during the period when the vaccine is effective. The vaccine loses its effectiveness at a rate v. Hence, the differential equation governing the model of transmission with the blocking vaccination can be given by

$$\begin{aligned}
\dot{x}_u &= \delta(1-p) - (\delta + h)x_u - vx_u \\
\dot{x}_v &= \delta p - (\delta + h)x_v - vx_v \\
\dot{y}_u &= (r + \delta)y_u + vy_u \\
\dot{y}_v &= hx_v - (r + \delta)y_v - vy_v \\
\dot{z} &= r(y_u + y_v) - \delta z
\end{aligned} \tag{2.17}$$

where the subscript v denotes the vaccinated, and the subscript u is the unvaccinated category. The basic reproduction number of the system 2.17 is given by

$$R_0^{tt} = R_0\left(1 - \frac{\delta(r+\delta)}{(\delta+v)(r+\delta+v)}p\right) \tag{2.18}$$

where R_o denotes the basic reproduction number of the model with no vaccination (see system 2.14). Hence, malaria cannot invade a population if $R_o^{tt} < 1$, and this condition would be satisfied if

$$p > \left(1 - \frac{1}{R_o}\right)\left(1 + \frac{v}{r+\delta}\right)\left(1 + \frac{v}{\delta}\right) \tag{2.19}$$

2.3.8 Relationship between DDE and ODE malaria model

Using the malaria model in Figure 2.7, the relationship between delay differential equation (DDE) and ordinary differential equation (ODE) is as follows.

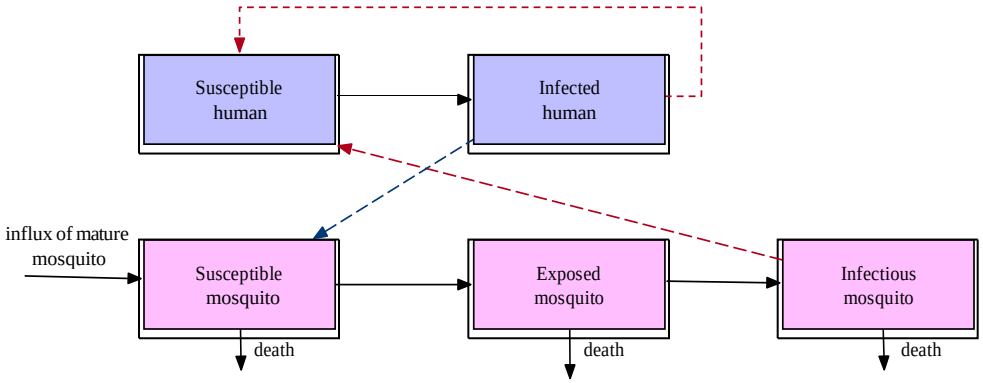

Figure 2.7: A malaria model with a SIS and SEI structure in human and mosquito populations respectively.

2.3.8.1 Model formulation with delay

Assuming the humans dynamics are typified as susceptible-infectious-susceptible (SIS) and the dynamics in the mosquito population is characterised by susceptible-exposed-infectious (SEI), as shown in Figure 2.7. The classes are described as: S_H : susceptible humans, I_H : infectious humans S_V : susceptible mosquitoes, E_V : exposed mosquitoes, I_V : infectious mosquitoes. Then, each state variables in Figure 2.7 is defined using differential equation (see system 2.20). The definition of parameters under subsection 2.3.8.1 is provided in Table 2.4.

$$\begin{aligned} x(t) &= \frac{I_H(t)}{(S_H(t) + I_H(t))} \\ y(t) &= \frac{E_V(t)}{(S_V(t) + E_V(t) + I_V(t))} \\ z(t) &= \frac{I_v(t)}{(S_V(t) + E_V(t) + I_V(t))} \end{aligned} \tag{2.20}$$

where $x(t)$ denotes the proportion of infectious humans at time t; $y(t)$ means the proportion of infectious but not yet infectious mosquitoes at time t and $z(t)$ is the proportion of infectious mosquitoes at time t. Hence, by using system 2.20, the following differential equation governing Figure 2.7 is expressed as:

$$\begin{aligned} \dot{x}(t) &= mabz(t)(1 - x(t)) - \gamma x(t) \\ \dot{y}(t) &= acx(t)(1 - y(t) - z(t)) - \mu y(t) - \Omega_{y(t)} \\ \dot{z}(t) &= acx(t - \tau)(1 - y(t - \tau) - z(t - \tau))e^{-\mu\tau} - \mu z(t) \end{aligned} \tag{2.21}$$

$$\text{where: } \Omega_{y(t)} = acx(t - \tau)(1 - y(t - \tau) - z(t - \tau))e^{-\mu\tau}$$

The differential equation in system 2.21 has two equilibrium points; the disease-free equilibrium point is given by $x_0 = 0$, $y_0 = 0$ and $z_0 = 0$, and endemic equilibrium point expressed by

$$\begin{aligned} x^* &= \frac{ma^2bce^{-\mu\tau} - \gamma\mu}{ma^2bce^{-\mu\tau} + ac\gamma} \\ y^* &= \left(\frac{1 - e^{-\mu\tau}}{e^{-\mu\tau}}\right)\left(\frac{ma^2bce^{-\mu\tau} - \gamma\tau}{ma^2bc + mab\mu}\right) \\ z^* &= \frac{ma^2bce^{-\mu\tau} - \gamma\mu}{ma^2bc + mab\mu} \end{aligned} \tag{2.22}$$

The system of DDE, shown in 2.21 can further be form a system of ODE by ignoring the time delay and assuming that the ratio of the proportion of exposed mosquitoes to infectious mosquitoes is at an equilibrium. This implies:

$$y(t) = \left(\frac{1 - e^{-\mu\tau}}{e^{-\mu\tau}}\right) z(t) \tag{2.23}$$

By substituting equation 2.23 in system 2.21 and by ignoring the delay, it becomes:

$$\begin{cases} \dot{x}(t) = mabz(1-x) - \gamma x \\ \dot{z}(t) = acx\left(1 - \left(\frac{1 - e^{-\mu\tau}}{e^{-\mu\tau}}\right) z\right) - z\, e^{-\mu\tau} - \mu z \end{cases} \tag{2.24}$$

which further simplifies as:

$$\begin{cases} \dot{x}(t) = mabz(1-x) - \gamma x \\ \dot{z}(t) = acx(e^{-\mu\tau} - z) - \mu z \end{cases} \tag{2.25}$$

Remark: The ODE Ross-Macdonald model has the equivalent equilibrium points as the DDE Ross-Macdonald model. Using the method of [165], the basic reproductive number is denoted by R_o, where the disease-free equilibrium point loses stability. However, most frequently, the basic reproductive number for malaria from the Ross-Macdonald model is given by

$$\hat{R}_o = \frac{ma^2bce^{-\mu\tau}}{\gamma\mu} \tag{2.26}$$

The disease-free equilibrium point is locally asymptotically stable when $R_o < 1$, and unstable when $R_o > 1$. The expression for R_o shows that malaria can be eliminated by increasing the mosquito death rate, μ or reducing the mosquito biting rate, a.

Table 2.4: Definition of the parameters used for the model in Figure 2.7

Parameters	Description
m	Number of female mosquitoes per human host.
a	Number of bites per mosquito per unit of time.
b	Prob. of transm. from infectious mosquitoes to human to human per bite.
c	Prob. of transm. from infectious human to mosquitoes per bite.
γ	Recovery rate of humans.
μ	Death rate of mosquitoes.
τ	Extrinsic incubation period.

2.4 Summary

Chapter 2 presents the theoretical framework of a malaria parasite transmission cycle, both in humans and mosquitoes. This was addressed by using the Ross-Macdonald model and by discussing its properties, such as the variable ramifications and basic reproduction number. Chapter 3 presents a climate dependent malaria model with a delay in mosquito dynamics.

Chapter 3

A Climate Dependent Malaria Model with Delay in Mosquito Dynamics

This chapter provides a mathematical model of the dynamics of malaria with temperature-dependent incubation periods in the mosquito exposed and juvenile mosquito maturation. The impact of delay due to the extrinsic incubation period and juvenile maturation on the dynamics of the malaria transmission was investigated, and performed a sensitivity analysis.

3.1 Introduction

A number of studies have been recently carried out to understand the impact of climate change on malaria transmission. The climate factors, such as temperature, rainfall and humidity, influence malaria's transmission in different ways. Temperature is the large-scale driver of malaria, which directly influences aquatic development process [33], parasite development [37], survival [51] and biting rate [172] of the mosquito. In dif-

ferent circumstances, rainfall can create breeding sites for mosquito reproduction, and consequently increases the likelihood of malaria prevalence. In contrast, high rainfall negatively distorts the aquatic mosquito's metamorphosis to adult. Nonetheless, study has further shown that a causal relationship exists between rainfall and the risk of malaria [101]. For that reason, understanding the population dynamics of mosquitoes and the relationship with climate factors is crucial to the study of mosquito-borne diseases [48]. The control of mosquito abundance in a human population remains a key challenging issue affecting the persistence of vector-borne diseases [184]. Therefore, it is important to include aquatic dynamics of mosquito in malaria models to enable effective and realistic control and preventive measures.

Recent studies have, respectively, focused on the impact of temperature on malaria risk with diurnal temperature fluctuation, vector population and the shift in optimal temperature [35, 36, 112]. There have also been efforts to study the effects of temperature and rainfall on the spatio-temporal risk of malaria [120], and the impact of variability in temperature and rainfall on the transmission of malaria [131]. Similarly, a non-autonomous deterministic model for assessing the effect of temperature variability on malaria transmission dynamics was studied [21]. A climate-based malaria model with periodic birth rate and age structure in the vector population was also analyzed [98]. Moreover, the climate-dependent modelling of malaria transmission dynamics to predict a malaria outbreak season was studied [164] to guide public health policies. Compartmental epidemic models in periodic environments were explicitly analyzed to study the impact of periodic contacts or migrations on the disease transmission [171]. Furthermore, several other models have been developed to study the impact of mosquito maturation delay on vector-borne diseases. For instance, the role of non-linear birth and maturation delay [121] and the sensitivity of maturation delay

on temperature [147] applied a Richer birth function; furthermore the stability analysis of a population model with maturation delay [186] was studied, as were the roles of maturation delay and vaccination with optimal control [123].

Previous studies [21, 116, 131, 163] have not dealt with the impact of temperature on *extrinsic incubation period (EIP)* delay, referring to the time taken for malaria parasites to complete its development in a mosquito. This period hinders the development of efficient and effective prevention and control mechanisms, which should not only focus on controlling the human infection, but also address the vector control. In particular, a common drawback [116] amongst studies is the focus on transmission in humans by considering the so-called *intrinsic incubation period (IIP)*. This refers to the time taken by malaria parasites to complete their development in humans. Furthermore, the incubation period in humans is not a requirement for the prevention and control of malaria in the light of climate factors, but is rather used for clinical diagnostic purposes or outbreak detection [151]. For this reason, EIP is foremost considered a mechanism for the prevention and control of malaria transmission. Moreover, the climate-dependent maturation delay of the aquatic mosquito is a determining factor for the abundance of adult female mosquitoes, thus its incorporation in the model of malaria transmission would offer well-grounded dynamics. Likewise, studies [21, 116, 131, 163] have considered the climate-dependent rate of transfer due to maturation. This rate is density-dependent, but does not take into account the spatio-temporal variability of the aquatic mosquito maturation.

This chapter addresses these limitations by studying malaria transmission from the perspective of mosquito dynamics. Based on the framework formulated in [116], the impact of EIP (in mosquito) on human infectiousness due to malaria will be analyse. A new mathematical model will be develop and introduce a temperature-dependent

incubation state into the exposed mosquito population. In order to describe the effects of temperature variability on spatio-temporal infectiousness in humans, therefore, two categories of incubation period for the exposed mosquito population, namely a short and a long period are considered. These periods are described by exponential distribution and step function respectively. An aquatic state is further introduced in the proposed model to capture the influx of adult mosquitoes through maturation from the immature stage. This provides an insight into rates at which the parasite and maturation of immature mosquitoes develop at different temperature scenarios. The model proposed in this chapter closes the knowledge gaps to those presented in the previous studies [21, 116, 131, 163]. Its building blocks can together provide possible measures to mitigate, prevent, control and predict the likelihood of malaria outbreak.

3.2 Model Formulation

In this section, and using a schematic diagram 3.2, firstly, a detailed description of how malaria pathogens transmit within its hosts, via human and mosquito. Secondly, the formulation of differential equations governing the proposed model and discuss its epidemiological positivity and bounded properties will be present.

The proposed model in this chapter is motivated by the malaria models used in [21, 116, 131, 163]. A significant difference is given by the derivation of a new malaria model that incorporates delay due to mosquito maturation and extrinsic incubation. This accounts for the spatial and temporal risk of malaria on human infectiousness since both mosquito maturation and extrinsic incubation depends on temperature and rainfall [18].

The human population at time t denoted by $N_h(t)$ is split into four mutually ex-

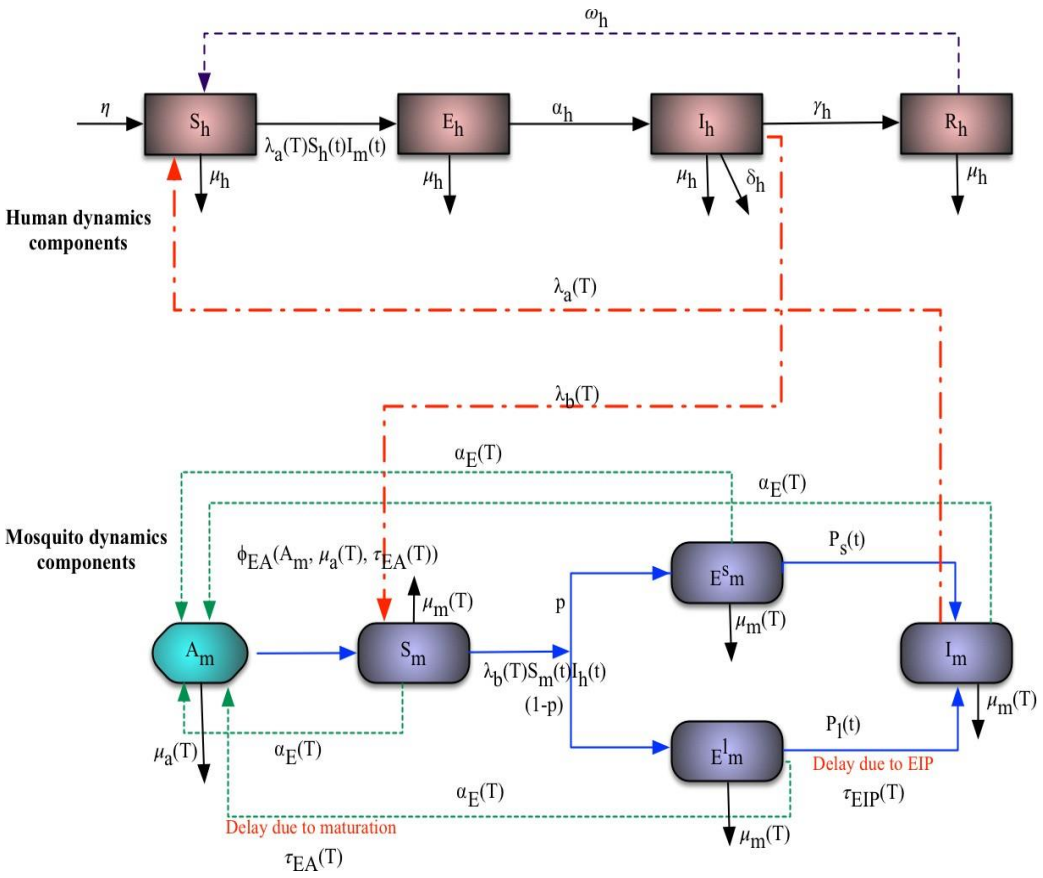

Figure 3.1: Schematic diagram of the model (3.9), showing malaria transmission dynamics.

clusive epidemiological categories represented by the following state variables: the susceptible class $S_h(t)$, exposed class $E_h(t)$, infective class $I_h(t)$ and recovered class $R_h(t)$, such that:

$$N_h(t) = S_h(t) + E_h(t) + I_h(t) + R_h(t). \tag{3.1}$$

Similarly, the mosquito population at time t denoted by $N_m(t)$, is divided into two categories, namely mature mosquitoes $N_A(t)$ and immature mosquitoes (eggs, larvae

and pupae) $A_m(t)$. Hence, the mosquito population at time t is given by

$$N_m(t) = A_m(t) + N_A(t). \tag{3.2}$$

Furthermore, the $N_A(t)$ is split into four sub-categories, i.e., susceptible mosquitoes $S_m(t)$, short-term exposed mosquitoes $E_m^s(t)$, long-term exposed mosquitoes $E_m^l(t)$ and infectious mosquitoes $I_m(t)$. Thus, the total mosquito population at time t becomes:

$$N_m(t) = A_m(t) + S_m(t) + E_m^s(t) + E_m^l(t) + I_m(t). \tag{3.3}$$

In Figure 3.1, adult mosquitoes in an exposed state are divided into two classes, namely $E_m^s(t)$ and $E_m^l(t)$, to distinguish the short and long term EIPs. Herein the EIP refers to the time taken for a *plasmodium species* (malaria parasites) to develop within its host mosquito while the development time taken in the human host is called the intrinsic incubation period (IIP), for instance, when a susceptible mosquito bites an infectious human, and sips human blood that contains *gametocytes* (cell in *plasmodium species*). *Gametocytes* transits the *plasmodium* to the mosquito from which the mosquito will become infected. At this stage, the infected mosquito is not capable of spreading malaria to susceptible humans until the EIP cycle is completed (usually taking around 10-14 days [36]). A form of the parasite, called *sporozoite*, will migrate to the mosquito salivary gland, and thereafter the mosquito will become virulent. The time in which the *plasmodium* species develops within the mosquito is driven largely by environmental temperature [36]. For instance, in a region where the the temperature pattern is predominantly lower, the EIP usually takes a longer time than a region with a relatively higher temperature. Consequently, the EIP is a large-scale risk factor for determining malaria's infectiousness in the human population, and it dynamically varies over space

and time.

In the proposed model, the states $E^s_m(t)$ and $E^l_m(t)$ is introduced in the exposed mosquito dynamics to account for temporal changes in human infectiousness. While both short and long term incubation was introduced in exposed human class by [116] which captured the dynamics of malaria in a human host, while noting that the proposed model is exclusively different as temperature-dependent extrinsic incubation period is introduce to the mosquito dynamics.

3.2.1 Aquatic Mosquito State

A matured female mosquito needs blood from humans and animals as meal for nourishment and egg development. Most often, mosquitoes lay their eggs directly into water, and sometimes lay near water bodies or moist areas on a farmland. The development from egg to adult usually takes about 1-2 weeks depending on the environmental niche; this is significantly large compared to the average lifespan of an adult mosquito, which is about 3 weeks [147]. In Figure 3.1, the parameter $\alpha_E(T)$ denotes the temperature-dependent eggs deposition rate for adult female mosquitoes. All the compartments would then have an equal egg deposition rate of $\alpha_E(T)$ in the aquatic mosquito state $A_m(t)$. This state is limited to certain resources, such as the breeding sites and nutrients available to sustain the population of immature mosquitoes. Assuming that a carrying capacity of K_c for the $A_m(t)$ and propose $K_c > A_m(t)\ \forall\ t$, which accounts for the possibility of unbounded growth in the immature mosquito population [130, 131]. Thus, the logistic growth rate (see Figure 3.1) for the immature mosquito population is given as follows

$$\alpha_E(T)\left(1-\frac{A_m(t)}{K_c}\right)[S_m(t)+E^s_m(t)+E^l_m(t)+I_m(t)].$$

The development of mosquitoes from egg to adult is density dependent, thus a Richer function is taken for the birth rate into the adult mosquitoes [123, 169]. Consider the maturation of the aquatic stages from egg to adult mosquitoe, and let the finite and positive constant $\tau_1(T)$ as shown in Figure 3.1 be the maturation time of the mosquito (meaning the average time required for an egg to develop into an adult mosquito). Assuming $\mu_a(T)$ is the natural death rate of mosquitoes in the immature stages, then the term $e^{-\mu_a(T)\tau_1(T)}$ represents the survival probability of the aquatic stages [75]. Thus, the immature mosquito's maturation into an adult female mosquito with delay $\tau_1(T)$ can be written as

$$p_{EA}(\hat{T})A_m(t-\tau_1(T))e^{-\mu_a(T)\tau_1(T)}e^{-A_m(t-\tau_1(T))},$$

where the term $A_m(t-\tau_1(T))$ represents the immature mosquito at $\tau_1(T)$ and $e^{-A_m(t-\tau_1(T))}$ reflects the survival probability of becoming an adult over the maturation period. The malaria infection is transmitted to susceptible adult female mosquito (called *anopheles*) following effective contact with an infectious human (via blood meals), at a temperature rate $\lambda_b(T)$. When the *anopheles* successfully obtain blood meals from an infectious human host, they will then rest for a few days while the blood ingested will undergo the process of egg development and infectiousness [35]. In essence, mosquitoes suffer a natural death at a temperature-dependent rate $\mu_m(T)$, and moreover unlike humans, experience no recovery after infection. The link between malaria's prevalence and climate factors, particularly with rainfall, has shown to coincide with the malaria season, which increases its incidence by creating breeding sites for mosquito abundance. However, excessive rainfall can negatively affect the reproduction cycle of the mosquito by flushing out the breeding space.

3.2.2 Exposed Mosquito State

Suppose that the exposed mosquitoes will remain in the exposed class after entering the compartment until they are infectious. The population of mosquito decreases exponentially by natural death within the compartment as time t increases. Thus, the exposed mosquitoes' compartment is defined by

$$E_m(t) = \int_0^{\infty} \lambda_b S_m(t-u) I_h(t-u) e^{-\mu_m u} P(u) du \tag{3.4}$$

where $P(u)$ is the Probability Distribution Function (pdf) that describes the exposed mosquitoes undergoing a latent period at time t after entering the exposed compartment. Also, $P(u)$ has the following properties:

i. $P : [0, \infty) \rightarrow [0, 1]$ is a non-increasing piece-wise continuous function with many possible finite jumps and satisfying: $P(0^+) = 1$ and $\lim_{t\to\infty} P(u) = 0$

ii. The integrand of $P(u)$ in the interval (0,1) is positive definite.

Noting that the exposed mosquito compartment is classified as short and long-term EIP to describe the influence of temperature. Let $p \in (0, 1)$ be the probability that an exposed mosquito experiences a short and long term incubation period upon successful contact with an infected human. Then, $P(u)$ can be represented by $P(u) = pP_s(u) + (1 - p)P_l(u)$, where P_s and P_l are explicitly defined in equation (3.5) and (3.6). The short-term EIP is modelled according to the exponential distribution with an average $\theta(T)^{-1}$, and thus the random variable U represents the probability at which the exposed mosquitoes survive the short-term EIP. The exposed mosquito population that survives the EIP is characterized by the survival function of the exponential distribution. The parametrized pdf is given by

$$P_s(u;\theta(T)) = \theta(T)e^{-\theta(T)u},\ u > 0, \tag{3.5}$$

For the long-term EIP, a step-function is assumed for the mosquito with a long term incubation and denoted by a fixed time τ [116]:

$$P_l(u) = 1, \text{if} u \in [0,\tau] \text{ and } P_l(u) = 0, \text{if} u \in (0,\infty) \tag{3.6}$$

The step-function here is to describe the long term EIP pattern in the exposed mosquitoes due to the influence of temperature. By substituting properties (i) and (ii) into the mathematical representation in equation (3.4), the short and long term EIP in the exposed mosquitoes compartment now becomes

$$\begin{aligned} E_m^s(t) &= \int_0^\infty p\lambda_b S_m(t-u)I_h(t-u)e^{-\mu_m u}P_s(u)du \\ E_m^l(t) &= \int_0^\tau (1-p)\lambda_b S_m(t-u)I_h(t-u)e^{-\mu_m u}P_l(u)du \end{aligned} \tag{3.7}$$

By differentiating (3.7) with respect to time t, then obtain

$$\begin{aligned} \dot{E_m}^s(t) &= p\lambda_b(T)S_m(t)I_h(t) - (\mu_m(T) + \theta_m(T))E^s(t) \\ \dot{E_m}^l(t) &= (1-p)\lambda_b(T)S_m(t)I_h(t) - (1-p)\lambda_b(T)\psi(\tau_2,\mu_m,T) - \mu_m(T)E^l(t) \end{aligned} \tag{3.8}$$

The malaria model depicted by the schematic diagram in Figure 3.1 can thus be described using differential equations, as shown in system (3.9). This involves four states of the human population, four states of the adult female mosquito population and a combined state of aquatic mosquitoes containing eggs, larva and pupa.

$$
\begin{aligned}
\dot{S}_h(t) &= \eta - \lambda_a(T)S_h(t)I_m(t) - \mu_h S_h(t) + \omega_h R_h(t)\\
\dot{E}_h(t) &= \lambda_a(T)S_h(t)I_m(t) - (\alpha_h + \mu_h)E_h(t)\\
\dot{I}_h(t) &= \alpha_h E_h(t) - (\mu_h + \delta_h + \gamma_h)I_h(t)\\
\dot{R}_h(t) &= \gamma_h I_h(t) - (\mu_h + \omega_h)R_h(t)\\
\dot{A}_m(t) &= \varphi_E(T)\phi(\tau_1, \mu_a, \hat{T})\left(1 - \frac{A_m(t)}{K}\right) - \mu_A(T)A_m(t) - \varphi_{EA}(T)\phi(\tau_1, \mu_a, \hat{T})\\
\dot{S}_m(t) &= \varphi_{EA}(T)\phi(\tau_1, \mu_a, \hat{T}) - \lambda_b(T)S_m(t)I_h(t) - \mu_m(T)S_m(t)\\
\dot{E}^s_m(t) &= p\lambda_b(T)S_m(t)I_h(t) - (\mu_m(T) + \theta_m(T))E^s_m(t)\\
\dot{E}^l_m(t) &= (1-p)\lambda_b(T)S_m(t)I_h(t) - (1-p)\lambda_b(T)\psi(\tau_2, \mu_m, T) - \mu_m(T)E^l_m(t)\\
\dot{I}_m(t) &= \theta_m(T)E^s_m(t) + (1-p)\lambda_b(T)\psi(\tau_2, \mu_m, T) - \mu_m(T)I_m(t)
\end{aligned}
\tag{3.9}
$$

The initial conditions of the system (3.9) state variables are given as: $S_h(0) > 0, E_h(0) \geq 0, I_h(0) \geq 0, R_h(0) \geq 0, A_m(0) > 0, S_m(\theta) = \varphi_{sm}(t) > 0, E^s_m(0) \geq 0, E^l_m(0) \geq 0, I_m(\theta) = \varphi_{im}(t) > 0$, where $\varphi_{sm}(\theta)$ and $\varphi_{im}(\theta)$ are positive continuous functions for $\theta \in [-\tau, 0]$. For convenience, the system (3.9) is scale down by replacing the terms $\phi(\tau_1, \mu_a, \hat{T})$ and $\psi(\tau_1, \mu_a, T)$ such that:

$$
\begin{aligned}
\phi(\tau_1, \mu_a, \hat{T}) &= A_m(t - \tau_1(\hat{T}))e^{-\mu_a(T)\tau_1(\hat{T})}e^{-A_m(t-\tau_1(\hat{T}))}\\
\psi(\tau_1, \mu_m, T) &= S_m(t - \tau_1(T))I_h(t - \tau_1(T))e^{-\mu_m(T)\tau_1(T)}
\end{aligned}
\tag{3.10}
$$

where $\varphi_{EA}(T)$ denotes the probability that an egg can survive to become an adult mosquito, and is defined functionally by $\varphi_{EA}(T) = -0.00924\hat{T}^2 + 0.453\hat{T} - 4.77$ [112]. On the other hand, $\lambda_a(T) = bc_m(T) = abm$ and $\lambda_b(T) = ac_m(T)$ correspond to forces of the malaria infection transmission. The meaning of the state variables, parameters and their corresponding values can be found in Table 3.1 and Table 4.3. In

equation (3.9) and (3.10), all the temperature dependent parameters are assumed to be continuous, bounded, positive and ω-periodic. As in [21], we also let $T = T(t)$ and $\hat{T} = T(t) + \delta_T$ to denote air and water temperature at time t, respectively.

3.2.3 Temperature-Dependent Model

The system (3.9) has temperature-dependent parameters, including the adult mosquito biting rate $c_m(T)$, adult mosquito mortality rate $\mu_m(T)$ and adult mosquito egg deposition rate $\alpha_E(T)$. A quadratic function is used to describe the relationship between temperature and these parameters as defined in [112]:

$$\begin{aligned} c_m(T) &= -0.00014T^2 + 0.027T - 0.322 \\ \mu_m(T) &= -\ln(-0.000828T^2 + 0.0367T + 0.522) \\ \alpha_E(T) &= -0.153T^2 + 8.61T - 97.7 \end{aligned} \tag{3.11}$$

However, rainfall is also another factor that influences the dynamic of mosquitos [131], which is largely related to their immature stages. In the event of excessive rainfall, aquatic mosquitoes usually suffer negative effects because the mosquito eggs, larva and pupa are flushed from the water body to a potentially unfavourable environment, which might eventually affect their survival and growth rates. Consequently, there would be a reduction in the population of immature mosquitoes emerging as adults.

Hence, the per-capita temperature-dependent death rate of the immature mosquitoes, $\mu_a(T)$ is defined in [133], and given by

$$\mu_a(T) = 1/[8.560 + 20.654[1 + (T/19.759)^{6.827}]^{-1}] \tag{3.12}$$

Therefore, each of these temperature-dependent parameter functions in equation (3.11) and equation (3.12) are visualised in Figures 3.2–3.5.

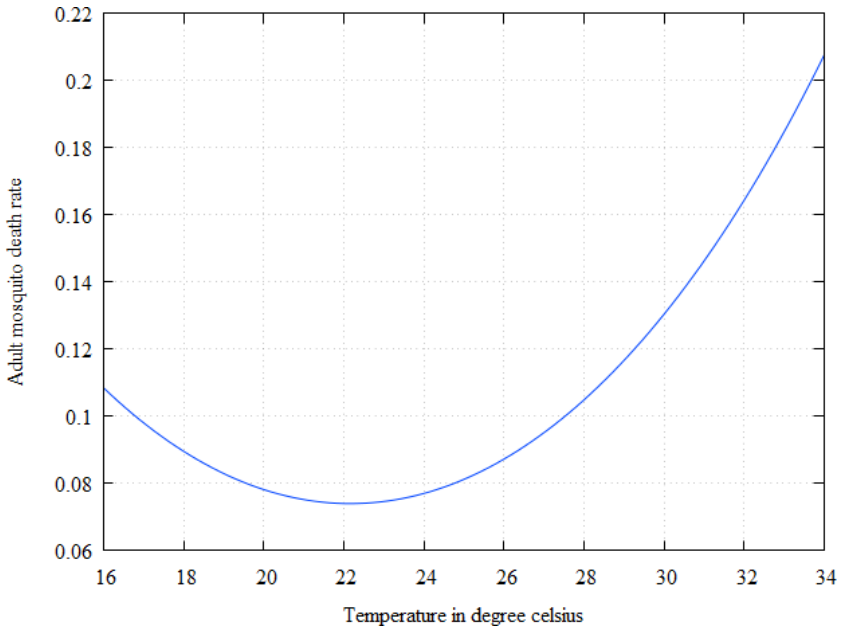

Figure 3.2: Adult mosquito death rate ($\mu_m(T)$) as a function of temperature (in °C).

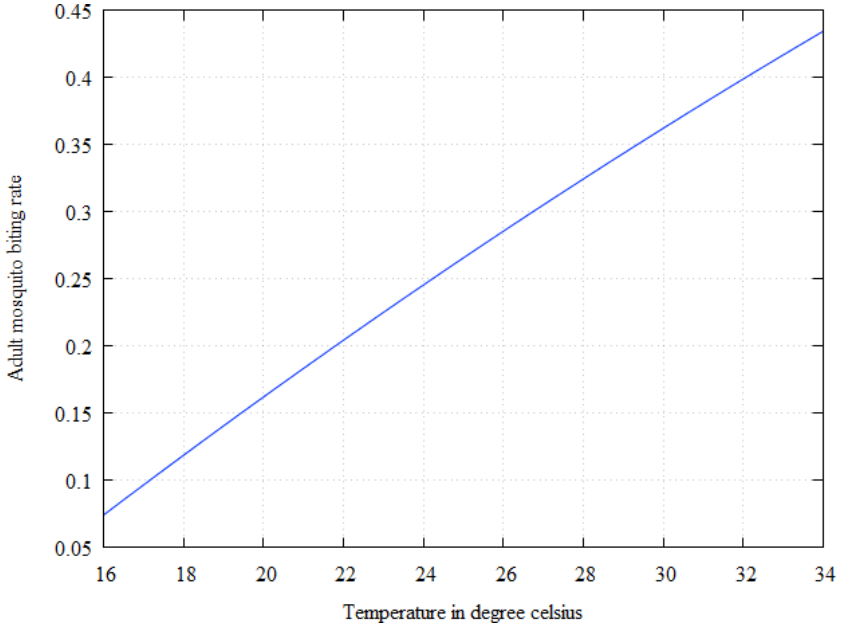

Figure 3.3: Adult mosquito biting rate ($c_m(T)$) as a function of temperature (in ˚C).

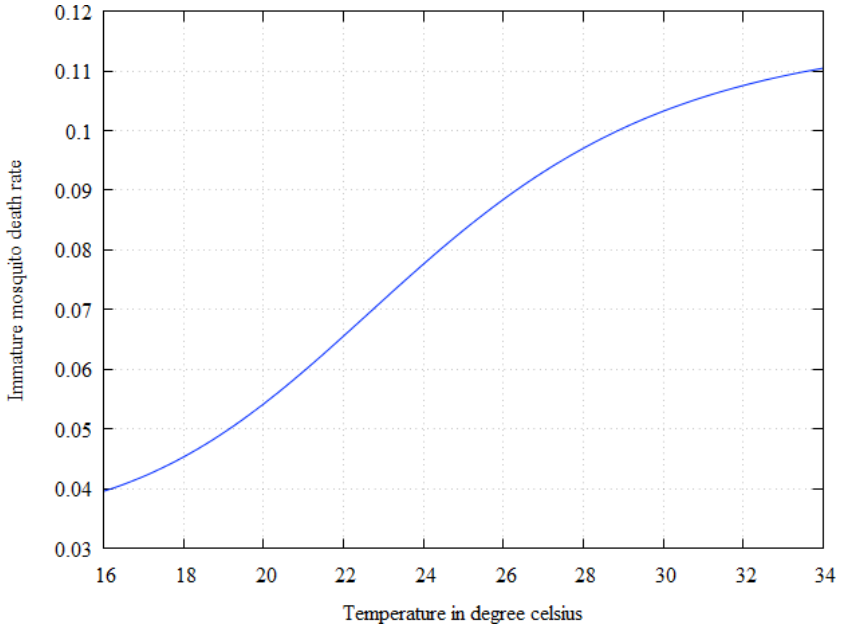

Figure 3.4: Immature mosquito death rate ($\mu_a(T)$) as a function of temperature (in ˚C).

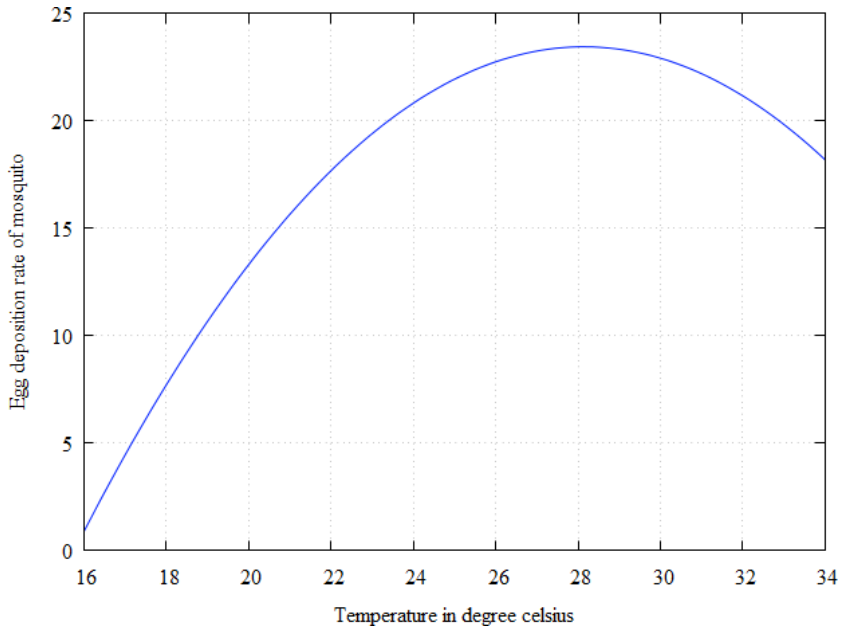

Figure 3.5: Egg deposition rate of mosquito ($\alpha_E(T)$) as a function of temperature (in $^\circ$C).

Table 3.1: Descriptions of the state variables and parameters used in the malaria transmission model (3.1).

Variable	**Definition**
$S_h(t)$	Population of susceptible humans.
$E_h(t)$	Population of infected humans undergoing latency.
$I_h(t)$	Population infectious humans.
$R_h(t)$	Population of recovered/removed humans.
$A_m(t)$	Population of aquatic mosquitoes.
$S_m(t)$	Population of susceptible adult female mosquitoes.
$E^s_m(t)$	Population of infected adult female mosquitoes undergoing short-term latency.
$E^l_m(t)$	Population of infected adult female mosquitoes undergoing long-term latency.
$I_m(t)$	Population of infectious mosquitoes.
Parameter	**Definition**
η	Recruitment rate of humans.
ω	Per capita loss of immunity.
μ_h	Per capita death rate of humans.
μ_m	Per capita death rate of adult female mosquitoes.
μ_a	Per capita death rate of aquatic mosquitoes.
α_E	Per capita egg deposition rate by adult female mosquitoes.
p	Transition probability of an exposed mosquitoes undergoing short-term latency.
α_h	Transfer rate of exposed humans to the infected class.
γ_h	Recovery rate of infectious humans without prior immunity to malaria infection.
θ	Transfer rate of infected mosquitoes undergoing latent period.
τ_2	Delay due to malaria parasite incubation period.
a	Probability of malaria transmission from infected mosquitoes to susceptible humans.
b	Probability of malaria transmission from infected humans to susceptible mosquitoes.
c_m	Per capita biting rate of mosquitoes on susceptible humans population.
m	Ratio of mosquito population to human population.
$\varphi_{EA}(T)$	Recruitment rate of adult female mosquitoes.
K_c	Carrying capacity of immature mosquitoes.
τ_1	Delay due to maturation rate of immature mosquitoes.
δ_h	Per capita malaria induced death rate of infectious humans.
λ_a	Infection rate for susceptible human population.
λ_b	Infection rate for susceptible adult female mosquito population.

3.3 Mathematical Analysis

This section presents the theoretical results of the system (3.9) and discuss its analytical properties by invoking biological interpretations. In the sequel, all the temperature-dependent parameters of the system (3.9) are expressed as a function of time *t*. Then, it follows from system (3.9) that the rate of change in the total number of humans, $N_h^{|}(t) = S_h^{|}(t) + E_h^{|}(t) + I_h^{|}(t) + R_h^{|}(t)$ satisfies:

$$\begin{aligned} N_h^{|}(t) &= \eta - \mu_h S_h - \mu_h E_h - \mu_h I_h - \mu_h R_h - \delta_h I_h \\ &= \eta - \mu_h N_h - \delta_h I_h \\ &\leq \eta - \mu_h N_h \end{aligned} \tag{3.13}$$

Similarly, $N_m^{|}(t) = A_m^{|}(t) + S_m^{|}(t) + E_m^{s\,|}(t) + E_m^{l\,|}(t) + I_m^{|}(t)$ is rate of change of the total number of mosquitoes (comprises of immature and adult female mosquitoes) satisfies:

$$\begin{aligned} N_m^{|}(t) &= \alpha_E(t)\left(1 - \frac{A_m(t)}{K_c}\right) N_A(t) - \mu_a(t)A_m(t) - \\ &\quad \mu_m(t)(S_m(t) - E_m^s(t) - E_m^l(t) - I_m(t)) \\ &= \alpha_E(t)\left(1 - \frac{A_m(t)}{K_c}\right) N_A(t) - \mu(t)N_m(t) \end{aligned} \tag{3.14}$$

where $\mu(t) = \min\{\mu_a(t), \mu_m(t)\}$ [131]. As system (3.9) is non-autonomous, it is then paramount to study the malaria transmission dynamics in a periodic environment. For that reason, we assume the vector (mosquito) population stabilizes at a periodic state [98]. Furthermore, by invoking the assumption in [98] for the time periodic function, a positive number h_0 exists, such that we can accordingly obtain equation (3.15):

$$N_m^{|}(t) = \alpha_E(t)\left(1 - \frac{A_m(t)}{K_c}\right) N_A(t) - \mu(t)L < 0 \;\forall\; L \geq h_0 \tag{3.15}$$

Lemma 3.3.1. *The system (3.9) with non-negative initial conditions satisfies* $N_h(t) > 0, \forall\ t \geq 0$. *Hence, the system has a unique non-negative solution in* $C([0], \mathbb{R}^9_+)$, *and thus, all solutions are ultimately-bounded and uniformly-bounded.*

Proof. Following the system in [171], the system (3.9) can further be written in an abstract form as:

$$\dot{X}(T) = B(X)X + H \tag{3.16}$$

with $X = (S_h, E_h, I_h, R_h, A_m, S_m, E^s_m, E^l_m, I_m)$, and

$$B(X) = \begin{bmatrix} -\lambda_a(t)I_m - \mu_h & 0 & 0 & \omega_h & 0 & 0 & 0 & 0 & 0 \\ \lambda_a(t)I_m & -\alpha_h - \mu_h & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha_h & -\mu_h - \delta_h - \gamma_h & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_h & -\mu_h - \omega_h & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -a_{45} & \alpha_E(t) & \alpha_E(t) & \alpha_E(t) & \alpha_E(t) \\ 0 & 0 & 0 & 0 & a_{55} & -\lambda_b(t)I_h - \mu_m(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & p\lambda_b(t)I_h & -\mu_m(t) - \theta_m(t) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{76} & 0 & -\mu_m(t) & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{86} & \theta_m(t) & 0 & -\mu_m(t) \end{bmatrix}$$

where a_{45}, a_{55}, a_{76} and a_{86} are entries of $B(X)$ and $H = (\eta, 0, 0, 0, 0, 0, 0, 0, 0)^T$. Note that $B(X)$ is a Metzler Matrix, i.e., matrix such that the off-diagonal elements are non-negative for all $X \in \mathbb{R}^9_+$. Thus, using the fact that $H \geq 0$, the system (3.9) is positively invariant in $\mathbb{R}^9_+$, which means that any trajectory of the model starting from an initial state in the positive orthant $\mathbb{R}^9_+$ remains forever in $C([0], \mathbb{R}^9_+)$. By applying the comparison principle [92] to equation (3.13) and equation (3.14), it follows that the solution exists $\forall\ t \geq 0$. Then,

$$\lim_{t\to\infty} sup[S_h(t) + E_h(t) + I_h(t) + R_h(h)] \leq \frac{\eta}{\mu_h}$$

and

$$\lim_{t\to\infty} sup[A_m(t) + S_m(t) + E^s_m(t) + E^l_m(t) + I_m(t) - N^*_m(t)] \leq 0$$

where $N_m^*(t)$ is the unique ω-periodic positive solution of equation (3.14) in $C([0], \mathbb{R}_+) \setminus \{0\}$, and thus given by

$$N_m^*(t) = e^{-\int_0^t \mu_m(s)ds} \times \frac{\int_0^t \alpha_E(s)\left(1 - \frac{A_m(s)}{K_c}\right) N_A(s) e^{\int_0^s \mu_m(\omega)d\omega} ds}{1 - e^{-\int_0^\tau \mu_m(s)ds}} + \int_0^t \alpha_E(s)\left(1 - \frac{A_m(s)}{K_c}\right) N_A(s) e^{\int_0^s \mu_m(\omega)d\omega} ds$$

Follows from equation (3.13) and equation (3.14) that $N_h^{\prime}(t) < 0$ and $N_m^{\prime}(t) < 0$, provided that $N_h(t) > \eta/\mu_h$ and $N_m(t) > h_0$ respectively. This implies that all solutions of the system (3.9) are uniformly bounded [98]. □

3.3.1 Malaria Dynamics Threshold

In Figure 3.1, the disease compartments due to malaria infections in both the human and mosquito populations are define. These are classified as either exposed, infectious, or recovered, which includes variables such as E_h, I_h, R_h, E_m^s, E_m^l, I_m. Since the system (3.9) has two disease-free solutions, which are a trivial disease-free equilibrium and a non-trivial disease-free periodic solution. The non-trivial version of the disease-free equilibrium will be analysed, as the trivial disease-free equilibrium is ecologically unrealistic due to the fact that the mosquito population will not be included. Therefore, for convenience, the threshold quantity $Q_n(t)$ can be defined using Figure 3.1 as

$$Q_n(t) = \frac{\alpha_E(t)\varphi_{EA}(t)}{\mu_m(t)(\varphi_{EA}(t) + \mu_a(t))} \tag{3.17}$$

To find the non-trivial disease-free state, the states $E_h = I_h = R_h = E^s{}_m = E^l_m = I_m = 0$ in the system (3.9) which implies a disease-free equilibrium condition as

$$\begin{aligned} \mathsf{E}_o &= (S_h^*,\ E_h^*,\ I_h^*,\ R_h^*,\ A_m^*,\ S_m^*,\ E^{s*}_m,\ E^{l^*}_m,\ I_m^*) \\ &= (\eta/\mu_h,\ 0,\ 0,\ 0,\ A_m^*(t),\ S_m^*(t),\ 0,\ 0,\ 0) \end{aligned} \tag{3.18}$$

where $A_m^*(t)$ and $S_m^*(t)$ are the unique positive periodic solution of the system (3.9) (satisfying for $Q_n(t) > 1\ \forall\ t \geq 0$) and given by

$$\begin{aligned} \dot{A_m}(t) &= \alpha_E(T)\left(1 - \frac{A_m(t)}{K_c}\right) N_A(t) - \mu_a(T)A_m(t) - \varphi_{EA}(T)\phi(\tau_1, \mu_a, \hat{T}) \\ \dot{S}_m(t) &= \varphi_{EA}(T)\phi(\tau_1, \mu_a, \hat{T}) - \lambda_b(T)S_m(t)I_h(t) - \mu_m(T)S_m(t) \end{aligned} \tag{3.19}$$

Hence, the local, asymptotically stability of the disease-free solution of equation (3.18) would be computed using the approach of the next generation operator [165].

3.3.2 Computation of Reproduction Ratio

The computation of basic reproduction ratio of the system (3.9) will be explored according to the theory developed in [98, 171], which is a generalization of the work in [165] for non-autonomous case. By linearizing the system (3.9) at the disease-free periodic state $\mathsf{E}_o = (\eta/\mu_h,\ 0,\ 0,\ 0,\ A_m^*(t),\ S_m^*(t),\ 0,\ 0,\ 0)$, then a reduce system (3.20) is obtained, and given by

$$
\begin{aligned}
\dot{E}_h(t) &= \lambda_a(T)S_h(t)I_m(t) - (\alpha_h + \mu_h)\,E_h(t)\\
\dot{I}_h(t) &= \alpha_h E_h(t) - (\mu_h + \delta_h + \gamma_h)\,I_h(t)\\
\dot{R}_h(t) &= \gamma_h I_h(t) - (\mu_h + \omega_h)\,R_h(t)\\
\dot{E}^s_m(t) &= p\lambda_b(T)S_m(t)I_h(t) - (\mu_m(T) + \theta_m(T))\,E^s_m(t)\\
\dot{E}^l_m(t) &= (1-p)\lambda_b(T)S_m(t)I_h(t) - (1-p)\lambda_b(T)\psi(\tau_2, \mu_m, T) - \mu_m(T)E^l_m(t)\\
\dot{I}_m(t) &= \theta_m(T)E^s_m(t) + (1-p)\lambda_b(T)\psi(\tau_2, \mu_m, T) - \mu_m(T)I_m(t)
\end{aligned}
\tag{3.20}
$$

Hence, the matrices $F(t)$ and $V(t)$ denoting the new infection and the infection transfer respectively are computed below.

$$
F(t) = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & f_{16}\\
0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0\\
0 & f_{42} & 0 & 0 & 0 & 0\\
0 & f_{52} & 0 & 0 & 0 & 0\\
0 & f_{62} & 0 & 0 & 0 & 0
\end{bmatrix}
\quad \text{and} \quad
V(t) = \begin{bmatrix}
v_{11} & 0 & 0 & 0 & 0 & 0\\
-\alpha & v_{22} & 0 & 0 & 0 & 0\\
0 & -\gamma_h & v_{33} & 0 & 0 & 0\\
0 & 0 & 0 & v_{44} & 0 & 0\\
0 & 0 & 0 & 0 & v_{55} & 0\\
0 & 0 & 0 & -\theta_m(t) & 0 & v_{66}
\end{bmatrix}
$$

The matrix $F(t)$ is a next generation matrix of the system (3.20) showing the new infection terms of malaria. Denoting the matrix $F(t)$ by $F(t)|_{6\times6}$, where entries $f_{16} = \eta/\mu_h\lambda_a(t)$, $f_{42} = p\lambda_a(t)S^*_m$, $f_{52} = (1-p)\lambda_b(t)S^*_m$ and $f_{62} = (1-p)\lambda_b(t)S^*_m e^{-\mu_m(t)\tau_2(t)}$. The remaining entries of $F(t)|_{6\times6}$ are all zeroes except those mentioned (f_{16}, f_{42}, f_{52}, f_{62}). Similarly, the matrix $V(t)$ is a Metzler Matrix [42] showing the transition terms of the malaria infection of the system (3.9). The Metzler Matrix is denoted mathematically by $V(t)_{6\times6}$, having diagonal entries diag$(v_{11}, v_{22}, v_{33}, v_{44}, v_{55}, v_{66})$, where $v_{11} = \alpha_h + \mu_h$, $v_{22} = \mu_h + \delta_h + \gamma_h$, $v_{33} = \mu_h + \omega_h$, $v_{44} = \mu_m + \theta_m$ and $v_{55} = v_{66} = \mu_m$. Also,

the $V(t)_{6\times6}$ has left off-diagonal entries, off-diag(v_{21}, v_{32}, v_{64}), where $v_{21} = -\alpha_h$, $v_{32} = -\gamma_h$ and $v_{64} = -\theta_m$. The remaining entries of the $V(t)_{6\times6}$ are all zeroes, except those in diagonal and left off-diagonal as mentioned.

The linearized system of the disease-free periodic state solution in equation (3.18) can be written in abstract form, given by

$$y^{\cdot}(t) = (F(t) - V(t))x(t) \tag{3.21}$$

where $x(t) = (E_h(t), I_h(t), R_h(t), \underset{m}{E^s}(t), \underset{m}{E^l}(t), I_m(t))^T$.

Let $Y(t, s)$, $t \geq s$, be the evolution operator of the linear ω-periodic system

$$y^{\cdot}(t) = -V(t)y \tag{3.22}$$

That is, for each $s \in \mathbb{R}$, the 6×6 matrix $Y(t, s)$ satisfies [98, 171]

$$\dot{Y}(t, s) = -V(t)Y(t, s), \ \forall\ t \geq s, \ Y(s, s) = I \tag{3.23}$$

where I is an identity matrix of order 6×6. Let C_T be the ordered Banach space of all ω-periodic functions from $\mathbb{R}$ to $\mathbb{R}^6$, which is equipped with the maximum norm and positive cone $C_\omega^+\{\varphi \in C_\omega : \varphi(t) \geq 0, \forall t \in C\}$ [98, 171]. Suppose that $\varphi(s) \in C_T$ (ω-periodic in s) is the initial distribution of infectious individuals in this periodic environment. Thus, $F(s)\varphi(s)$ is the rate of new infections produced by the infected individuals who were introduced into population at time s [98, 171]. At time $t \geq s$, then it follows that $Y(t, s)F(s)\varphi(s)$ denotes the distribution of those individuals who were newly infected at time s and remain in the infected compartments at time t for

$t \geq s$.

$$\psi(t) = \int_{-\infty}^{t} Y(t,s)F(s)\varphi(s)ds = \int_{0}^{\infty} Y(t,t-a)F(t-a)\varphi(t-a)da \qquad (3.24)$$

Therefore, equation (3.24) describes the cumulative distribution of new cases of infections at time t produced by all those infected individuals $\varphi(s)$ and introduced at a prior time. Let us consider a linear operator $L : C_T \rightarrow C_T$ [98, 171], such that

$$(L\varphi)(t) = \int_{0}^{\infty} Y(t,t-a)F(t-a)\varphi(t-a)da \;\forall\; t \in \mathbb{R},\ \varphi \in C_T. \qquad (3.25)$$

Hence, the reproduction ratio R_0*, is then defined by the spectral radius of L, thus represented by $\rho(L)$. This is given as $R_0* := \rho(L)$; however it can be verified for the system (3.9). Then, the result below follows from Theorem 2.2 in [171]. Let $W(t,\lambda)$ be the monodromy matrix of the following ω-periodic system with parameter $\lambda \in (0, \infty)$.

$$\dot{W}(t) = \left(-V(t) + \frac{1}{\lambda}F(t)\right) W(t), t \in \mathbb{R} \qquad (3.26)$$

Since $F(t)$ is non-negative and $-V(t)$ is cooperative, it follows that $\rho(W(\omega, \lambda))$ is continuous and non-increasing for $\lambda \in (0, \infty)$, and $\lim_{\lambda\rightarrow\infty}\rho(W(\omega, \lambda)) < 1$ [171].

Theorem 1. *The disease-free-state (DFS) of the non-autonomous system (3.9), given by equation (3.19), is locally asymptotically stable if R_0*, is less than unity. The disease-free-state is unstable if R_0* exceeds unity.*

Remark: The epidemiological quantity R_0* measures the average number of secondary infections generated by a typical infected individuals introduced into completely susceptible population. The following result can be used to numerically compute the reproduction ratio R_0* of the equation (3.9) (R_0 is used instead of R_0* in

our notation). Following from Theorem 1, the spread of malaria can be effectively controlled in a population when $R_0* < 1$.

The basic reproduction number associated with the autonomous version of the non-autonomous system (3.9), denoted by $R_0 = \rho(FV^{-1})$, where ρ = spectral radius of the next generation matrix. Its computation follows first by setting $\lambda_a(T) = \lambda_a$, $\lambda_b(T) = \lambda_b$, $\mu_a(T) = \mu_a$, $\alpha_E(T) = \alpha_E$, $\varphi_{EA}(T) = \varphi_{EA}$, $\tau_1(T) = \tau_1$, $\tau_2(T) = \tau_2$, $\mu_m(T) = \mu_m$ and $\theta_m(T) = \theta_m$ in the system (3.9). Then, the associated R_0 of the autonomous version of the system (3.9) is defined below

$$R_0 = \frac{\lambda_a \eta \left(\lambda_b (1-p) e^{-\mu_m \tau_2} + \lambda_b p \frac{\theta_m}{\theta_m + \mu_m} \right)}{\mu_h \mu_m (\alpha_h + \mu_h)(\mu_h + \delta_h + \gamma_h)} \tag{3.27}$$

3.4 Results Presentation and Discussion

This section presents numerical simulations of the system 3.9 to illustrate the impact of temperature-dependent maturation and extrinsic incubation delay on the dynamics of human malaria. All the simulations were run using a package in *R-software* called *deSolve*. Subsequently, a sensitivity analysis was performed to determine the most dominant parameters influencing the spread of malaria against the proposed model. The set of parameters and values used for running the simulations can be found in Table 4.3 and their references therein. Furthermore, the ranges and baseline values of the temperature-dependent parameters of the system (3.9) were computed using their respective functional forms presented in subsection 3.2.3.

3.4.1 Numerical simulations

The spread of malaria in a human population is largely leveraged by the density of immature mosquitoes that emerge as adults. In this, the maturation delay is the antecedent that drives the abundance of adult mosquitoes. Over and above, the maturation delay depends on climatic factors, and particularly the ambient and water temperature.

By using the temperature-dependent maturation delay function in [112] given by $\tau_1 = (-0.00094T^2 + 0.049T - 0.522)^{-1}$, and computed the total maturation delays as precedents between 16°C to 34°C. The delay values obtained were then used to assess their impact on the system (3.9) and observed the pattern presented (see Figures 3.6–3.18). The plots shown in Figure 3.6 and 3.7 depicts the dynamics of immature mosquitoes with a maturation delay. In Figure 3.6, a simulation was performed for τ_1 = 107 days (at 17°C equivalent) while the remaining parameters (see Table 4.3) were held constant, except α_E = 5 eggs. A damping oscillation pattern to the point of stability was observed as the time increased. Similarly, another simulation was ran as shown in Figure 3.7, but for τ_1 = 12 days (at 27°C equivalent), and with α_E = 5 eggs. A pattern similar to Figure 3.6 was observed, but they were different in amplitude and stability as the time increased. Furthermore, a simulation was ran repeatedly using τ_1 values other than those considered in Figures 3.6 and 3.7.

In summary, it was observed that, at temperatures around 23°C to 29°C, the survival and development rate of immature mosquitoes was optimum (indicating that immature mosquitoes would emerge as adult at an average of 12 days). This result gives an insight into the temperature-dependent maturation delays with a greater influence on the abundance of adult mosquitoes. Moreover, these results further reaffirmed the impact of climate factors, and notably temperature by showing compelling evidence for

their influence on the spread of malaria and other related vector-borne diseases, such as dengue, West Nile Virus (WNV) among others. However, as part of the preventive mechanisms to control immature mosquitoes, the use of larvicide, sanitation and sewage control will strengthen the efforts to curb breeding sites.

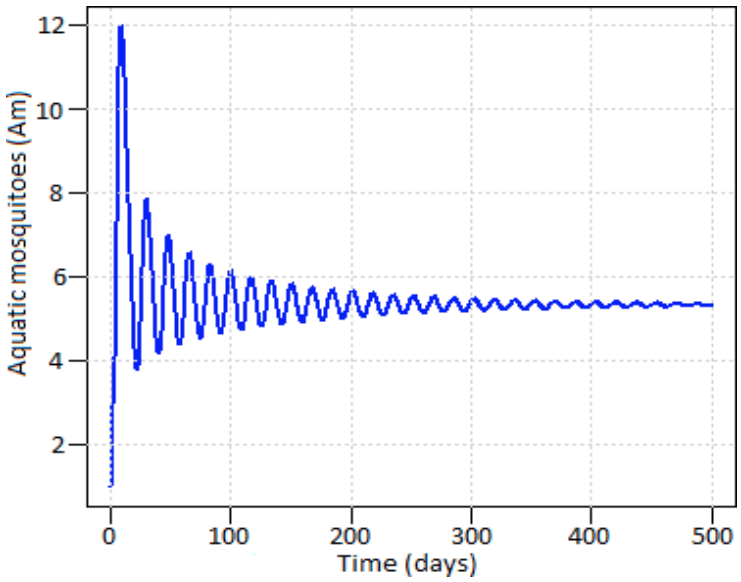

Figure 3.6: The effect of maturation delay on the dynamics of immature mosquitoes, $A_m(t)$ when $\tau_1 = 107$days, $\alpha_E = 5$ eggs, $K_c = 4\times 10^4$, while all other parameters (see Table 4.3 of the model 3.1 were held constant during the simulation.

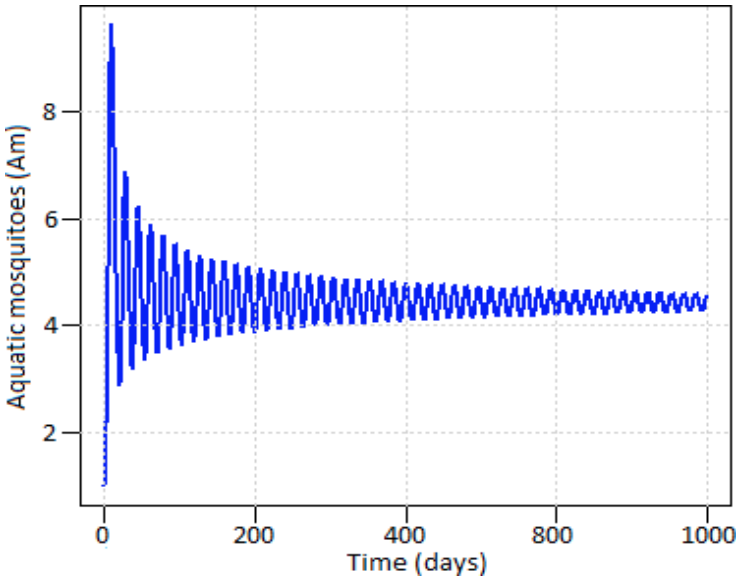

Figure 3.7: The effect of maturation delay on the dynamics of immature mosquitoes, $A_m(t)$ when $\tau_1 = 12$days, $\alpha_E = 5$ eggs, $K_c = 4\times10^4$, while all other parameters (see Table 4.3 of the model 3.1 were held constant during the simulation.

The plots shown in Figures 3.8 and 3.9 have metamorphosed from Figures 3.6 and 3.7, illustrating the dynamics of the recruited susceptible adult mosquito population. This shows that a large number of immature mosquitoes emerging as adults increases the likelihood of an abundance of mosquitoes susceptible to becoming transmitters of malaria. Hence, this depends on temperature dependent maturation delay, which is the risk factor leveraging malaria's transmission. Subsequently, the abundance of adult mosquitoes has a direct consequence on the rate of the malaria's transmission in a population. The pattern shown in Figure 3.8 was produced by using the values for Figure 3.6. Thus, as the maturation delay is, $\tau_1 = 107$ days (at 17°C equivalent), the population of susceptible mosquitoes stabilizes faster. This indicates that at lower temperatures, say less than 17°C, the maturation of immature mosquitoes takes longer, and thus many would eventually die. Similarly, another simulation was performed using the same pa-

rameter values for Figure 3.7 and this produced Figure 3.9. We fixed τ_1 = 12 days (at 27°C, equivalent), and observed the pattern of the susceptible mosquitoes. Hence, the pattern shown illustrates that, at 27°C, the immature mosquito dynamics were suitably supported. As a result, malaria would be likely to be transmitted due to the abundance of susceptible mosquitoes described by the long term damping oscillation.

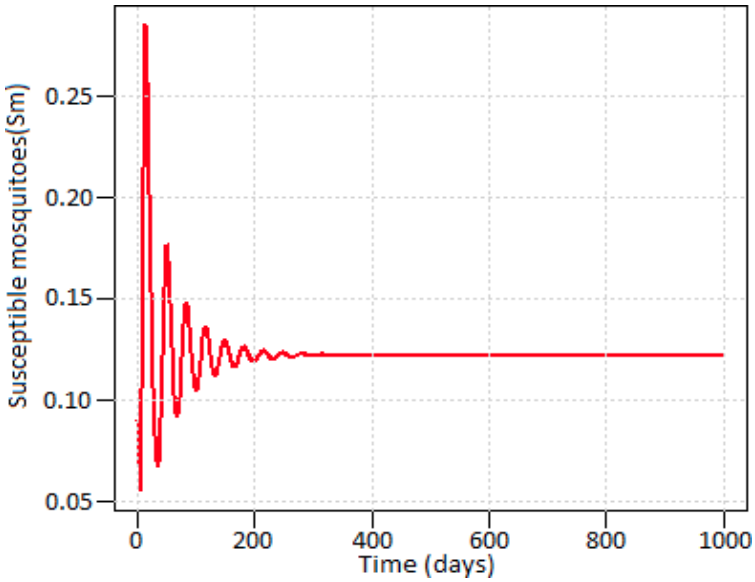

Figure 3.8: The dynamics of the susceptible mosquito pattern showing short term oscillation in the population, when τ_1 = 107 days at approximately temperature of 16°C.

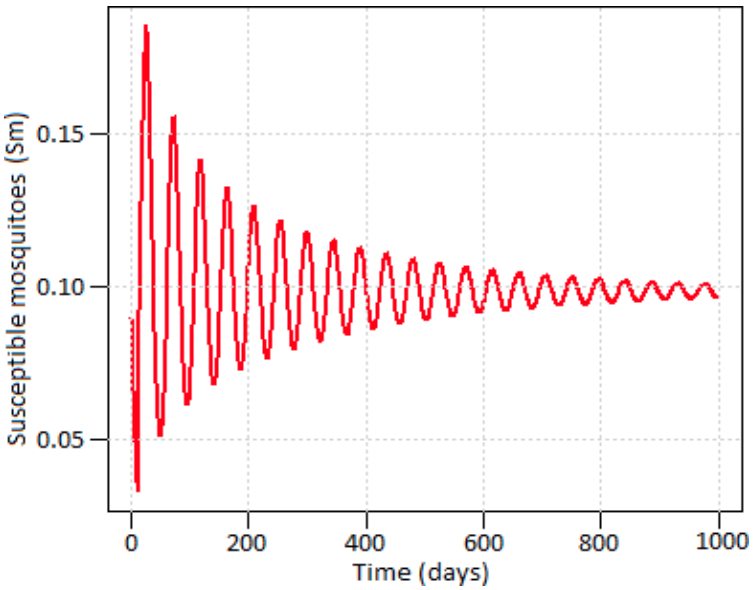

Figure 3.9: The dynamics of the susceptible mosquito pattern showing long term oscillation in the population, when τ_1 = 12 days at approximately temperature of 27°C.

In Figures 3.10 and 3.11, it was further performed a simulation and presented the pattern of infectious humans and mosquitoes. The impact of a temperature-dependent extrinsic incubation period on infectious human dynamics was examined, as shown in Figure 3.10. The parameter τ_2 = 10, 11, 12, 13and14 while the other parameters (see Table 4.3) were fixed during the simulation. Thus, it was observed that as, τ_2 increases, the number of infectious humans decrease and vice-versa. In addition, the result further substantiates that, at temperatures between 20°C to 28°C, the pathogens causing malaria hasten the extrinsic incubation period. Thereby, the spreading of the malaria infection would be effective in a population experiencing this temperatures regime. On the other hand, Figures 3.11 and 3.12 depict the pattern of infectious mosquito dynamics over time. This result shows the sensitivity of temperature on the abundance of infectious mosquitoes. The increase in value τ_2 decreases the abundance of infectious mosquitoes to the point of stability, as shown in Figure 3.11. However, decreasing

τ_2 would enable a large number of mosquitoes to become potential candidates for the spread of malaria (see Figure 3.12).

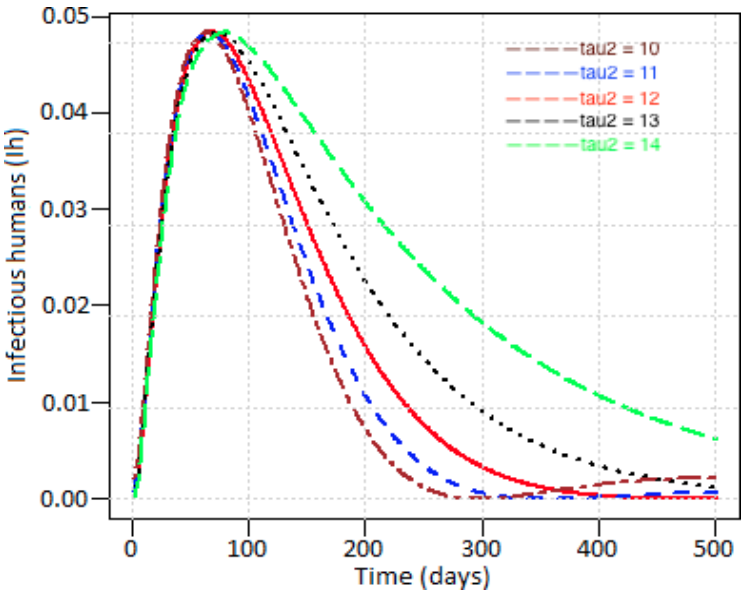

Figure 3.10: The pattern of infectious humans, for τ_2 = 10, 11, 12, 13and14 are indicated using the colours green, black, red, blue and brown respectively.

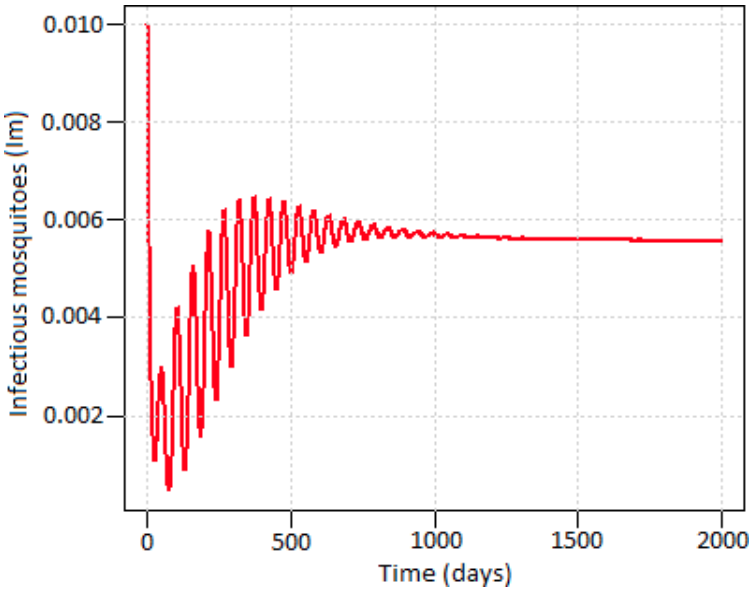

Figure 3.11: The pattern of infectious mosquitoes over time t when τ_2 increases.

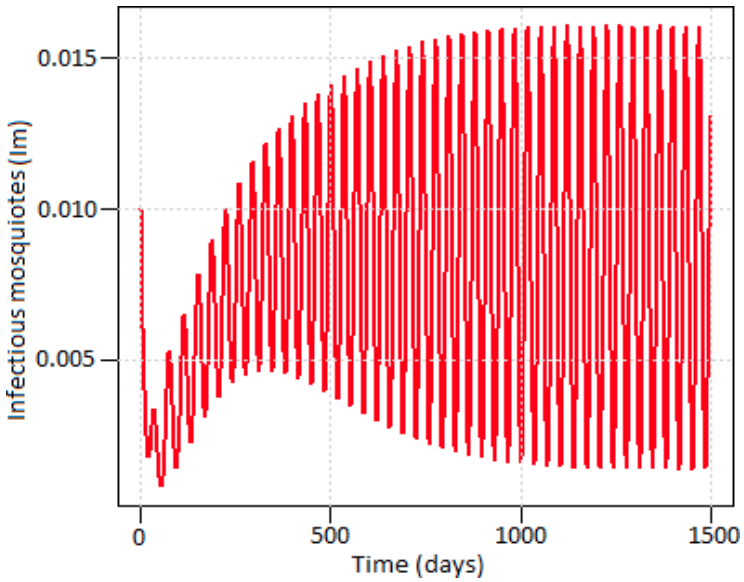

Figure 3.12: The pattern of infectious mosquitoes over time t when τ_2 decreases.

Figures 3.13 and 3.14 presents the simulation results of exposed mosquitoes undergoing a short and long term extrinsic incubation period, respectively. The pattern depicted in Figure 3.13 describes that, at ambient temperatures between 23°C to 28°C, the malaria parasites develop faster inside their host mosquito. However, at this range of temperatures $16°C \leq T < 23°C$, the parasites develop at a slow phase due to a lower temperature regime. While, in this scenario, $28°C \leq T < 34°C$ the parasites also develop slower as the temperatures is more extreme and lethal to their survival.

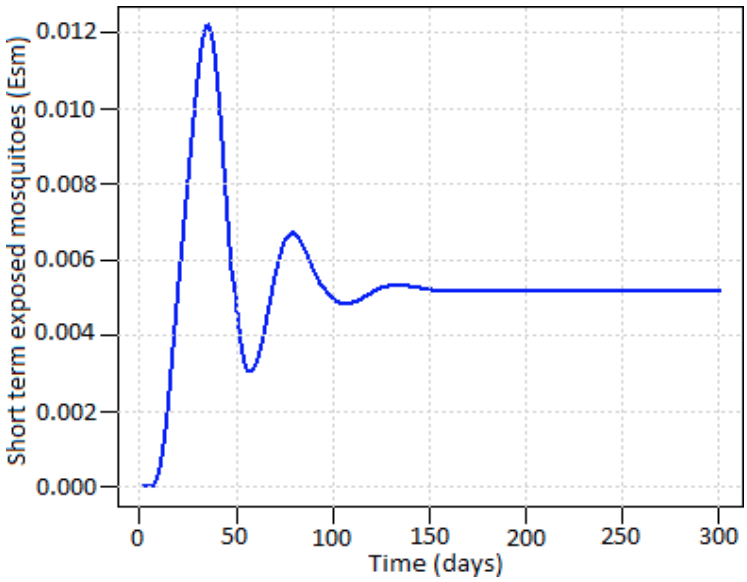

Figure 3.13: The pattern of exposed mosquitoes undergoing a short-term extrinsic incubation period; this scenario has been simulated for temperatures between 23°C to 28°C.

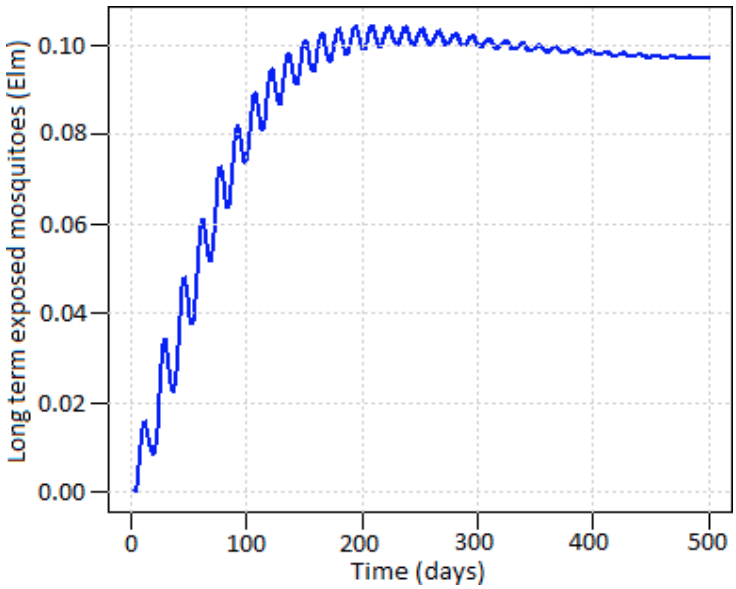

Figure 3.14: The pattern of exposed mosquitoes undergoing a long-term extrinsic incubation period. We performed the simulation using temperatures greater than 28˚C but less than or equal to 34˚C, and greater than or equal to 16˚C but less than 23˚C.

Figures 3.15 and 3.16 illustrate the pattern of susceptible human dynamics and exposed human dynamics. The pattern in Figure 3.15 reveals that an increase in τ_2 increases the population of susceptible humans; however, a decrease in τ_2 increases the population of infectious mosquitoes that are able to infect more susceptible humans. Thus, this decreases the population of susceptible humans. As the human population are infected with malaria by infectious mosquitoes, humans have different chances to contact the infection. Hence, the variability in susceptible human populations against the dynamics of malaria transmission is presented in Figure 3.16.

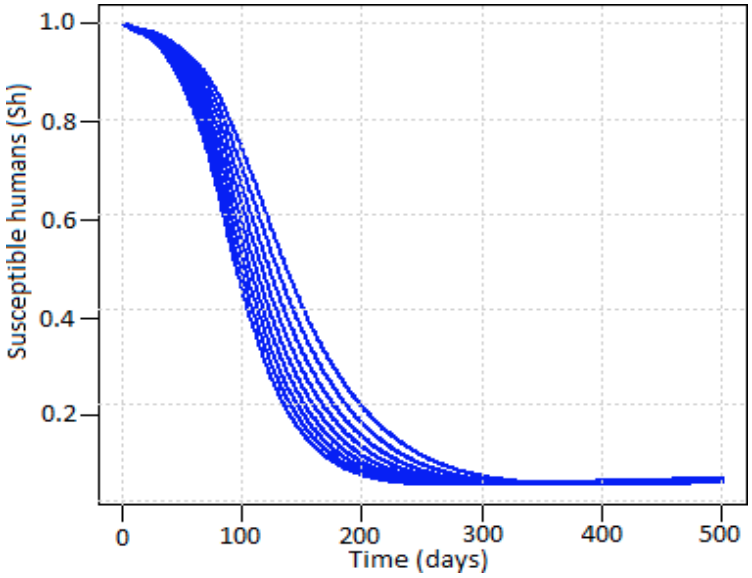

Figure 3.15: The pattern of susceptible humans dynamics.

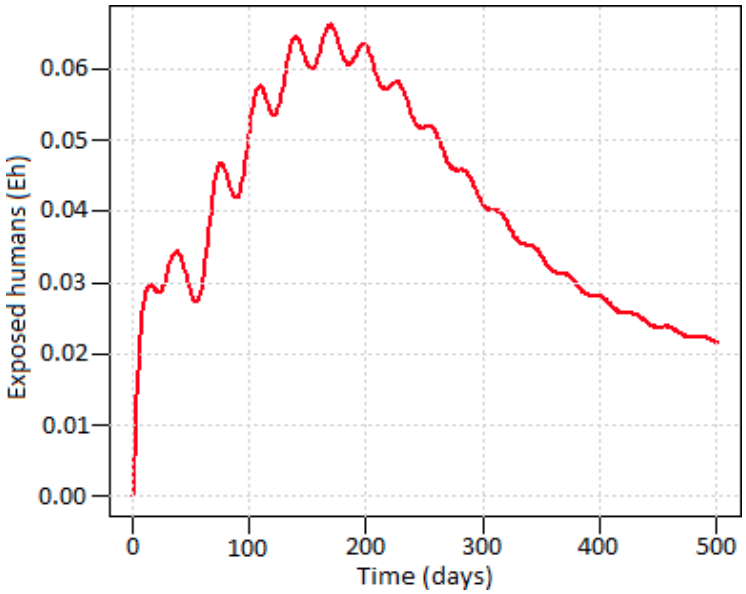

Figure 3.16: The pattern of exposed humans dynamics.

The Figures 3.17 and 3.18 presents the human recovery pattern increase as the increase in τ_2.

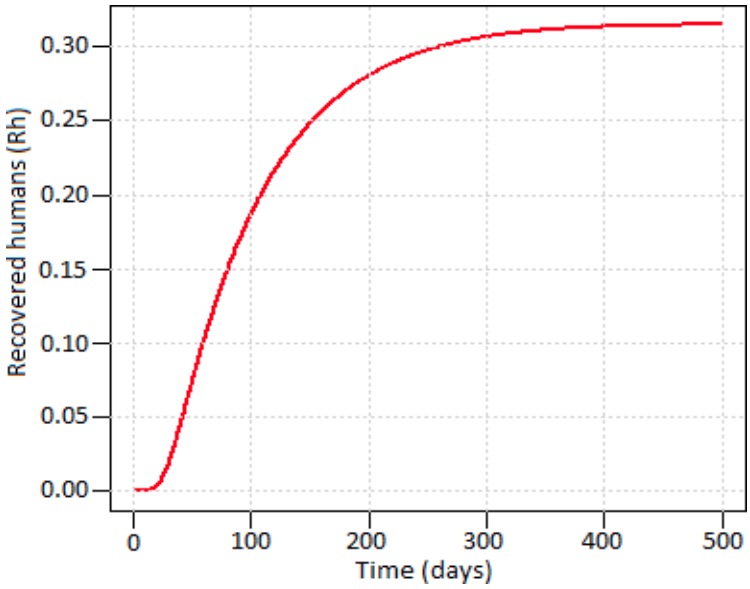

Figure 3.17: The human recovery pattern from malaria infection.

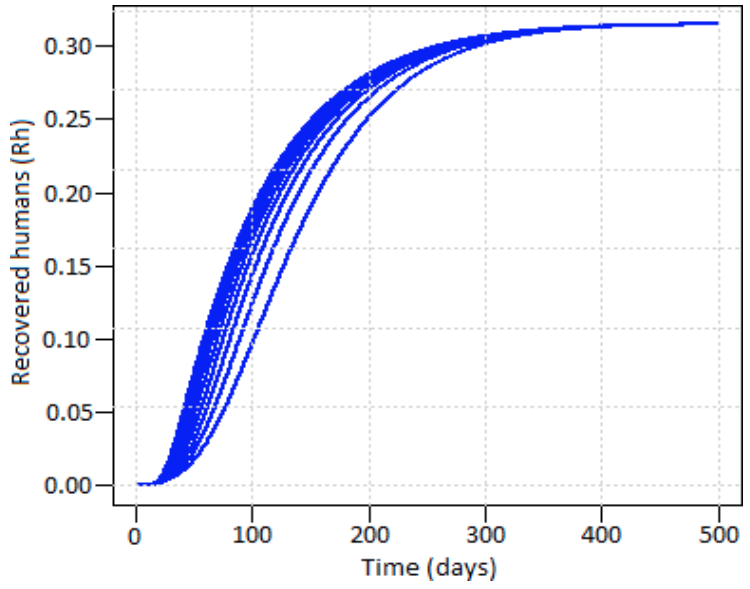

Figure 3.18: The human recovery pattern from malaria infection increases by small shifts.

3.5 Sensitivity Analysis

The central idea for infectious disease modelling is to investigate and understand it transmission potential by determining the necessary factors responsible. This is to reduce human infectiousness by strategising effective prevention and control mechanisms. This chapter analyses a model for malaria transmission (see Figure 3.1), involving 17 parameters, and each of these parameters have a causative influence on human infectiousness to varying degrees. The aim is to search and identify the most influential parameters of the system (3.9), and accordingly, a sensitivity analysis (SA) will be performed against the model output. Then, the output variable of interest that will be used to perform the SA in the system (3.9) is the infected human dynamics, $I_h(t)$. The SA methods are presented in the following subsections:

3.5.1 Latin Hypercube Sampling Partial-Rank Correlation Coefficient

The two techniques called Latin Hypercube Sampling and Partial Rank Correlation Coefficient (LHS-PRCC) [188] would be both deployed to perform the SA. These are two independent methodologies and are used for different purposes, but are often combined when applying to SA [188].

LHS is a stratified Monte Carlo sampling scheme [180, 188], in which each parameter range is divided into N distinct equal intervals and this randomly draws one sample from each interval. In this approach, the entire range of each parameter and each interval for each parameter would be randomly sampled once. The LHS is more efficient than the Monte Carlo sampling approach [78], due to stratified sampling without replacement. This enables a rapid convergence as the number of samples increase

over the parameter space.

PRCC is an efficient and reliable technique that uses LHS. This provides a measure of monotonicity between the parameters and output of interest after excluding the linear effects of all parameters excepting the parameter of interest [180]. PRCC is a robust sensitivity measure that assists in the identification of most influential parameters and optimizes the model structure. The population parameter $\rho_{x_j y}$, denotes a *correlation coefficient* between two variables, x_j and y and can defined as:

$$\rho_{x_j y} = \frac{Cov(x_j, y)}{Var(x_j)Var(y)}, \text{ and } j = 1, 2, \cdots, m \tag{3.28}$$

where $\rho_{x_j y} = \pm 1$ is a constraint on the *correlation coefficient*, while $Cov(x_j, y)$, $Var(x_j)$ and $Var(y)$ are covariance of x_j and y, variance of x_j and variance of y respectively. The *correlation coefficient* gives the measure of strength of a linear association between input x_j and output y. Hence, the partial *correlation coefficient* between x_j and y is the *correlation coefficient* between the two residuals $(x_j - \hat{x}_j)$ and $(y - \hat{y})$, where $\hat{x}_j$ and $\hat{y}$ are defined using the following linear regression models:

$$\hat{x}_j = c_0 + \sum_{\substack{p=1 \\ p \neq j}}^{k} c_p x_p \text{ and } \hat{y} = b_0 + \sum_{\substack{p=1 \\ p/=j}}^{k} b_p x_p \tag{3.29}$$

The PRCC values should lie between -1 and +1, such that the values of PRCC are sometimes called the sensitivity index. When a parameter has an absolute value of its PRCC close to one, then it indicates the parameter is sensitive or has a strong impact on the model. In general, the computation of LHS-PRCC involves seven steps that are well described in [104].

3.5.2 Processing of Parameter Space

In the system (3.9), there are 17 parameters, hence by doing the sensitivity analysis the robustness of the model output ($I_h(t)$) using the parameter space will be explored. All the parameter values and their ranges can be found in Table 3.2 (most of the values used were adopted from the literature and their references therein). First, as a baseline, the parameters τ_1 and τ_2 were chosen as control parameters to investigate their effects on the dynamics of human infectiousness $I_h(t)$ at time t. Then, the LHS will be apply to explore a certain-dimensional parameter space. The package *lhs* in *R* software was used to simulate 10000 sample points in the 17-dimensional unit cube. Each point in the unit cube sampling of the 17-dimensional parameter space was mapped using the minimum and maximum values for each. By doing so, a parameter set would be generated by rescaling the simulated LHS using the factor:

$$\upsilon_k = lhs[\,,.] * (\upsilon_k max - \upsilon_k min) + \upsilon_k min,\ k = 1, 2, \cdots, 17 \tag{3.30}$$

where $lhs[\,,.]$ is the index of the *lhs* corresponding to the parameter column and where υ_k, $\upsilon_k min$ and $\upsilon_k max$ are the parameter default, minimum and maximum values, respectively.

3.5.3 Sensitivity Analysis Results

The Table 3.2 presents the results of the sensitivity analysis for the system (3.9) using LHS-PRCC. It follows that none of the associated confidence intervals of the PRCC values contain zero, hence all estimates of the model parameters are robust. In order to depict the PRCC values pattern presented in Table 3.2, and visualised in Figure 3.19.

From the results presented in Table 3.2, it is observed that the most influential parameters of the system (3.9) are $\alpha_E|PRCC| = 0.80664$, $K_c|PRCC| = 0.57280$, $b|PRCC| = 0.93165$, $c_m|PRCC| = 0.80127$ and $\tau_2|PRCC| = 0.98257$. All parameters with values crossing the upper bound (greater than 0.5, as indicated in Figure 3.19), are strong and positively correlated with human infectiousness. However, the parameters $\mu_a|PRCC| = |-0.79813|$ and $\tau_1|PRCC| = |-0.63662|$, with values crossing the lower bound (less than -0.5, as indicated in Figure 3.19), have a strong negative influence on human infectiousness. Furthermore, parameters, such as $\gamma_h|PRCC| = |-0.49634|$, $\varphi_{EA}(T)|PRCC| = 0.36423$, $\mu_m|PRCC| = -0.47952$, $\theta_m|PRCC| = 0.43543$ and $a|PRCC| = 0.33642$ have a moderate correlation (since all their values lie between -0.5 to 0.5, as shown in Figure 3.19). Hence, the remaining parameters include: $\eta|PRCC| = 0.02489$, $\mu_h|PRCC| = 0.01056$, $\omega_h|PRCC| = 0.02424$ and $\alpha_h|PRCC| = 0.18587$ and are insensitive to causing variation in human infectiousness.

An effective strategy to reduce the malaria infections would be suggested based on the associated PRCC values of the system (3.9) parameters. A high PRCC value for the egg deposition rate, $\alpha_E|PRCC| = 0.80664$ shows the possibility of a large number of adult female mosquitoes. This owes to the abundance of nutrients and the availability of breeding sites to enable the mosquitoes' survival and reproduction, as indicated by the PRCC value parameter, $K_c|PRCC| = 0.57280$. Moreover, the absolute value of the total maturation delay of aquatic mosquitoes, $\tau_1|PRCC| = 0.63662$ also shows a high association with the density of adult mosquitoes. Consequently, the abundance of adult female mosquitoes will trigger the biting of available humans, as indicated by $c_m|PRCC| = 0.80127$. The high and relatively suitable temperature, supports The survival of mosquitoes, and the biting rate of humans will eventually increase. Hence,

once mosquitoes bite an infectious human and become infected, the temperature also supports the development of malaria parasites, which has been shown by the influence of $\tau_2|PRCC| = 0.98257$. Subsequently, the transmission of malaria continues as mosquitoes are infectious and capable of transmitting the infection, as indicated by the transmission probability, $b|PRCC| = 0.93165$. From the results of the model's sensitivity analysis, hence suggest the strategy to prevent the spread of malaria due to the control of the following parameters c_m, α_E, K_c, μ_a, and μ_m.

Table 3.2: Partial rank correlation coefficients of the system (3.9) parameters.

Label	Parameter	PRCC	Bias	Standard Error	95% Confidence Interval
a	η	0.02489	0.00054	0.00751	0.00982, 0.03888
b	μ_h	0.01056	0.00032	0.00720	0.00263, 0.02530
c	ω_h	0.02424	0.00033	0.00649	0.01116, 0.03609
d	α_h	0.18587	0.00021	0.00741	0.17098, 0.20026
e	δ_h	-0.04278	0.00022	0.00717	-0.05738, -0.02981
f	γ_h	-0.49634	0.00060	0.00161	-0.62310, -0.39943
g	α_E	**0.80664**	0.00017	0.00330	0.80027, 0.81304
h	K_c	**0.57280**	0.00024	0.00707	0.41687, 0.61038
i	μ_a	**-0.79813**	0.00054	0.00335	-0.80365, -0.79066
j	φ_{EA}	0.36423	0.00029	0.00761	0.34947, 0.37936
k	μ_m	-0.47952	0.00011	0.00571	-0.49112, -0.46860
l	θ_m	0.43543	0.00054	0.00703	0.42202, 0.54890
m	a	0.33642	0.00005	0.00746	0.12220, 0.45218
n	b	**0.93165**	0.00003	0.00100	0.92971, 0.93368
o	c_m	**0.80127**	0.00003	0.00157	0.79780, 0.90450
p	τ_1	**-0.63662**	0.00004	0.00805	-0.85264, -0.52021
q	τ_2	**0.98257**	0.00003	0.00035	0.98185, 0.98320

The PRCC values in bold indicates the sensitivity of the malaria model's parameters.

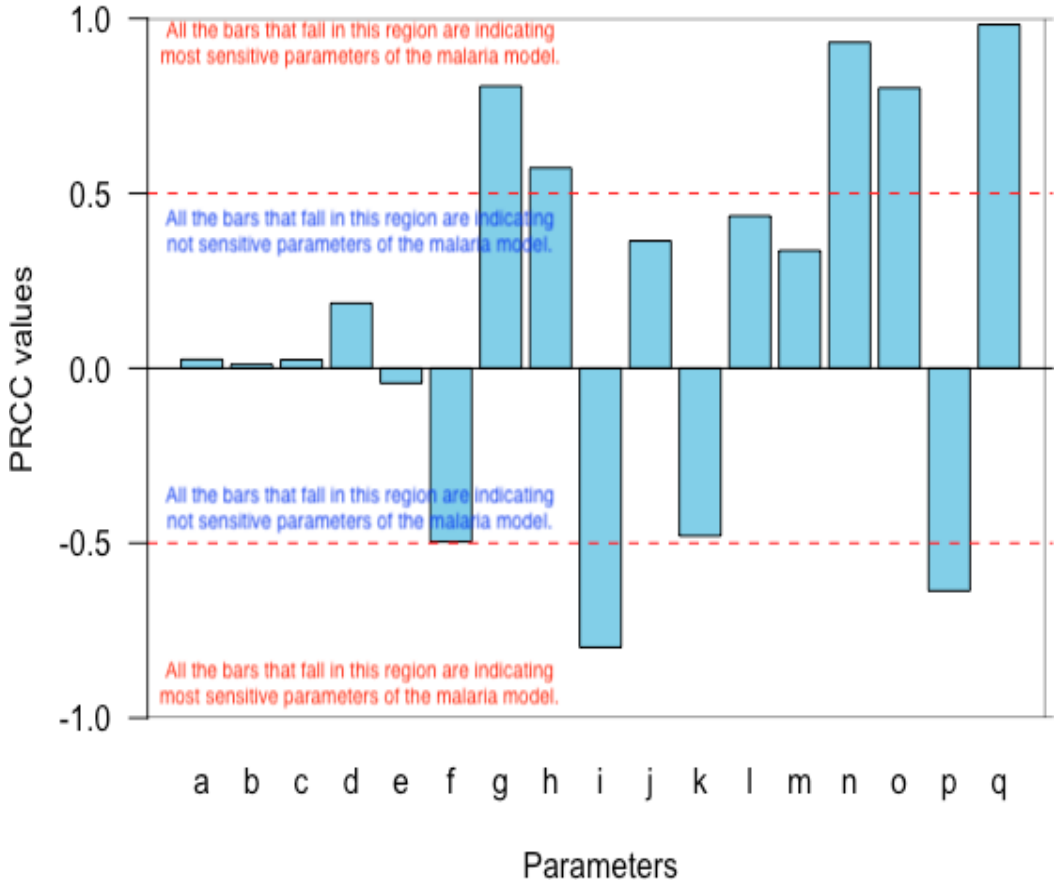

Figure 3.19: Sensitivity of the parameters for the system (3.9) obtained using the LHS-PRCC sensitivity analysis. The Partial Rank Correlation Coefficients show the influence of each parameter on the dynamics of the infectious human population, $I_h(t)$.

3.6 Summary

In this chapter, a mathematical model for malaria's transmission was developed, and the impact of temperature on delay due to the extrinsic incubation and maturation in mosquito dynamics was analysed. The model was analysed both theoretically and numerically to investigate human infectiousness. Also, as part of the preventive mechanisms, a sensitivity analysis was performed on the model and the most dominant control parameters were identified. However, the mathematical model has a limitation

due to its inability to effectively investigate heterogeneity arising in human characteristics in a population. Hence in Chapter 4, this limitation will be address by deploying Agent-based modelling.

Chapter 4

Agent-Based Modelling of Malaria Transmission Dynamics

In this chapter, an agent-based modelling approach was developed to investigate malaria transmission dynamics and its appropriateness for future use was examined.

4.1 Introduction

In an effort to end the threat of malaria several studies (e.g., [21, 67, 116, 123, 129, 131, 132, 148]) have been carried out using (mathematical) compartmental modelling techniques to model and understand the disease transmission dynamics, which helped researchers to investigate the progression of disease from a population perspective by characterising individuals according to their health status (e.g., susceptible, infected and recovered). Compartmental models are more often used compared to other types of models, such as stochastic models, complex network models, statistical process control models, spatial models and machine learning-based models [150]. This is mainly

because of their effectiveness in tracking disease progression and thereby transforming compartments into differential equations. Although compartmental models are powerful tools to investigate the disease dynamics, they impose a severe constraint, i.e. assuming homogeneity across the human population [79]. This means that each and every individual in the population has constant rates of infection, recovery and immunity loss.

As a matter of fact, the human population is heterogeneous, and influenced by a broad spectrum of factors, including age, sex, spatial and temporal changes, human movement patterns, and social network patterns [90] amongst others. This limitation makes compartmental models inadequate for capturing the heterogeneities arising from the population dynamics of malaria. To address the limitations, we propose an agent-based modelling approach that alleviates the limitations imposed by compartmental models and permits modelling and analysing malaria dynamics for heterogeneous populations.

Agent-based models (ABMs) are computational modelling tools consisting of agents, which communicate to each other within their environment and behave according to pre-defined rules. ABMs are powerful due to their stochasticity, spatial explicitness, and discrete-time-based simulation where each agent interacts in space and time [76]. ABMs often work as a bottom-up modelling approach, as population-based behaviour emerges from interactions amongst autonomous agents [178]. This characteristic means that ABMs are more flexible because of their ability to consolidate heterogeneous variables (e.g., host movement, heterogeneous implementation interventions) and stochastic (e.g., inter-patient variability at the time of infection, time to recovery, and the location of infection) [153]. Furthermore, ABMs allow a high degree of heterogeneity in the creation, disappearance and movement of a finite collection of dis-

crete interacting individuals [178]. The stochasticity of ABMs permits variation due to randomness, and thus more accurately mimics the transmission of malaria; it thereby reduces the effects of systematic preference amongst the agents [153].

The agent-based malaria transmission models developed in this study enables us to investigate not only individual agents behaviour, but also how they communicate to each other according to predefined rules and their responses to climate factors. The agent-based models will be utilised to simulate the actual malaria cases in Tripura, Limpopo and Benin using the climate and demographic data obtained for these cities. Hence, the emerging results will be used to validate against the actual reported cases these cities. We also perform some statistical tests, such as *t-test* for 2-independent samples and *correlation* analysis, to evaluate the accuracy of the proposed model.

4.2 Model Development

This section provides a theoretical background for compartmental modelling, agent-based modelling and their application to the study of malaria. We first present a malaria transmission model, as shown in Figure 3.1, and describe its dynamics in human and mosquito populations; we also discuss the resultant complexity from the impact of temperature.

Malaria spreads into the human population through bites by the *anopheles* mosquito (female type), seeking human blood for nourishment and egg production. Figure 3.1 presents the compartments describing the human and mosquito dynamics, in which the human dynamics are structured using SEIR (susceptible, exposed, infected and recovered) attributes, and mosquito dynamics are structured using ASEI (aquatic, susceptible, exposed and infected) attributes. Temperature is incorporated into the model to

account for its influence on the biting rate, survival rate, parasite development, juvenile maturation rate and mortality rate.

4.2.1 Mathematical Model

A compartmental model is a mathematical modelling technique that has long been used to investigate epidemics and public health policies. The 'classical model' [103] was the first mathematical technique that appeared in the literature that studied malaria transmission. Since then, several remarkable extensions have been made (e.g.,[21, 27, 28, 60, 67, 100, 116, 123, 129, 131, 132, 148]) that build upon this model, addressing various emerging problems. In the mathematical modelling of malaria, a compartmental model (see Figure 3.1) is used to describe the disease status transition using a set of differential equations [80]. The population within a particular compartment in a disease transmission model is assumed to be homogeneous, well-mixed and split (for instance, the SIR model) into compartments based on health status, for example., susceptible, infectious and recovered. Hence, each compartment of the disease transmission model is defined by its own differential equations [62]. As shown in Figure 3.1, the human population is split into four compartments depicted by the rectangular boxes labelled S_h, E_h, I_h, R_h indicating susceptible, exposed, infectious and recovered humans, respectively. Similarly, the mosquito population is split into five compartments that include its juvenile stages, labelled A_m, S_m, E^s_m, E^l_m, I_m indicating aquatic, susceptible, short-term exposed, long-term exposed and infectious, respectively. The differential equation governing the compartmental model shown in Figure 3.1 is given in system 3.9. Table 3.1 detailed the definition of the parameters in Figure 3.1 while their values including their references (see Table 4.3). This model can be used to better understand

the dynamics in a malaria transmission. However, mathematical models have proven suitable for modelling systems at a macro-scale rather than at fine granular levels.

4.2.2 Agent-Based Modelling

In contrast to a mathematical model, an agent-based model is generally characterised by a bottom-up modelling approach. The agent-based approach enables individual agents to interact within their environment and behave according to predefined rules [173]. In addition, individual entities are represented by discrete autonomous agents communicating among themselves in a space to produce non-intuitive emergent patterns at the population level [119]. Moreover, agent-based modelling represents purely rule-based algorithms, which start from scratch and continue until the desired model represents the real-world phenomena of interest. In general, agent-based modelling is characterised by its ability to capture heterogeneity, spatial and complex interactions, a micro-scale perspective, discrete time considerations and non-intuitiveness. All agents involved in the agent-based modelling of the malaria transmission dynamics in Figure 3.1 (including their detailed descriptions) are presented in the following subsections.

4.2.2.1 Human Agents

An infected human agent can transmit the malaria infection to susceptible mosquitoes provided that the incubation period of the malaria parasite in the human is complete. This period is called *intrinsic incubation period* (IIP), which literally refers to the starting time when an infected mosquito has successfully infected a susceptible human, and the pathogen has started to greatly increase in number inside the host body [85]. Subsequently, when the pathogen has multiplied and reached a certain threshold, malaria

symptoms would then manifest. Furthermore, the chances of infection transmission would be very high during this period. The IIP is accounted for in Figure 3.1 and particularised as an exposed compartment in the human dynamics. In essence, agent-based modelling tries to realistically imitate the behaviour of individual agents and their interactions in the environment. The transmission of malaria is characterised by spatial and temporal considerations, in which the movement of people, whether short or long term, supports the spread of the malaria infection. For obvious reasons, people move from place to place within their environment for different purposes. This study considers the movement of people within their environment regardless of where they are heading, and thus observe the emerging patterns. Although mosquitos usually bite at night, other species bite during the day; however this study track the pattern of the infection spread irrespective of day or night.

4.2.2.2 Mosquito Agents

The mosquito agent is a carrier of the malaria parasite, and moves freely through a space in search of humans to bite, and thus transmits the parasite. Naturally, mosquitoes have certain characteristics that compose their life-cycle, which includes: biting rate, mortality rate, egg-deposition rate, birth rate and immature mortality rate.

A malaria infection starts, if a susceptible mosquito bites an infected human, following which it becomes infected at a probability of b. Conversely, when such an infected mosquito bites a susceptible human, it infects the human with a probability of a. In Figure 3.1 we have illustrated the individual transition in the compartments (SEIR) and their respective probabilities.

4.2.2.3 Pathogen Agents

The pathogen is a parasite causing the malaria infection, which in a biological sense is called the *plasmodium species*. *Plasmodium* transfers to the susceptible mosquito through biting an infectious human and by sipping blood that contains the pathogen. Hence, the mosquito will become infected but not capable of transmitting the malaria infection to susceptible humans until the ingested blood containing the pathogen has developed. This will then follow some developmental stages before the mosquito becomes infectious. The time it takes for the pathogen to complete its development inside the mosquito is called the *extrinsic incubation period* (EIP). This period is sensitive to environmental temperature [112], meaning at a relatively high or low temperature spectrum the EIP could be shorter or longer, respectively.

4.2.2.4 Environment Agents

The environment as an agent leverages the spread of malaria and is ever-changing in space and time as the climate changes. The environment plays a vital role in the transmission of malaria, and enables the movement of people and mosquitoes, providing mosquitoes with breeding sites for egg deposition and the maturation of juvenile mosquitoes. In order to realistically represent the spatial movement of humans and mosquitoes, an artificial environment was created in the *Netlogo* platform (see Figure 4.8), in which both humans and mosquitoes are displayed in the environment representing the spatial distribution of agents in a town, city, or community settlement.

4.3 Materials and Methods

This section presents a vivid description of the methodologies used to achieve the aim of this study. Three cities were selected, each from a different country; these were Tripura district in India, Limpopo province in South Africa and Benin city in Nigeria. World map data [10] were deployed to produce the maps of the three cities. However, since the entire city regions were difficult to handle or study, we resorted to the reduction of complexity in order to alleviate the computational task.

A statistical technique was adopted to reduce the computational task and guide the selection of the best performing model. The tools for simulation and computational functions of the temperature-dependent parameters of the malaria transmission model in Figure 3.1 are also presented.

4.3.1 Case Study

This paper aims to investigate the dynamics of malaria in human and mosquito populations through agent-based modelling and mathematical modelling. The model developed will be validated against reported cases of malaria for different populations. To do this, data from three cities in different countries were considered, this including Tripura district in India, Limpopo province in South-Africa and Benin City in Nigeria. These countries are known for their malaria endemic status [72], and apparently have different seasons for transmission, climate patterns, parasites and vector species. For instance, Tripura district is known for its high malaria incidence, and its predominant species of malaria parasites are the *plasmodium falciparum*. This species alone is accountable for about 90% of the cases reported, and *plasmodium vivax* comprises the remaining 10% [59] of cases in the district. In Limpopo province, malaria is still

endemic, as incidence in the area is characterised by the low altitude and climate [38]; moreover, it is connected to other regions in sharing boundaries with some parts of Zimbabwe and Mozambique where malaria incidence is similarly high. Furthermore, the malaria season in Limpopo coincides with its warm and rainy summer that starts in September and goes through to May of the following year [55]. According to the World Health Organization (WHO), Nigeria is rated among the highest malaria endemic countries across the globe [57], and Benin City is the capital of a state called Edo located in the southern part of Nigeria. The City has a tropical climate, which is characterised by a longer rainy season over a 12-month period and the average annual temperature is 26.1°C. This rainfall pattern (with an average 2025mm annual precipitation) is regarded as the most likely influential climate factor leading to the high incidence of malaria cases.

4.3.2 Sources of Data

The reported cases of malaria for Tripura district, Limpopo province and Benin City were taken from published sources [8, 20, 59] respectively. Since the occurrence of malaria is connected to climate factors, as such temperature data will be use for the computation of temperature-dependent parameters in the malaria model. This study uses the average monthly temperature data, as temperature is the large-scale driver of malaria transmission since it influences the mosquito survival, the parasite development, its biting rate and the aquatic development of the juvenile mosquito. The following is the record of the average monthly temperature, for Tripura district [59], Limpopo province and Benin City [7]. The temperature data were collected for the same period of time, namely a year, in which the cases of malaria were reported (see

Table 4.1 for detailed information).

Table 4.1: Information on the cases of malaria and temperature distribution for a period of one year as reported in the three cities

Months	Cities					
	Tripura district, 2011		Limpopo province, 2015		Benin City, 2011	
	Reported cases	Mean temperature	Reported cases	Mean temperature	Reported cases	Mean temperature
January	240	18	863	27.4	58	26.4
February	298	29.95	1843	26.6	110	27.2
March	552	27.5	1588	25	199	27.4
April	254	27.9	411	22.3	258	27.5
May	1398	29.4	85	18.2	534	27
June	1817	29.1	38	15	512	25.6
July	1833	29.05	25	15	396	24.5
August	1760	29.35	49	17.9	787	24.5
September	1181	29.3	123	21.5	1092	24.9
October	684	27.55	192	24.2	129	25.9
November	614	23.15	144	25.9	201	26.7
December	431	17.95	83	26.9	54	26

4.3.3 Relationship Between Malaria Occurrence and Temperature

As temperature is a large-scale driver of malaria transmission, it is crucial to check and substantiate the existing relationship. This would enable to understand the spatial and temporal dynamics of malaria incidence. Figures 4.1–4.3 presents the relationships between the average monthly temperature distribution and the occurrences of malaria for Tripura district, Limpopo province and Benin City. As seen in Figure 4.1, the relationship existing between the plots is relatively positive, showing that an increase in the mean monthly temperature causes a significant increase in the pattern of malaria incidence in Tripura. As shown in Figure 4.1, the occurrences of malaria remain at high levels in the months starting from May until September. This shows a relatively unceasing pattern of malaria that spreads in Tripura, and could be possibly due to the average monthly temperature distribution that falls around the optimal temperature conditions (ranging between 25°C–27°C [148]) which favour *plasmodium* parasite development. In Figure 4.2 the relationship is different from that observed in

Figure 4.1; this indicates the negative impact of a lower temperature regime due to the abundance of mosquitoes and parasite development. The parasite causing malaria will cease to develop when the temperature is below 14.5°C for *plasmodium vivax* and *plasodium malariae*, and 16°C for *plasmodium falciparum* [128]. Occurrences of malaria in Limpopo have lag effects of 1-2 months around the average monthly temperature distribution. This demonstrates the relationship between the malaria parasite incubation period and temperature [117]. The malaria season in Limpopo starts during the last quarter of a given year and continues until the first quarter of the following year. In Figure 4.3, the pattern of malaria occurrence in Benin City is optimally driven by the average monthly temperature, which ranges from 25°C–27°C [148]. Moreover, the pattern of malaria incidence in Benin is perennial; its peak season starts February and lasts until November of the same year.

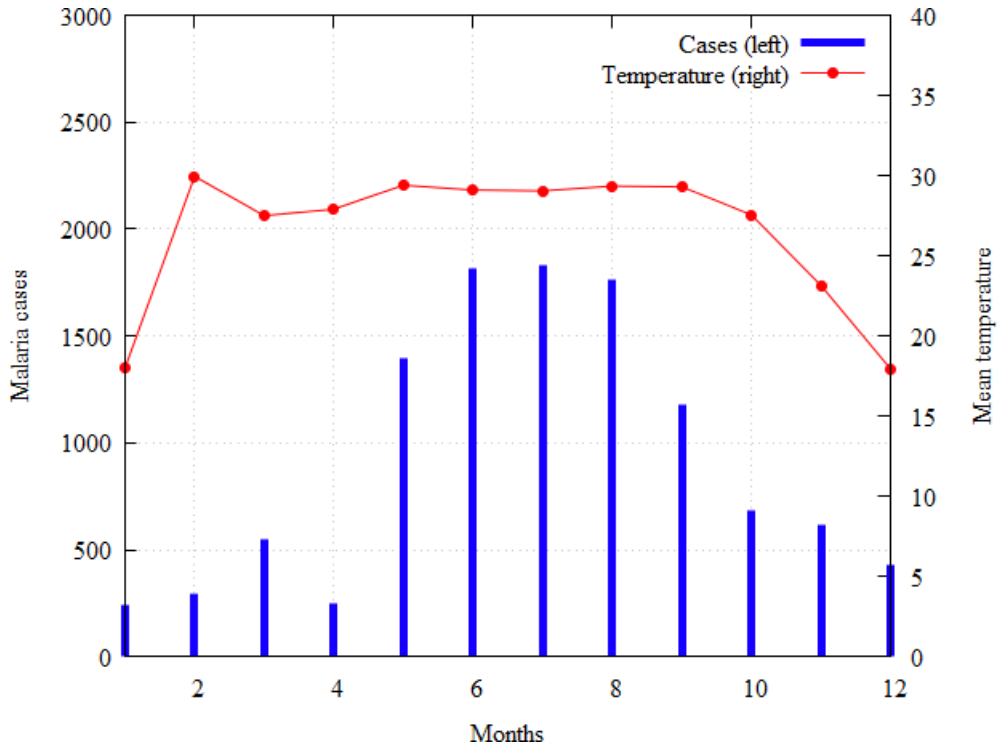

Figure 4.1: The plot describes the cases of malaria and temperature distribution reported in Tripura district, India [59].

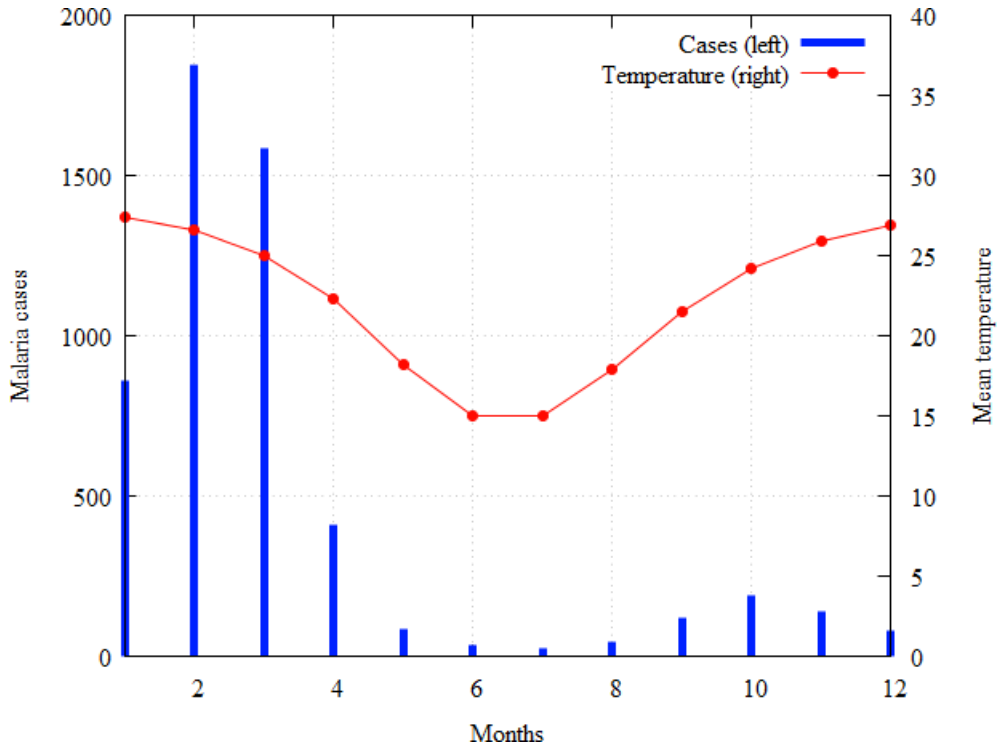

Figure 4.2: The plot describes the cases of malaria temperature distribution reported in Limpopo province, South Africa [7].

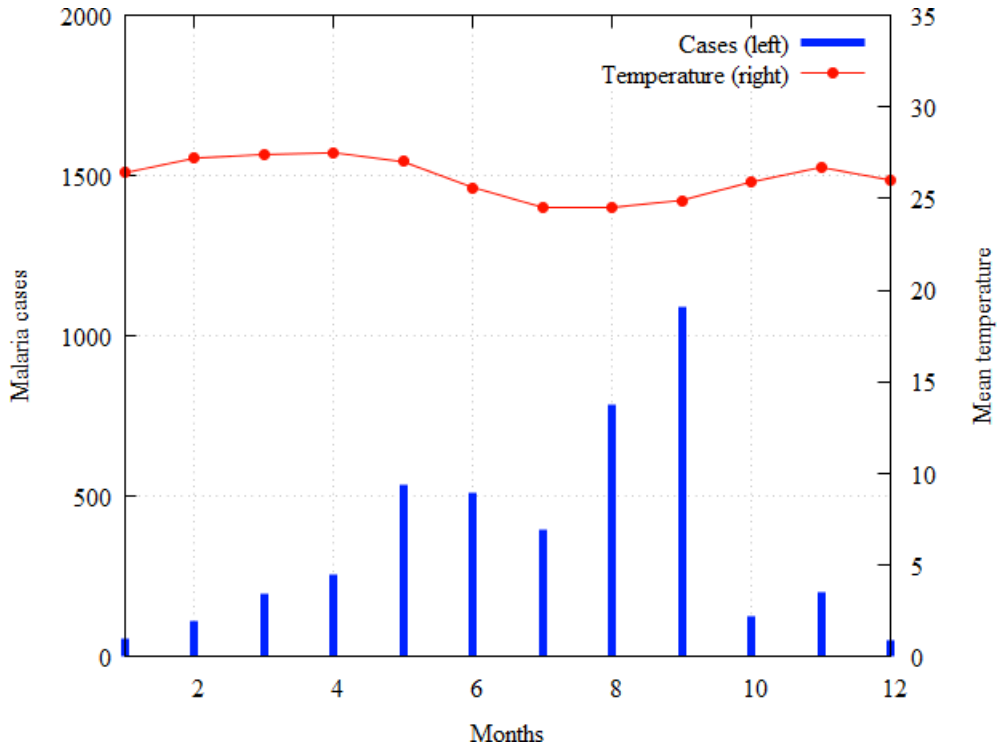

Figure 4.3: The plot describes the cases of malaria temperature distribution reported in Benin City, Nigeria [7].

4.3.4 Complexity Reduction

The entire regions covering the cities shown in Figures 4.4(a–c) are too large in terms of population size and land size. To alleviate the computational difficulties, the region under study will be scaled down. Using the population size of the cities and land mass, the population density of the cities can be obtained from $\rho = N/A$, where ρ = population density, N = total human population and A = total land area. Table 4.2 presents the population density of the cities, showing that all cities have different densities. Hence, the cities were scaled down according to their densities, and the consideration was to study the areas covering the dimensions: 1km×1km in Tripura, 3km×3km in Limpopo and 1km×1km in Benin (as indicated in Figures 4.4(a–c)). Based on the reduced dimensions of the cities, as shown in Figures 4.4(a–c), the sub-population, or target population, of the demarcated areas were as follows: 1:35 (representing 350 people), 1:4 (representing 400 people) and 1:13 (representing 1300 people) respectively.

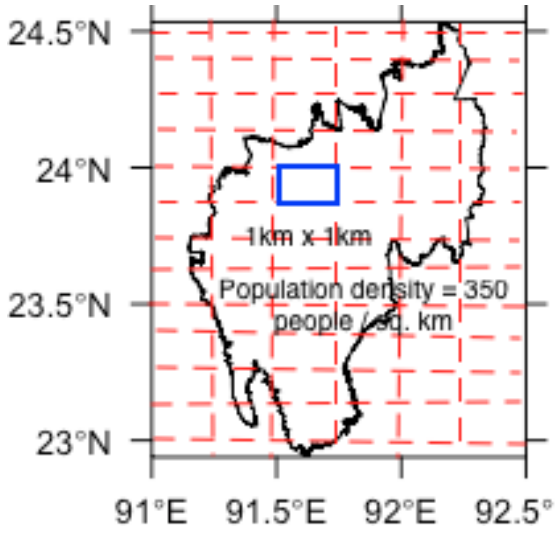

Figure 4.4: Illustrates the scaling down of the Tripura district.

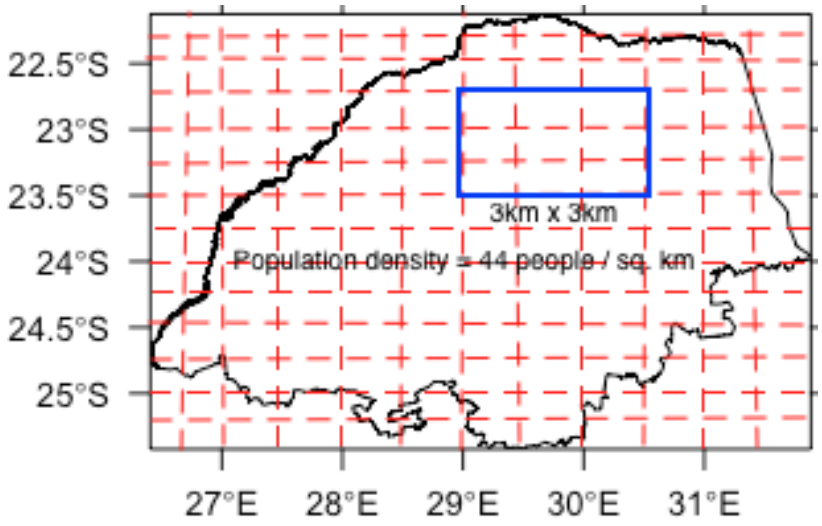

Figure 4.5: Illustrates the scaling down of the Limpopo province.

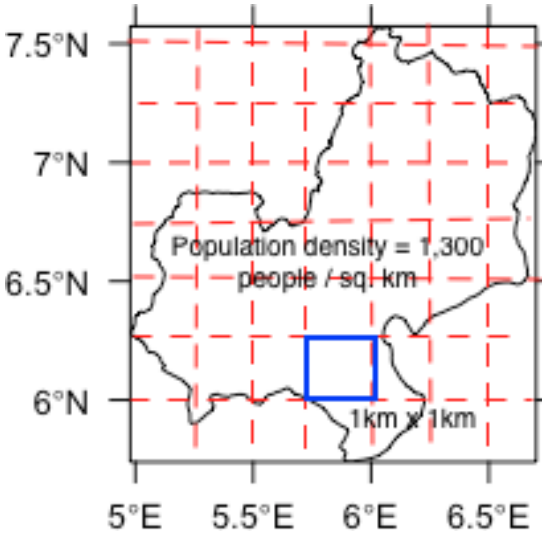

Figure 4.6: Illustrates the scaling down of the Benin City.

Table 4.2: Demographic information of the study areas.

	Cities		
	Tripura district	Limpopo province	Benin City
Population size	3,673,917	5,554,657	1,495,800
Land area	10,492km^2	125,754Km^2	1,204Km^2
Population density	350/km^2	44/Km^2	1242/Km^2
Target population	350	400	1300
Life Expectancy	69	62.77	54.5

4.3.5 Statistical tests

It is a verifiable fact that synthetic data are often generated to represent original data even though they are relatively different. However, George Box says [40] that *all models are wrong, but some are useful*. This study focuses on the application of agent-based modelling and mathematical modelling to study malaria dynamics in a population; and thus compare the results against the real cases. In fact, it is difficult to pinpoint the model that is best representation of the reported cases of malaria in each of the cities by merely looking at plotted patterns. To address this challenges, a *t-test* for two independent samples [88] and a *product-moment* correlation [109] were utilised. These statistical tests are robust for testing whether there is a significant difference between the means of the two independent samples (model generated cases and reported cases of malaria). The statistical tests results are discussed in Section 4.5 while its summary is presented in Table 4.5.

4.3.6 Parametrization

Fundamentally, malaria transmission is leveraged by climate factors and, in particular, temperature is the large-scale driver of its transmission. As shown in Figure 3.1, the mosquito related parameters, like the mosquito biting rate $c_m(T)$, adult mosquito mortality rate $\mu_m(T)$, immature mosquito mortality rate $\mu_a(T)$ and adult mosquito egg deposition rate $\alpha_E(T)$, all depend on temperature. Since there was no empirical evidence or values of these temperature-dependent parameters for the cities selected, the functional relationship between these parameters [112] were utilised to determine the precise values corresponding to the demographic and climate information in Table 4.2. The temperature-dependent function of the parameters were described using the

polynomial of degree two; their mathematical representations are shown in equation 3.11. Similarly, the temperature-dependent linear function describing the immature mosquito mortality rate was defined in [133], and is shown in equation 3.12.

Using the temperature records of these cities (see [59] for Tripura district and [7] for Limpopo province and Benin City), the values of the temperature-dependent parameters can be obtain, and summarised in Table 4.3. The values are used in our simulations. Other parameters, e.g. the human birth or recruitment rate η, and the per capita human death rate μ_h, can be calculated using the population size and human life expectancy of the cities (see Table 4.2). This information was accessed in the census database of the cities through online published sources (following the references [11, 12, 13, 14, 15, 63]). The formulas used for computing the human birth and death rates are $\eta = \mu_h \times N$ and $\mu_h = 1/(LE \times I)$, where N is the total human population size, LE is the human life expectancy and I is the index denoting the rate per month or day [98].

Table 4.3: The parameter values and their ranges.

Symbol	Baseline	Range	Reference
η	4×10^{-5}/day	$(3.91 - 5) \times 10^{-5}$/day	[116]
ω_h	1.7×10^{-5}/day	$5.5 \times 10^{-5} - 1.1 \times 10^{-2}$/day	[21]
μ_h	4×10^{-5}/day	$(3.42 - 3.91) \times 10^{-5}$/day	[21]
μ_m	5×10^{-2}/day	$(4.76 - 7.14) \times 10^{-2}$/day	[21, 124]
μ_a	1.04×10^{-1}/day	$1 \times 10^{-3} - 2 \times 10^{-1}$/day	[21, 98, 131]
α_E	1.84/day	$1 - 500$/day	[21, 98, 131]
p	0.25	-	[116]
α_h	5×10^{-3}	$(2 - 7) \times 10^{-3}$	[131]
γ_h	2.3×10^{-3}/day	$1.4 \times 10^{-3} - 1.7 \times 10^{-2}$/day	[21]
θ_m	9.1×10^{-2}/day	$2.9 \times 10^{-2} - 3.3 \times 10^{-1}$/day	[21]
τ_2	10	$10 - 14$/days	[132]
a	2.4×10^{-1}/day	$7.2 \times 10^{-2} - 6.4 \times 10^{-1}$/day	[21, 50, 124, 131]
b	2.2×10^{-2}/day	$2.7 \times 10 - 3 - 6.4 \times 10^{-1}$/day	[21, 50, 131]
c_m	0.29/day	0.10 – 1.0/day	[21, 50, 124, 131]
m	2	-	[116]
$\vartheta(\hat{T})$	0.343/day	$0.333 - 1.0$	[131]
K_c	4×10^4	$50 - 3.3 \times 10^6$	[21, 131]
τ_1	12	$10 - 37$days	[112, 131]
δ_h	3.454×10^{-4}	$0 - 4.1 \times 10^{-4}$/day	[21]

4.3.7 Simulation Toolkits

Two simulation platforms were used: *VenSim* [16] is for mathematical modelling and *NetLogo* [176] is for agent-based modelling. However, agent-based models can also be implemented through programming languages, such as *C*, *Java* and *Python*. The *C* programming language was used to develop *Repast*, *Soar* and *Swarm* platforms. These platforms are primarily designed for social sciences, general learning problems and general purpose agent-based systems, respectively [17]; however, *Java* is a versatile programming language used in building many platforms, including: *AnyLogic*, *Cougaar*, *JADE*, *MASON*, *Repast*, *SARL*, *Soar*, *Sugarscape* and *Swarm*. In *Python*, agent-based models are implemented in a framework called *Mesa*, a modular framework for building, analysing and visualising agent-based models [46].

4.4 Experimentation

In this section, we present our experiments by running the malaria model shown in Figure 3.1 using a mathematical modelling and agent-based modelling. A system dynamic modeller in the *VenSim* platform was used to design the causal loop diagram shown in Figure 4.7, and the simulation of the malaria transmission model in Figure 3.1. The *NetLogo* platform was subsequently utilised for agent-based modelling and simulation by tuning the values of the parameters in Table 4.3, e.g. average temperatures, population sizes, land masses, densities and human life expectancy. A detailed discussion of the experiments and processes is given below.

4.4.1 VenSim Simulation

The digram in Figure 4.7 is a causal loop representation of the model in Figure 3.1 using the *VenSim* system dynamic modeller. The values of the parameters in Table 4.3 were picked and referenced to the cities' climate and demographic information and supplied within the system dynamic modeller's causal loop diagram in the *VenSim* platform (see Figure 4.7). Subsequently, the model was calibrated and the results were simulated within the ranges of the parameters; thus the dynamics of malaria transmission were generated for each of the cities. The results obtained for each of the cities are presented in Table 4.4, and further depicted in Figures 4.13–4.15.

A numerical solution of mathematical modelling is a deterministic or non-probabilistic outcome in which, within a particular set of parameters, the results of the simulation remain consistent for any number of trials. In agent-based modelling, the results of performing simulation from the same set of parameters is considerably different for every trail.

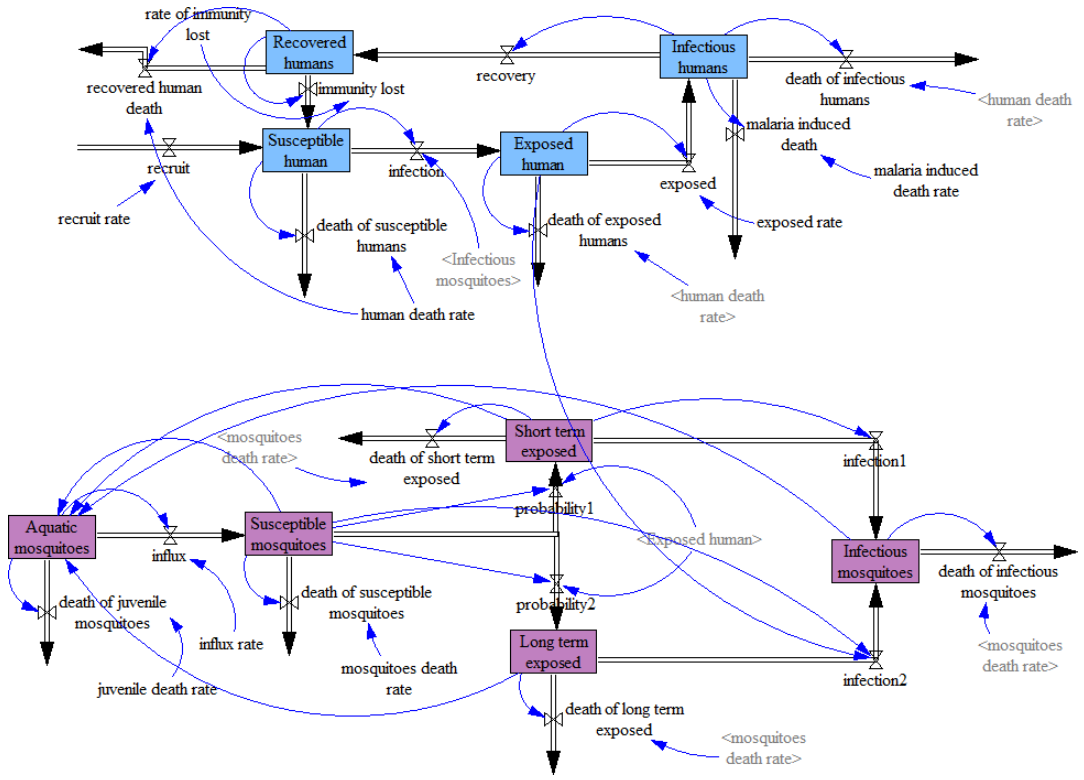

Figure 4.7: Simulation of the model shown in 3.1 using *Vensim* system dynamic modeller. The upper transition shows human population dynamics to malaria transmission while the lower transition shows mosquito population dynamics.

4.4.2 NetLogo Simulation

The *NetLogo* platform recognises all mobile agents, like humans and mosquitos, as *turtle*, and static agent, like environment, as *patches* [176]. The visualisation dashboard provided in this platform is not sufficient to illustrate the number of human and mosquito agents. For this reason, the rescaled agent populations within the demarcated areas of the cities, as indicated in Figures 4.4(a–c), will now be used for this study. Similarly, the mosquito population used for the simulation is actually difficult to determine as they are uncountable and are difficult to control. However, 1:2 ratio is often used for the human to mosquito population (as suggested by [85, 116]). For this reason, 1:40 mosquitoes (one real shaped mosquito in the *NetLogo* environment represents 40 virtual mosquitoes) were represented in the *NetLogo* space. Figure 4.8 presents the initial setup of the agent-base model simulation interface of the malaria transmission model in Figure 3.1. In order to realistically mimic human and mosquito disposition in the world, we spatially distributed all agents in the *NetLogo* environment, while the sliders and plotting spaces were used for the parameters and outputs of the simulation, respectively.

4.4.2.1 Creating Environment

The setup of the agent-based model of malaria transmission in Figure 3.1 using the *NetLogo* platform is presented in Figure 4.8. A *NetLogo* environment was created, and the following code is used for defining a hypothetical world (`ask patches [set pcolor green]`), then populate the space with human and mosquito agents. The human and mosquito agents are depicted by *person* and *butterfly* shapes, respectively. Subsequently, the human and mosquito agents were initialised in accordance with their

spatial dispositions in the environment by invoking [`random-xcor,random-ycor`], to account for systematic preferences during the simulation. In order to start malaria transmission within the created small environment, it is presumed that few of the human agents have the malaria infection in their blood, and moreover, that humans are hosts of the pathogens causing the malaria [138], called *Plasmodium species*. Hence, for the malaria infection to effectively spread in the environment, the following assumptions would be considered:

i. 10% of the human agents have the malaria infection.

ii. All mosquitoes are adults and susceptible at the initial time. With the assumption that 10% of the human agents have malaria infection, the red colour will be use to distinguish infected human agents from healthy, this is to avoid bewildering the susceptible human counterparts.

Then, by invoking the following command, (`if random-float initial-number-of-humans < 10`) to ask that 10% of the human agents to be infected and assigned a red colour. Meanwhile, the remaining human agents were healthier and maintained their white colour, as shown in Figure 4.8. Since all agents were displaced randomly in the created *NetLogo* environment, the next step is to define the sets of rules or instructions for the agents to follow and behave accordingly.

4.4.2.2 Agents Procedure

The spread of malaria infection depends on how well the agents involved interact. People move from one place to another for several reasons, for example school, work, business or tourism. Thus, this movement increases the chance of becoming infected.

Therefore, the agents have to move within the *NetLogo* environment for the malaria infection to take effect (see Figure 4.12(a)), and the procedure (`right-turn random 360°, left-turn random 360°and forward 1`) accounts for agents movements. Subsequently, the instruction follows that, if any of the mosquitos come into contact with red coloured humans, as the interaction progresses the particular mosquito will then be changed to a blue colour (see Figure 4.12(a)) indicating infected. The infected mosquito will not be capable of transmitting malaria to the susceptible human until it completes its extrinsic incubation period (EIP). This period mostly lasts for about 7-14 days [128]; sometimes the mosquito and parasite species could also be another factor affecting the EIP. When the EIP is completed, the infected mosquito will then change colour to black (see Figure 4.12(a)) indicating infectious, and thus becomes a potential candidate for the spread of malaria. The aftermath is the spread of the malaria infection to the humans coloured white upon successful contact and bites by infectious mosquitoes. Consequently, infectious mosquitoes will pick human blood for nourishment and egg production. Regardless of whether the mosquito is infected or infectious, both will deposit eggs around swamp areas or water bodies, and thus new mosquitoes will be recruited (`hatch new-mosquito [set colour yellow forward 1]`).

To better understand the procedural stages involved in the simulation cycles of the malaria transmission model (Figure 3.1) using an agent-based model (as shown in Figure 4.12(a)). The **Algorithm** 1 and **Algorithm** 2 are developed to explain the human and mosquito procedures.

Algorithm 1 Summary of the human procedures

```
1: let initial human population be i
2: select 10% random-float of i and assign colour red
3: if 10% of i with colour red then
4:     set status infected
5: else
6:     if i coloured white then
7:         set status susceptible
8:         for(i in 1:n)
9:         set i to move random left or random right at 360°
10:        repeat step 3, for each iteration
11:        then step 8
12:    end if
13: end if
```

Algorithm 2 Summary of the mosquito procedures

1: **let** initial mosquito population be *j*
2: all j^{ls} are assumed susceptible
3: **for** (*j* in 1:n)
4: **set** *j* to fly **random left** or **random right** at 360° seeking i^{ls} blood
5: **if** any (*j* in 1:*m*) bite any (*i* in 1:*n*) with colour red **then**
6: **set** status infected for *j*
7: **set** colour blue
8: **repeat step 5**
9: **else**
10: **if** status is susceptible **then**
11: otherwise **step 12**
12: **for** (*j* in 1:*m*) with status infected undergo extrinsic incubation period
13: **else**
14: **if step 12** is completed **then**
15: **set** status infectious
16: **set** colour black
17: **repeat step 12**
18: **repeat step 3 to 4** through **step 14** until all j^{ls} are infectious
19: **let** λ be the average lifespan for j^{ls}
20: **else**
21: **if** lifespan for each j^{ls} in **step 18** exceed λ **then**
22: **set** status die
23: **repeat step 3**
24: **else**
25: **if** all j^{ls} execute **step 21 to 22 then**
26: stop
27: **end if**
28: **end if**
29: **end if**
30: **end if**
31: **end if**

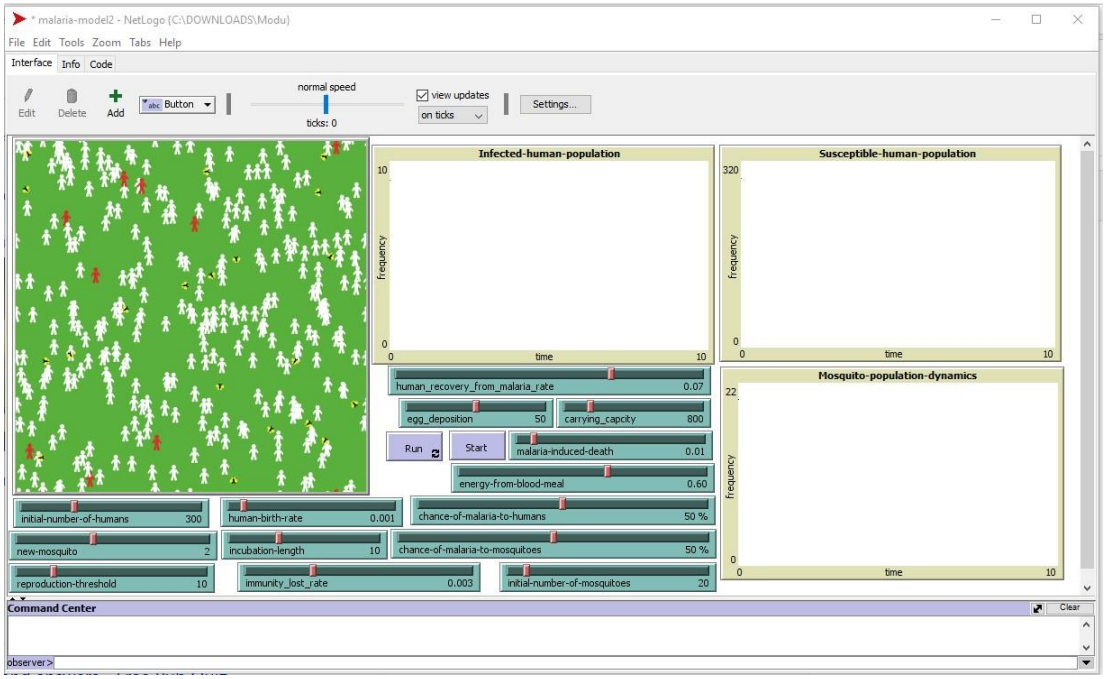

Figure 4.8: A screenshot showing the interface of theABM using the *NetLogo* platform.

The setup has three components, which are:

i. The main environment, which is the world where all agents would be spatially distributed accordingly.

ii. Three plane frames are meant to plot the simulation outputs in real-time as the agents interact in the environment according to predefined rules.

iii. Then widgets, of two kinds namely the sliders and buttons, are used to interact with the agents through the calibration of model parameters and start/run.

4.4.2.3 Veracity of the Model

To test the agent-based model shown in Figure 4.8, a trial simulation was performed and ran it 50 times by varying the referenced parameters in the sliders. The pattern it produces was visualised in the plotting spaces while all agents communicate with each other in the environment according to the predefined rules. The plots shown in Figures 4.12(a–c) describes the pattern of infectiousness in humans, in mosquitoes, and susceptible human dynamics, respectively. The hypothetical values of the parameters were calibrated in the sliders, to test the veracity of the agent-based model to the malaria transmission model in Figure 3.1. The pattern produced in Figures 4.12(a–c) clearly manifested the general characteristics of an epidemic model's behaviour [32].

As the season of malaria transmission usually falls within a year depending on the length of the rainfall season, hence the simulation was pre-set to run for a year period in order to spot the peak malaria season. In Figure 4.12(a) a uni-modal peak season is connected with a period of abundance for mosquitoes and Figure 4.12(b) corroborates this. The results produced in Figures 4.12(a and b) further proved that the incidence of malaria is largely leveraged by the availability of adult mosquitoes [133].

In a population with a relatively low fertility and mortality rate, the pattern of malaria susceptibility in humans will decrease as the infection increases. Thus, Figure 4.12(c) confirms that a susceptible human population shows a decreasing pattern as time increases. This shows the robustness of an agent-based model in its ability to study the characteristics of individual agents involved in the phenomena and mimics its real-world scenario.

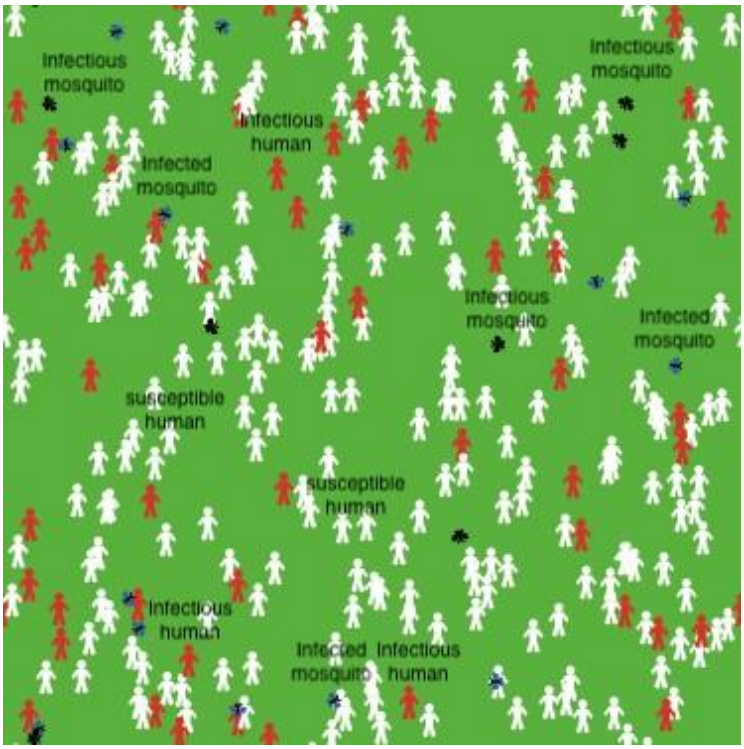

Figure 4.9: This plot shows the spatial distribution of the malaria agents in the *NetLogo* environment.

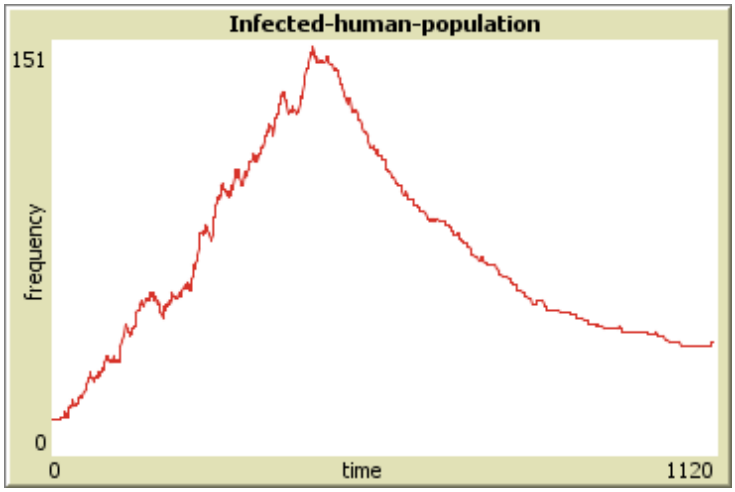

Figure 4.10: This plot shows the pattern of malaria's infectiousness in human population with its season indicated by the topknot.

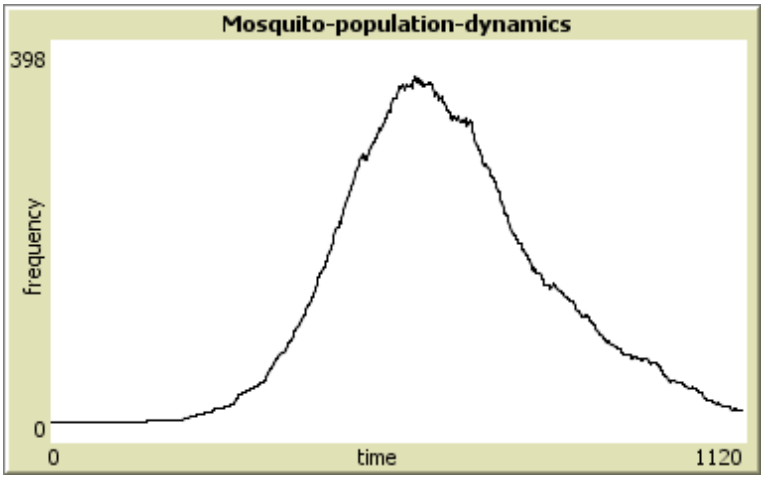

Figure 4.11: Describes the dynamics of the mosquito population in the spread of malarial infection to the human population. The plot also presented some embedded features of the mosquito like biting rate, the extrinsic incubation period, egg deposition, birth rate and death rate.

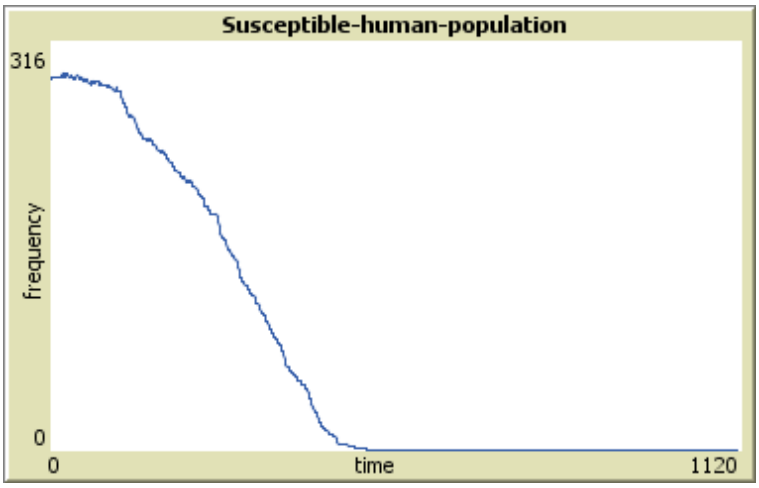

Figure 4.12: The simulation plot showing the pattern of human susceptibility due to malaria infection in a population with relatively low fertility and mortality.

By turning the values of the parameters within their ranges (Table 4.3) and with reference to the cities' climates and demographic information, the pattern of malaria transmission will then be resolved. For each of the cities, the simulation was ran 100, and produced results on an average of the entire simulation cycle.

4.5 Results Presentation and Validation

This section presents the results of the simulating model in Figure 3.1 through an agent-based and mathematical modelling approach. For each of the cities, the simulation outputs is provided (see Table 4.4) and visualised them in Figures 4.13–4.15. The reported cases of malaria within the cities and the model's results are combined in Table 4.4. By looking at the patterns of the results compared with the cases, a wide variance is noticed, as indicated by the magnitude of the values. This made it difficult to critically comprehend the patterns. However, to make them consistent, and since no zero instances was observed in the results, hence, logarithmic transformation [47] technique was applied to normalise the results in Table 4.4. Subsequently, the log-transformed values in Table 4.4 are plotted, as shown in Figures 4.13–4.15, and a detailed explanation follows.

Figure 4.13 presents the cases of malaria reported in the Tripura district against the model results produced using an agent-based model and mathematical model (see Table 4.4). The plots were produced using the two modelling approach, and indicate a certain strength of relationship within the occurrences of malaria in the district. The peak season of malaria incidence was predicted, as evidenced by the trough and node pattern produced in both the modelling approach, which is near-similar to the cases reported. The agent-based modelling and mathematical modelling performed well in

predicting not only the pattern of malaria transmission in the district but also the season.

Similarly, Figure 4.14 shows a comparison between the cases of malaria reported in Limpopo province and the model generated results using an agent-based model and mathematical model (see Table 4.4). The pattern produced by the plots in Figure 4.14 clearly shows that malaria occurrences in the province is represented well using the agent-based model results. However, the results produced using the mathematical model mimicked the pattern of the province malaria cases at the beginning of the simulation but later drifts down. This is because mathematical models are suitable for trend determinations and the investigation of continuous phenomena. Hence, between the two modelling approaches, the agent-based model performed well in predicting the pattern of malaria incidence in the province.

Figure 4.15 presents the plots of the reported cases of malaria in Benin City together with model outcomes generated through an agent-based model and mathematical model (see Table 4.4). The plots in Figure 4.15 show a moderate relationship between the cases reported in Benin and the simulated results. In particular, the mathematical model produces a trend that follows a fluctuating pattern of malaria occurrences in the City. This shows that mathematical modelling is a good candidate for trend determination.

Table 4.4: Reported cases of malaria in the three cities and the results of the simulation that used mathematical modelling and agent-based modelling.

Months	Cities																	
	Tripura district, 2011						Limpopo province, 2015						Benin City, 2011					
	C	log(C)	A	log(A)	M	log(M)	C	log(C)	A	log(A)	M	log(M)	C	log(C)	A	log(A)	M	log(M)
Jan	240	2.38	320	2.51	91	1.96	863	2.94	293	2.47	33	1.52	58	1.76	133	2.12	28	1.45
Feb	298	2.47	337	2.53	300	2.48	1843	3.27	594	2.77	61	1.79	110	2.04	191	2.28	134	2.13
Mar	552	2.74	493	2.69	384	2.58	1588	3.20	341	2.53	43	1.63	199	2.30	219	2.34	251	2.40
Apr	254	2.40	210	2.32	152	2.18	411	2.61	143	2.16	15	1.18	258	2.41	268	2.43	377	2.58
May	1398	3.15	538	2.73	388	2.59	85	1.93	143	2.16	09	0.95	534	2.73	324	2.51	523	2.72
Jun	1817	3.26	699	2.84	374	2.57	38	1.58	113	2.05	06	0.78	512	2.71	251	2.40	686	2.84
Jul	1833	3.26	1520	3.18	812	2.91	25	1.40	51	1.71	04	0.60	396	2.60	134	2.13	848	2.93
Aug	1760	3.25	750	2.88	543	2.73	49	1.69	142	2.15	03	0.48	787	2.90	610	2.79	986	2.99
Sep	1181	3.07	630	2.80	329	2.52	123	2.09	120	2.08	02	0.30	1092	3.04	1567	3.20	1078	3.03
Oct	684	2.84	575	2.76	315	2.50	192	2.28	192	2.28	02	0.30	129	2.11	720	2.86	1118	3.05
Nov	614	2.79	554	2.74	302	2.48	144	2.16	201	2.30	02	0.30	201	2.30	675	2.83	1111	3.05
Dec	431	2.63	347	2.54	289	2.46	83	1.92	132	2.12	02	0.30	54	1.73	112	2.05	1074	3.03

where: C = Reported cases of malaria, A = Agent-based method results, M = Mathematical method results and log(C), log(A), log(M) are their corresponding logarithmic transformations.

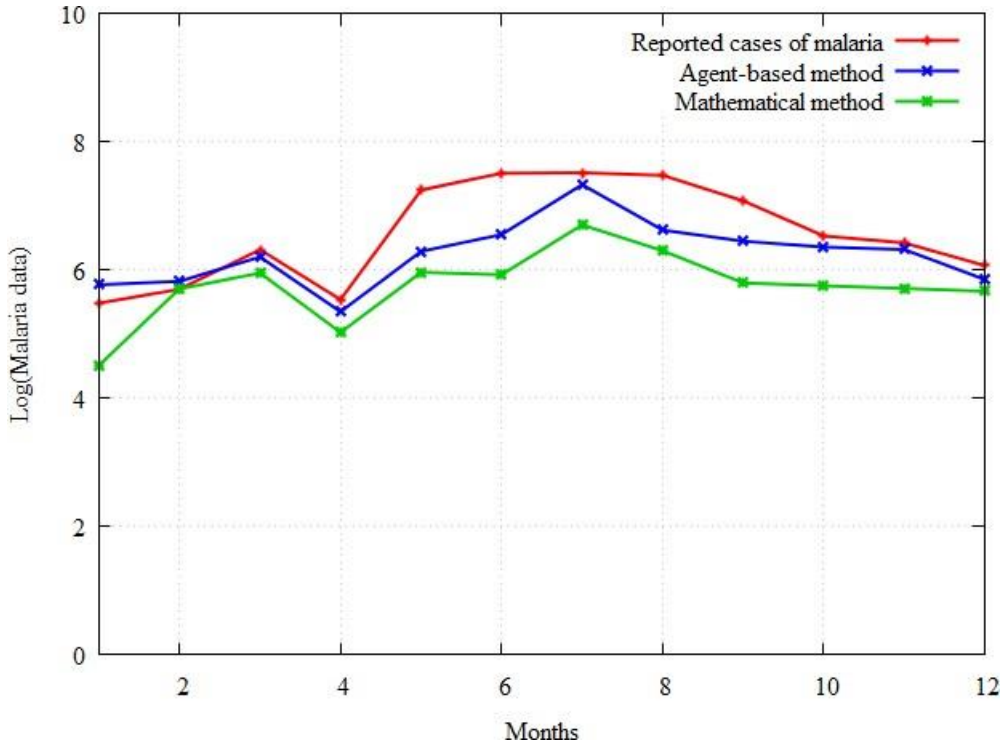

Figure 4.13: The reported cases of malaria in Tripura district and the simulated results produced using mathematical modelling and agent-based modelling were plotted together. This aimed to recognize the best approach between the two that described the pattern of incidence in the district.

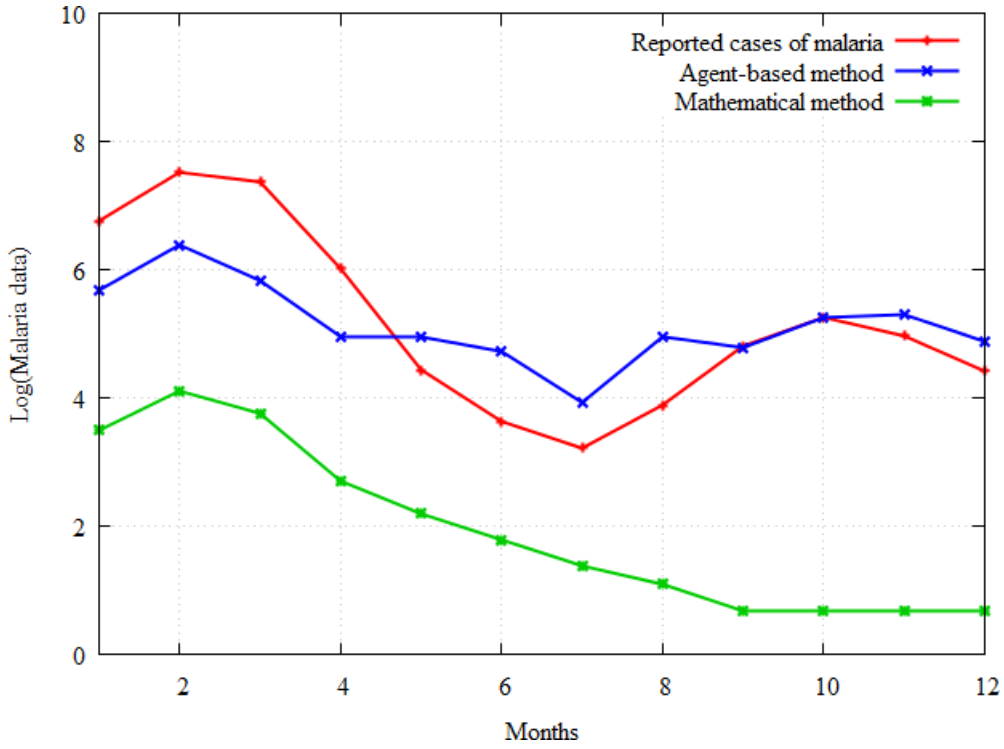

Figure 4.14: The reported cases of malaria in Limpopo province and the simulated results produced using mathematical modelling and agent-based modelling were plotted together. This aimed to recognize the best approach between the two that described the pattern of incidence in the province.

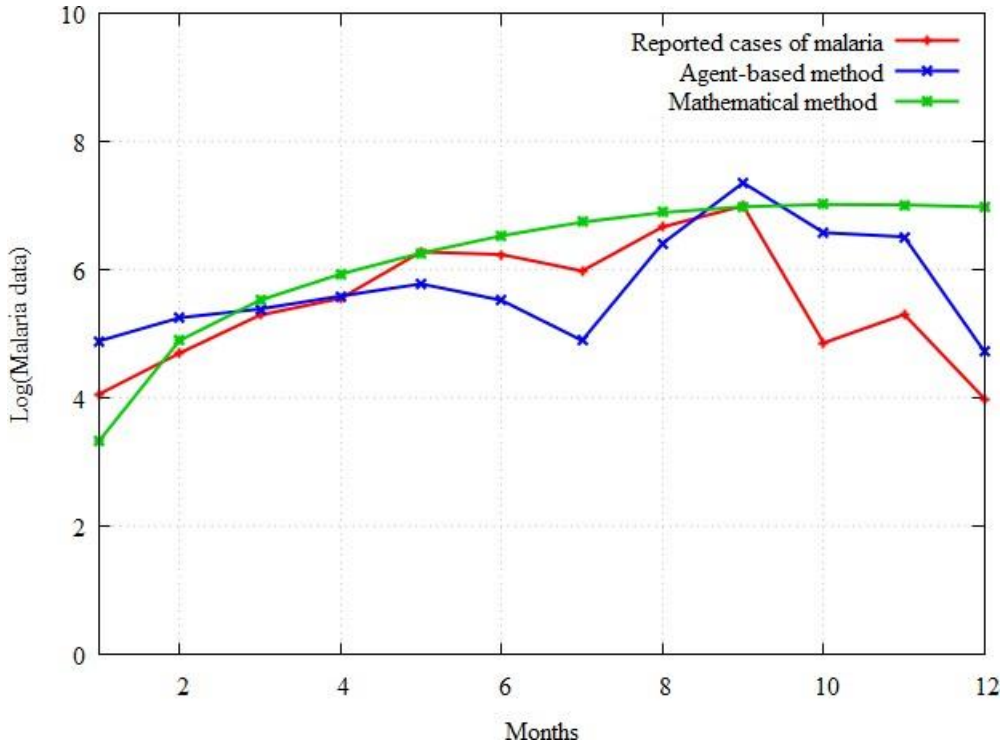

Figure 4.15: The reported cases of malaria in Benin City and the simulated results produced using mathematical modelling and agent-based modelling were plotted together. This aimed to recognize the best approach between the two that described the pattern of incidence in the City.

4.5.1 Statistical Tests and Inferences

The plots in Figures 4.13–4.15 demonstrates a pattern that means it is difficult to recognise the best performing model in predicting the occurrence of malaria between the agent-based and mathematical models. A *t-test* for two independent samples [88] together with a *correlation coefficient* [109] were utilised and the best performing model was selected. The computational results of the parameters for the two statistical techniques were summarised and presented in Table 4.5.

Table 4.5: Present the summarised results of the *t-test* for two independent samples.

	Cities					
	Tripura district		Limpopo province		Benin City	
t-statistic	C vs A	C vs M	C vs A	C vs M	C vs A	C vs M
t-value	1.6378	2.9560	1.3444	2.3861	0.0850	3.4243
p-value	0.1157	0.0073	0.1925	0.0261	0.9330	0.0084
r	0.7732	0.7768	0.9784	0.9318	0.7176	0.5420
Decision	$p_{value} > 0.05$	$p_{value} < 0.05$	$p_{value} > 0.05$	$p_{value} < 0.05$	$p_{value} > 0.05$	$p_{value} < 0.05$
Remark	Not sig	Sig	Not sig	Sig	Not sig	Sig

where: C = Reported cases of malaria, A = Agent-based method, M = mathematical method, t-value= test-statistic, p-value = highest threshold probability for rejecting the null hypothesis, r = coefficient of correlation use to measure a degree of association, Not sig = indicating there is no significant difference and Sig = indicating there is significant difference.

Based on the results presented in Table 4.5, Figure 4.13 shows that there is no significant difference on average between the reported cases of malaria and the results produced through an agent-based model ($p_{value} = 0.1157$, is greater than $\alpha = 0.05$). Similarly, it is also observed that there is a significant difference in the results produced through the mathematical model compared to the reported cases of malaria in Tripura ($p_{value} = 0.0073$). A moderate degree of association was found (see Table 4.5, $r = 0.7732$ and $r = 0.7768$) between the reported cases of malaria and the model-based results produced using the agent-based model and mathematical model, respectively. The results in Table 4.5 affirmed no significant difference between the reported cases

of malaria and the results produced using the agent-based model in Limpopo and Benin (p_{value} = 0.1925 and p_{value} = 0.9330). However, the mathematical model results were found to be significantly different with the malaria cases reported in the two cities (p_{value} = 0.0261 and p_{value} = 0.0084). Furthermore, a strong correlation between the cases of malaria in Limpopo and the agent-based model cases was found (r = 0.9784 shown in Table 4.5), and this result was also similar to those produced by the mathematical model (r = 0.9318 shown in Table 4.5). However, in Benin the degree of association between the cases of malaria and the results produced using the agent-based model was moderate good compared to the mathematical model (r = 0.7176 and r = 0.5420 respectively).

In all the cities studied, the agent-based model performed well in predicting the occurrences of malaria incidence as opposed to mathematical model. However, the mathematical model was good for trend analysis and performed particularly very well for Tripura and Benin (see Figure 4.13 and 4.15) compared to Limpopo (see Figure 4.15).

4.6 Summary

This chapter presents the application of agent-based modelling to investigate the dynamics of malaria, and compared the results with a compartmental modelling approach. The results show that agent-based modelling is a good candidate for studying malaria in a heterogeneous human population. Moreover, compartmental modelling is appropriate for trend analysis as it suitably dealt with continuous phenomena and a homogeneous population. In Chapter 5, a data-driven model using machine learning methods will be deployed to develop an intelligent system capable of informing healthcare providers with the means of anticipating a high malaria epoch or outbreak.

Chapter 5

Towards a Predictive Analytics-Based Intelligent Malaria Outbreak Warning System

This chapter provides a data-driven modelling approach to investigate the dynamics in malaria transmission. Firstly, a regression model was explored with an autoregressive structure and later extended the methodology to account for the confounding the relationship among the climate factors. Hence, the most influential climate factors for malaria was identified. Secondly, an intelligent malaria outbreak prediction model was developed using machine-learning methods and an app was developed.

5.1 Introduction

In Chapter 3 and 4, mathematical and agent-based modelling was deployed to investigate the dynamics and predict the pattern of malaria transmission. However, these

methods cannot conduct a predictive analysis because they depend on parameters that may not be known. In addition, although mathematical and agent-based model utilises temperature and precipitation, but the inability of these methods to include humidity could mean the dynamics of malaria transmission would be under-explored. Though, there is no established functional relationship between humidity and the occurrence of malaria. Hence, In Chapter 5, the data-driven method will be deployed to offer a complementary approach to model-based methods by incorporating all the climatic predictors of malaria.

Many works (e.g., [81, 101, 112, 132]) showed that malaria's prevalence is connected to climate factors and, most importantly, to temperature, precipitation and humidity. This relationship has been fused to mathematical models [18, 21, 116, 129, 131] and agent-based models [73, 83, 110, 137] to explore malaria transmission dynamics, and hence suggest prevention mechanisms. These methods performed very well in providing strategies for prevention, intervention and policy formulation. However, the modelling of temperature was used substantially, whilst the remaining factors contribute moderately to the spread of malaria. However, temperature is the large-scale driver of malaria's spread as it influences biting, speeds the parasite's development and above all provides a conducive atmosphere for the mosquito's survival. Hence, the lesser influence of precipitation and humidity in the methods will be under-explored when examining the general pattern of malaria's transmission dynamics. Chapter 5 fills this gap by deploying a robust technique that uses data analytics to consolidate the aforementioned climate factors with malaria cases in order to develop a predictive model. To bring this to a successful conclusion, there is a need to examine which of the climate factors is dominant in the incidence of malaria.

In section 5.2, a hybrid-approach will be use that is combining time-series mod-

elling and lagged-regression analysis by utilising climate factors data with reported cases of malaria. The results showed that malaria's incidence in the area studied have a significant association with relative humidity, whereas temperature and precipitation were found to have negligible effects. This finding might particularly reveal that malaria's incidence can be strongly influenced by relative humidity alone. However, this methodology is limited by its inability to capture the pre-determined existing causal relationship among the climate factors. Hence, this investigation will be extended further to deploy a technique capable of providing an insight into the lack of confounding effects that exist among the climate factors on the incidence of malaria since all contribute. These are called the hidden ecological factors of malaria's incidence.

The fundamental concept behind the hidden factors emanated from the fact that a causal relationship exists among the climate factors [39]. Some recent studies [91, 120] combined meteorological variables with data on the incidence of malaria and established time series models to predict malaria's incidence. The regression and correlation analysis modelling was applied and meteorological variables used to determine the trend of malaria's incidence [118]. In section 5.3, the partial least squares path modelling (PLS-PM) [160] methodology was deployed to analyse the causal relationships among the meteorological variables, e.g., the minimum average temperature, maximum average temperature, relative humidity, wind speed, precipitation and solar radiation, and explored their impact on the incidence of malaria. By doing so, an integrated model will be develop to provides an insight within which the lacking predetermined confounding effects can be identified as hidden ecological factors. Hence, understanding the dominant climate factors influencing malaria's transmission will enable the inclusion of variables that contribute significantly to the predictive model.

This requires the use of sufficient data. Unfortunately, most of these data were incomplete, particularly for the climate factors. The data have to be completed for the climate variables by using a satellite-based meteorological database, CFSR (Climate Forecast System Reanalysis) before deploying a predictive model technique.

In section 5.4, the machine learning methods was used to identify a pattern or model that will be used to make an accurate prediction of a malaria outbreak. The prediction precisions of the methods was evaluated, and obtained a very high accuracy rate. Machine learning has been used for the prediction and diagnosis of several diseases, e.g., Parkinson's [155], cancer [69] and heart disease [108]. Among the machine learning methods, Support Vector Machines [170] have been used to predict malaria's incidence [149], although this study has several shortcomings:

i. The dataset used was extremely small (the size was only 60), which makes the accuracy of prediction questionable;

ii. The dataset was used without analysing the ecological factors, which could result in including statistically insignificant variables in the prediction model, and hence could cause overfitting;

iii. There is no systematic methodology to transform this predictor into a smart healthcare system.

The machine learning methods to predict malaria's incidence was developed in section 5.4, hence this methods was extended to deploy *an intelligent malaria outbreak early warning system*, which is a mobile application that predicts malaria outbreaks based on climatic factors using the algorithms. The system will help hospitals, healthcare providers, health organisations to take precautions in time and utilise their

resources in the case of emergencies. Section 5.5, presents the mobile application by embedding the best predictor generated in sections 5.2 and 5.3. The application reads climatic information, i.e., temperature, relative humidity, wind speed, solar radiation and precipitation, from free weather and geographical APIs. It then predicts the possibility of malaria's outbreak several days in advance (based on the available forecasting data). As well as deploying a smart healthcare application, Chapter 5 offers a remarkable contribution by identifying the hidden ecological factors of malaria (e.g., temperature, humidity, wind, location, drought, floods, etc.). Since the confounding effects of climatic factors have a greater influence in the incidence of malaria, hence, further research was conducted on a new ecosystem model for the assessment of hidden ecological factors and identified three confounding factors that significantly contribute to the outbreak of malaria.

5.2 Data Analytics of Climate Factors Influence

Despite the apparent correlation between climate factors and malaria's prevalence, the exact contribution of at least one particular factor that significantly influences its transmission is still not fully understood. However, efforts have been made by deploying time series modelling to investigate the trend of malaria's occurrence driven by climate factors.

For instance, the works by [25, 29] use a time series analysis to determine trends in the reported malaria cases and deaths by considering the incidence data for Ethiopia and Gabon. Furthermore, studies [91, 120, 185] used meteorological and malaria incidence data, and predicted future causes of these incidences. However, the works presented [25, 29] applied univariate trend analysis on malaria incidence data and pre-

dicted the future causes of malaria, while [185] used support vector regression and random forests and compared their prediction capability from malaria incidence and climate data. However, the work presented [91, 120] studied the physical influence of malaria's incidence and its climatic predictors, and predicts the future incidence. All studies mentioned in this section presented good results for understanding the pattern of malaria and its prediction using time series and machine learning, respectively. However, malaria's incidence has lagged effects with climate factors, and has not been considered in the modelling. In section 5.2, an improvement is propose to previous works [25, 29, 91, 120] by deploying regression analysis with a times series structure in the stochastic term (white noise). This aims to incorporate climate factors with cases of malaria reported to develop a predictive model that accounts for the lag effect. Hence, the model will help to identify the particular climate factors that significantly contribute to a high malaria incidence in a given geographical area. To proceed with this modelling, monthly malaria incidence cases and climatic data obtained from the Aboh Mbaise region of Imo State–Nigeria, which has tropical and rain forest climate characteristics, will be use.

5.2.1 Materials and methods

This comprises of the data source, study area, data condensation techniques and overview of disease model cradle (DMC).

5.2.1.1 Study area

The study area lies within Latitudes $5^{\circ}\ 10^{\prime}$ N and $5^{\circ}\ 51^{\prime}$ N and with Longitudes $6^{\circ}\ 15^{\prime}$ E and $7^{\circ}\ 28^{\prime}$ E, occupying a land area of 184 km^2 [52]. The Aboh Mbaise region is

one of 27 local government areas of Imo State Nigeria. Its community lives within a 15 km radius from the local government headquarters, with rain forest climate Characteristics, an average annual temperature above 20°C (68° F) and an annual relative humidity of 75% [64]. The rainy season begins in April and lasts until October with an annual rainfall varying from 1,500 mm to 2,200 mm (60 to 80 inches). The dry season experiences two months of Harmattan, from late December to late February, and the hottest months are between January and March [9].

5.2.1.2 Data collection

The data used in this study were extracted from the data presented in existing literature [56], and originally collected from Aboh Mbaise General Hospital. A total of 2,148 confirmed diagnosed cases of malaria were collected from a period of 18 years, from January 1996 to December 2013. In order to have an insight into the possible climatic causes of a high malaria incidence in the study area, a meteorological data will be use. However, these data were not readily available at the time needed, within the reference weather station of Aboh Mbaise. As an alternative, a satellite meteorological database was explored through: http://globalweather.tamu.edu/. The boundary metrics used in generating the data were Latitude 5.6556° N to 5.3494° S and Longitude 5.6291° W to 6.3776° E. Within the demarcation of the study area, one weather station was found. A daily minimum temperature, maximum temperature, precipitation and relative humidity for the period of study was generated. The daily climate variable time series dimension was very large (6575) compared to the monthly malaria incidence data, which had fewer dimensions (216). To ease the analysis, the daily recorded climate data dimension was convered to a monthly time series using the code *ts()* function in *R*, in order to pair the dimension with the monthly malaria

incidence data.

5.2.1.3 Data preprocessing

The methods of data analysis include: descriptive statistics, *cross-correlation*, pre-whitening and regression time series analysis, respectively. To determine the lagged effect of the meteorological predictor variables on malaria's incidence, we used a *cross-correlation* analysis (see Figure 5.2) together with a pre-whitening (see Figure 5.3) scheme to investigate the most significant predictor of malaria's incidence. The *auto-correlation function* (ACF) and *partial autocorrelation function* (PACF) were used to identify the lag order of the time series models.

In Figure 5.1, the cases of malaria reported in the study area was depicted, while the pattern of climate factor distribution, particularly of precipitation, humidity and temperature, are presented in **Appendices** D.1, D.2 and D.3 respectively.

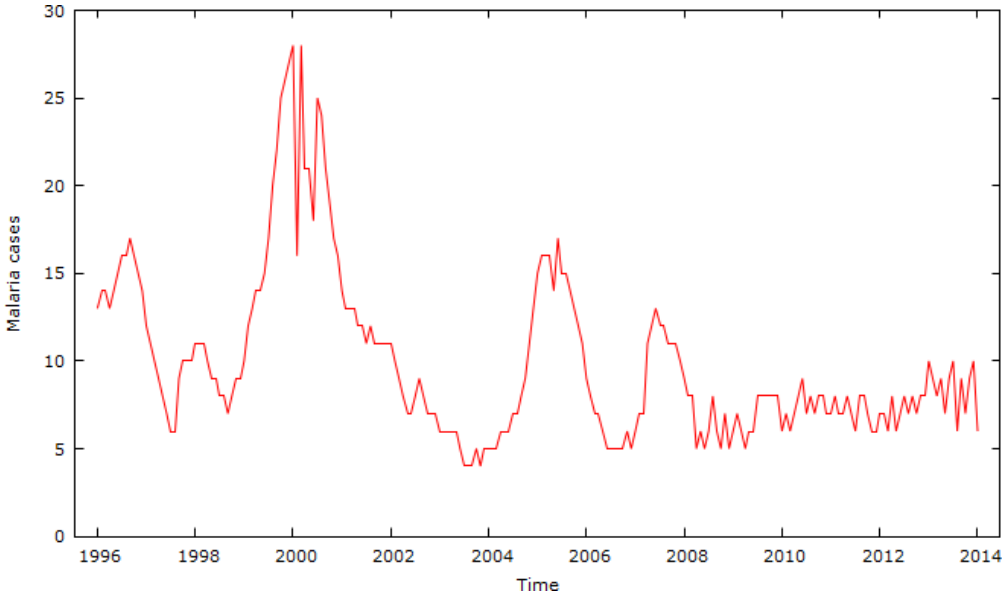

Figure 5.1: The plot shows the pattern of monthly malaria distribution for the study area (Aboh Mbaise, Imo State Nigeria).

The summary statistics of malaria incidence cases and the climate predictors used in this section are presented in Table 5.1. Hence, this summary will enable to develop an understanding of the distributional pattern in the data components, such as the variability, mean, minimum and maximum values. Moreover, the summary statistics provide quick information, like data seasonality, irregular or even deterministic patterns in the occurrences of the general time series.

Table 5.1: The summary statistics for the monthly reported cases of malaria and its climate predictors.

Variables	Mean	Standard deviation	Minimum value	Maximum value
Malaria Data	9.94	4.80	4.00	28.00
Temperature	31.29	3.73	23.66	39.98
Relative humidity	7.92	9.53	0.00	73.31
Precipitation	3.00	1.00	2.00	4.00

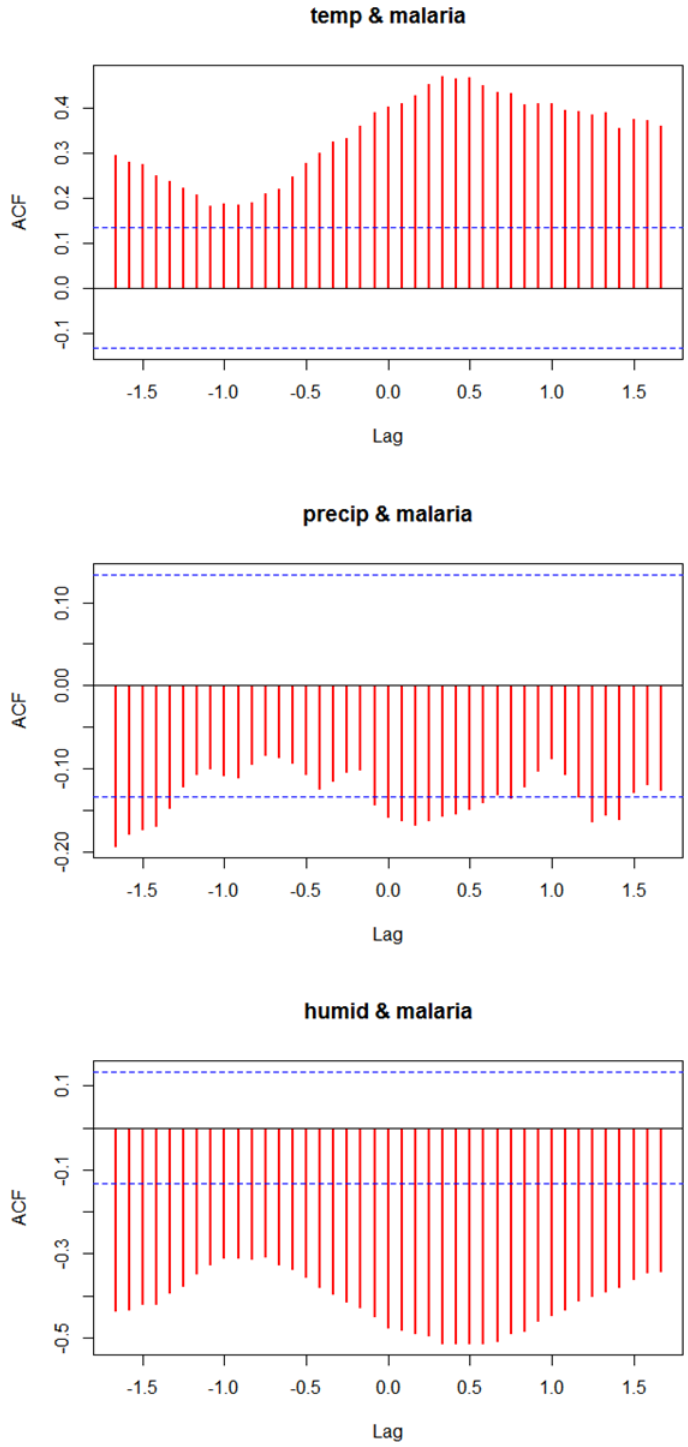

Figure 5.2: The plots show the *cross-correlation function* between temperature and malaria incidence, precipitation, respectively.

5.2.1.4 Overview of Disease model cradle (DMC)

Forecasting malaria's incidence requires not only the use of incidence data and recorded climate information, but also the investigation and understanding of the transmission at the micro-scale level.

This can better achieved by using the Disease Model Cradle (DMC) [9, 101], which uses a daily times series of temperature and precipitation, and explicitly simulates the *gonotrophic* cycle, *sporogonic* cycle and interaction between mosquitoes and humans. This is epidemiological software designed to investigate and validate results with respect to field measurements, such as malaria incidence and the number of infected mosquitoes, using the Liverpool Malaria Model (LMM) together with meteorological datasets. The DMC interface only provides space for temperature and precipitation, and explicitly simulates the pattern of malaria's incidence including micro-scale modelling [101].

In section 5.2, as a benchmark, the meteorological datasets, temperature and precipitation for Aboh Mbaise as an input into the DMC software and simulated the outputs of malaria's incidence and prevalence, its *sporogonic* cycle, and its *gonotrophic* cycle.

A key element of the DMC is the temperature-dependent mosquitoes' survival options. The potential candidate for malaria's transmission, namely the adult female mosquito, has three survival options within which it survives under the influence of a temperature regime. The first survival option is called the Martens scheme [65, 106] in which the daily survival probability (P) is linked to the temperature (T) as captured by the following second-order polynomial equation:

$$P(T) = -0.0016T^2 + 0.054T + 0.45. \tag{5.1}$$

The second survival option is called the Lindsay and Birlay scheme [96], which uses a fixed probability per *gonotrophic* cycle as:

$$P = \exp\left(\frac{P_{lb}}{T_g}\right) \tag{5.2}$$

where: P_{lb} is the survival per cycle, and T_g is the length of *gonotrophic* cycle. The third survival option is called the Craig scheme [55], which links the survival probability (P) with an exponential function of the temperature (T) as [106]

$$P = \exp\left(\frac{-1}{-4.4 + 1.31T - 0.03T^2}\right) . \tag{5.3}$$

5.2.2 Regression Analysis

As previous studies [65, 66, 101] have shown that the climate predictors of malaria, particularly temperature and precipitation, have lagged effects on the occurrence of malaria transmission, while the humidity does not have that effects. Another study [187] showed that humidity may have an impact on malaria transmission as it supports the vector (mosquito) by providing a suitable atmosphere to survive longer and increases biting competence. These different observations suggest that the strength of each climate factor may vary from one geographical area to another. In the context of the study area, the significant (lagged and non-lagged) contributing factors was determined by invoking a *cross-correlation function* (CCF) as follows.

5.2.2.1 *Cross-Correlation Function* (CCF)

CCF can be considered a useful tool for determining the most influential climatic variable to predict the occurrence of malaria and is mathematically expressed in [64], and

is given by

$$c_{uy}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (u_t - \bar{u})(y_{t+k} - \bar{y}) \tag{5.4}$$

for $k = 0, 1, \cdots, (N-1)$

$$c_{uy}(k) = \frac{1}{N} \sum_{t=1-k}^{N-k} (u_t - \bar{u})(y_{t+k} - \bar{y}) \tag{5.5}$$

for $k = -1, -2, \cdots, -(N-k)$ as the *product-moment correlation* of a time-offset or a function of lag between two time series u_t and y_t. Herein N is the series length, $\bar{u}$ and $\bar{y}$ are the sample means, and k is the CCF lag. Hence the *cross-correlation* function is also an *auto-covariance function* when scaled by the variances of the two series as:

$$r_{uy}(k) = \frac{c_{uy}(k)}{\sqrt{c_{uu}(0)c_{yy}(0)}} \tag{5.6}$$

where $c_{uu}(0)$ and $c_{yy}(0)$ are the sample variances of u_t and y_t, respectively.

Between the climatic variables and malaria incidence data, CCF is depicted in Figure 5.2, which aims to help identify the lags of the climatic predictors that might be a useful predictor of malaria's incidence. However, the three plots show a pattern in which it is difficult to identify any lagged effects of the climate predictor of malaria. This difficulty happens due to the fact that CCF values are sometimes affected by the time series structure of the independent variable against the dependent variable series over time.

5.2.2.2 CCF with Pre-Whitened Climatic Data

In order to alleviate the difficulty of identifying the significant lags of the climatic predictors in Figure 5.2, a pre-whitening technique was invoked in order to stationarize

the climatic input variables.

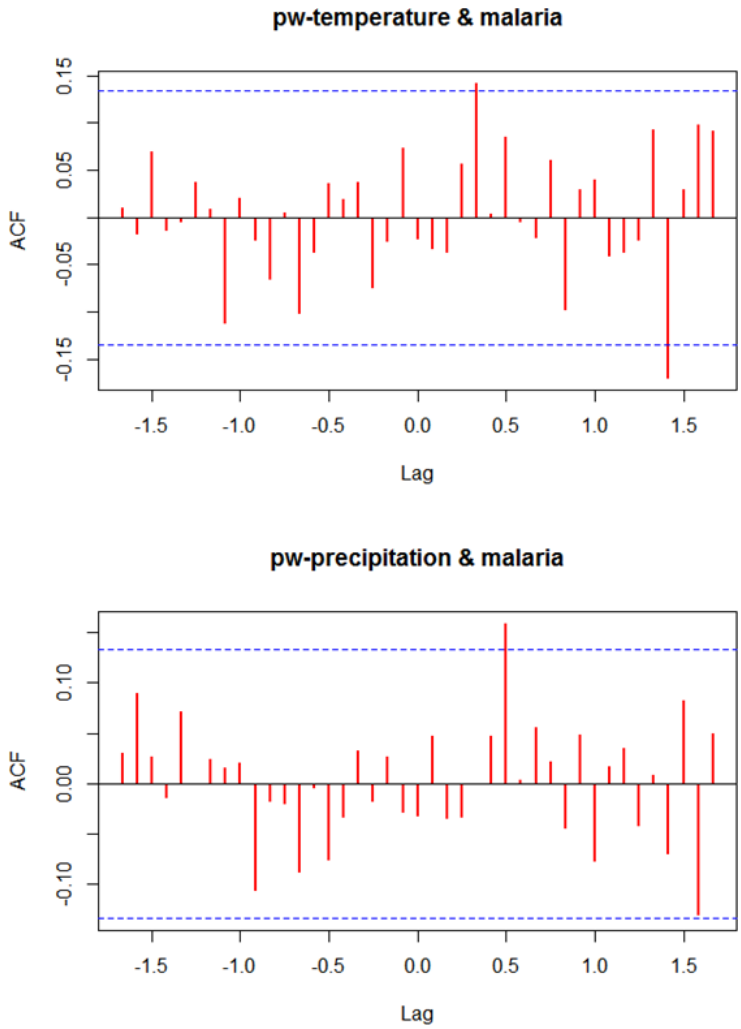

Figure 5.3: The plots indicate pre-whitened lagged correlations between temperature with malaria incidence and precipitation with malaria incidence.

When the input series behaves like white noise, and the pattern of the CCF between climate variables and malaria incidence is a linear combination of lags of the input variables. The incorporation of pre-whitening for the time series of interest involved the following steps. Firstly, the time series model for the malaria predictor variables was determined and stored the residuals from the model. Secondly, the malaria incidence variable was filtered using the model of predictors. Thirdly, the CCF between residuals of the predictors was examined and the filtered values. Hence, the resultant CCF can be used to identify the possible pattern for lagged effects that would be used in a regression model.

Using an autocorrelation function (ACF) and the partial ACF (PACF) of climatic-predictors after the differencing operation, the time series models for temperature and precipitation data was estimated and tabulated in Table 5.2. The resulting time-series model of temperature is estimated to be ARIMA (1,1,0), which is an autoregressive moving average model for order 1, with a differencing order of 1. The estimated coefficient parameter for this ARIMA (1,1,0) model is given by $\hat{\theta} = -0.343$, which leads to the estimated model:

$$\Phi_T(B)y_t = E_t \tag{5.7}$$

where

$$\Phi_T(B) = (1 - 0.657B - 0.343B^2). \tag{5.8}$$

Similarly, aided by the plots of ACF and PACF after the differencing operation, the precipitation time series is approximated as the ARIMA (2,1,0) model, which can be represented by

$$\Phi_P(B)z_t = E_t \tag{5.9}$$

where

$$\Phi_P(B) = (1 - 0.6709B - 0.0536B^2 - 0.2755B^3). \quad (5.10)$$

The time series defined by the polynomial Φ_T (B) and Φ_P (B) have been found stationary by inspection, which confirm that the temperature and precipitation time series become stationary after taking the first differences.

Once the time series models for temperature and precipitation were determined, the next stage was the filtering process, which involved filtering the malaria incidence data using the aforementioned time series models. Hence, continued by examining the CCF between the residuals from the time series models for the climatic input variables and the filtered malaria incidence to identify the significant lagged terms of the regression model. From the pre-whitened CCF plots in Figure 5.3 of the temperature time series with malaria incidence and precipitation time series with malaria's incidence. It is observe that the most significant spike in both plots appeared on the positive lag segments of the *cross-correlation function* indicating overlapping effects. Therefore, this shows evidence that there is no significant lagged effect of temperature and precipitation to effectively predict the occurrence of malaria in the study area.

Table 5.2: Pre-whitened models for the temperature and precipitation time series.

Pre-whitened model	Estimates	Standard Error
ARIMA(1,1,0)	-0.343	0.064
Log-likelihood	-473.65	
AIC	951.3	
ARIMA(2,1,0)	-0.3291	0.0654
	-0.2755	0.0652
Log-likelihood	-784.08	
AIC	1574.14	

5.2.2.3 Regression Model

By taking into account the three climate predictors (temperature, precipitation and relative humidity), a regression model that predicts the pattern of malaria incidence occurring in Aboh Mbaise is proposed as:

$$MI_t = \beta_o + \beta_1 f(TM_{t-1} TM_{t-2}) + \beta_2 g(PR_{(t-1)}) + \beta_3 RH + E_t \quad (5.11)$$

where MI denotes the incidence of malaria as the output of the model, and TM, PR and RH denote temperature, precipitation and relative humidity, respectively as the inputs of the model. Herein β_1, β_2 and β_3 are the coefficients of the regressed variables, whilst β_o is intercept, and E_t is error the term.

From the results of the pre-whitening analysis of the predictor variables presented in Table 5.2, it was found that temperature and precipitation have insignificant lagged effects on malaria's incidence. Based on this, their effects on the model could be dropped to avoid redundancy. Hence, the regression model can be compactly expressed as:

$$MI_t = \beta_o + \beta_3 RH_t + E_t \quad (5.12)$$

Using the least squares method, the estimated model as follows:

$$MI_t = 31.212 + 25.467RH_t \tag{5.13}$$

In regression modelling that involved time series variables, the residuals component was assumed to have a time series model, for example., AR or MA, meaning autoregressive or moving average [44]. However, using the conventional approach (least squares method), such an assumption of independent error is violated. Hence, the consequence is the wrong coefficient estimate and standard errors if the time series structure of the errors is ignored. Hence, the model regression coefficients and standard errors have to be adjusted in order to ensure a well-fit model, including the component of the AR error structure.

Suppose the following equation to illustrate the adjustment of regression coefficients and standard errors for a simple case and use it to advance our analysis

$$y_t = \beta_o + \beta_1 x_t + E_t \tag{5.14}$$

where the autoregressive structure of error is captured by

$$E_t = \vartheta_1 E_{t-1} + \vartheta_2 E_{t-2} + \ldots + \omega_t \tag{5.15}$$

where $\omega_t \sim$ i.i.d. $N(0, \sigma^2)$. Then, let

$$\phi(L) = 1 - \vartheta_1 BL - \vartheta_2 B^2 - \ldots, \tag{5.16}$$

denotes the AR model for the errors, which reduces to $\phi(L)E_t = \omega_t$. Assuming the in-

verse of $\phi(L)$ exists, hence $E_t = \phi^{-1}(L)\omega_t$, where $\omega_t \sim$ i.i.d. $N(0, \sigma^2)$. Then, equation (5.14) is given by:

$$y_t = \beta_o + \beta_1 x_t + \phi^{-1}(L)\omega_t \tag{5.17}$$

such that $\phi(L)$ gives the AR polynomial for the errors.

By multiply equation (5.17) by $\phi(L)$, then it becomes

$$\phi(L)y_t = \phi(L)\beta_o + \beta_1\phi(L)x_t + \omega_t. \tag{5.18}$$

Let

$$y_t^* = \phi(L)y_t = y_t - \vartheta_1 y_{t-1} - \ldots - \vartheta_p y_{t-p} \tag{5.19}$$

$$x_t^* = \phi(L)x_t = x_t - \vartheta_1 x_{t-1} - \ldots - \vartheta_p x_{t-p} \tag{5.20}$$

$$\beta_o^* = \phi(L)\beta_o = (1 - \vartheta_1 - \ldots - \vartheta_p)\beta_o. \tag{5.21}$$

The model can written as:

$$y_t^* = \beta_o^* + \beta_1 x_t^* + \omega_t \tag{5.22}$$

where $\omega_t \sim$ i.i.d. $N(0, \sigma^2)$. Note that the unknown constant β_o in equation (5.17) does not depend on the time and is also independent of the shifting operation. Then β_o can be approximated by

$$\hat{\beta}_o = \frac{\hat{\beta}_o^*}{1 - \hat{\vartheta}_1 - \ldots - \hat{\vartheta}_p}. \tag{5.23}$$

Similarly, the standard error for $\hat{\beta}_o$ is given as

$$s.e(\hat{\beta}_o) = \frac{s.e(\hat{\beta}^*_o)}{1 - \hat{\vartheta}_1 - \ldots - \hat{\vartheta}_p} \tag{5.24}$$

Reference [53] developed this procedure, and the process is repeated until the estimates converge. Following the methodology of [53], the adjustment of the regression model for predicting malaria's incidence can be presented as:

$$y_t^* = -0.0896 - 0.8902 y_{t-1}^* \tag{5.25}$$

$$x_t^* = -0.0896 - 0.8902 x_{t-1}^* \tag{5.26}$$

Thus, the estimated relationship between malaria's incidence and the relative humidity in the study area is given by

$$MI_t = 30.095 + 25.387 RH_t + E_t, \tag{5.27}$$

where the error term can be expressed as

$$E_t = 0.9236 E_{t-1} + \omega_t, \tag{5.28}$$

where $\omega_t \sim$ i.i.d. $N(0, 3.773)$. The model accuracies are presented in Table 5.3, for both the model without the AR structure errors and the adjusted model with the AR structure errors.

Table 5.3: Model Accuracy

Model	RMSE	MAE	MAPE	MASE
Without AR errors	4.2100	2.9664	32.5958	0.8158
With AR errors	3.7556	2.6482	32.2130	0.8176

The statistical significance of three climatic factors, namely temperature, precipitation and relative humidity is examined by pre-whitening the climatic explanatory data and performing a *cross-correlation* analysis with the malaria incidence data. Among these three factors, the relative humidity was found statistically significant and associated (at probability, $p_{value} = 0.0002 < 0.05$) with the malaria incidence, whereas temperature and precipitation have a negligible lagged correlation with the incidence.

A linear regression with an autoregressive error structure AR(1) is then developed to precisely specify the relationship between the incidence and relative humidity time series. This finding contrasts with some previous results [65, 66, 101] that highlight the lagged contributions of temperature and precipitation, and suggest variability in the strength of the climate factors affecting malaria's incidence in different geographical areas. This finding together with references [74, 156] can further motivate improvements to the existing physical models of malaria-risk prediction, such as DMC [101], by incorporating comprehensive climate variables. Since the regression analysis with a time series structure in the error term results in identifying humidity as the dominants climate predictor of malaria (in subsection 5.2.1.1). However, this factor cannot occur independently without the contribution of other factors, like temperature and precipitation. Therefore, in Section 5.3, a new methodology will be deployed to further investigate the existence of a causal relationship among the climate predictors of malaria.

5.3 Assessment of hidden ecological factors

Climate factors are the drivers of malaria transmission [134]; however a study analysing the causal ecological relationship among the climate factors that affect the incidence of malaria is still lacking, particularly in Sub-Saharan African countries.

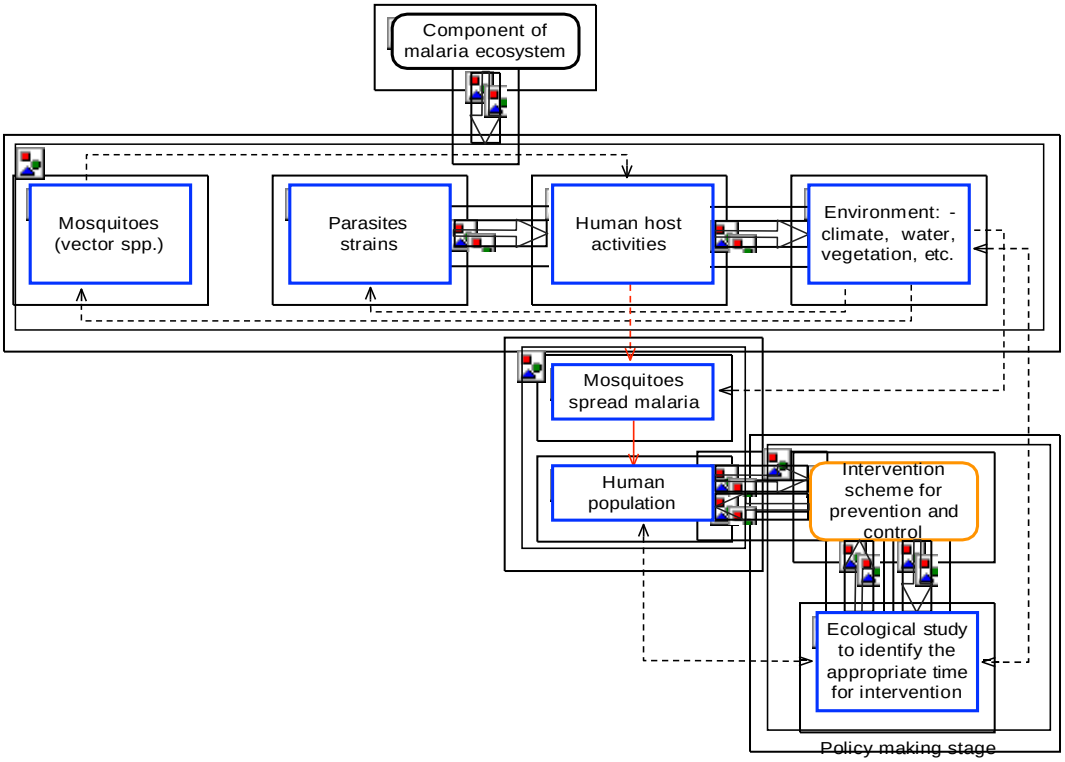

Figure 5.4: Conceptual framework of a malaria ecosystem describing the dynamic stages of malaria transmission from human and mosquito under the influence of environmental factors.

The malaria ecosystem comprises four main components, namely human host, mosquitoes vector, parasites and environmental condition (see Figure 5.4). These components are very dynamic in nature due to the inherent characteristics of ecology and the anticipatory change to biodiversity from global warming. Studies [34, 114, 115] reported that ecological changes would adversely affect human health in some ways that are both obvious and obscure. However, growing evidence also suggests that, due to the rise in temperature as a result of the anticipated global warming, some previously unexposed regions would have a 50% chance of malaria transmission; this is attributed to the link between malaria's incidence and ecological factors [68]. The relationship between environmental changes and human health cannot be overemphasised because of the inherent variability and complexity of human nature. In many circumstances, grasslands and forest are converted for agriculture to reduce communicable disease, which includes wetland drainage for the prevention and control of malaria [34]. These activities can either succeed in their designed purpose or lead to unintended negative health effects. Also, transforming forests to augment food production, may in the long run lead to the creation of environments conducive to disease-causing agents, such as mosquitoes for malaria transmission [58].

5.3.1 Study site and population

Ejisu-Juaben Municipal has a population of 143,762 [127]; it lies within latitudes $1^0 15^{|}$N and $1^0 45^{|}$N also with longitudes $6^0 15^{|}$W and $7^0 00^{|}$W, and occupies a land area of 582.5 km^2 [146]. The vegetation of the municipal is a typical semi-deciduous forest (see Figure 5.5), with an undulating topography and low altitude of about 240m–300m above the sea level [146]. Also, the rainfall pattern of the area is bi-modal (i.e., two dis-

tinct seasons in a year), characterized by major and minor rainfall. The major rainfall begins from March to July with an average annual rainfall between 1200mm–1500mm, while the minor rainfall begins in September and tapers off in November with annual average of 900mm–1120mm. Usually, December through to February is hot, dry and dusty with a mean annual temperature of 25^0C-32^0C, and the relative humidity is moderately high during the rainy season [146]. Figure 5.5 presents the map of Ejisu-Juaben Municipal, which lies within the red-squared portion labelled Kumasi, the capital city of the Ashanti Region in Southern Ghana.

Figure 5.5: The picture on the left shows the map of Ghana and the portion of Kumasi city, where the study area - Ejisu-Juaben - lies. The picture on the right illustrates the climate vegetational belt characterized by a typical semi-deciduous forest.

5.3.2 Data collection and source

A total of 85,627 confirmed diagnosed cases of malaria for the period of five years from 2009 to 2013, were retrieved [158]. The distributional pattern of malaria cases reported in the study area indicates a high incidence. Although, climate factors data was sought in the designated weather station of the study area location [127], unfor-

tunately very few data were available and also a lot were missing. Since the data available is not sufficient for the analysis, hence, to overcome this challenge by using satellite-based meteorological data obtained via [2]. The boundary metric dimensions of [158] at latitude 6.7989^0N to 6.6823^0S and longitude -1.5656^0W to -1.4186^0E and demarcated the location of the study area on the satellite globe map was used. Within the demarcated area, a weather station was identified. The data of the climate variables of interest was generated. The Ghana malaria incidence data were sufficient for the application of PLS-PM due to their suitability for handling small sample data, non-normality, multi-dimension and multicollinearity [31, 125]. However, the sample set was not sufficient to obtain high precision accuracy when applying machine-learning algorithms. A small dataset might also cause the overfitting of the data. For that reason, the malaria incidence data used in [56] and [158] was combined to proceed the analysis.

5.3.3 Factor analysis

Exploratory factor analysis (EFA) is one of the techniques for factor analysis (FA). It is primarily used in statistics to describe the variance among observed correlated variables in terms of a potential smaller number of unobserved variables, usually referred to as factors [142]. In this study, EFA was employed to search for latent confounding ecological factors [142, 160] from the set of observed meteorological variables.

The FA technique is demonstrated using simple mathematical sketches; the observed variables can be expressed as a linear combinations of the potential factors plus the residual terms. Consider the following observed variables $Y_1, Y_2, \cdots, Y_M$ of size M, and assume they are linearly related to a small number of unobservable (latent

variables) factors $F_1, F_2, \cdots, F_N$, with $N \ll M$ such that:

$$\begin{cases} Y_1 = \psi_{10} + \psi_{11}F_1 + \cdots + \psi_{1N}F_N + e_1 \\ Y_2 = \psi_{20} + \psi_{21}F_1 + \cdots + \psi_{1N}F_N + e_2 \\ \vdots \\ Y_M = \psi_{M0} + \psi_{M1}F_1 + \cdots + \psi_{MN}F_N + e_M \end{cases} \tag{5.29}$$

where $e_1, \cdots, e_M$ are the residual terms, assuming that $E(e_i) = 0$, and $Var(e_i) = \delta_i^2$. While the unobservable factors F_i are independent from each other and $E(F_j) = 0$ and $Var(F_j) = 1$. These two assumptions stand as the robust pre-conditions for the application of structural equation modelling (SEM). The loadings scores can be obtained from covariance and the variance of any two observed variables using the following formula presented in equation (5.30)

$$Cov(Y_i, Y_j) = \sum_{i}^{N} \sum_{j}^{N} \psi_{iN}\psi_{jN} \; Var(Y_i) = \sum_{i=1}^{N} \psi_{iK} + \delta^2_{\;i} \tag{5.30}$$

where the summation sign in equation 5.30 denotes the communality of the variables, and the variance is explained by the common factors F_N.

5.3.4 Structural equation modelling

SEM is a popular technique that has multidisciplinary applications, which combine both measurement and structural models [86, 89, 111]. Figure 5.6a presents a complex hypothetical SEM showing the causal relationship between malaria's incidence and latent ecological factors together with their observed variables. The ellipse shapes is used to represents the latent factors, while the observed variables were represented by

rectangular shapes.

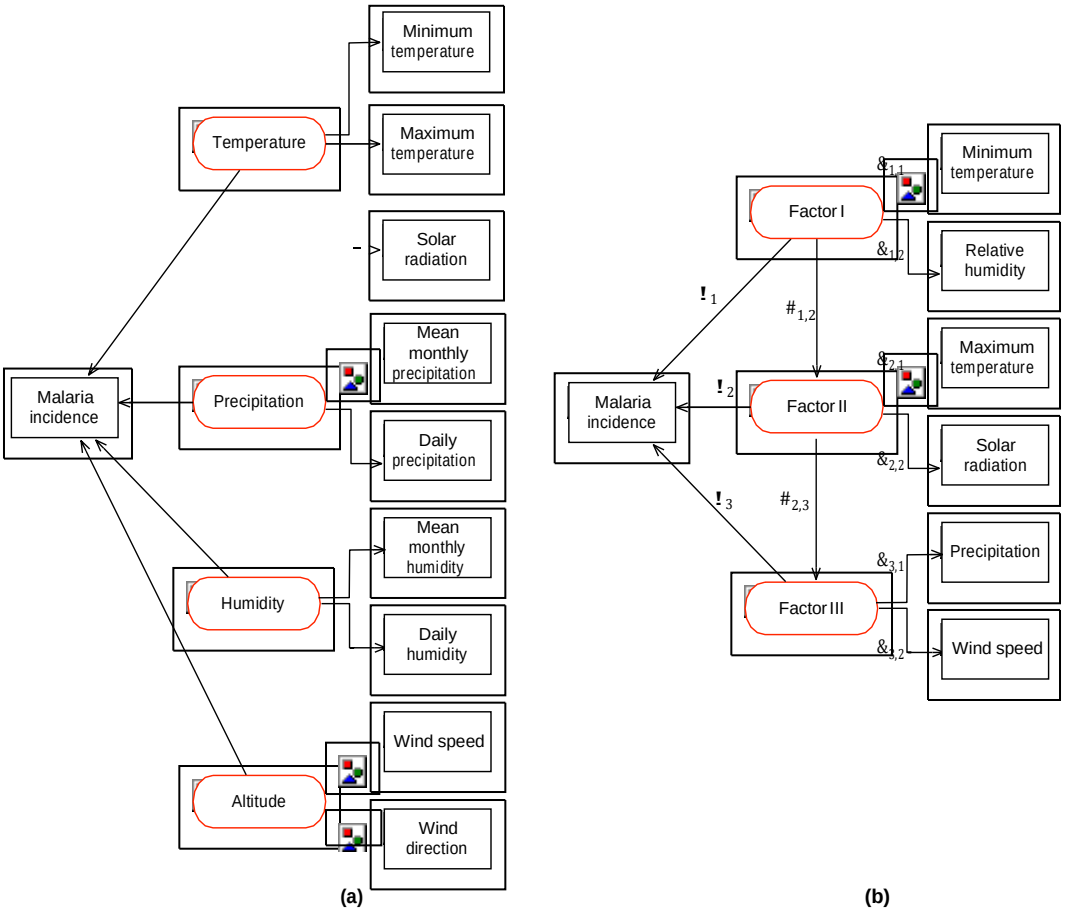

Figure 5.6: Structural equation model showing the hypothetical relationship between malaria's incidence and meteorological variables.

The following system 5.31 describes the SEM technique in which the observed variables can be expressed as a linear combination of the potential factors plus the residual terms. The SEM mathematical representations shown in Figure 5.6(b) as fol-

lows:

$$\begin{aligned}
&\text{Factor I} = \lambda_{1,1}(\text{minimum temperature}) + \\
&\lambda_{1,2}(\text{relative humidity}) + \beta_{1,2}(\text{Factor II}) + \gamma_1(\text{malaria incidence}) + e_1 \\
&\text{Factor II} = \lambda_{2,1}(\text{maximum temperature}) + \\
&\lambda_{2,2}(\text{solar radiation}) + \beta_{2,3}(\text{Factor III}) + \gamma_2(\text{malaria incidence}) + e_2 \\
&\text{Factor III} = \lambda_{3,1}(\text{precipitation}) + \\
&\lambda_{3,2}(\text{wind speed}) + \gamma_3(\text{malaria incidence}) + e_3
\end{aligned} \tag{5.31}$$

5.3.5 Estimation of PLS-PM

The technique called PLS-PM, or PLS-SEM, was developed by [179] and chosen due to its characteristics, namely its small sample size, non-normality, multi-dimension, and multicollinearity [125]. Three hidden factors was identified using EFA, and subsequently applied SEM for the construction of the model (see Figure 5.6(b)). The PLS-PM is basically divided into three components: The estimation of the LVs, the estimation of the inner and outer models, and the estimation of the structural relations. The PLS algorithm is essentially represented as a sequence of regression in terms of weight vectors [61] and estimates the LV values (factor scores) iteratively until a convergence is achieved. The fundamental PLS algorithm was suggested by [111] (see **Appendix** E for the detail procedural descriptions).

5.3.6 Weighting scheme

The weighting schemes are used to estimate the inner weight in (E.2) of the PLS algorithm. Generally, there are three weighting schemes, centroid [70], and later [97]

factorial and path weighting. The PLS-PM is a component-based estimation technique that uses an iteration algorithm, separately analyzes the blocks of the measurement model, and estimates the path coefficients in the structural model [45]. A package called *semPLS* in *R* was used for the estimation of the PLS-SEM parameters including the analysis presented in Table 5.5. To estimate the SEM parameters, the PLS technique was invoked, furthermore, 10,000 samples was used for the bootstrapping analysis instead of the default number of samples (set at 500) [45]. Also, the PLS-PM latent variable scores were expressed as a linear combination of their observed variables and treated as an error-free substitute for the observed variables [45].

5.3.6.1 Measurement model

The model presented in Figure 5.6(b) shows how the observed variables (MVs) are related to their LVs. Hence without any loss of generality, for a good representation of the inner model, the following assumptions must hold:

i. The matrix of MVs $\mathbf{Y}$ are scaled for zero mean and unit variance.

ii. Each block of MVs $\mathbf{Y}_g$ is already transformed for a positive correlation for all LVs $\mathbf{x}_g, g = 1, \cdots, G$.

The measurement model is broadly classified as either reflective (Mode A) or formative (Mode B) [56], which depends on the relationship between the LV and MV formation.

5.3.6.2 Mode A

In this form, each block of MVs reflects its LV and can be represented in a multivariate regression form as:

$$\mathbf{Y}_g = \mathbf{x}_g \mathbf{w}_g^{\mathsf{T}} + \mathbf{F}_g \tag{5.32}$$

where $\mathbf{w}_g^{\mathsf{T}}$ can be estimated using least squares method.

5.3.6.3 Mode B

Also, in this form, it is considered that the LV is formed by its MVs and represented by a multiple regression:

$$\mathbf{x}_g = \mathbf{Y}_g \mathbf{w}_g + \delta_g \tag{5.33}$$

using the same method of least squares, the estimate for $\mathbf{w}_g$ can be obtained.

5.3.7 Presentation of results

In the application of PLS-SEM, three weighting schemes - centroid, factorial and path weighting - are conceptually used for the model specifications and estimations. The conceptual SEM, presented in Figure 5.6a, shows the hypothetical causal relationship between the latent (hidden) variables and observed meteorological (manifest) variables to the occurrence of malaria. For the identification of the confounding hidden variables, a factor analysis was performed using an exploratory factor analysis (EFA) [95]. From the results, three hidden factors were identified, namely: Factor I (related to minimum temperature and relative humidity), Factor II (related to maximum temperature and solar radiation) and Factor III (related to precipitation and wind speed). These factors accounted for 64% of the total variance, and at $\alpha = 5\%$ level of significance, $\chi^2 = 13.91$, $df = 8$, $p_{value} = 0.0841$. This result provides sufficient evidence to explain malaria's incidence in the study area. The Guttman-Kaiser Criterion [183] and Cattel scree plots [94] was explored to determine the number of factors to extract. The result reconfirmed the existence of three hidden ecological factors to the incidence of malaria. In the Guttman-Kaiser Criterion, the eigenvalues 2.71, 1.53, 1.02, 0.82, 0.57,

0.29, 0.05 were computed using the correlation matrix (see Table5.4); however, the rule for extraction is based on the factors whose eigenvalues are greater than unity. Those eigenvalues less than unity are discarded, which then left three eigenvalues indicating the number of factors to be considered. Similarly, the Cattel scree plot presented in Figure 5.7 facilitates the decision regarding the number of factors to retain. By analysing Table 5.4, the scree plot shown in Figure 5.7 was obtained, which represents the relative proportion of variance accounted for by the components. In the scree plot, the eigenvalues of the first three components greater than unity can be seen from the parallel indicator, while the subsequent components below unity also line-up beneath the parallel indicator. However, it is important to evaluate the variance accounted for by a few of the eigenvalues regarded as sufficient so that attention will be paid on them and discard the remaining insufficient factors as noise.

Table 5.4: Correlation matrix of the climate drivers and malaria incidence.

	1	2	3	4	5	6	7
1	1.00	-	-	-	-	-	-
2	0.28	1.00	-	-	-	-	-
3	0.68	0.04	1.00	-	-	-	-
4	-0.21	-0.36	0.22	1.00	-	-	-
5	0.51	-0.24	0.90	0.38	1.00	-	-
6	0.19	0.54	-0.33	-0.10	-0.44	1.00	-
7	-0.16	0.07	0.45	0.17	0.39	0.01	1.00

Note: (1) Malaria incidence, (2) Maximum temperature, (3) Minimum temperature, (4) Precipitation, (5) Relative humidity, (6) Solar radiation and (7) Wind speed.

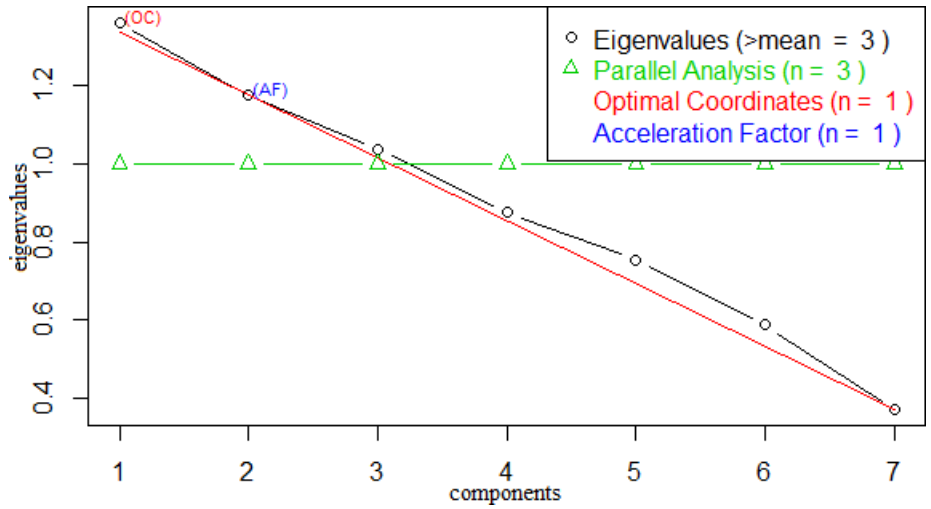

Figure 5.7: The Cattel scree plot presents the eigenvalues of the components and the threshold for identifying the number of hidden ecological factors for consideration using the information in Table 5.4.

Table 5.5 presents *Pearson's cross-correlation* between the meteorological variables and the occurrence of malaria at various lag effects from 0 to 3 months. Lag0, Lag1 and Lag2 (e.g., 0 month, 1 month and 2 month) presented in Table 5.5 indicates the lagged correlation effects between the climates variables and the incidence of malaria in the study area. It was observed that, with the exception of precipitation, the maximum temperature, minimum temperature and relative humidity have positive

Table 5.5: *Cross-correlation* between meteorological variables and malaria incidence.

Variables	Lag 0	Lag 1	Lag 2	VIF	Kurtosis	Standard error
Maximum temperature	0.284	0.321[b]	0.092	2.4096	5.48	0.38
Minimum temperature	-0.122	0.215[b]	-0.237	8.7919	2.07	0.33
Precipitation	-0.214	-0.292[a]	-0.155	1.4194	20.73	0.27
Relative humidity	-0.134	0.254[b]	-0.198	9.0065	1.42	0.02
Solar radiation	-	-	-	1.9000	6.73	0.50
Wind speed	-	-	-	1.3452	-0.58	0.04

[a] negative association at lag 1.
[b] positive association at lag 1.

lag effects in 1 month, which are given as 0.321, 0.215 and 0.254, respectively. This explained that the influence of climate drivers at lags of 1 month would result in sufficient mosquitoes for reproductive capability and to complete their incubation periods (EIP) to become fully active in transmitting the malaria infection. It was found that the preceding result was consistent with other relevant studies on the influence of meteorological variables on the incidence of malaria [181]. The 1 month time lag in the study area is sufficient to capture the pattern of malarial transmission for various strains of *plasmodium* parasites with definite lengths of EIP. This period usually takes about 10-15 days [154] and temporally varies over location, parasite species and climatic resolution. At Lag0 and Lag2, the minimum temperature, precipitation and relative humidity had negative lag effects at 0 month and 2 months, whilst the maximum temperature was 0.284 and 0.092. These results revealed some clear indications that the transmission of malaria in the study area at Lag0 and Lag2 suffered a negative effect, which might be attributed due to the bi-annual rainfall pattern, low relative humidity (say less than 50%) and the inability of mosquitoes to complete the EIP cycle. In general, the result showed that the maximum temperature, minimum temperature, and relative humidity were related to the incidence of malaria at lagged effects of 1 month (i.e. a month in advance) except for precipitation, which had a negative association in

the study area.

Some important summary statistics are presented in Table 5.5, which describes the distributional pattern of the climate indicators of malaria's incidence and the variance inflated factor (VIF). In the factor analysis, multicollinearity can be used as a diagnostics check prior to the application of a regression analysis, whereby variables with high factor loadings are typically multicollinear. The VIF of the climate's variable was computed to measure the degrees of multicollinearity that exist and identify the factors that are independent by the magnitude of their VIF. In Table 5.5, the minimum temperature and relative humidity have a VIF of 8.7919 and 9.0065, which gives a high degree of multicollinearity. The results revealed the highly independent predictor of malaria's incidence in the study area, and the degree of independence provides evidence to accurately be the major factors. However, the kurtosis values (see Table 5.5) indicate a high peak amongst the climate variables with positive values across most indicators, excepting wind speed, which indicates a flat distribution. Positive values are generally listed in Table 5.5, and indicate a peaked distribution amongst the climate variables, which particularly influences the incidence of malaria. Also, the standard error estimates provide information on the accuracy of the statistics of climate variables; thus, the larger the standard error, the wider the confidence interval about the statistic and vice-versa.

The non-normality of the dataset is one necessity when adopting PLS-SEM; moreover, it is very robust when used on extremely non-normal data [77]. The degree to which the data on malaria's incidence were non-normal was examined by using the Shapiro-Wilk tests, this was implemented in *R* software. The result showed that the null hypothesis (Ho) was rejected indicating the malaria incidence dataset is non-normal, as suggested by the following indices W = 0.9486, p_{value} = 0.0134. This

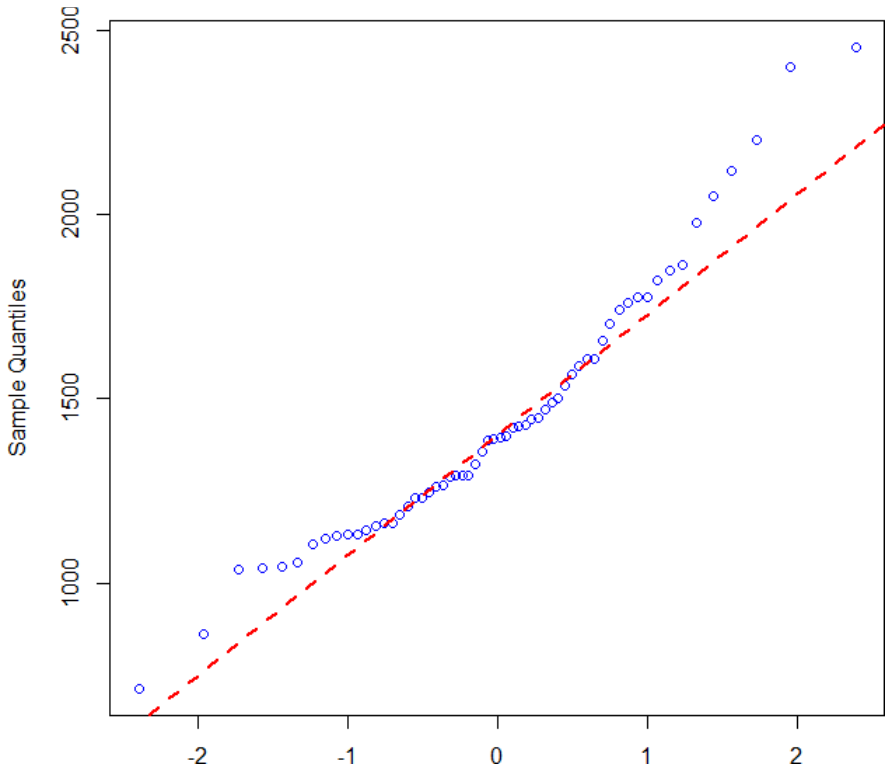

Figure 5.8: Graphical representation of Q-Q plot normality tests.

method is particularly useful and therefore chosen for smaller samples size of less than 2000 [84], and the null hypothesis is explained by the fact that the data are from a normal distribution. Similarly, a graphical approach called quantile-quantile (Q-Q) plot [177] was further used for testing the normality of the dataset. This approach creates a plot from the ranked samples of the dataset against a similar number of ranked theoretical samples from a normal distribution. The plot shown in Figure 5.8, clearly indicates that the data points for malaria's incidence deviate from the straight line. Hence, the malarial incidence dataset is therefore not normally distributed (using Shapiro-Wilk tests) as well as using the Q-Q plot.

Table 5.6: Factor scores for the path coefficients in the PLS-PM using three weighting schemes.

Measurement / structural model	Parameter	Estimate	Centroid (A)	Factorial (B)	Path weighting (C)
Minimum temperature ←− FactorI	$\lambda_{1,1}$	0.9479	0.9479	0.9495	0.9495
Relative humidity ←− FactorI	$\lambda_{1,2}$	0.9910	0.9910	0.9903	0.9903
Maximum temperature ←− FactorII	$\lambda_{2,1}$	0.8816	0.8816	0.8675	0.8675
Solar radiation ←− FactorII	$\lambda_{2,2}$	0.8735	0.8735	0.8873	0.8873
Precipitation ←− FactorIII	$\lambda_{3,1}$	0.9849	0.9849	0.9852	0.9852
Wind speed ←− FactorIII	$\lambda_{3,2}$	0.0017	0.0017	0.0031	0.0031
FactorI −→ FactorII	$\beta_{1,2}$	-0.3248	-0.3248	-0.3302	-0.3302
FactorII −→ FactorIII	$\beta_{2,3}$	-0.2774	-0.2774	-0.2690	-0.2690
FactorI −→ Malaria incidence	γ_1	0.9700	-	-	-
FactorII −→ Malaria incidence	γ_2	0.7700	-	-	-
FactorIII −→ Malaria incidence	γ_3	0.4900	-	-	-
Maximum number of iterations	-	-	12	15	15

In Table 5.6, the results of the factors score estimates for the path coefficients of SEM estimated using PLS path modelling alongside three different structural model weighting schemes are presented. It was observed that Centroid (A) converges faster after 12 iterations, while factorial (B) and path weighting (C) converge after 15 iterations. The best weighting scheme was determined by the maximum number of iterations that will be used to calculate the PLS results and this algorithm did not stop until the maximum number of iterations were reached due to the stop criterion. From Table 5.6, it was observed that the B and C weighting schemes converged at the same maximum number of iterations by estimating the SEM parameters. The weighting scheme provides the highest R^2 value for endogenous latent variables in the PLS path model specifications and estimations. These results show that C weighting scheme is better than A and B as suggested by [97], in terms of its robustness and when the path model includes higher-order constructs.

Table 5.7: Bootstrapping test of outer loadings and path coefficients in the PLS-PM with 95% confidence interval

Measurement / structural model	Parameter	Estimate	Bias	Standard error	Lower	Upper
Minimum ←− *FactorI*	$\lambda_{1,1}$	0.9479	-0.0057	0.0467	0.8240	0.9890
Relative humidity ←− *FactorI*	$\lambda_{1,2}$	0.9910	-0.0055	0.0347	0.9823	1.0000
Maximum temperature ←− *FactorII*	$\lambda_{2,1}$	0.8816	-0.0329	0.1289	0.4769	0.9810
Solar radiation ←− *FactorII*	$\lambda_{2,2}$	0.8735	-0.0343	0.1748	-0.0705	0.9550
Precipitation ←− *FactorIII*	$\lambda_{3,1}$	0.9849	-0.1748	0.4044	0.7666	1.0000
Wind speed ←− *FactorIII*	$\lambda_{3,2}$	0.0017	0.1356	0.4059	-0.6593	0.7300
FactorI −→ *FactorII*	$\beta_{1,2}$	-0.3248	-0.0333	0.1692	-0.4974	0.4260
FactorII −→ *FactorIII*	$\beta_{2,3}$	-0.2774	-0.0264	0.2191	-0.4963	0.3810

Table 5.7 presents the results of the bootstrap sampling for the outer loadings of the observed variables and path coefficient of the latent variables; these were estimated using PLS-PM. The results also show that all outer loadings and path coefficients are significant at α = 5%, except for the solar radiation with *Factor II* and wind speed with *Factor III*, which contain zero-points in the bootstrap confidence interval. Furthermore, the interaction effects of the *Factors* (between I and II, II and III) were also investigated and the results revealed that none of the Factor combinations were significant to the incidence of malaria in the study area. These results provide sufficient evidence that high malaria incidence in the study area was attributed to the occurrence of minimum temperatures and relative humidity, which were identified as *Factor I*.

The decision to select the most influential hidden ecological factor on the incidence of malaria was based on the communality and Dillon-Goldstein's indices. Furthermore, Table 5.8 summarizes the results indicating some indices for selecting the hidden ecological factors to the high incidence of malaria in the study area. Among the three factors identified by EFA, we find that *Factor I*, indicated by a minimum temperature and relative humidity, influences malaria's transmission with the communality index (0.94) and Dillon-Goldstein's ρ (0.97). This result is also consistent with other findings [181], where a positive association exists between temperature and the occurrence

of dengue. *Factor II* and *Factor III* appear to have less influence on the influence of malaria.

Table 5.8: Indices for selecting the ecological hidden factor of a high malaria incidence in the study area.

Factor	**Reflective variables**	**Communality**	**Dillon-Goldstein's ρ**
I	2	0.94[c](94%)	0.97[c](97%)
II	2	0.77(77%)	0.87(87%)
III	2	0.49(49%)	0.49(49%)

[c] the most significant hidden factor.

5.4 Intelligent Malaria Outbreak Warning System

In sections 5.2 and 5.3, the most influential and hidden ecological factors of malaria incidence were identified by deploying regression and partial least squares path modelling, respectively. This section discusses in detail the implementation of the malaria outbreak system, based on the identified hidden ecological factors. The deployment comprises three stages: data preprocessing, generating the predictive model using machine learning, and the development of a mobile application.

5.4.1 Data preprocessing

It has been a tradition prior to the application of machine-learning algorithms that datasets need to be preprocessed to enable a faster learning process and greater accuracy. The heuristic approach most commonly used discretisation techniques in data mining. This involves transforming continuous-value datasets into discrete datasets by creating a set of contagious intervals [99]. To start the preprocessing, the dataset on climate variables was chosen as the input (independent variable), while the reported

cases of malaria's incidence is the output (dependent variable). Hence, the concern is to develop a predictive model using supervised machine learning algorithms that will predict the likelihood of malaria's incidence. Since the output variable appeared to have a high-magnitude in-terms of the reported number of malaria cases, using it directly may cause over-fitting to the predictive model. Therefore, it is pertinent to transform the dataset using discretisation techniques in order to build an efficient model.

The output variable is discretised to form a target variable using a *k*-means clustering algorithm [107]. This methodology is chosen over equal width (EW) and equal frequency (EF) because it is less sensitive to outliers, and the number of clusters (partitions) can be optimised by analysis rather than pre-determination. In general, the choice of discretisation method and choice of *k* can be guided by the objectives of the discretisation task. The *R* software was invoked to determine the optimum number of clusters to enable us to partition the output variable.

From the analysis, the optimum number of clusters obtained was $k = 4$ and the algorithm converged after 9 iterations with 89.9% percent variation. Also, it is observed that, for $k = 5$, the number of iterations exceeded the maximum number of tolerable iterations to achieve convergence; in this case, it diverged even though the variation percentage was still good at 93%. For $k = 2$ and 3, the algorithm converges after 3 and 4 iterations with 66.4% and 82% percent variation, respectively. This gives sufficient evidence to choose the optimum number of clusters as $k = 4$. Similarly, the *NbClust* package in *R* software was also tried to determine the optimal number of clusters. Using *k* values ranging from 2 to 5 allows the algorithm to select the optimum cluster for use in order to partition the output variable. The algorithm ran and selected $k = 4$ as the optimum number of clusters to partition the output variables. Hence, both methodologies give the same number of optimum clusters and subsequently prove consistency.

The output variable was partitioned into four classes according to the results of the *k*-means algorithm, which was relabelled as: *low*, *medium*, *high* and *very high* incidence status of malaria. Table 5.9 presents a summary analysis of the *k*-means algorithm clustering.

Table 5.9: Summary of the data descritisation using the k-means algorithm.

Number of clusters (k)	**2**	**3**	**4**	**5**
iteration	3	4	9	6
convergence	yes	yes	yes	no
$\frac{SSB}{SST}$	66.4%	82%	89.9%	93%

5.4.2 Machine Learning

The next stage is to identify a pattern or model from the data preprocessed in subsection 5.4.1 that could be used to make an accurate prediction of malaria's incidence. Evolving from traditional pattern recognition approaches, machine-learning methods explore the algorithms that can learn from the data and overcome prediction tasks by building a mathematical model with a data sample input. A learning algorithm will mark each given malaria epidemic data sample as one category, then, after being trained using the training dataset, it will build a model to predict which category the forthcoming data sample falls into.

Several machine learning algorithms was applied, including *SMV*, *KNN*, *Naive Bayes* and *Decision Trees*, to find the best prediction algorithm for the *scikit* framework [4] in *Python*.

To evaluate the prediction of machine learning algorithms on the training set, a 10-fold cross-validation technique was used for selecting a training set and test sets that

were mutually independent. Table 5.10 shows the prediction results when comparing 7 different machine learning methods.

Table 5.10: Accuracy comparison of model checking algorithms.

Algorithm	LiR	LoR	DT	SVM	SVM (o)	NB	KNN1	KNN5	KNN10	K-M (3)
Accuracy	83.8%	75.0%	63.8%	80.6%	99.0%	63.9%	58.3%	80.6%	80.6%	47.2%

Based on the performance accuracy of the algorithms presented in Table 5.10 and their interpretations are given as follows.

5.4.2.1 Linear Regression (LiR)

This method gives overall good prediction results, but failed to produce any *medium* predictions.

5.4.2.2 Logistic Regression (LoR)

This method predicts the probability of occurrence for an event by fitting the dataset, as a set of independent variables, into a logic function. In other words, for a correlated data set, LoR may not be able to find the intrinsic-relationships between events.

5.4.2.3 Decision Tree (DT)

This algorithm works very well for both categorical and continuous dependent variables; however, this dataset cannot be separated as distinct groups since the edges of the samples are fuzzy. Therefore, DTs gave a bad prediction.

5.4.2.4 Support Vector Machine (SVM)

SVM is one of the most efficient supervised machine learning algorithms, and is mainly used to solve classification and regression problems. The best part of this algorithm is that data training and testing can be plotted as a point on an n-dimensional plane, with a feature being the value of a particular coordinate. Without optimising the parameters, SVM gave a 80.56% prediction result. After a parameter optimisation, especially on the penalty parameter and gamma coefficient adjustment, SVM (o) gave a 99.0% prediction result.

5.4.2.5 Naive Bayes (NB)

NB is a well-known classification method, which is based on 'Bayes Theorem' with an oversimplified assumption of independence between classifiers. Moreover, NB is a conditional probability model, which means that the method needs to be assigned a series of certain events. For this dataset, NB did not produce a good prediction overall.

5.4.2.6 K-Nearest Neighbours (KNN)

The KNN method is able to deal with both classification and regression problems. In comparison to KNN5 (where k=5) and KNN10 (where k=10), KNN1 (where k=1) failed to make a good prediction. This means that the data could need more pre-processing and/or noise removal within a theory; however, most data from the real world are incomplete, which is why KNN5 and KNN10 offer better predictions.

5.4.2.7 K-Means (K-M)

K-M is a type of unsupervised clustering method. In this case, 3 clusters have been set at the beginning; however, a convergence did not perfectly land, and therefore it could not provide a good overall prediction.

Thus, the results, presented in Table 5.10, show that the best performing algorithm is SVM. Therefore, the SVM model will be integrated into our system.

5.5 Mobile Application

This section presents the development of mobile application, the Malaria Outbreak Warning System, with a built-in SVM model, published at Google Play. The tool can be accessed via [3].

The application was based on the theoretical experiences and practical experiments of the SVM algorithm and model, which was tested for the development of systematic and effective strategies to predict the outbreak of a malaria epidemic. Meanwhile, the parameters of the model kernel were optimised and set into this application.

The application consists of 3 processes: preprocessing the weather forecasting data, processing the data by applying them into the model and implementing the model's interface, and post-processing the prediction data by presenting results onto the app's UI front layer. It is a well-suited implementation for location detection.

Figure 5.9 shows a screenshot of the tool. The application does not only support the automatic gathering of weather forecasting data, but also supports manual data input. The application reads climatic information, i.e., temperature, relative humidity, wind speed, solar radiation and precipitation, from the weather and geographical APIs. When the units of the weather and atmosphere were different from the dataset used to

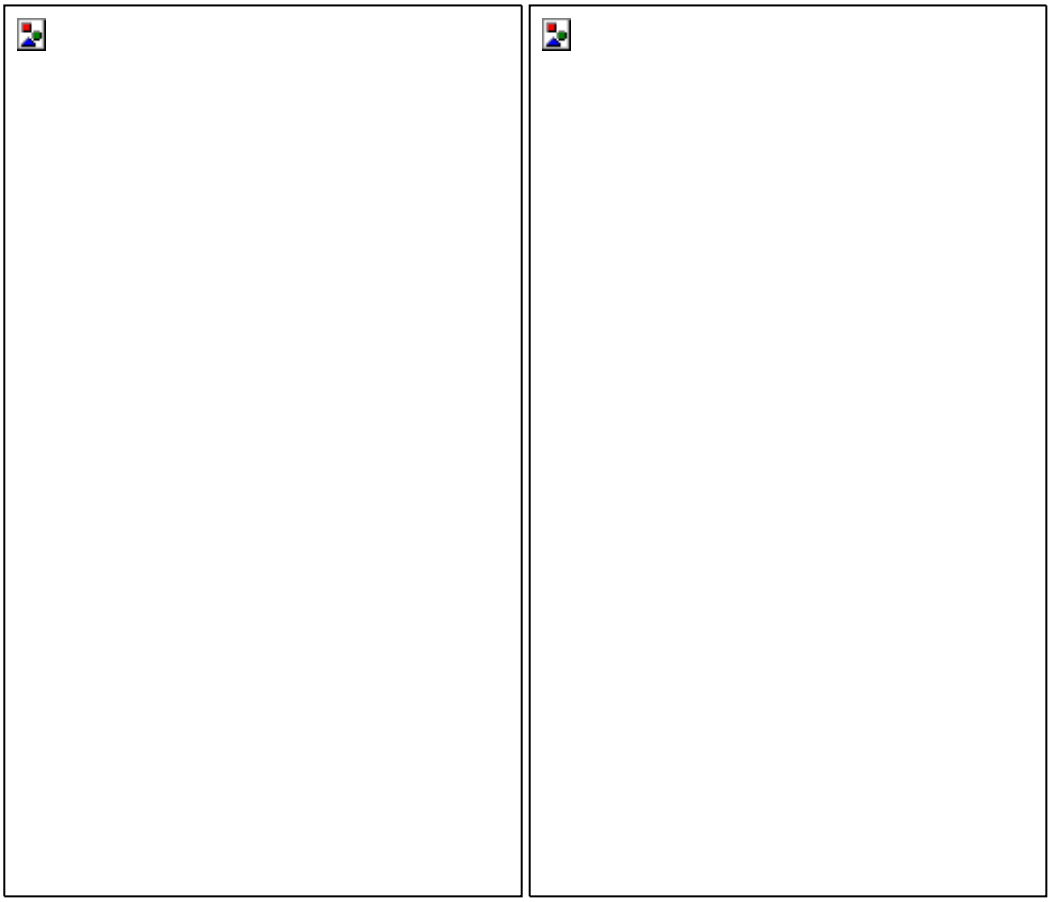

Figure 5.9: A screen shot for the mobile application.

construct the predictor, the required normalisation was carried out or feature scaling or similar preprocessing. The tool then predicted the malaria outbreak in a couple of days advance based on available forecast information acquired from the APIs. The user can slide the screen to see the available outbreak predictions for current and future days. The additional button on the bottom of screen lets users manually enter a set of weather measurements to gather prediction results for customised parameters.

The trained *SVM* model has been implemented in *Java* by taking advantage of the *LIBSVM* [1]. *LIBSVM* is an integrated software for *SVM*, regression and distribution estimation. The mobile application was developed for Android using Android Studio.

The weather forecasting data is powered by OpenWeatherMap API [5], which is an online service provider for weather data. OpenWeatherMap provides an API to search forecasting data for up to 5 days by coordinates; the responses are served with *JSON*, *XML* and *HTML* endpoints. All of the data provided are under CC BY-SA 4.0 license.

5.5.1 Discussion

The current prototype of the intelligent malaria outbreak warning system relies on a batch machine learning process. Thus, the learning algorithm is trained and tested offline using the available data set, and the prediction model is embedded within the tool. Hence, the prediction process relies on the offline training of the prediction model.

A more effective approach is to make the learning process online. That is, whenever new data are available, they are automatically updated, and the learning process is run again to incorporate the new data. This will not only allow an automatic and dynamic learning process, but also increase the accuracy of the prediction by adapting to new patterns in the data.

The online learning approach requires a mechanistic data collection mechanism, which is very challenging to perform as hospitals and health service providers do not make the relevant data available online. Even acquiring permission for access to the available data is a long and bureaucratic process. On the other hand, as discussed in subsection 5.3.2, most available data cannot be directly used in this system as they are incomplete and/or not processed.

To alleviate these issues and to support the online learning process, the malaria outbreak warning application can be extended to collect online data from its users. These users include hospitals, healthcare providers, individuals, etc., who would re-

port malaria cases to the system. Using the geographical location of the incident, the application will acquire all the necessary information for the ecological factors. In this way, new data will be collected at the run time, and the learning process will be instantiated each time new data are available.

5.6 Summary

Chapter 5 presented a data-driven model to investigate and understand the dynamics in the transmission of malaria. An intelligent malaria prediction model was developed by using embedded machine learning algorithms to read and process the likelihood of malaria's incidence from climate data. This model will supply information like a warning signal to healthcare providers so as to prepare and strategise prevention and control mechanisms. Prior to developing the model, climate factors and malaria data were investigated and the hidden factors that bring about a high incidence within the area studied were identified. The next chapter presents the conclusion of this thesis.

Chapter 6

General Discussion

This chapter provides a summary of the entire work presented in this thesis; furthermore, recommendations for future study are also are given.

6.1 Consequences for climate change to malaria

Malaria creates serious health problems for people, and about two-third of the world population are at risk of infection. Globally, malaria is spreading, mostly in tropical and sub-tropical regions. This is because the region's climate supports the survival of mosquitos and enabling its faster reproduction by causing high malaria transmission in a human population.

Recent campaigns on the consequences of anticipating global warming, particularly on the likelihood of malaria's prominence in previously unexposed areas, has attracted the attention of the World Health Organization (WHO) and allied healthcare providers to further strengthen efforts towards intervention, prevention and control. The Inter-governmental Panel of Climate Change (IPCC) says gradual rises in temperature over time can alter the natural habitats of the mosquito by changing its prevalence

and prolonging the season of malaria's spread [145]. A marginal 0.5°C increase in temperature could mean a 30–100% increase in the mosquito population [162]. Similarly, a small shift in temperature, from 2°C to 3°C, could increase the number of humans vulnerable to malaria by up to 5% [167].

6.2 Summary of the thesis findings

As this study focuses on investigating the influence of climate factors in malaria transmission dynamics, the causal relationship between climate factors and malaria occurrences [30] has been explored to develop an understanding of the epidemic pattern, and thus develop an efficient predictive model. This enabled the development of three modelling approaches, which fully addressed and exhausted the scope of the established research question (as stated in Subsection 1.2.1). The main contributions of this study are presented in Chapter 3, 4 and 5; each captures the specific modelling techniques used (to address the aim and objectives outlined under subsection 1.2.2) when investigating and understanding the dynamics of malaria transmission in the light of climate factors. In Chapter 1, a succinct introduction of malaria, how it is transmitted and the factors responsible for its global distribution were presented alongside the motivation, research question and the significance of this study. In Chapter 2, a theoretical background of the malaria life cycle, mathematical models and their application to malaria studies were also presented.

In Chapter 3, a mathematical model of malaria transmission dynamics was investigated by incorporating temperature dependent delay through the extrinsic incubation and maturation in adult and aquatic mosquito states, respectively. The theoretical results of the non-autonomous model show that a disease-free equilibrium is locally

asymptotically stable whenever the basic reproduction number is less than unity or otherwise. This means the malaria infection persists whenever the basic reproduction number is greater than the unity. However, it has also emerged from the non-autonomous model that the spread of malaria is highly influenced by the dynamics of immature mosquitoes. The numerical results indicate that the maturation time delay reflects the effects of the temperature variation on the development of the aquatic stage of the mosquito. This result suggests that, within the limits of lower and upper development thresholds, immature mosquitoes develop faster as the temperature increases. In addition, the increase in temperature shortens the maturation time from an egg to an adult mosquito, and thus triggers a greater mosquito population. As part of the control and prevention strategy, a sensitivity analysis was carried out using LHS-PRCC and this identified the most influential parameters of the model (see Figure 3.19). The suggested parameters for control are the mosquito's biting rate (c_m); its egg deposition rate (α_E); its carrying capacity (K_c); the mortality rate of an immature mosquito (μ_a) and the adult mosquito mortality (μ_m). Subsequently, our study has shown that malaria transmission can be effectively controlled in a population if the incubation period of mosquitoes is considered part of the prevention strategies. This results have complemented outcomes from previous studies (see [161]) where the incubation period in humans was the main focus. Thus, the work presented in this Section 3 provides an important contribution when developing a preventive mechanism that can assist in diagnosing malaria infection and the detection of outbreaks. Moreover, using mathematical approaches to investigate malaria dynamics is good but has a limitation when studying a homogeneous population [79]. Since both human and mosquito populations are naturally heterogeneous (e.g., sex, age, spatial and changes, movement pattern amongst other), it is inadequate to capture the heterogeneity arising in a population's dynamics.

In Chapter 4, agent-based modelling was utilised to address the limitations of mathematical modelling, which arise from the heterogeneity of both human and mosquito populations. This contribution addresses the population dynamics of malaria transmission using agent-based modelling. Using this method, the pattern of a malaria epidemic was predicted and validated the outcome against the reported cases (see Table 4.1) obtained from three cities: Tripura district in India (see map 4.4), Limpopo province in South-Africa (see map 4.5) and Benin city in Nigeria (see map 4.6). Both agent-based modelling and mathematical modelling are used in epidemiology to investigate and understand the dynamics in disease transmission. Agent-based modelling complements the mathematical technique due to the limitations on heterogeneity, hence the results was compared by studying population dynamics of malaria epidemics in the cities. Agent-based modelling is identified as the most robust technique as it was able to predict a malaria season in all three cities with very good precision, as confirmed by the *t-test* and *correlation coefficient* (see Table 4.5). However, mathematical modelling was slightly better in predicting the occurrence and season of malaria in Tripura than in Limpopo and Benin. Moreover, the mathematical technique is suitable for simulating continuous, rather than discrete, phenomena. When devising prevention and control mechanisms, both mathematical and agent-based techniques provide a good platform to alleviate malaria infections. The transmission of malaria has been linked with climatic factors; in particular temperature plays the most significant role among other factors. In Chapter 3 and 5, temperature was the only climate factor used in the dynamic modelling of malaria. Therefore, the inclusion of other malaria predictors, such as rainfall, humidity etc., in the model will significantly increase the robustness of the model by approximating the dynamics.

In Chapter 5, an intelligent early warning system for malaria outbreaks was devel-

oped based on climate factors and using machine learning algorithms. Prior to that, climate predictors of malaria together with reported cases were preprocessed to identify at least one of the most influential factors to influence its high incidence (see subsection 5.2.1.1). From the results of the *cross-correlation* function (CCF) between malaria cases and pre-whiten climate variables, it was found that temperature and precipitation have negligible lagged effects on the occurrence of malaria. Temperature supports the development speed of the parasites and their morphological stage, while precipitation creates breeding sites for mosquitoes to lay their eggs and supports its survival in the water body. The study area experiences the rainy season, between April and October with annual rainfall varying from 1,500 mm–2,000 mm (60 to 80 inch). The annual temperature is above 20°C (68°F) and the annual relative humidity is about 75%.The temperature and precipitation data in the study area have exceeded the threshold to become predictors of the incidence of malaria, however they have negligible influence. Hence, it is believed that a huge amount of rain is likely to wash out the ground and kill the eggs and mosquito larvae [126, 143]. However, it was understood that sustained rainfall provides breeding sites for mosquitoes, and thereby increases its population. Furthermore, an analysis of variance was performed on the regression model, between malaria incidence and climate variables, and finds the probability ($p_{value} < 0.05$). This indicates a statistically significant level that relative humidity seems to be a dominant climate predictor of malaria. The relative humidity may be determined by precipitation. From the available information, the relative humidity in the study area is 90% greater during the rainy season than the dry season. Relative humidity is a climatic variable that does not support the mosquitoes' *sporogonic* cycle or *gonotrophic* cycle but strengthens the vector longevity and provides a good atmosphere for biting [187], especially in the night hours. The study area has a rain forest vegetation belt, which in-

dicates a high level of relative humidity and is the dominant climate factor. However, it is natural that an inherent relationship exists among the climate factors on the incidence of malaria. Hence, in section 5.3 further analysis was conducted from this, the hidden factors were identified representing the combination of at least two among the climate predictors of malaria. This provides an overview of the ecosystem for malaria modelling and proposes a new framework for the study of malaria transmission ecosystems including the prevention and control of its scourge (see Figure 5.4).

Hidden factors have been identified that lead to high incidence of malaria. Hence, the analysis of the results shows that the minimum temperature and relative humidity, which are related to Factor I, have positive associations with the incidence of malaria in the study area. The other observed variables, like maximum temperature, solar radiation, precipitation and wind speed, which are related to hidden Factor II and Factor III, appear to have mildly influenced the incidence malaria as shown in Table 5.8. The primary results have demonstrated the power of the proposed predicative analytics-based malaria outbreak warning. This system will help hospitals, healthcare providers, and health organisations to take precautions in time and to best utilise their resources in the case of emergencies. To our best knowledge, the system developed shown in Figure 5.9 is the first publicly available application.

6.3 Limitation and Future Work

Although, this thesis addressed the research question stated in subsection 1.2.1, the results have raised new issues of importance indicating areas where more work needs to be done. Instead of considering the details of these possible extensions, hence this section provides a discussion on some general ideas for the extension of the presented

results. The possible directions for extending the main results presented in Chapter 3, 4 and 5 are as follows. Chapter 3 has the following limitations: the implicit case of the aquatic mosquito was considered in the modelling, namely combining the egg, larva and pupa in a single compartment. However, it is important to investigate the dynamics by considering the explicit cases of the aquatic stage. The model uses embedded temperature-dependent models for the biting rate, mortality rate, maturation rate and survival rate of the mosquito in the modelling. Although temperature is a large-scale driver of malaria transmission, it is imperative to include other factors, including precipitation and humidity. In Chapter 4, the modelling was achieved by considering the spatial and temporal movement of people in an arbitrary environment without recourse to reasons such as school, work, specific locations, or farms, among others. As human behaviour is complex and dynamic, it is imperative to investigate this limitation in order to better approximate the diffusion of malaria in a population. In Chapter 5, the further development of the system will incorporate the automatic data gathering from a variety of sources. This will not only allow for an automatic and dynamic learning process, but also increase the accuracy of predictions by adapting to new patterns in the data.

6.4 Overall Conclusion

Malaria remains endemic in Africa and some parts of the world. It seasonal resurgences is imposing health burden on the populace and to the economy. Despite malaria being preventable and treatable, globally, it accounts for highest index of morbidity and mortality among other infectious diseases. Although, interventions have been continuously provided by the World Health Organisations and other donor agencies, but

malaria still persists. The effective tools for management and planning of the interventions particularly in the developing countries is the key challenge.

Towards developing tools for guiding the interventions, the link between malaria spread and climate factors was utilised. These are factors, temperature, precipitation and humidity. In particular, temperature is a large scale driver that supports mosquito parasite development, biting and survival. This thesis investigates the impact of temperature-dependent extrinsic incubation and juvenile maturation delays on malaria spread. The results showed that malaria transmission is greatly influenced by the juvenile mosquito maturation and extrinsic incubation. In other words, the maturation delay from an egg to adult mosquito is sensitive to ambient temperature as well as the malaria parasites development. This results complemented the study [39], where the intrinsic incubation was investigated. On that account, this study has proven to have the potential towards developing a preventive mechanism by guiding policy makers to better planning of intervention as well as prior detection of outbreaks.

References

[1] "A library for support vector machines." Available online: http://www.csie.ntu.edu.tw/~cjlin/libsvm/., accessed on 24 June 2017.

[2] "Global weather data for swat." Available online: http://globalweather.tamu.edu.

[3] "Mlsvm for research." Available online: https://play.google.com/store/apps/details?id=project.lanydr.mlsvm&hl=en.

[4] "Scikit-learn." Available online: http://www.scikit-learn.org.

[5] "Weather api." Available online: http://openweather.org/api.

[6] "World health organization.(2018)," *World Malaria Report*, vol. 2018, 2018.

[7] "Climate data for cities worldwide-climate-data.org." *Retrieved from: https://en.climate-data.org/* , 2019.

[8] "Communicable diseases communique," *Retrieved from: http:// www.nicd.ac.za/wp-content/uploads/2017/06/Malaria-in-SA.pdf* , 2019.

[9] "Disease model cradle (dmc) practical," *[online] Available at: https: //www.liverpool.ac.uk/media/livacuk/qweci/dmc_practical_english.pdf [Accessed 2 Nov. 2019]*, 2019.

[10] "Gadm." *Retrieved 10 October 2019, from: https://gadm.org/data.html*, 2019.

[11] "Gauteng," *Retrieved 17 July 2019, from: https:// en.wikipedia.org/ wiki/ Gauteng*, 2019.

[12] "Life expectancy in benin," *Retrieved 17 July 2019, from: https:// www. worldlifeexpectancy.com/ benin-life-expectancy*, 2019.

[13] "Limpopo population," *Retrieved 17 July 2019, from: http:// population.city/ south-africa/adm/limpopo/*, 2019.

[14] "North tripura district population census 2011-2019, tripura literacy sex ratio and density." *[online] Available at: https:// census2011.co.in/ census/ district/ 460-north-tripura.html [Accessed 17 Jul. 2019].*, 2019.

[15] "Outline of tripura," *Retrieved 17 July 2019, from: https:// en.wikipedia.org/ wiki/ Outline-of-Tripura*, 2019.

[16] "Vensim software," *Retrieved from: https:// vensim.com/ vensim-software/* , 2019.

[17] S. Abar, G. K. Theodoropoulos, P. Lemarinier, and G. M. O'Hare, "Agent based modelling and simulation tools: A review of the state-of-art software," *Computer Science Review*, vol. 24, pp. 13–33, 2017.

[18] A. Abdelrazec and A. B. Gumel, "Mathematical assessment of the role of temperature and rainfall on mosquito population dynamics," *Journal of mathematical biology*, vol. 74, no. 6, pp. 1351–1395, 2017.

[19] M. A. Acevedo, O. Prosper, K. Lopiano, N. Ruktanonchai, T. T. Caughlin, M. Martcheva, C. W. Osenberg, and D. L. Smith, "Spatial heterogeneity, host movement and mosquito-borne disease transmission," *PloS one*, vol. 10, no. 6, p. e0127552, 2015.

[20] F. Adeyemo, O. Makinde, L. Chukwuka, and E. Oyana, "Incidence of malaria infection among the undergraduates of university of benin (uniben), benin city, nigeria," *The Internet Journal of Tropical Medicine*, vol. 9, no. 1, pp. 1–8, 2013.

[21] F. Agusto, A. Gumel, and P. Parham, "Qualitative assessment of the role of temperature variations on malaria transmission dynamics," *Journal of Biological Systems*, vol. 23, no. 04, p. 1550030, 2015.

[22] H. Alout, B. Roche, R. K. Dabire, and A. Cohuet, "Consequences of insecticide resistance on malaria transmission," *PLoS pathogens*, vol. 13, no. 9, p. e1006499, 2017.

[23] W. H. O. and others, "Who guide for standardization of economic evaluations of immunization programmes," World Health Organization, Tech. Rep., 2019.

[24] R. M. Anderson and R. M. May, "The population dynamics of microparasites and their invertebrate hosts," *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, vol. 291, no. 1054, pp. 451–524, 1981.

[25] M. Aregawi, M. Lynch, W. Bekele, H. Kebede, D. Jima, H. S. Taffese, M. A. Yenehun, A. Lilay, R. Williams, M. Thomson *et al.*, "Time series analysis of trends in malaria cases and deaths at hospitals and the effect of antimalarial interventions, 2001–2011, ethiopia," *PLoS One*, vol. 9, no. 11, p. e106359, 2014.

[26] J. L. Aron, "Acquired immunity dependent upon exposure in an sirs epidemic model," *Mathematical biosciences*, vol. 88, no. 1, pp. 37–47, 1988.

[27] J. L. Aron, "Mathematical modelling of immunity to malaria," *Mathematical Biosciences*, vol. 90, no. 1-2, pp. 385–396, 1988.

[28] J. L. Aron and R. M. May, "The population dynamics of malaria," in *The population dynamics of infectious diseases: theory and applications*. Springer, 1982, pp. 139–179.

[29] V. Assele, G. E. Ndoh, D. Nkoghe, and T. Fandeur, "No evidence of decline in malaria burden from 2006 to 2013 in a rural province of gabon: implications for public health policy," *BMC Public health*, vol. 15, no. 1, p. 81, 2015.

[30] C. M. Baak-Baak, D. A. Moo-Llanes, N. Cigarroa-Toledo, F. I. Puerto, C. Machain-Williams, G. Reyes-Solis, Y. J. Nakazawa, A. Ulloa-Garcia, and J. E. Garcia-Rejon, "Ecological niche model for predicting distribution of disease-vector mosquitoes in yucatán state, méxico," *Journal of medical entomology*, vol. 54, no. 4, pp. 854–861, 2017.

[31] R. P. Bagozzi and Y. Yi, "Specification, evaluation, and interpretation of structural equation models," *Journal of the academy of marketing science*, vol. 40, no. 1, pp. 8–34, 2012.

[32] E. Bakare, "On the optimal control of vaccination and treatments for an sir-epidemic model with infected immigrants," *Applied & Computational Mathematics*, vol. 4, no. 4, pp. 01–09, 2015.

[33] A. M. Barreaux, P. Barreaux, K. Thievent, and J. Koella, "Larval environment influences vector competence of the malaria mosquito anopheles gambiae," *MWJ*, vol. 7, p. 8, 2016.

[34] B. R. Bayles, K. A. Brauman, J. N. Adkins, B. F. Allan, A. M. Ellis, T. L. Goldberg, C. D. Golden, D. S. Grigsby-Toussaint, S. S. Myers, S. A. Osofsky *et al.*, "Ecosystem services connect environmental change to human health outcomes," *EcoHealth*, vol. 13, no. 3, pp. 443–449, 2016.

[35] L. M. Beck-Johnson, W. A. Nelson, K. P. Paaijmans, A. F. Read, M. B. Thomas, and O. N. Bjørnstad, "The effect of temperature on anopheles mosquito population dynamics and the potential for malaria transmission," *PLOS one*, vol. 8, no. 11, p. e79276, 2013.

[36] L. M. Beck-Johnson, W. A. Nelson, K. P. Paaijmans, A. F. Read, M. B. Thomas, and O. N. Bjørnstad, "The importance of temperature fluctuations in understanding mosquito population dynamics and malaria risk," *Royal Society open science*, vol. 4, no. 3, p. 160969, 2017.

[37] J. I. Blanford, S. Blanford, R. G. Crane, M. E. Mann, K. P. Paaijmans, K. V. Schreiber, and M. B. Thomas, "Implications of temperature variation for malaria parasite development across africa," *Scientific reports*, vol. 3, p. 1300, 2013.

[38] L. Blumberg and J. Frean, "Malaria control in south africa-challenges and successes," *South African Medical Journal*, vol. 97, no. 11, pp. 1193–1197, 2007.

[39] G. B. Bonan and H. H. Shugart, "Environmental factors and ecological processes in boreal forests," *Annual review of ecology and systematics*, vol. 20, no. 1, pp. 1–28, 1989.

[40] G. E. Box, "Science and statistics," *Journal of the American Statistical Association*, vol. 71, no. 356, pp. 791–799, 1976.

[41] A. J. Brazier, M. Avril, M. Bernabeu, M. Benjamin, and J. D. Smith, "Pathogenicity determinants of the human malaria parasite plasmodium falciparum have ancient origins," *mSphere*, vol. 2, no. 1, pp. e00 348–16, 2017.

[42] C. Briat, "Sign properties of metzler matrices with applications," *Linear Algebra and its Applications*, vol. 515, pp. 53–86, 2017.

[43] A. R. Brock, C. A. Gibbs, J. V. Ross, and A. Esterman, "The impact of antimalarial use on the emergence and transmission of plasmodium falciparum resistance: a scoping review of mathematical models," *Tropical medicine and infectious disease*, vol. 2, no. 4, p. 54, 2017.

[44] J. Brownlee, *Introduction to time series forecasting with python: how to prepare data and develop models to predict the future*. Machine Learning Mastery, 2017.

[45] B. M. Byrne, *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press, 2013.

[46] M. Cardinot, C. O'Riordan, J. Griffith, and M. Perc, "Evoplex: A platform for agent-based modeling on networks," *SoftwareX*, vol. 9, pp. 199–204, 2019.

[47] F. Changyong, W. Hongyue, L. Naiji, C. Tian, H. Hua, L. Ying *et al.*, "Log-transformation and its implications for data analysis," *Shanghai archives of psychiatry*, vol. 26, no. 2, p. 105, 2014.

[48] L. F. Chaves, Y. Higa, S. H. Lee, J. Y. Jeong, S. T. Heo, M. Kim, N. Minakawa, and K. H. Lee, "Environmental forcing shapes regional house mosquito synchrony in a warming temperate island," *Environmental entomology*, vol. 42, no. 4, pp. 605–613, 2013.

[49] D. Chen, B. Moulin, J. Wu *et al., Analyzing and modeling spatial and temporal dynamics of infectious diseases*. Wiley Online Library, 2015.

[50] N. Chitnis, J. M. Cushing, and J. Hyman, "Bifurcation analysis of a mathematical model for malaria transmission," *SIAM Journal on Applied Mathematics*, vol. 67, no. 1, pp. 24–45, 2006.

[51] C. Christiansen-Jucht, P. E. Parham, A. Saddler, J. C. Koella, and M.-G. Basáñez, "Temperature during larval development and adult maintenance influences the survival of anopheles gambiae ss," *Parasites & vectors*, vol. 7, no. 1, p. 489, 2014.

[52] U. M. Chukwuocha and I. N. Dozie, "Malaria transmission and morbidity patterns in holoendemic areas of imo river basin of nigeria," *BMC research notes*, vol. 4, no. 1, p. 514, 2011.

[53] D. Cochrane and G. H. Orcutt, "Application of least squares regression to relationships containing auto-correlated error terms," *Journal of the American statistical association*, vol. 44, no. 245, pp. 32–61, 1949.

[54] E. Collins, N. M. Vaselli, M. Sylla, A. H. Beavogui, J. Orsborne, G. Lawrence, R. E. Wiegand, S. R. Irish, T. Walker, and L. A. Messenger, "The relationship between insecticide resistance, mosquito age and malaria prevalence in anopheles gambiae sl from guinea," *Scientific Reports*, vol. 9, no. 1, p. 8846, 2019.

[55] M. H. Craig, R. Snow, and D. le Sueur, "A climate-based distribution model of malaria transmission in sub-saharan africa," *Parasitology today*, vol. 15, no. 3, pp. 105–111, 1999.

[56] E. D. Dan, O. Jude, and O. Idochi, "Modelling and forecasting malaria mortality rate using sarima models (a case study of aboh mbaise general hospital, imo state nigeria)," *Science Journal of Applied Mathematics and Statistics*, vol. 2, no. 1, pp. 31–41, 2014.

[57] S. Dawaki, H. M. Al-Mekhlafi, I. Ithoi, J. Ibrahim, W. M. Atroosh, A. M. Abdulsalam, H. Sady, F. N. Elyana, A. U. Adamu, S. I. Yelwa *et al.*, "Is nigeria winning the battle against malaria? prevalence, risk factors and kap assessment among hausa communities in kano state," *Malaria journal*, vol. 15, no. 1, p. 351, 2016.

[58] M. C. de Castro, R. L. Monte-Mor, D. O. Sawyer, and B. H. Singer, "Malaria risk on the amazon frontier," *Proceedings of the National Academy of Sciences*, vol. 103, no. 7, pp. 2452–2457, 2006.

[59] V. Dev, T. Adak, O. P. Singh, N. Nanda, and B. K. Baidya, "Malaria transmission in tripura: Disease distribution & determinants," *The Indian journal of medical research*, vol. 142, no. Suppl 1, p. S12, 2015.

[60] K. Dietz, L. Molineaux, and A. Thomas, "A malaria model tested in the african savannah," *Bulletin of the World Health Organization*, vol. 50, no. 3-4, p. 347, 1974.

[61] T. K. Dijkstra, "Latent variables and indices: Herman wold's basic design and partial least squares," in *Handbook of partial least squares*. Springer, 2010, pp. 23–46.

[62] W. Duan, Z. Fan, P. Zhang, G. Guo, and X. Qiu, "Mathematical and computational approaches to epidemic modeling: a comprehensive review," *Frontiers of Computer Science*, vol. 9, no. 5, pp. 806–826, 2015.

[63] N. P. C. P. N. F. p. L. . H. P. Earth, P., "Population of benin city in 2018-statistics," *Retrieved 17 July 2019, from: https:// all-populations.com/ en/ ng/ population-of-benin-city.html*, 2019.

[64] G. C. Ejezie, E. Ezedinachi, E. Usanga, E. Gemade, N. Ikpatt, and A. Alaribe, "Malaria and its treatment in rural villages of aboh mbaise, imo state, nigeria," *Acta Tropica*, vol. 48, no. 1, pp. 17–24, 1990.

[65] V. Ermert, A. H. Fink, A. E. Jones, and A. P. Morse, "Development of a new version of the liverpool malaria model. i. refining the parameter settings and mathematical formulation of basic processes based on a literature review," *Malaria journal*, vol. 10, no. 1, p. 35, 2011.

[66] V. Ermert, A. H. Fink, A. E. Jones, and A. P. Morse, "Development of a new version of the liverpool malaria model. ii. calibration and validation for west africa," *Malaria journal*, vol. 10, no. 1, p. 62, 2011.

[67] G. Fan, J. Liu, P. Van den Driessche, J. Wu, and H. Zhu, "The impact of maturation delay of mosquitoes on the transmission of west nile virus," *Mathematical Biosciences*, vol. 228, no. 2, pp. 119–126, 2010.

[68] T. P. I. for Climate Impact Research and C. Analytics, "Turn-down the heat-why a 4 degree warmer world must be avoided," *Biometrika*, 2012.

[69] N. Ganesan, K. Venkatesh, M. Rama, and A. M. Palani, "Application of neural networks in diagnosing cancer disease using demographic data," *International Journal of Computer Applications*, vol. 1, no. 26, pp. 76–85, 2010.

[70] S. Gang, "Soft modeling: Intermediate between traditional model building and data analysis," in *Mathematical statistics*. Citeseer, 1980.

[71] G. Geleta and T. Ketema, "Severe malaria associated with plasmodium falciparum and p. vivax among children in pawe hospital, northwest ethiopia," *Malaria research and treatment*, vol. 2016, 2016.

[72] P. W. Gething, A. P. Patil, D. L. Smith, C. A. Guerra, I. R. Elyazar, G. L. Johnston, A. J. Tatem, and S. I. Hay, "A new world malaria map: Plasmodium falciparum endemicity in 2010," *Malaria journal*, vol. 10, no. 1, p. 378, 2011.

[73] N. M. Gharakhanlou, M. S. Mesgari, and N. Hooshangi, "Developing an agent-based model for simulating the dynamic spread of plasmodium vivax malaria: A case study of sarbaz, iran," *Ecological Informatics*, p. 101006, 2019.

[74] A. Gomez-Elipe, A. Otero, M. Van Herp, and A. Aguirre-Jaime, "Forecasting malaria incidence based on monthly case reports and environmental factors in karuzi, burundi, 1997–2003," *Malaria Journal*, vol. 6, no. 1, p. 129, 2007.

[75] K. Gopalsamy, *Stability and oscillations in delay differential equations of population dynamics*. Springer Science & Business Media, 2013, vol. 74.

[76] J. Hackl and T. Dubernet, "Epidemic spreading in urban areas using agent-based transportation models," *Future Internet*, vol. 11, no. 4, p. 92, 2019.

[77] J. F. Hair, M. Sarstedt, T. M. Pieper, and C. M. Ringle, "The use of partial least squares structural equation modeling in strategic management research: a review of past practices and recommendations for future applications," *Long range planning*, vol. 45, no. 5-6, pp. 320–340, 2012.

[78] J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering & System Safety*, vol. 81, no. 1, pp. 23–69, 2003.

[79] S. A. Herzog, S. Blaizot, and N. Hens, "Mathematical models used to inform study design or surveillance systems in infectious diseases: a systematic review," *BMC infectious diseases*, vol. 17, no. 1, p. 775, 2017.

[80] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM review*, vol. 42, no. 4, pp. 599–653, 2000.

[81] M. B. Hoshen and A. P. Morse, "A weather-driven model of malaria transmission," *Malaria journal*, vol. 3, no. 1, p. 32, 2004.

[82] J. H. Huber, G. L. Johnston, B. Greenhouse, D. L. Smith, and T. A. Perkins, "Quantitative, model-based estimates of variability in the generation and serial intervals of plasmodium falciparum malaria," *Malaria journal*, vol. 15, no. 1, p. 490, 2016.

[83] C. Illangakoon, R. D. McLeod, and M. R. Friesen, "Agent based modeling of malaria," in *2014 IEEE Canada International Humanitarian Technology Conference-(IHTC)*. IEEE, 2014, pp. 1–6.

[84] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *International Statistical Review/Revue Internationale de Statistique*, pp. 163–172, 1987.

[85] A. Jindal and S. Rao, "Agent-based modeling and simulation of mosquito-borne disease transmission," in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2017, pp. 426–435.

[86] E. K. Kelloway, *Using LISREL for structural equation modeling: A researcher's guide*. Sage, 1998.

[87] S. Kibret, G. G. Wilson, D. Ryder, H. Tekie, and B. Petros, "Environmental and meteorological factors linked to malaria transmission around large dams at three ecological settings in ethiopia," *Malaria journal*, vol. 18, no. 1, p. 54,2019.

[88] T. K. Kim, "T test as a parametric statistic," *Korean journal of anesthesiology*, vol. 68, no. 6, p. 540, 2015.

[89] R. B. Kline, *Principles and practice of structural equation modeling*. Guilford publications, 2015.

[90] L. Kong, J. Wang, W. Han, and Z. Cao, "Modeling heterogeneity in direct infectious disease transmission in a compartmental model," *International journal of environmental research and public health*, vol. 13, no. 3, p. 253,2016.

[91] V. Kumar, A. Mangal, S. Panesar, G. Yadav, R. Talwar, D. Raut, and S. Singh, "Forecasting malaria cases using climatic factors in delhi, india: a time series analysis," *Malaria research and treatment*, vol. 2014, 2014.

[92] V. Lakshmikantham and S. Leela, *Differential and Integral Inequalities: Theory and Applications: Volume I: Ordinary Differential Equations*. Academic press, 1969.

[93] D. G. Lalloo, D. Shingadia, D. J. Bell, N. J. Beeching, C. J. Whitty, P. L. Chiodini *et al.*, "Uk malaria treatment guidelines 2016," *Journal of Infection*, vol. 72, no. 6, pp. 635–649, 2016.

[94] R. D. Ledesma, P. Valero-Mora, and G. Macbeth, "The scree test and the number of factors: a dynamic graphics approach," *The Spanish journal of psychology*, vol. 18, 2015.

[95] X.-X. Li, L.-X. Wang, J. Zhang, Y.-X. Liu, H. Zhang, S.-W. Jiang, J.-X. Chen, and X.-N. Zhou, "Exploration of ecological factors related to the spatial heterogeneity of tuberculosis prevalence in pr china," *Global health action*, vol. 7, no. 1, p. 23620, 2014.

[96] S. Lindsay and M. Birley, "Climate change and malaria transmission," *Annals of Tropical Medicine & Parasitology*, vol. 90, no. 5, pp. 573–588, 1996.

[97] J.-B. Lohmöller, *Latent variable path modeling with partial least squares*. Springer Science & Business Media, 2013.

[98] Y. Lou and X.-Q. Zhao, "A climate-based malaria transmission model with

structured vector population," *SIAM Journal on Applied Mathematics*, vol. 70, no. 6, pp. 2023–2044, 2010.

[99] J. L. Lustgarten, V. Gopalakrishnan, H. Grover, and S. Visweswaran, "Improving classification performance with discretization on biomedical datasets," in *AMIA annual symposium proceedings*, vol. 2008. American Medical Informatics Association, 2008, p. 445.

[100] G. Macdonald *et al.*, "The epidemiology and control of malaria." *The Epidemiology and Control of Malaria.*, 1957.

[101] D. MacLeod and A. Morse, "Visualizing the uncertainty in the relationship between seasonal average climate and malaria risk," *Scientific reports*, vol. 4, p. 7264, 2014.

[102] B. Makanga, P. Yangari, N. Rahola, V. Rougeron, E. Elguero, L. Boundenga, N. D. Moukodoum, A. P. Okouga, C. Arnathau, P. Durand *et al.*, "Ape malaria transmission and potential for ape-to-human transfers in africa," *Proceedings of the National Academy of Sciences*, vol. 113, no. 19, pp. 5329–5334, 2016.

[103] S. Mandal, R. R. Sarkar, and S. Sinha, "Mathematical models of malaria-a review," *Malaria journal*, vol. 10, no. 1, pp. 1–19, 2011.

[104] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner, "A methodology for performing global uncertainty and sensitivity analysis in systems biology," *Journal of theoretical biology*, vol. 254, no. 1, pp. 178–196, 2008.

[105] J. M. Marshall, S. L. Wu, S. S. Kiware, M. Ndhlovu, A. L. Ouédraogo, M. B. Touré, H. J. Sturrock, A. C. Ghani, N. M. Ferguson *et al.*, "Mathematical models

of human mobility of relevance to malaria transmission in africa," *Scientific reports*, vol. 8, no. 1, p. 7713, 2018.

[106] W. Martens, T. Jetten, J. Rotmans, and L. Niessen, "Climate change and vector-borne diseases: a global modelling perspective," *Global environmental change*, vol. 5, no. 3, pp. 195–209, 1995.

[107] D. M. Maslove, T. Podchiyska, and H. J. Lowe, "Discretization of continuous features in clinical datasets," *Journal of the American Medical Informatics Association*, vol. 20, no. 3, pp. 544–553, 2012.

[108] A. Methaila, P. Kansal, H. Arya, P. Kumar *et al.*, "Early heart disease prediction using data mining techniques," *Computer Science & Information Technology Journal*, pp. 53–59, 2014.

[109] B. Modu, A. T. Asyhari, and Y. Peng, "Data analytics of climatic factor influence on the impact of malaria incidence," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2016, pp. 1–8.

[110] P. N. L. Y. . K. S. Modu, B., "Machine learning analysis and agent-based modelling of malaria transmission," *Frontiers in Artificial Intelligence and Applications*, vol. 309, no. 2, pp. 465–472, 2018.

[111] A. Monecke and F. Leisch, "sempls: structural equation modeling using partial least squares," 2012.

[112] E. A. Mordecai, K. P. Paaijmans, L. R. Johnson, C. Balzer, T. Ben-Horin, E. de Moor, A. McNally, S. Pawar, S. J. Ryan, T. C. Smith *et al.*, "Optimal

temperature for malaria transmission is dramatically lower than previously predicted," *Ecology letters*, vol. 16, no. 1, pp. 22–30, 2013.

[113] A. Y. Mukhtar, J. B. Munyakazi, and R. Ouifki, "Assessing the role of climate factors on malaria transmission dynamics in south sudan," *Mathematical biosciences*, vol. 310, pp. 13–23, 2019.

[114] S. S. Myers, L. Gaffikin, C. D. Golden, R. S. Ostfeld, K. H. Redford, T. H. Ricketts, W. R. Turner, and S. A. Osofsky, "Human health impacts of ecosystem alteration," *Proceedings of the National Academy of Sciences*, vol. 110, no. 47, pp. 18 753–18760, 2013.

[115] S. S. Myers and J. A. Patz, "Emerging threats to human health from global environmental change," *Annual Review of Environment and Resources*, vol. 34, pp. 223–252, 2009.

[116] K. Nah, Y. Nakata, and G. Röst, "Malaria dynamics with long incubation period in hosts," *Computers & Mathematics with Applications*, vol. 68, no. 9, pp. 915–930, 2014.

[117] N. Nanvyat, C. Mulambalah, Y. Barshep, J. Ajiji, D. Dakul, and H. Tsingalia, "Malaria transmission trends and its lagged association with climatic factors in the highlands of plateau state, nigeria," *Tropical parasitology*, vol. 8, no. 1, p. 18, 2018.

[118] D. C. Nath and D. D. Mwchahary, "Association between climatic variables and malaria incidence: A study in kokrajhar district of assam, india: Climatic variables and malaria incidence in kokrajhar district," *Global journal of health science*, vol. 5, no. 1, p. 90, 2013.

[119] A. Nejat and I. Damnjanovic, "Agent-based modeling of behavioral housing recovery following disasters," *Computer-Aided Civil and Infrastructure Engineering*, vol. 27, no. 10, pp. 748–763, 2012.

[120] E. Ngarakana-Gwasira, C. Bhunu, M. Masocha, and E. Mashonjowa, "Assessing the role of climate change in malaria transmission in africa," *Malaria research and treatment*, vol. 2016, 2016.

[121] G. A. Ngwa, A. M. Niger, and A. B. Gumel, "Mathematical assessment of the role of non-linear birth and maturation delay in the population dynamics of the malaria vector," *Applied Mathematics and Computation*, vol. 217, no. 7, pp. 3286–3313, 2010.

[122] G. A. Ngwa and W. S. Shu, "A mathematical model for endemic malaria with variable human and mosquito populations," *Mathematical and Computer Modelling*, vol. 32, no. 7-8, pp. 747–763, 2000.

[123] L.-F. Nie and Y.-N. Xue, "The roles of maturation delay and vaccination on the spread of dengue virus and optimal control," *Advances in Difference Equations*, vol. 2017, no. 1, p. 278, 2017.

[124] A. M. Niger and A. B. Gumel, "Mathematical analysis of the role of repeated exposure on malaria transmission dynamics," *Differential Equations and Dynamical Systems*, vol. 16, no. 3, pp. 251–287, 2008.

[125] C. Nitzl and W. W. Chin, "The case of partial least squares (pls) path modeling in managerial accounting research," *Journal of Management Control*, vol. 28, no. 2, pp. 137–156, 2017.

[126] A. A. Nobre, A. M. Schmidt, and H. F. Lopes, "Spatio-temporal models for mapping the incidence of malaria in pará," *Environmetrics: The official journal of the International Environmetrics Society*, vol. 16, no. 3, pp. 291–304, 2005.

[127] P. Nyarko, "Population and housing census, district analytical report, ejisu-juaben municipal." *https:// www.citypopulation.de/ php/ ghana-admin.php? adm2id\=0117(accessed on 12 January, 2017)*, 2015.

[128] J. R. Ohm, F. Baldini, P. Barreaux, T. Lefevre, P. A. Lynch, E. Suh, S. A. Whitehead, and M. B. Thomas, "Rethinking the extrinsic incubation period of malaria parasites," *Parasites & vectors*, vol. 11, no. 1, p. 178, 2018.

[129] K. Okuneye, A. Abdelrazec, and A. B. Gumel, "Mathematical analysis of a weather-driven model for the population ecology of mosquitoes," *Mathematical Biosciences & Engineering*, vol. 15, no. 1, pp. 57–93, 2018.

[130] K. Okuneye, S. E. Eikenberry, and A. B. Gumel, "Weather-driven malaria transmission model with gonotrophic and sporogonic cycles," *Journal of biological dynamics*, pp. 1–37, 2019.

[131] K. Okuneye and A. B. Gumel, "Analysis of a temperature-and rainfall-dependent model for malaria transmission dynamics," *Mathematical biosciences*, vol. 287, pp. 72–92, 2017.

[132] K. P. Paaijmans, A. F. Read, and M. B. Thomas, "Understanding the link between malaria risk and climate," *Proceedings of the National Academy of Sciences*, vol. 106, no. 33, pp. 13 844–13 849, 2009.

[133] P. E. Parham, D. Pople, C. Christiansen-Jucht, S. Lindsay, W. Hinsley, and

E. Michael, "Modeling the role of environmental variables on the population dynamics of the malaria vector anopheles gambiae sensu stricto," *Malaria Journal*, vol. 11, no. 1, p. 271, 2012.

[134] P. E. Parham and E. Michael, "Modeling the effects of weather and climate change on malaria transmission," *Environmental health perspectives*, vol. 118, no. 5, pp. 620–626, 2009.

[135] T. A. Perkins, T. W. Scott, A. Le Menach, and D. L. Smith, "Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission," *PLoS computational biology*, vol. 9, no. 12, p. e1003327, 2013.

[136] T. A. Perkins, A. S. Siraj, C. W. Ruktanonchai, M. U. Kraemer, and A. J. Tatem, "Model-based projections of zika virus infections in childbearing women in the americas," *Nature microbiology*, vol. 1, no. 9, p. 16126, 2016.

[137] F. Pizzitutti, W. Pan, B. Feingold, B. Zaitchik, C. A. Álvarez, and C. F. Mena, "Out of the net: An agent-based model to study human movements influence on local-scale malaria transmission," *PloS one*, vol. 13, no. 3, p. e0193493, 2018.

[138] L. C. Pollitt, J. T. Bram, S. Blanford, M. J. Jones, and A. F. Read, "Existing infection facilitates establishment and density of malaria parasites in their mosquito vector," *PLoS pathogens*, vol. 11, no. 7, p. e1005003, 2015.

[139] R. C. Reiner Jr, A. Le Menach, S. Kunene, N. Ntshalintshali, M. S. Hsiang, T. A. Perkins, B. Greenhouse, A. J. Tatem, J. M. Cohen, and D. L. Smith, "Mapping residual transmission for malaria elimination," *Elife*, vol. 4, p. e09520, 2015.

[140] R. Ross, "The prevention of malaria. 1911," *London: Murray*, vol. 2.

[141] R. Ross, *Report on the Prevention of Malaria in Mauritius*. Waterlow, 1908.

[142] J. Ruscio and B. Roche, "Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure." *Psychological assessment*, vol. 24, no. 2, p. 282, 2012.

[143] M. Salehi, K. Mohammad, M. M. Farahani, H. Zeraati, K. Nourijelyani, and F. Zayeri, "Spatial modeling of malaria incidence rates in sistan and baluchistan province, islamic republic of iran," *Saudi Med J*, vol. 29, no. 12, pp. 1791–1796, 2008.

[144] C. D. Scarbrough Lefebvre, A. Terlinden, and B. Standaert, "Dissecting the indirect effects caused by vaccines into the basic elements," *Human vaccines & immunotherapeutics*, vol. 11, no. 9, pp. 2142–2157, 2015.

[145] H. J. Schellnhuber, W. Hare, O. Serdeczny, S. Adams, D. Coumou, K. Frieler, M. Martin, I. M. Otto, M. Perrette, and A. Robinson, "Turn down the heat: why a 4 c warmer world must be avoided," 2012.

[146] G. S. Services, "Population and housing census. district analytical report, accra metropolitan," 2015.

[147] C. Shan, G. Fan, and H. Zhu, "Periodic phenomena and driving mechanisms in transmission of west nile virus with maturation time," *Journal of Dynamics and Differential Equations*, pp. 1–24, 2019.

[148] L. L. Shapiro, S. A. Whitehead, and M. B. Thomas, "Quantifying the effects of temperature on mosquito and parasite traits that determine the transmission potential of human malaria," *PLoS biology*, vol. 15, no. 10, p. e2003489, 2017.

[149] V. Sharma, A. Kumar, D. Lakshmi Panat, G. Karajkhede *et al.*, "Malaria outbreak prediction model using machine learning," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 4, no. 12, 2015.

[150] C. I. Siettos and L. Russo, "Mathematical modeling of infectious disease dynamics," *Virulence*, vol. 4, no. 4, pp. 295–306, 2013.

[151] B. Singh and C. Daneshvar, "Human infections and detection of plasmodium knowlesi," *Clinical microbiology reviews*, vol. 26, no. 2, pp. 165–184, 2013.

[152] D. L. Smith, T. A. Perkins, R. C. Reiner Jr, C. M. Barker, T. Niu, L. F. Chaves, A. M. Ellis, D. B. George, A. Le Menach, J. R. Pulliam *et al.*, "Recasting the theory of mosquito-borne pathogen transmission dynamics and control," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 108, no. 4, pp. 185–197, 2014.

[153] N. R. Smith, J. M. Trauer, M. Gambhir, J. S. Richards, R. J. Maude, J. M. Keith, and J. A. Flegg, "Agent-based models of malaria transmission: a systematic review," *Malaria journal*, vol. 17, no. 1, p. 299, 2018.

[154] N. Srinivasulu, B. Gujju Gandhi, R. Naik, and S. Daravath, "Influence of climate change on malaria incidence in mahaboobnagar district of andhra pradesh, india," *International Journal of Current Microbiology and. Applied Science*, vol. 2, no. 5, pp. 256–266, 2013.

[155] T. V. Sriram, M. V. Rao, G. S. Narayana, D. Kaladhar, and T. P. R. Vital, "Intelligent parkinson disease prediction using machine learning algorithms," *Int. J. Eng. Innov. Technol*, vol. 3, pp. 212–215, 2013.

[156] W. Sriwattanapongse, S. Me-ead, and S. Khanabsakdi, "Forecasting malaria incidence based on monthly case reports and climatic factors in ubon ratchathani province, thailand, 2000-2009," *Science Journal Ubon Ratchathani University*, vol. 2, pp. 17–24, 2011.

[157] W. Stone, B. P. Gonçalves, T. Bousema, and C. Drakeley, "Assessing the infectious reservoir of falciparum malaria: past and future," *Trends in parasitology*, vol. 31, no. 7, pp. 287–296, 2015.

[158] S. Takyi Appiah, H. Otoo, and I. Nabubie, "Times series analysis of malaria cases in ejisu-juaben municipality," *Int. J. Sci. Technol. Res*, vol. 4, pp. 220–226, 2015.

[159] N. Tejedor-Garavito, N. Dlamini, D. Pindolia, A. Soble, N. W. Ruktanonchai, V. Alegana, A. Le Menach, N. Ntshalintshali, B. Dlamini, D. L. Smith *et al.*, "Travel patterns and demographic characteristics of malaria cases in swaziland, 2010–2014," *Malaria journal*, vol. 16, no. 1, p. 359, 2017.

[160] M. Tenenhaus, V. E. Vinzi, Y.-M. Chatelin, and C. Lauro, "Pls path modeling," *Computational statistics & data analysis*, vol. 48, no. 1, pp. 159–205, 2005.

[161] B. Tesla, L. R. Demakovsky, E. A. Mordecai, M. H. Bonds, C. N. Ngonghala, M. A. Brindley, and C. C. Murdock, "Impacts of temperature on zika virus transmission potential: combining empirical and mechanistic modeling approaches," *bioRxiv*, p. 259531, 2018.

[162] M. X. Tong, A. Hansen, S. Hanson-Easey, S. Cameron, J. Xiang, Q. Liu, X. Liu, Y. Sun, P. Weinstein, G.-S. Han *et al.*, "Perceptions of malaria control and pre-

vention in an era of climate change: a cross-sectional survey among cdc staff in china," *Malaria journal*, vol. 16, no. 1, p. 136, 2017.

[163] B. Traoré, B. Sangaré, and S. Traoré, "A mathematical model of malaria transmission with structured vector population and seasonality," *Journal of Applied Mathematics*, vol. 2017, 2017.

[164] A. Vaidya, A. D. Bravo-Salgado, and A. R. Mikler, "Modeling climate-dependent population dynamics of mosquitoes to guide public health policies," in *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*. ACM, 2014, pp. 380–389.

[165] P. Van den Driessche and J. Watmough, "Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission," *Mathematical biosciences*, vol. 180, no. 1-2, pp. 29–48, 2002.

[166] G. M. Vazquez-Prokopec, T. A. Perkins, L. A. Waller, A. L. Lloyd, R. C. Reiner Jr, T. W. Scott, and U. Kitron, "Coupled heterogeneities and their impact on parasite transmission and control," *Trends in parasitology*, vol. 32, no. 5, pp. 356–367, 2016.

[167] P. Vignaroli, "Building resilience to drought in the sahel by early risk identification and advices," in *Renewing Local Planning to Face Climate Change in the Tropics*. Springer, Cham, 2017, pp. 151–167.

[168] E. Vynnycky and R. White, *An introduction to infectious disease modelling*. OUP oxford, 2010.

[169] H. Wan and H. Zhu, "A new model with delay for mosquito population dynamics," *Mathematical Biosciences and Engineering*, vol. 11, no. 6, pp. 1395–1410, 2014.

[170] L. Wang, *Support vector machines: theory and applications*. Springer Science & Business Media, 2005, vol. 177.

[171] W. Wang and X.-Q. Zhao, "Threshold dynamics for compartmental epidemic models in periodic environments," *Journal of Dynamics and Differential Equations*, vol. 20, no. 3, pp. 699–717, 2008.

[172] X. Wang and X.-Q. Zhao, "A periodic vector-bias malaria model with incubation period," *SIAM Journal on Applied Mathematics*, vol. 77, no. 1, pp. 181–201, 2017.

[173] Z. Wang and T. S. Deisboeck, "Agent-based modeling," *Encyclopedia of Systems Biology*, pp. 13–13, 2013.

[174] S. Westra, L. V. Alexander, and F. W. Zwiers, "Global increasing trends in annual maximum daily precipitation," *Journal of climate*, vol. 26, no. 11, pp. 3904–3918, 2013.

[175] M. White and J. Watson, "Malaria: Age, exposure and immunity," *eLife*, vol. 7, p. e40150, 2018.

[176] U. Wilensky and W. Rand, *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. MIT Press, 2015.

[177] M. B. Wilk and R. Gnanadesikan, "Probability plotting methods for the analysis for the analysis of data," *Biometrika*, vol. 55, no. 1, pp. 1–17, 1968.

[178] L. Willem, F. Verelst, J. Bilcke, N. Hens, and P. Beutels, "Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015)," *BMC infectious diseases*, vol. 17, no. 1, p. 612, 2017.

[179] H. Wold, "Soft modeling: the basic design and some extensions," *In Systems Under Indirect Observation: Causality-Structure-Prediction*, pp. 1–54, 1982.

[180] J. Wu, R. Dhingra, M. Gambhir, and J. V. Remais, "Sensitivity analysis of infectious disease models: methods, advances and their application," *Journal of The Royal Society Interface*, vol. 10, no. 86, p. 20121018, 2013.

[181] L. Xu, L. C. Stige, K.-S. Chan, J. Zhou, J. Yang, S. Sang, M. Wang, Z. Yang, Z. Yan, T. Jiang *et al.*, "Climate variation drives dengue dynamics," *Proceedings of the National Academy of Sciences*, vol. 114, no. 1, pp. 113–118, 2017.

[182] Y. Ye, *Environmental factors and malaria transmission risk: modelling the risk in a holoendemic area of Burkina Faso*. Routledge, 2017.

[183] K. A. Yeomans and P. A. Golder, "The guttman-kaiser criterion as a predictor of the number of common factors," *The Statistician*, pp. 221–229, 1982.

[184] E.-H. Yoo, D. Chen, C. Diao, and C. Russell, "The effects of weather and environmental factors on west nile virus mosquito abundance in greater toronto area," *Earth Interactions*, vol. 20, no. 3, pp. 1–22, 2016.

[185] O. P. Zacarias and H. Boström, "Comparing support vector regression and random forests for predicting malaria incidence in mozambique," in *2013 Interna-*

tional Conference on Advances in ICT for Emerging Regions (ICTer). IEEE, 2013, pp. 217–221.

[186] C. Zheng, F. Zhang, and J. Li, "Stability analysis of a population model with maturation delay and ricker birth function," in *Abstract and Applied Analysis*, vol. 2014. Hindawi, 2014.

[187] G. Zhou, N. Minakawa, A. K. Githeko, and G. Yan, "Association between climate variability and malaria epidemics in the east african highlands," *Proceedings of the National Academy of Sciences*, vol. 101, no. 8, pp. 2375–2380, 2004.

[188] Y. Zhu, Q. A. Wang, W. Li, and X. Cai, "Uncertainty and sensitivity analysis to complex systems," *International Journal of Modern Physics C*, vol. 28, no. 08, p. 1750109, 2017.

Appendix A

Source Code for Mathematical Modelling Simulation

A.1 source code for mathematical modelling simulation

Figure A.1: Malaria Transmission Model Simulation

```
library(deSolve)
Revised_Model2 <- function(t, y, parms, tau1, tau2) {
        tlag1 <- t-tau1
        tlag2 <- t-tau2
  if (tlag2 < 0) {
    smlag <- 1
    }
  else if (tlag2 == 0) {
    smlag <- 0.99
    }
  else {
    smlag <- lagvalue(t-tau2,6)
    }

  if (tlag2 < 0) {
    ihlag <- 0
    }
  else if (tlag2 == 0) {
        ihlag <- 0
  }
  else {
    ihlag <- lagvalue(t-tau2,3)
  }

  if (tlag1 < 0) {
        Amlag <- 0
  }
  else if (tlag1 == 0) {
        Amlag <- 1
  }
  else {
        Amlag <- lagvalue(t-tau1,5)
  }

  with(as.list(c(y, parms,tau1, tau2)),{
    dsh <- a-b*sh*im-c*sh+d*rh
    deh <- b*sh*im-e*eh-c*eh
    dih <- e*eh-c*ih-f*ih-g*ih
    drh <- g*ih-c*rh-d*rh
    dAm <- h*(1-Am/i)*(sm+es+el+im)-j*Am
           -k*Amlag*exp(-j*tau1)*exp(-Amlag)
    dsm <- k*Amlag*exp(-j*tau1)*exp(-Amlag)-m*sm*ih-n*sm
```

Figure A.2: Malaria Transmission Model Simulation Continued

```
  }
  )
}
yinit <- c(sh = 1, eh = 0, ih = 0, rh = 0,
           Am = 1, sm = 0.09, es = 0,
           el = 0, im = 0.01)

parms <- c(a = 0.00004, b = 0.00638, c = 0.00004, d = 0.04,
           e = 0.05, f = 0.0003454, g = 0.02, h = 200,
           i = 40000, j = 0.1, k = 0.344, m = 0.0696, n = 0.1,
           r = 0.25, p = 0.091)

times <- seq(0,1000)

result2 <- dede(y = yinit, times = times, func = Revised_Model2,
             parms = parms, tau1 = 30, tau2 = 10)
plot(result2[,'Am'], col = 'red', type = 'l', lwd = 3, las = 1, lty = 6)
```

Figure A.3: Temperature-Dependent Parameters Plots

```
> curve(-0.153*x^2 + 8.61*x + 97.7, 16, 34)
> curve(-0.153*x^2 + 8.61*x - 97.7, 15, 35)
> T <- c(16, 18, 20, 22, 24, 26, 28, 30, 32, 34)
> immature_death <- c(0.040, 0.045, 0.054, 0.066, 0.078, 0.088, 0.097,
                      0.103, 0.108, 0.110)
> plot(T, immature_death, xlim = c(15,35), ylim = c(0.01,0.12), las = 1,
      type = 'o', lty = 5, lwd = 1.2, col = 'red')
> grid()
> adult_death <- c(0.121, 0.103, 0.093, 0.091, 0.095, 0.107, 0.127,
                   0.155, 0.191, 0.237)
> plot(T, adult_death, xlim = c(15,35), las = 1, type = 'o', lty = 5,
      lwd = 1.2, col = 'red', ylim = c(0.05,0.25))
> grid()
> biting_rate <- c(0.074, 0.119, 0.162, 0.204, 0.245, 0.285, 0.324,
                   0.362, 0.399, 0.434)
> plot(T, biting_rate, xlim = c(15,35), las = 1, type = 'o', lty = 5,
      lwd = 1.2, col = 'red', ylim = c(0.05, 0.5))
> grid()
> egg_laid <- c(0.892, 7.708, 13.300, 17.668, 20.812, 22.732, 23.428,
                22.900, 21.148, 18.172)
> plot(T, egg_laid, xlim = c(15,35), las = 1, type = 'o', lty = 5,
      lwd = 1.2, col = 'red', ylim = c(0,25))
```

Figure A.4: Sensitivity Analysis Simulation

```
library(deSolve)
lambda_a <- function (beta_b, biting_rate){
        beta_b * biting_rate}
lambda_b <- function(beta_a, biting_rate){
        beta_a * biting_rate}
Revised_Model2 <- function(t, y, parms, tau1, tau2) {
  tlag1 <- t-tau1
  tlag2 <- t-tau2
  if (tlag2 < 0) {
    smlag <- 1
  }
  else if (tlag2 == 0) {
    smlag <- 0.09
  }
  else {
    smlag <- lagvalue(t-tau2,6)
  }

  if (tlag2 < 0) {
    ihlag <- 0
  }
  else if (tlag2 == 0) {
    ihlag <- 0
  }
  else {
    ihlag <- lagvalue(t-tau2,3)
  }

  if (tlag1 < 0) {
    Amlag <- 0
  }
  else if (tlag1 == 0) {
    Amlag <- 1
  }
  else {
    Amlag <- lagvalue(t-tau1,5)
  }
```

Figure A.5: Sensitivity Analysis Simulation Continued

```
  with(as.list(c(y, parms,tau1, tau2)),{
        lambda_a <- beta_b * biting_rate
        lambda_b <- beta_a * biting_rate
             dsh <- a-lambda_a*sh*im-c*sh+d*rh
             deh <- lambda_a*sh*im-e*eh-c*eh
             dih <- e*eh-c*ih-f*ih-g*ih
             drh <- g*ih-c*rh-d*rh
             dAm <- h*(1-Am/i)*(sm+es+el+im)-j*Am
                    -k*Amlag*exp(-j*tau1)*exp(-Amlag)
             dsm <- k*Amlag*exp(-j*tau1)*exp(-Amlag)
                    -lambda_b*sm*ih-n*sm
             des <- 0.25*lambda_b*sm*ih-n*el-p*es
             del <- (1-0.25)*lambda_b*sm*ih-
                    (1-0.25)*lambda_b*smlag*ihlag*exp(-n*tau2)-n*el
             dim <- p*es+(1-0.25)*lambda_b*smlag*ihlag*exp(-n*tau2)-n*im
        list(c(dsh, deh, dih, drh, dAm, dsm, des, del, dim))
  }
  )
}
yinit <- c(sh = 1, eh = 0, ih = 0, rh = 0,
           Am = 1, sm = 0.09, es = 0,
           el = 0, im = 0.01)

parms <- c(
        a = 0.00004,
        lambda_a = 0.00638,
        c = 0.00004,
        d = 0.003,
           e = 0.02,
           f = 0.0003454,
           g = 0.02,
           h = 5,
           i = 40000,
           j = 0.1,
           k = 0.344,
           lambda_b = 0.0696,
           n = 0.1,
           p = 0.091,
           beta_a = 0.24,
           beta_b = 0.022,
           biting_rate = 0.29,
           tau1 = 5,
```

Figure A.6: Sensitivity Analysis Simulation Continued

```
           tau2 = 14)

times <- seq(0,1000)

result2 <- dede(y = yinit, times = times, func = Revised_Model2,
                parms = parms, tau1 = 5, tau2 = 14)

require(lhs)
h <- 10000
lhs <- maximinLHS(h, 16)

a.min <- 0.0000391
a.max <- 0.00004
c.min <- 0.0000342
c.max <- 0.0000391
d.min <- 0.000055
d.max <- 0.011
e.min <- 0.002
e.max <- 0.005
f.min <- 0.00000
f.max <- 0.00041
g.min <- 0.0014
g.max <- 0.017
h.min <- 1
h.max <- 500
i.min <- 50
i.max <- 3300000
j.min <- 0.001
j.max <- 0.2
k.min <- 0.333
k.max <- 1
n.min <- 0.0476
n.max <- 0.0714
p.min <- 0.029
p.max <- 0.33
tau2.min <- 7
tau2.max <- 21
beta_a.min <- 0.072
beta_a.max <- 0.64
beta_b.min <- 0.0027
beta_b.max <- 0.64
```

Figure A.7: Sensitivity Analysis Simulation Continued

```
biting_rate.min <- 0.10
biting_rate.max <- 1.00

parms.set <- cbind(
                a = lhs[,1]*(a.max - a.min) + a.min,
                c = lhs[,2]*(c.max - c.min) + c.min,
                d = lhs[,3]*(d.max - d.min) + d.min,
                e = lhs[,4]*(e.max - e.min) + e.min,
                f = lhs[,5]*(f.max - f.min) + f.min,
                g = lhs[,6]*(g.max - g.min) + g.min,
                h = lhs[,7]*(h.max - h.min) + h.min,
                i = lhs[,8]*(i.max - i.min) + i.min,
                j = lhs[,9]*(j.max - j.min) + j.min,
                k = lhs[,10]*(k.max - k.min) + k.min,
                n = lhs[,11]*(n.max - n.min) + n.min,
                p = lhs[,12]*(p.max - p.min) + p.min,
                beta_a = lhs[,13]*(beta_a.max - beta_a.min) + beta_a.min,
                beta_b = lhs[,14]*(beta_b.max - beta_b.min) + beta_b.min,
                biting_rate = lhs[,15]*(biting_rate.max - biting_rate.min)
                              + biting_rate.min,
                tau2 = lhs[,17]*(tau2.max - tau2.min) + tau2.min

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
parms.set <- cbind(
                a = lhs[,1]*(a.max - a.min) + a.min,
                c = lhs[,2]*(c.max - c.min) + c.min,
                d = lhs[,3]*(d.max - d.min) + d.min,
                e = lhs[,4]*(e.max - e.min) + e.min,
                f = lhs[,5]*(f.max - f.min) + f.min,
                g = lhs[,6]*(g.max - g.min) + g.min,
                h = lhs[,7]*(h.max - h.min) + h.min,
                i = lhs[,8]*(i.max - i.min) + i.min,
                j = lhs[,9]*(j.max - j.min) + j.min,
                k = lhs[,10]*(k.max - k.min) + k.min,
                n = lhs[,11]*(n.max - n.min) + n.min,
                p = lhs[,12]*(p.max - p.min) + p.min,
                beta_a = lhs[,13]*(beta_a.max - beta_a.min) + beta_a.min,
                beta_b = lhs[,14]*(beta_b.max - beta_b.min) + beta_b.min,
                biting_rate = lhs[,15]*(biting_rate.max - biting_rate.min)
                              + biting_rate.min,
                tau1 = lhs[,16]*(tau1.max - tau1.min) + tau1.min
```

Figure A.8: Sensitivity Analysis Simulation Continued

```
l <- 21
h2 <- 1000

j <- 1
data <- data.frame(matrix(rep(NA, l*h2*17), nrow = l*h2))
for(i in 1:h2){
        for (T in seq(7, 21, length = l)){
                yinit <- c(sh = 1, eh = 0, ih = 0, rh = 0,
                        Am = 1, sm = 0.09, es = 0,
                        el = 0, im = 0.01)
                parms <- as.list(c(parms.set[i,], tau2 = T))
                result2 <- as.data.frame(dede(y = yinit, times = times,
                           func = Revised_Model2,
                parms = parms,  tau1 = 5, tau2 = T))
                data[j,1:17] <- parms
                infectious_humans <- sum(data.frame(result2$ih))
                data[j,18] <- infectious_humans
                j <- j+1
        }
}

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

parms.set <- cbind(
                a = lhs[,1]*(a.max - a.min) + a.min,
                c = lhs[,2]*(c.max - c.min) + c.min,
                d = lhs[,3]*(d.max - d.min) + d.min,
                e = lhs[,4]*(e.max - e.min) + e.min,
                f = lhs[,5]*(f.max - f.min) + f.min,
                g = lhs[,6]*(g.max - g.min) + g.min,
                h = lhs[,7]*(h.max - h.min) + h.min,
                i = lhs[,8]*(i.max - i.min) + i.min,
                j = lhs[,9]*(j.max - j.min) + j.min,
                k = lhs[,10]*(k.max - k.min) + k.min,
                n = lhs[,11]*(n.max - n.min) + n.min,
                p = lhs[,12]*(p.max - p.min) + p.min,
                beta_a = lhs[,13]*(beta_a.max - beta_a.min) + beta_a.min,
                beta_b = lhs[,14]*(beta_b.max - beta_b.min) + beta_b.min,
                biting_rate = lhs[,15]*(biting_rate.max - biting_rate.min)
                             + biting_rate.min,
                tau2 = lhs[,16]*(tau2.max - tau2.min) + tau2.min
```

Figure A.9: Sensitivity Analysis Simulation Continued

```
l <- 14
h2 <- 1000

j <- 1
data <- data.frame(matrix(rep(NA, l*h2*17), nrow = l*h2))
for(i in 1:h2){
        for (T in seq(2, 14, length = l)){
                yinit <- c(sh = 1, eh = 0, ih = 0, rh = 0,
                        Am = 1, sm = 0.09, es = 0,
                        el = 0, im = 0.01)
                parms <- as.list(c(parms.set[i,], tau1 = T))
                result2 <- as.data.frame(dede(y = yinit, times = times,
                           func = Revised_Model2,
                parms = parms,  tau1 = T, tau2 = 14))
                data[j,1:17] <- parms
                infectious_humans <- sum(data.frame(result2$Am))
                data[j,18] <- infectious_humans
                j <- j+1
        }
}


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
names(data) <- c(names(parms), 'infecetious_humans')
save(data, file = 'data.Rdata')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

tau1 = lhs[,16]*(tau1.max - tau1.min) + tau1.min)
tau1.min <- 2
tau1.max <- 14



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
require(sensitivity)
prcc <- pcc(data[,1:17], data[,18], nboot = 100, rank=TRUE)

plot(prcc)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Figure A.10: Sensitivity Analysis Simulation Continued

```
plot(result2[,'ih'], col = 'blue', type = 'l', lwd = 2, las = 1, lty = 1)
grid()

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


l <- 21
h3 <- 5
j <- 1
data <- data.frame(matrix(rep(NA, l*h3*18), nrow = l*h3))
for(i in 1:h3){
        for (T in seq(7:21)){
                yinit <- c(sh = 1, eh = 0, ih = 0, rh = 0,
                        Am = 1, sm = 0.09, es = 0,
                        el = 0, im = 0.01)
                parms <- as.list(c(parms.set[i,], tau2 = T))
                result2 <- as.data.frame(dede(y = yinit, times = times,
                           func = Revised_Model2,
                parms = parms,  tau1 = 5, tau2 = T))
                data[j,1:18] <- parms
                j <- j+1
        }
}


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

incubation.periods <- seq(5,21)
j <- 4
data <- data.frame(matrix(rep(NA, levels*h2*18), nrow = levels*h2))
for (i in 1:h2){
        for (T in incubation.periods){
                data[j, 1:18] <- parms
                parms <- as.list(c(parms.set[i,], T = T))
                result2 <- as.data.frame(dede(y = yinit, times = times,
                           func = Revised_Model2,
                parms = parms,  tau1 = 5, tau2 = 14))
                data[j, 18] <- get.size(result2)
                j <- j+1
        }
}
```

Figure A.11: Sensitivity Analysis Simulation Continued

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

infectious.humans.baseline <- c()
for (T in seq(7,21, by = 1)){
        parms <- c(
        a = 0.00004,
        lambda_a = 0.00638,
        c = 0.00004,
        d = 0.003,
           e = 0.02,
           f = 0.0003454,
           g = 0.02,
           h = 5,
           i = 40000,
           j = 0.1,
           k = 0.344,
           lambda_b = 0.0696,
           n = 0.1,
           p = 0.091,
           beta_a = 0.24,
           beta_b = 0.022,
           biting_rate = 0.29,
           tau1 = 5,
           tau2 = 14)
result2 <- as.data.frame(dede(y = yinit, times = times, func = Revised_Model2,
                parms = parms,  tau1 = 5, tau2 = 14))
                number.infection <- plot(result2[,'ih'])

}

Call:
pcc(X = data[, 1:17], y = data[, 18], rank = TRUE, nboot = 500)

Partial Rank Correlation Coefficients (PRCC):


                                          prcc for tau2
       original          bias  std. error    min. c.i.    max. c.i.
X1   0.024894563  5.354390e-04 0.007514897  0.009824972  0.038869702
X2   0.010559965 -3.226914e-04 0.007204945 -0.002632983  0.025304253
X3   0.024243175  3.304834e-04 0.006492467  0.011158419  0.036085610
X4   0.185865947  2.095901e-04 0.007406250  0.170978656  0.200158331
```

Figure A.12: Sensitivity Analysis Simulation Continued

```
X5  -0.042780450  2.230327e-04 0.007171836 -0.057375557 -0.029814277
X6  -0.896335023 -5.977869e-05 0.001613124 -0.899425363 -0.893104111
X7  -0.103676995  5.815515e-06 0.007106229 -0.117136301 -0.089287751
X8  -0.002797706 -2.434767e-04 0.007067410 -0.016870435  0.010382585
X9   0.099240183  1.576523e-04 0.007072193  0.083497750  0.112855637
X10 -0.031580170 -3.282733e-04 0.007420851 -0.045959994 -0.015203657
X11 -0.479524870  1.103502e-04 0.005714085 -0.491116348 -0.468599293
X12  0.035432129  5.355905e-04 0.007027232  0.022020792  0.048900816
X13  0.036417805 -4.607437e-05 0.007465180  0.022204036  0.052178073
X14  0.931654546  3.188178e-05 0.001003265  0.929708513  0.933680401
X15  0.901266188  2.599936e-05 0.001565134  0.897801907  0.904504058
X16 -0.012080445 -8.566321e-05 0.007076205 -0.025737881  0.002791445
X17 -0.324108907 -1.462623e-04 0.006523702 -0.337133555 -0.311098591



Call:
pcc(X = data[, 1:17], y = data[, 18], rank = TRUE, nboot = 500)

Partial Rank Correlation Coefficients (PRCC):

                                          prcc for tau1
         original          bias   std. error    min. c.i.    max. c.i.
X1  -0.0031219070 -4.126251e-04 0.0092492980 -0.019565483  0.017592460
X2  -0.0006570649 -1.355944e-04 0.0091901413 -0.018891174  0.018057696
X3  -0.0046904130  6.727717e-04 0.0088787059 -0.023031200  0.011958728
X4   0.0076760303  2.805390e-06 0.0088333978 -0.008599827  0.023934431
X5   0.0043661210 -2.829765e-04 0.0086887158 -0.014470113  0.021658842
X6  -0.0414425324 -2.563443e-04 0.0088149234 -0.058663281 -0.024724225
X7   0.8066443851  1.738023e-04 0.0033014263  0.800266502  0.813042709
X8  -0.0033205923  5.495855e-05 0.0086135985 -0.020367362  0.013577285
X9  -0.7981304527 -5.356295e-04 0.0033466293 -0.803654728 -0.790656542
X10  0.3642323939 -2.909949e-04 0.0076125379  0.349467831  0.379364514
X11  0.2000970862  1.976331e-04 0.0093635701  0.182085179  0.218774830
X12 -0.0109740082 -2.603531e-04 0.0087763436 -0.029079057  0.006262002
X13  0.0136949635 -4.660599e-04 0.0091823830 -0.003824351  0.032543018
X14  0.0249570487  2.156335e-04 0.0084300692  0.008301513  0.041550551
X15  0.0253839691  6.917826e-05 0.0087016361  0.008342611  0.041718386
X16 -0.4366205283 -4.037775e-04 0.0080452879 -0.452635472 -0.420207813
X17  0.9825695341  2.640122e-05 0.0003523282  0.981851491  0.983203662
```

Figure A.13: PRCC plot

```
> x <- c(0.0249, 0.0106, 0.0242, 0.1859, -0.0428, -0.8963, 0.8066, 0.7728,
         -0.7981, 0.3642, -0.4795, 0.6354, 0.3364, 0.9317, 0.9013, -0.4366,
          0.9826)
> M <- c('a', 'b',  'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k', 'l', 'm',
         'n', 'o', 'p', 'q')
> barplot(x, names.arg = M, xlab = "Parameters", ylab = "PRCC values",
          col = 'skyblue', width = 1, ylim = c(-1,1), las = 1)

> abline(h = 0.5, lwd = 0.5, col = 'red', lty = 2)
> abline(h = 0.5, lwd = 1, col = 'red', lty = 2)
> abline(h = -0.5, lwd = 1, col = 'red', lty = 2)
```

Appendix B

Source Code for Data Analytics Modelling

Figure B.1: Data used for SEM-PLS Modelling

```
> confdata
   malaria_data maxtemp mintemp   precipdata  relhumid  solarrad windspeed
1          1427  23.038  21.991  0.013732898 0.9718059  0.000000  1.373885
2          1471  37.299  21.631  0.511550892 0.6543729 17.690064  1.404738
3          1267  35.493  21.497  0.106430040 0.6180652 19.365021  1.600264
4          1288  37.367  21.080  0.089263908 0.6789334 14.477004  1.445897
5          1431  34.505  21.289  2.156065142 0.6829567 19.728779  1.599487
6          1161  35.107  20.975  0.166511736 0.6818188 18.829634  1.448409
7          1209  37.996  20.919  0.000000000 0.5829214 21.278807  1.289580
8          1608  37.653  21.623  0.192260750 0.5897548 20.192340  1.487888
9          1444  36.633  23.844  0.190544220 0.6360240 13.896901  1.943205
10         1130  33.913  21.918  4.123302638 0.7798013 15.359230  1.635222
11         1142  34.319  20.457  0.267791472 0.7059637 13.730515  1.342044
12         1400  39.491  20.533  0.000000000 0.5823346 20.726222  1.649133
13         1759  37.775  21.863  0.027465818 0.6889962 15.145878  1.624055
14         1233  37.980  22.244  0.127029398 0.6453784 17.060791  1.209895
15         1848  37.255  23.206  0.556183058 0.6222124 19.875263  1.296791
16         1449  36.767  18.543  0.296973936 0.4888293 19.798709  1.524810
17         1536  38.178  19.639  0.000000000 0.5665459 22.219146  1.567832
18          712  37.440  21.115  0.000000000 0.5687231 21.636116  1.735703
19         1977  36.676  20.320  0.027465833 0.5950711 18.084152  1.615036
20         1324  36.213  15.156  0.000000000 0.2833195 22.455795  1.180813
21         1823  34.855  13.796  0.000000000 0.1864515 22.751601  1.395827
22         1122  33.369  15.680  0.000000000 0.1390097 11.048119  1.420346
23         1056  35.309  14.014  0.000000000 0.1246412 21.351601  1.827167
24         1356  34.427  12.634  0.000000000 0.1860739 22.902575  1.953594
25         1658  36.681  15.735  0.000000000 0.2274495 22.782880  1.025375
26         1743  38.683  15.537  0.000000000 0.3475673 22.550735  1.219112
27         2118  39.255  17.526  0.000000000 0.4098196 22.856572  1.368659
28         2051  39.417  18.878  0.000000000 0.5508621 22.613228  1.265696
29         1704  37.234  20.266  0.000000000 0.5899579 22.389075  1.580708
30         1607  40.208  21.109  0.010299679 0.5782447 21.888115  1.618182
31         2456  39.759  20.858  0.000000000 0.6038565 20.165513  1.727813
32         2402  37.129  21.243  0.003433226 0.6682392 18.148559  1.996225
33         1107  38.927  21.176  0.000000000 0.6209560 18.301894  2.060307
34         1775  40.412  21.065  0.830840760 0.6312732 20.177604  1.830456
35         1492  39.686  21.655  0.000000000 0.6236662 18.392114  2.091920
36         1502  37.841  21.916  2.557754100 0.7252407 14.330021  1.866451
37          863  40.026  22.868  0.250625340 0.6311006 17.157093  1.770643
38         1386  40.349  22.994  0.861739164 0.7008553 18.530783  2.418214
39         1231  36.426  22.643  0.048065184 0.6836129 18.519968  2.576098
40         1294  32.500  22.271 13.556097360 0.8285222 19.277718  2.184496
```

Figure B.2: Data used for SEM-PLS Modelling Continued

```
41          1131  33.496  21.772  4.236599556 0.7737849 21.742538  1.806159
42          1131  39.304  22.153  0.411987600 0.6976882 16.334604  1.831639
43          1865  37.655  22.368  0.030899030 0.6899483 17.149330  1.906785
44          1164  36.944  22.764  0.264358253 0.7446740 14.488984  2.093225
45          1042  34.179  22.628  1.078033428 0.7748053 17.271841  2.212386
46          1186  37.768  22.331  0.161361734 0.7555730 16.859586  2.111509
47          1396  34.153  22.468  0.559616126 0.7777768 20.579906  2.349005
48          1393  30.805  21.397  3.374862480 0.8347659  9.743568  1.676898
49          1260  35.961  21.248  3.663253800 0.7961933 16.492217  1.522344
50          1156  32.130  22.391  4.744724342 0.8011281 14.606949  1.673668
51          1293  31.188  22.214  3.381728112 0.8484703 18.889990  1.710422
52          1293  33.605  21.060  0.106429975 0.7022577 22.266618  1.961750
53          1248  36.183  20.801  0.000000000 0.6515638 21.401100  1.895682
54          1777  38.077  21.437  0.044631936 0.6561550 20.251126  1.664907
55          2203  39.655  22.416  0.175094352 0.7298102 18.646035  2.107378
56          1422  37.112  22.687  1.194763320 0.7490903 22.304752  2.094981
57          1045  36.608  22.771  1.435088340 0.7569737 19.218991  2.301860
58          1568  35.533  22.698  0.130462668 0.6998090 20.595827  2.269967
59          1588  39.404  22.092  0.000000000 0.6756907 19.135749  1.950841
60          1035  38.151  21.830  2.509683120 0.7217011 20.347812  1.834415

> Modelsm <- read.csv("C:/Users/BMODU/Dropbox/Ghana Ecological
variables/Modelsm.txt", sep="")
>   View(Modelsm)
> Modelsm
  source target
1  fact1  fact2
2  fact2  fact3
3  fact3  fact1
> Modelmm
       source       target
1       fact1      mintemp
2       fact1     relhumid
3     mintemp malaria_data
4    relhumid malaria_data
5       fact2      maxtemp
6       fact2     solarrad
7     maxtemp malaria_data
8    solarrad malaria_data
9       fact3   precipdata
10 precipdata malaria_data
```

Figure B.3: SEM-PLS Modelling

```
> sm <- as.matrix(Modelsm)
> sm
     source  target
[1,] "fact1" "fact2"
[2,] "fact2" "fact3"
[3,] "fact3" "fact1"
> mm <- as.matrix(Modelmm)
> mm
      source       target
 [1,] "fact1"      "mintemp"
 [2,] "fact1"      "relhumid"
 [3,] "mintemp"    "malaria_data"
 [4,] "relhumid"   "malaria_data"
 [5,] "fact2"      "maxtemp"
 [6,] "fact2"      "solarrad"
 [7,] "maxtemp"    "malaria_data"
 [8,] "solarrad"   "malaria_data"
 [9,] "fact3"      "precipdata"
[10,] "precipdata" "malaria_data"
> data(confdata)
> Modelsm <- read.csv("C:/Users/BMODU/Dropbox/Ghana Ecological
  variables/Modelsm.txt", sep="")
>   View(Modelsm)
> sm <- as.matrix(Modelsm)
> sm
     source  target
[1,] "fact1" "fact2"
[2,] "fact2" "fact3"
> sweden.model <- plsm(data = confdata, strucmod = sm, measuremod = mm)
> sweden.model
$latent
[1] "fact1" "fact2" "fact3"

$manifest
[1] "mintemp"    "relhumid"   "maxtemp"    "solarrad"   "precipdata"

$strucmod
     source  target
[1,] "fact1" "fact2"
[2,] "fact2" "fact3"
```

Figure B.4: SEM-PLS Modelling Continued

```
$measuremod
     source  target
[1,] "fact1" "mintemp"
[2,] "fact1" "relhumid"
[3,] "fact2" "maxtemp"
[4,] "fact2" "solarrad"
[5,] "fact3" "precipdata"

$D
      fact1 fact2 fact3
fact1     0     1     0
fact2     0     0     1
fact3     0     0     0

$M
           fact1 fact2 fact3
mintemp        1     0     0
relhumid       1     0     0
maxtemp        0     1     0
solarrad       0     1     0
precipdata     0     0     1

$blocks
$blocks$fact1
[1] "mintemp"  "relhumid"
attr(,"mode")
[1] "A"

$blocks$fact2
[1] "maxtemp"  "solarrad"
attr(,"mode")
[1] "A"

$blocks$fact3
[1] "precipdata"
attr(,"mode")
[1] "A"
```

Figure B.5: SEM-PLS Modelling Continued

```
$order
[1] "generic"

attr(,"class")
[1] "plsm"
> summary(sweden.model)
           Length Class  Mode
latent      3     -none- character
manifest    5     -none- character
strucmod    4     -none- character
measuremod 10     -none- character
D           9     -none- numeric
M          15     -none- numeric
blocks      3     -none- list
order       1     -none- character
> mvpairs(model = sweden.model, data = confdata, LVs = "fact1")
> mvpairs(model = sweden.model, data = confdata, LVs = "fact2")
> mvpairs(model = sweden.model, data = confdata, LVs = "fact3")
> sweden_model <- sempls(model = sweden.model, data = confdata, wscheme = "A")
All 60 observations are valid.
Converged after 10 iterations.
Tolerance: 1e-07
Scheme: centroid
> summary(sweden_model)
                  Length Class      Mode
coefficients        2    data.frame list
path_coefficients   9    -none-     numeric
outer_loadings     15    -none-     numeric
cross_loadings     15    -none-     numeric
total_effects       9    -none-     numeric
inner_weights       9    -none-     numeric
outer_weights      15    -none-     numeric
blocks              0    -none-     NULL
factor_scores     180    -none-     numeric
data              300    -none-     numeric
scaled              1    -none-     logical
model               8    plsm       list
weighting_scheme    1    -none-     character
weights_evolution   4    data.frame list
sum1                1    -none-     logical
pairwise            1    -none-     logical
method              1    -none-     character
```

Figure B.6: SEM-PLS Modelling Continued

```
iterations          1    -none-     numeric
convCrit            1    -none-     character
verbose             1    -none-     logical
tolerance           1    -none-     numeric
maxit               1    -none-     numeric
N                   1    -none-     numeric
incomplete          0    -none-     numeric
Hanafi             33    -none-     numeric
> sweden_model <- sempls(model = sweden.model, data = confdata, wscheme = "B")
All 60 observations are valid.
Converged after 13 iterations.
Tolerance: 1e-07
Scheme: factorial
> sweden_model <- sempls(model = sweden.model, data = confdata, wscheme = "C")
All 60 observations are valid.
Converged after 13 iterations.
Tolerance: 1e-07
Scheme: path weighting
> sweden_model <- sempls(model = sweden.model, data = confdata, wscheme = "A")
All 60 observations are valid.
Converged after 10 iterations.
Tolerance: 1e-07
Scheme: centroid
> sweden_model
                      Path Estimate
lam_1_1     fact1 -> mintemp     0.95
lam_1_2    fact1 -> relhumid     0.99
lam_2_1     fact2 -> maxtemp     0.88
lam_2_2    fact2 -> solarrad     0.88
lam_3_1  fact3 -> precipdata     1.00
beta_1_2      fact1 -> fact2    -0.33
beta_2_3      fact2 -> fact3    -0.26
> sweden_model <- sempls(model = sweden.model, data = confdata, wscheme = "B")
All 60 observations are valid.
Converged after 13 iterations.
Tolerance: 1e-07
Scheme: factorial
```

Figure B.7: SEM-PLS Modelling Continued

```
> sweden_model
                        Path Estimate
lam_1_1     fact1 -> mintemp     0.95
lam_1_2    fact1 -> relhumid     0.99
lam_2_1     fact2 -> maxtemp     0.86
lam_2_2    fact2 -> solarrad     0.90
lam_3_1  fact3 -> precipdata     1.00
beta_1_2      fact1 -> fact2    -0.33
beta_2_3      fact2 -> fact3    -0.25
> sweden_model <- sempls(model = sweden.model, data = confdata, wscheme = "C")
All 60 observations are valid.
Converged after 13 iterations.
Tolerance: 1e-07
Scheme: path weighting
> sweden_model
                        Path Estimate
lam_1_1     fact1 -> mintemp     0.95
lam_1_2    fact1 -> relhumid     0.99
lam_2_1     fact2 -> maxtemp     0.86
lam_2_2    fact2 -> solarrad     0.90
lam_3_1  fact3 -> precipdata     1.00
beta_1_2      fact1 -> fact2    -0.33
beta_2_3      fact2 -> fact3    -0.25
> Modelmm <- read.csv("C:/Users/BMODU/Dropbox/Ghana Ecological
            variables/Modelmm.txt", sep="")
>   View(Modelmm)
> mm <- as.matrix(Modelmm)
> mm
     source  target
[1,] "fact1" "mintemp"
[2,] "fact1" "relhumid"
[3,] "fact1" "malaria_data"
[4,] "fact2" "malaria_data"
[5,] "fact2" "maxtemp"
[6,] "fact2" "solarrad"
[7,] "fact3" "malaria_data"
[8,] "fact3" "precipdata"
> sweden.model <- plsm(data = confdata, strucmod = sm, measuremod = mm)
> sweden.model
$latent
[1] "fact1" "fact2" "fact3"
```

Figure B.8: SEM-PLS Modelling Continued

```
$manifest
[1] "malaria_data" "mintemp"      "relhumid"     "malaria_data" "maxtemp"
[6] "solarrad"     "malaria_data" "precipdata"

$strucmod
     source  target
[1,] "fact1" "fact2"
[2,] "fact2" "fact3"

$measuremod
     source  target
[1,] "fact1" "malaria_data"
[2,] "fact1" "mintemp"
[3,] "fact1" "relhumid"
[4,] "fact2" "malaria_data"
[5,] "fact2" "maxtemp"
[6,] "fact2" "solarrad"
[7,] "fact3" "malaria_data"
[8,] "fact3" "precipdata"

$D
      fact1 fact2 fact3
fact1     0     1     0
fact2     0     0     1
fact3     0     0     0

$M
             fact1 fact2 fact3
malaria_data     1     1     1
mintemp          1     0     0
relhumid         1     0     0
malaria_data     1     1     1
maxtemp          0     1     0
solarrad         0     1     0
malaria_data     1     1     1
precipdata       0     0     1

$blocks
$blocks$fact1
[1] "malaria_data" "mintemp"      "relhumid"
attr(,"mode")
[1] "A"
```

Figure B.9: SEM-PLS Modelling Continued

```
$blocks$fact2
[1] "malaria_data" "maxtemp"      "solarrad"
attr(,"mode")
[1] "A"

$blocks$fact3
[1] "malaria_data" "precipdata"
attr(,"mode")
[1] "A"


$order
[1] "generic"

attr(,"class")
[1] "plsm"

> parallelplot(boot_model, pattern = "beta", reflinesAt = c(0,0.5),
  alpha = 0.3, type = "bca", main = "Path coefficient\nof 500
  bootstrap samples")
> boot_model <- bootsempls(fit_model, nboot = 1000, start = "ones",
                verbose = FALSE)
> boot_model
Call: bootsempls(object = fit_model, nboot = 1000, start = "ones",
    verbose = FALSE)

                   Estimate     Bias Std.Error
fact1 -> mintemp       0.948 -0.00436  4.77e-02
fact1 -> relhumid      0.991 -0.00640  3.84e-02
fact2 -> maxtemp       0.878 -0.03987  1.49e-01
fact2 -> solarrad      0.877 -0.03874  1.69e-01
fact3 -> precipdata    1.000  0.00000  6.36e-17
fact1 -> fact2        -0.326 -0.02098  1.70e-01
fact2 -> fact3        -0.263 -0.05650  1.36e-01
> parallelplot(boot_model, pattern = "beta", reflinesAt = c(0,0.5),
    alpha = 0.3, type = "bca", main = "Path coefficient\nof 500
    bootstrap samples")
> parallelplot(boot_model, pattern = "beta", reflinesAt = c(0,0.5),
    alpha = 0.3, type = "bca", main = "Path coefficient\nof 500
    bootstrap samples")
```

Figure B.10: SEM-PLS Modelling Continued

```
> parallelplot(boot_model, pattern = "beta", reflinesAt = c(0,0.5),
    alpha = 0.2, type = "bca", main = "Path coefficient\nof 500
    bootstrap samples")
> parallelplot(boot_model, pattern = "beta", reflinesAt = c(0,0.5),
    alpha = 0.01, type = "bca", main = "Path coefficient\nof 500
    bootstrap samples")
> parallelplot(boot_model, pattern = "beta", reflinesAt = c(0,0.5),
    alpha = -0.3, type = "bca", main = "Path coefficient\nof 500
    bootstrap samples")
> parallelplot(boot_model, pattern = "beta", reflinesAt = c(0,0.5),
    alpha = 0.3, type = "bca", main = "Path coefficient\nof 500
    bootstrap samples")
> parallelplot(boot_model, pattern = "beta", reflinesAt = c(0,0.3),
    alpha = 0.3, type = "bca", main = "Path coefficient\nof 500
    bootstrap samples")
> parallelplot(boot_model, pattern = "beta", reflinesAt = c(0,0.3),
    alpha = 0.5, type = "bca", main = "Path coefficient\nof
    500 bootstrap samples")
> densityplot(boot_model, pattern = "beta")
> densityplot(boot_model)
> densityplot(fit_model, use = "residuals")
> densityplot(fit_model)
> densityplot(fit_model, use = "residuals")
> vif(confdata)
     Variables      VIF
1 malaria_data 1.187590
2      maxtemp 2.574410
3      mintemp 9.122451
4   precipdata 1.445585
5     relhumid 9.378320
6     solarrad 1.900029
7    windspeed 1.345176
> Modelmm <- read.csv("C:/Users/BMODU/Dropbox/Ghana Ecological
            variables/Modelmm.txt", sep="")
```

Figure B.11: SEM-PLS Modelling Continued

```
>   View(Modelmm)
> mm <- as.matrix(Modelmm)
> mm
      source       target
 [1,] "fact1"      "mintemp"
 [2,] "fact1"      "relhumid"
 [3,] "mintemp"    "malaria_data"
 [4,] "relhumid"   "malaria_data"
 [5,] "fact2"      "maxtemp"
 [6,] "fact2"      "solarrad"
 [7,] "maxtemp"    "malaria_data"
 [8,] "fact3"      "precipdata"
 [9,] "solarrad"   "malaria_data"
[10,] "precipdata" "malaria_data"
[11,] "fact3"      "maxtemp"
> sweden.model <- plsm(data = confdata, strucmod = sm, measuremod = mm)
> sweden.model
$latent
[1] "fact1" "fact2" "fact3"

$manifest
[1] "mintemp"    "relhumid"   "maxtemp"    "solarrad"   "maxtemp"    "precipdata"

$strucmod
     source  target
[1,] "fact1" "fact2"
[2,] "fact2" "fact3"

$measuremod
     source  target
[1,] "fact1" "mintemp"
[2,] "fact1" "relhumid"
[3,] "fact2" "maxtemp"
[4,] "fact2" "solarrad"
[5,] "fact3" "maxtemp"
[6,] "fact3" "precipdata"
```

Figure B.12: SEM-PLS Modelling Continued

```
$D
      fact1 fact2 fact3
fact1     0     1     0
fact2     0     0     1
fact3     0     0     0

$M
           fact1 fact2 fact3
mintemp        1     0     0
relhumid       1     0     0
maxtemp        0     1     1
solarrad       0     1     0
maxtemp        0     1     1
precipdata     0     0     1

$blocks
$blocks$fact1
[1] "mintemp"  "relhumid"
attr(,"mode")
[1] "A"

$blocks$fact2
[1] "maxtemp"  "solarrad"
attr(,"mode")
[1] "A"

$blocks$fact3
[1] "maxtemp"    "precipdata"
attr(,"mode")
[1] "A"


$order
[1] "generic"

attr(,"class")
[1] "plsm"
```

Figure B.13: SEM-PLS Modelling Continued

```
> Modelmm <- read.csv("C:/Users/BMODU/Dropbox/Ghana Ecological
             variables/Modelmm.txt", sep="")
>   View(Modelmm)
> mm <- as.matrix(Modelmm)
> mm
      source       target
 [1,] "fact1"      "mintemp"
 [2,] "fact1"      "relhumid"
 [3,] "mintemp"    "malaria_data"
 [4,] "relhumid"   "malaria_data"
 [5,] "fact2"      "maxtemp"
 [6,] "fact2"      "solarrad"
 [7,] "maxtemp"    "malaria_data"
 [8,] "fact3"      "precipdata"
 [9,] "solarrad"   "malaria_data"
[10,] "precipdata" "malaria_data"
[11,] "fact3"      "windspeed"
> sweden.model <- plsm(data = confdata, strucmod = sm, measuremod = mm)
> sweden.model
$latent
[1] "fact1" "fact2" "fact3"

$manifest
[1] "mintemp"    "relhumid"   "maxtemp"    "solarrad"   "precipdata" "windspeed"

$strucmod
     source  target
[1,] "fact1" "fact2"
[2,] "fact2" "fact3"

$measuremod
     source  target
[1,] "fact1" "mintemp"
[2,] "fact1" "relhumid"
[3,] "fact2" "maxtemp"
[4,] "fact2" "solarrad"
[5,] "fact3" "precipdata"
[6,] "fact3" "windspeed"
```

Figure B.14: SEM-PLS Modelling Continued

```
$D
      fact1 fact2 fact3
fact1     0     1     0
fact2     0     0     1
fact3     0     0     0

$M
           fact1 fact2 fact3
mintemp        1     0     0
relhumid       1     0     0
maxtemp        0     1     0
solarrad       0     1     0
precipdata     0     0     1
windspeed      0     0     1

$blocks
$blocks$fact1
[1] "mintemp"  "relhumid"
attr(,"mode")
[1] "A"

$blocks$fact2
[1] "maxtemp"  "solarrad"
attr(,"mode")
[1] "A"

$blocks$fact3
[1] "precipdata" "windspeed"
attr(,"mode")
[1] "A"


$order
[1] "generic"

attr(,"class")
[1] "plsm"
```

Figure B.15: SEM-PLS Modelling Continued

```
> sweden_model <- sempls(model = sweden.model, data = confdata, wscheme = "A")
All 60 observations are valid.
Converged after 12 iterations.
Tolerance: 1e-07
Scheme: centroid
> sweden_model
                      Path Estimate
lam_1_1     fact1 -> mintemp   0.9479
lam_1_2    fact1 -> relhumid   0.9910
lam_2_1     fact2 -> maxtemp   0.8816
lam_2_2    fact2 -> solarrad   0.8735
lam_3_1  fact3 -> precipdata   0.9849
lam_3_2   fact3 -> windspeed   0.0017
beta_1_2      fact1 -> fact2  -0.3248
beta_2_3      fact2 -> fact3  -0.2774
> sweden_model <- sempls(model = sweden.model, data = confdata, wscheme = "B")
All 60 observations are valid.
Converged after 15 iterations.
Tolerance: 1e-07
Scheme: factorial
> sweden_model
                      Path Estimate
lam_1_1     fact1 -> mintemp   0.9495
lam_1_2    fact1 -> relhumid   0.9903
lam_2_1     fact2 -> maxtemp   0.8675
lam_2_2    fact2 -> solarrad   0.8873
lam_3_1  fact3 -> precipdata   0.9852
lam_3_2   fact3 -> windspeed   0.0031
beta_1_2      fact1 -> fact2  -0.3302
beta_2_3      fact2 -> fact3  -0.2690
> sweden_model <- sempls(model = sweden.model, data = confdata, wscheme = "C")
All 60 observations are valid.
Converged after 15 iterations.
Tolerance: 1e-07
Scheme: path weighting
```

Figure B.16: SEM-PLS Modelling Continued

```
> sweden_model
                        Path Estimate
lam_1_1     fact1 -> mintemp   0.9495
lam_1_2    fact1 -> relhumid   0.9903
lam_2_1     fact2 -> maxtemp   0.8675
lam_2_2    fact2 -> solarrad   0.8873
lam_3_1  fact3 -> precipdata   0.9852
lam_3_2   fact3 -> windspeed   0.0031
beta_1_2      fact1 -> fact2  -0.3302
beta_2_3      fact2 -> fact3  -0.2690
> rSquared(sweden_model)
      R-squared
fact1         .
fact2      0.91
fact3      0.72
> gof(sweden_model)
                    Value
Average R-squared   0.91
Average Communality 0.732
GoF                 0.258
> dgrho(sweden_model)
      Dillon-Goldstein's rho reflective MVs
fact1                   0.97              2
fact2                   0.87              2
fact3                   0.49              2
> communality((sweden_model))
      communality reflective MVs
fact1        0.94              2
fact2        0.77              2
fact3        0.49              2

        Average communality: 0.73
> redundancy(sweden_model)
      redundancy
fact1          .
fact2      0.084
fact3      0.035

        Average redundancy: 0.06
```

Figure B.17: SEM-PLS Modelling Continued

```
> densityplot(sweden_model)
> densityplot(sweden_model, use = "residuals")
> set.seed(123)
> swedenboot_model <- bootsempls(sweden_model, nboot = 500,
       start = "one", verbose = F)
Resample: 500 Done.
> swedenboot_model
Call: bootsempls(object = sweden_model, nboot = 500, start = "one",
    verbose = F)

                   Estimate     Bias Std.Error
fact1 -> mintemp    0.94952 -0.01071    0.0442
fact1 -> relhumid   0.99027 -0.00308    0.0132
fact2 -> maxtemp    0.86752 -0.00415    0.1037
fact2 -> solarrad   0.88730 -0.04922    0.1790
fact3 -> precipdata 0.98518 -0.17090    0.3975
fact3 -> windspeed  0.00314  0.13602    0.4018
fact1 -> fact2     -0.33019 -0.03462    0.1675
fact2 -> fact3     -0.26904 -0.03734    0.2269
> summary(swedenboot_model)
Call: bootsempls(object = sweden_model, nboot = 500, start = "one",
    verbose = F)

Lower and upper limits are for the 95 percent perc confidence interval

         Estimate     Bias Std.Error  Lower Upper
lam_1_1   0.94952 -0.01071    0.0442  0.841 0.994
lam_1_2   0.99027 -0.00308    0.0132  0.957 1.000
lam_2_1   0.86752 -0.00415    0.1037  0.601 0.988
lam_2_2   0.88730 -0.04922    0.1790  0.386 0.978
lam_3_1   0.98518 -0.17090    0.3975 -0.550 1.000
lam_3_2   0.00314  0.13602    0.4018 -0.600 0.935
beta_1_2 -0.33019 -0.03462    0.1675 -0.568 0.169
beta_2_3 -0.26904 -0.03734    0.2269 -0.626 0.314
> densityplot(swedenboot_model, pattern = "beta")
> parallelplot(swedenboot_model, pattern = "beta", reflinesAt = c(0,0.3),
          alpha = 0.5, type = "bca", main = "Path coefficient\nof 500
          bootstrap samples")
> parallelplot(swedenboot_model, pattern = "beta", reflinesAt = c(0,0.5),
          alpha = 0.3, type = "bca", main = "Path coefficient\nof 500
          bootstrap samples")
```

Figure B.18: SEM-PLS Modelling Continued

```
> parallelplot(swedenboot_model, pattern = "beta", reflinesAt = c(0,0.1),
          alpha = 0.3, type = "bca", main = "Path coefficient\nof 500
          bootstrap samples")
> parallelplot(swedenboot_model, pattern = "beta", reflinesAt = c(0,0.1),
          alpha = 0.2, type = "bca", main = "Path coefficient\nof 500
          bootstrap samples")
> parallelplot(swedenboot_model, pattern = "beta", reflinesAt = c(0,0.5),
          alpha = 0.3, type = "bca", main = "Path coefficient\nof 500
           bootstrap samples")
```

Figure B.19: Data Discretization using K-Means Algorithms

```
                    K-MEANS ALGORITHM
Dataset:
w <- c(1427, 1471, 1267, 1288, 1431, 1161, 1209, 1608, 1444,
       1130 , 1142, 1400, 1759, 1233, 1848, 1449, 1536, 712,
       1977, 1324, 1823, 1122, 1056, 1356, 1658, 1743, 2118,
       2051, 1704, 1607, 2456, 2402, 1107, 1775, 1492, 1502,
       863, 1386, 1231, 1294, 1131, 1131, 1865, 1164, 1042,
       1186, 1396, 1393, 1260, 1156, 1293, 1293, 1248, 1777,
       2203, 1422, 1045, 1568, 1588, 1035)

-----------------Iteration for k = 2----------------------------------
clust_2 <- kmeans(w, 2, iter.max = 10, nstart = 5)
print(clust_2)
str(clust_2)
plot(w, col = clust_2$cluster, las =1, pch = 3, cex = 1)

-----------------Iteration for k = 3----------------------------------
clust_3 <- kmeans(w, 3, iter.max = 10, nstart = 5, algorithm = "Forgy")
print(clust_3)
str(clust_3)
plot(w, col = clust_3$cluster, las = 1, pch = 3, cex = 1)

-----------------Iteration for k = 4----------------------------------
clust_4 <- kmeans(w, 4, iter.max = 10, nstart = 5, algorithm = "Forgy")
print(clust_4)
str(clust_4)
plot(w, col = clust_4$cluster, las = 1, pch = 3, cex = 1)

-----------------Iteration for k = 5----------------------------------
clust_5 <- kmeans(w, 5, iter.max = 10, nstart = 5, algorithm = "Forgy")
print(clust_5)
str(clust_5)
plot(w, col = clust_5$cluster, las = 1, pch = 3, cex = 1)


                    SUMMARY OF THE RESULTS
 k             2            3            4            5
iteration      3            4            9            6
convergence    Yes          Yes          Yes          No
SSB/SST        66.4%        82%          89.9%        93%
```

Figure B.20: Data Discretization using K-Means Algorithms Continued

```
From the result summary, the best cluster is k = 4 which converges after
9 iterations with 89.9%. For k = 5, the converges did not achieved after
10 iterations which exceed the maximum number of iteration for k-means
clustering algorithm. Alternative approach for seleting the best number
of clusters for our dataset presented below.

              METHOD OF SELECTING OPTIMUM CLUSTER
package: NbClust

nc <- NbClust(w, min.nc = 2, max.nc = 5, method = "kmeans")

the algorithm will run and selects the best number of cluster to be
consider for partitioning our dataset. According to the this algorithm
the best cluster is 4. Both the methodology suggested k = 4 number of
clusters for the dataset and this can be use for determining the
predictor variables after partition the dataset.

              PARTITION FOR DISRETIZATION
Here we partition the dataset into four (4) classes and label it as:
low, medium, high and very high incidence status of malaria.
The following R code provides algorithm for partitioning the dataset.

package: discretization

partition <- cut(w, breaks = 4,
   labels = c('low', 'medium', 'high', 'very high'), ordered = TRUE)
table(partition)

The discretized dataset presented as:
      w  partition
1  1427    medium
2  1471    medium
3  1267    medium
4  1288    medium
5  1431    medium
6  1161    medium
7  1209    medium
8  1608      high
9  1444    medium
10 1130       low
11 1142       low
12 1400    medium
```

Figure B.21: Data Discretization using K-Means Algorithms Continued

```
13 1759      high
14 1233    medium
15 1848      high
16 1449    medium
17 1536    medium
18  712       low
19 1977      high
20 1324    medium
21 1823      high
22 1122       low
23 1056       low
24 1356    medium
25 1658      high
26 1743      high
27 2118 very high
28 2051 very high
29 1704      high
30 1607      high
31 2456 very high
32 2402 very high
33 1107       low
34 1775      high
35 1492    medium
36 1502    medium
37  863       low
38 1386    medium
39 1231    medium
40 1294    medium
41 1131       low
42 1131       low
43 1865      high
44 1164    medium
45 1042       low
46 1186    medium
47 1396    medium
48 1393    medium
49 1260    medium
50 1156    medium
```

Figure B.22: Data Discretization using K-Means Algorithms Continued

```
51 1293    medium
52 1293    medium
53 1248    medium
54 1777      high
55 2203 very high
56 1422    medium
57 1045       low
58 1568    medium
59 1588      high
60 1035       low

where:
w = malaria inicedence dataset
partition = discretize status of malaria incidence
```

Appendix C

Source Code for Agent-Based Modelling

Figure C.1: Agent-based modelling source code

```
globals [
  ambient-temperature
  previous-cases-malaria-in-humans]

breed [
  humans
  human]

breed [
  mosquitoes
  mosquito]

humans-own [
  susceptible-humans?
  infected-humans?
  ;recovered-humans?
  recovery-time
  infection-length
  number-of-infected-humans]

mosquitoes-own [
  susceptible-mosquitoes?
  exposed-mosquitoes?
  infected-mosquitoes?
  extrinsic-incubation-complete?
  number-of-exposed-mosquitoes
  number-of-infected-mosquitoes
  energy
  age
  latent-period]
```

Figure C.2: Agent-based modelling source code continued

```
patches-own [temperature-resolution]

to setup
  clear-all
  setup-patches
  setup-humans
  setup-mosquitoes
  reset-ticks
end

to setup-patches
  ask patches [set pcolor green]
end

to setup-humans
  create-humans initial-number-of-humans [
    setxy random-xcor random-ycor
    set shape "person"
    set size 1.5
    set color white
    set susceptible-humans? true
    set infected-humans? false
    set infection-length 0
    ;set recovered-humans? false

    ;set recovery-time random-normal average-recovery-time average-recovery-time / 4
    ;if recovery-time > average-recovery-time * 2 [set recovery-time average-recovery-time * 2]
    ;if recovery-time < average-recovery-time * 0 [set recovery-time average-recovery-time * 0]

    if (random-float initial-number-of-humans < 10) [
      set susceptible-humans? false
      set infected-humans? true
      ;set infection-length random recovery-time
    ]
      ifelse (infected-humans?)
      [set color red]
```

Figure C.3: Agent-based modelling source code continued

```
      [set color white]]
end


to setup-mosquitoes
  create-mosquitoes initial-number-of-mosquitoes [
    setxy random-xcor random-ycor
    set shape "butterfly"
    set size 0.8
    set color yellow
    set susceptible-mosquitoes? true
    set exposed-mosquitoes? false
    set extrinsic-incubation-complete? false
    set infected-mosquitoes? false
    set energy random 5
    set latent-period random 5
    ;set mosquito-lifespan random 56
    set age 0]
end

to go
  ;if (all? mosquitoes [infected-mosquitoes?]) [stop]

  ask humans [
    move-around-the-world
    human-death
    human-birth]

  ask humans with [infected-humans?] [
    humans-infect-mosquitoes]

  ask mosquitoes [
    wander-around-the-world
    mosquito-death]
```

Figure C.4: Agent-based modelling source code continued

```
  ask mosquitoes with [ exposed-mosquitoes? or infected-mosquitoes?] [
    mosquito-reproduce
    infectious-mosquitoes-to-spread-malaria-to-susceptible-humans]

  tick
end

to move-around-the-world
  right random-float 180
  left  random-float 180
  forward 1
end

to wander-around-the-world
  right random-float 180
  left  random-float 180
  forward 1
end



to humans-infect-mosquitoes
  let target one-of mosquitoes-here
  if target != nobody
  [ ask target
    [ if (random-float initial-number-of-mosquitoes < chance-of-malaria-to-mosquitoes)
      [
        set susceptible-mosquitoes? false
        set exposed-mosquitoes? true
        set infected-mosquitoes? false
        set extrinsic-incubation-complete? false
        set color blue
        set energy energy + energy-from-blood-meal
        set latent-period latent-period + 1
        set age age + 1]
```

Figure C.5: Agent-based modelling source code continued

```
    ask mosquitoes with [exposed-mosquitoes?]
        [
         if latent-period > incubation-length [
            set extrinsic-incubation-complete? true
            set infected-mosquitoes? true
            set color black]]
  ]]
end

to mosquito-reproduce
  ask mosquitoes [
  if energy > reproduction-threshold
   [
     set energy energy / 2
      hatch new-mosquito [
        set color yellow
        forward 1]]]
end

to mosquito-death
  if age > 56 [die]
end

to infectious-mosquitoes-to-spread-malaria-to-susceptible-humans
  let prey one-of humans-here with [not infected-humans?]
  if (prey != nobody)
  [ ask prey
   [ if (random-float initial-number-of-humans < chance-of-malaria-to-humans)
     [ set susceptible-humans? false
       set infected-humans? true
       set color red


        ;set infection-length infection-length + 1
```

Figure C.6: Agent-based modelling source code continued

```
        ;ask humans with [color = red][
        ;if infection-length > 14 [
          ;set infected-humans? false
          ;set susceptible-humans? true
          ;set color white]]
  ]]]
end


to human-death
  ask humans with [infected-humans?][
    if random-float initial-number-of-humans < malaria-induced-death [die]]
end

to human-birth
  ask humans [
    if random-float initial-number-of-humans < human-birth-rate [
      hatch 1 [
        set color white
        forward 1 ]]]
end

;ask mosquitoes with [exposed-mosquitoes?]
        ;[
          ;set energy energy + energy-from-blood-meal
          ;set latent-period latent-period + 1
          ;set age age + 1]


;to clear-count
  ;set number-of-infected-humans 0
;end

;to clear-count1
  ;set number-of-exposed-mosquitoes 0
```

Figure C.7: Agent-based modelling source code continued

```
;end

;to clear-count1
  ;set number-of-exposed-mosquitoes 0
  ;set number-of-infected-mosquitoes 0
;end


;to humans-to-infect-mosquitoes
  ;let target one-of mosquitoes-here with [susceptible-mosquitoes?]
  ;if target != nobody
  ;[ ask target
    ;[ if (random-float initial-number-of-mosquitoes < chance-of-malaria-to-mosquitoes)
      ;[
       ; set susceptible-mosquitoes? false
       ; set exposed-mosquitoes? true
       ; set infected-mosquitoes? false
        ;set extrinsic-incubation-complete? false
        ;set number-of-exposed-mosquitoes (number-of-exposed-mosquitoes + 1)
        ;set color blue
        ;set energy energy + energy-from-blood-meal

        ;ask mosquitoes with [ color = blue][
            ;if (energy > capacity-of-reproduction)
       ;[ set energy energy / 2
            ;hatch new-mosquito
            ;[
            ; set color yellow
            ; set age 0
             ;forward 1 ]

       ; ]
         ; if (energy < 0) [die]
       ; ]
```

Figure C.8: Agent-based modelling source code continued

```
      ;ask mosquitoes with [ color = blue ][
        ;if (random-float initial-number-of-mosquitoes > length-of-incubation)
        ;[
         ; set susceptible-mosquitoes? false
         ; set exposed-mosquitoes? true
         ; set infected-mosquitoes? true
          ;set extrinsic-incubation-complete? true
          ;set number-of-infected-mosquitoes number-of-infected-mosquitoes + 1
          ;set color black
            ;ask mosquitoes with [ color = black ] [
              ;if (energy > capacity-of-reproduction)
              ;[ set energy energy / 2
               ; hatch new-mosquito
                ;[
                  ;set color yellow
                  ;set age 0
                  ;forward 1 ]

            ;]
              ;if (energy < 0) [die]
            ;]
;]]]]]
;end


;to infectious-mosquitoes-to-spread-malaria-to-susceptible-humans
  ;let prey one-of humans-here with [color = white]
  ;if (prey != nobody)
  ;[ ask prey
   ; [ if (random-float initial-number-of-mosquitoes < chance-of-transmitting-malaria-to-humans)
     ; [ set susceptible-humans? false
       ; set infected-humans? true
        ;set color red
        ;set number-of-infected-humans (number-of-infected-humans + 1)
     ;]]]
```

Figure C.9: Agent-based modelling source code continued

```
;to mosquitoes-to-reproduce
  ;if (energy > threshold-of-reproduction)
  ;[ set energy energy / 2
    ;hatch 1 [ forward 1 ]
  ;]
;end

;to mosquito-to-die
  ;if (energy <= 0)
  ;[die]
;end
```

Appendix D

Plot of Climate Data from 1996-2013

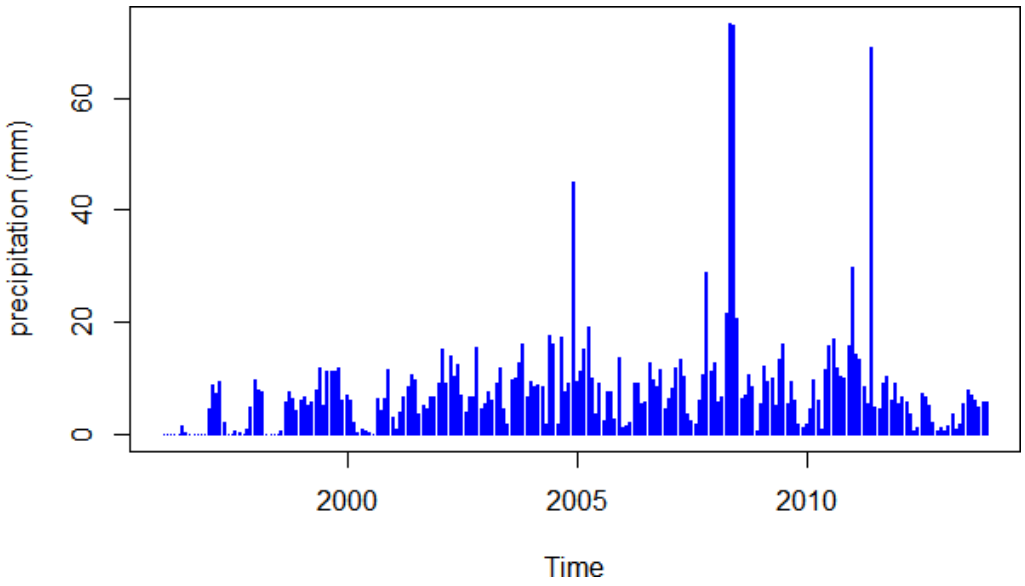

Figure D.1: Precipitation (mm) distribution plot

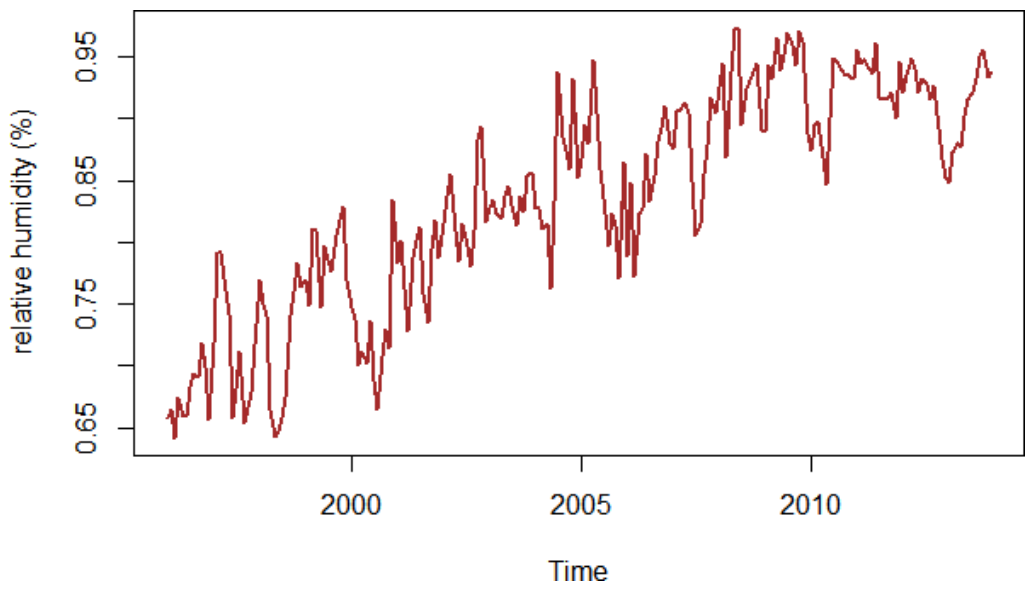

Figure D.2: Relative humidity (%) distribution plot

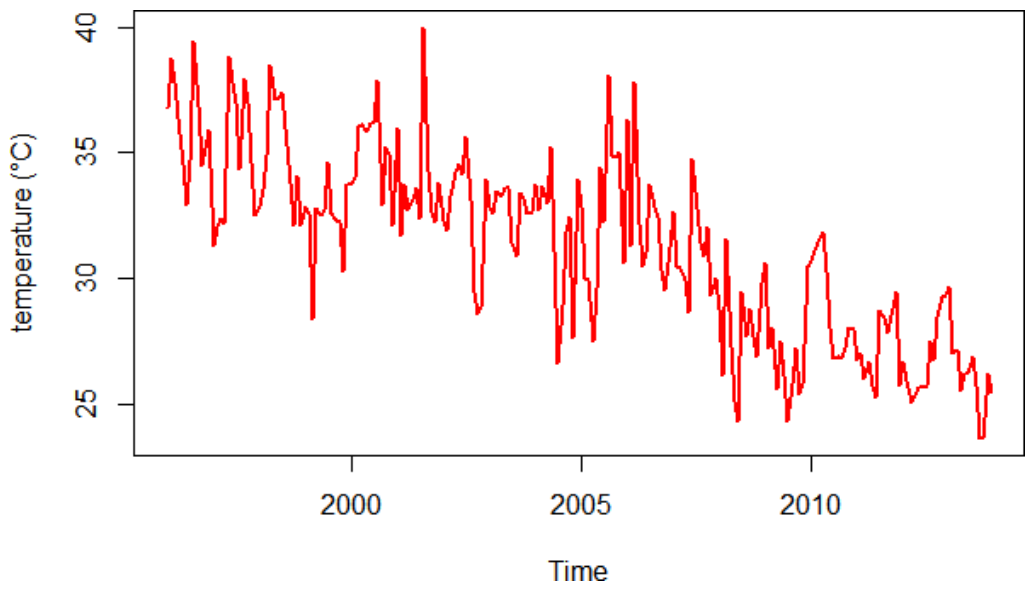

Figure D.3: Temperature (˚C) plot

Appendix E

Algorithms for PLS modelling

Step1 Initialization Suppose $\mathbf{Y}_1, \cdots, \mathbf{Y}_K$ be the respective MVs, and are scaled such that $\mathbf{E}(\mathbf{Y}_i) = 0$ and $\mathbf{V}(\mathbf{Y}_i) = 1$. We are interested to express each LV as a linear combination of MVs, and represented in compact form:

$$\begin{aligned} \hat{\mathbf{X}} &= \mathbf{YM} \\ \hat{\mathbf{x}}_g &= \frac{\hat{x}_g}{\sqrt{VAR(\hat{x}_g)}}, g = 1, \cdots, G \end{aligned} \tag{E.1}$$

Hence, the LVs are initialized as: $\hat{\mathbf{X}} = \hat{\mathbf{x}}_1, \cdots, \hat{\mathbf{x}}_G$.

Step2 Inner approximation

Within the inner model domain, the estimation of path parameter of each LV can be mathematically represented as the weighted sum of its neighbouring LVs.

$$\begin{aligned} \tilde{\mathbf{X}} &= \hat{\mathbf{X}}\mathbf{E} \\ \tilde{\mathbf{x}}_g &= \frac{\tilde{x}_g}{\sqrt{(VAR(\tilde{\mathbf{x}}_g))}}, g = 1, \cdots, G \end{aligned} \tag{E.2}$$

The approximate estimation of inner model path parameter takes: $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_1, \cdots, \tilde{\mathbf{x}}_G)$.

Step3 Outer approximation

The outer approximation are computed based on the weight of the LV loads from the inner approximation. This comes in two forms, Mode A and Mode B. For Mode A, a multivariate regression coefficient with the block of MVs as response and the LV as regressor:

$$\hat{\mathbf{w}}_g^{\mathsf{T}} = (\tilde{\mathbf{x}}_g^{\mathsf{T}} \tilde{\mathbf{x}}_g)^{-1} \tilde{\mathbf{x}}_g^{\mathsf{T}} \mathbf{Y}_g \tag{E.3}$$

While, Mode B is a multiple regression coefficient with the block of MVs as response and its block of MVs as regressor:

$$\hat{\mathbf{w}}_g = (\mathbf{Y}_g^{\mathsf{T}} \mathbf{Y}_g)^{-1} \mathbf{Y}_g^{t} \tilde{\mathbf{x}}_g \tag{E.4}$$

Step4 Outer weight vector

Let $k_g = \{k \in \{1, \cdots, K\} | y_k \; x_g\}$ be a set of indices for MVs related to LV x_g then w_g, $g = 1, \cdots, g$, is a column vector of length $|k_g|$. We can write down the matrix of outer weights, W as:

$$W = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_G \end{bmatrix}$$

The outer weights vectors, $w_1, \cdots, w_G$, in an outer weights matrix W, which we

are using now to estimate the factor scores by means of the MVs:

$$\begin{aligned} \hat{\mathbf{X}} &= \mathbf{YW} \\ \hat{\mathbf{X}}_g &= \frac{\hat{\mathbf{X}}_g}{\sqrt{VAR(\hat{\mathbf{X}}_g)}}, g = 1, \cdots, G, \end{aligned} \tag{E.5}$$

resulting in the outer estimation: $\mathbf{X} = (\hat{\mathbf{x}}_1, \cdots, \hat{\mathbf{x}}_G)$.

Step5 Iteration

If the relative change of all the outer weights from one iteration to the next are smaller than a predefined tolerance,

$$\frac{\hat{w}_{kg}^{old} - \hat{w}_{kg}^{new}}{\hat{w}_{kg}^{new}} < E, \forall, k = 1, \cdots, K \wedge g = 1, \cdots, G, \tag{E.6}$$

the estimation of factor scores done in (E.5) is taken to be final. Otherwise go back to (E.2).

E.0.1 Weighting scheme

The weighting schemes are used for estimation of the inner weight in (E.2) of the PLS algorithm. Generally, there are three weighting schemes, centroid [70], and later [97] introduced factorial and path weighting scheme.

E.0.1.1 Centroid (A)

The centroid weighting scheme, takes the form:

$$e_{ij} = \begin{cases} sign(r_{ij}), & \text{for } c_{ij} = 1, i, j = 1, \cdots, G \\ 0, & \text{else} \end{cases} \tag{E.7}$$

where $\mathbf{E}$ denotes the matrix of inner weights .

E.0.1.2 Factorial (B)

The factorial weighting scheme also takes the form:

$$e_{ij} = \begin{cases} r_{ij}, & \text{for } c_{ij} = 1, i, j = 1, \cdots, G \\ 0, & \text{else} \end{cases} \tag{E.8}$$

E.0.1.3 Path weighting (C)

In this weighting scheme, the predecessor and successor of a LV play a different role in the relation. When considering the relation for one specific LV $\mathbf{x}_i$ with its successor is determined by their correlation, for the predecessors it is determined by a multiple regression

$$\begin{aligned} \mathbf{x}_i &= \mathbf{x}_i^{pred}\gamma + \mathbf{z}_i \\ \mathbf{E}[\mathbf{z}_i] &= 0, i = 1, \cdots, G \end{aligned} \tag{E.9}$$

with $\mathbf{x}_i^{pred}$ the predecessor set of the LV $\mathbf{x}_i$. Denoting $\mathbf{x}_i^{succ}$ the successor set of the LV $\mathbf{x}_i$ the elements of the inner weight matrix $\mathbf{E}$ as:

$$e_{ij} = \begin{cases} \gamma_j, & \text{for } \boldsymbol{j} \in \mathbf{x}_i^{pred}, \\ COR(\mathbf{y}_i, \mathbf{x}_j), & \text{for } \boldsymbol{j} \in \mathbf{x}_i^{succ}, \\ 0, & \text{else} \end{cases} \tag{E.10}$$

E.0.2 Discriminant validity check

In the structural equation model, the factor scores are estimated by PLS algorithm, while the path coefficients are also estimated using ordinary least squares (OLS). Now, for each LV $\hat{\mathbf{x}}_g$, $g = 1, \cdots, G$, the path coefficient is the regression coefficient on its predecessor set $\hat{\mathbf{x}}_g^{pred}$ defined as:

$$\hat{\beta}_g = (\hat{\mathbf{x}}_g^{pred\top} \hat{\mathbf{x}}_g^{pred})^{-1} \hat{\mathbf{x}}_g^{pred\top} \hat{\mathbf{x}}_g \tag{E.11}$$

Using (E.11), we can compute the element $\hat{\mathbf{b}}_{ij}, i, \boldsymbol{j} = 1, \cdots, G$, of the estimated matrix of path coefficients $\hat{\beta}$

E.0.2.1 Path coefficients

$$\hat{\beta}_{ij} = \begin{cases} \hat{\beta}_{gj}, & \text{for } \boldsymbol{j} \in \mathbf{x}_i^{pred}, \\ 0, & \text{else} \end{cases} \tag{E.12}$$

Therefore, the matrix $\hat{\mathbf{B}}$ denotes a transition matrix for the structural equation model.

E.0.2.2 Total effects

We can calculate the matrix of the total effects $\hat{\mathbf{T}}$ as the sum of the 1 to $\mathbf{G}$ step transition matrices:

$$\hat{\mathbf{T}} = \sum_{g=1}^{G} \hat{\mathbf{B}}_g \tag{E.13}$$

Note, that $\hat{\mathbf{B}}^g$ expands to $\hat{\mathbf{B}} \cdot \hat{\mathbf{B}} \cdot \underbrace{\cdots}_{g-times} \cdot \hat{\mathbf{B}}$, e.g., $\hat{\mathbf{B}}^2$ contains all the indirect effects mediated by only one LV.

E.0.2.3 Outer loadings

The cross and outer loadings are estimated as:

$$\hat{\Lambda}^{cross} = COR(\mathbf{Y}, \hat{\mathbf{X}}) \tag{E.14}$$

$$\hat{\lambda}_{kg}^{outer} = \begin{cases} \hat{\lambda}_{kg}^{cross}, & \text{if } m_{kg} = 1 \\ 0, & \text{else} \end{cases} \tag{E.15}$$