

State-Transition Model for Malaria Symptoms

Drinold Aluda Mbete^{1*} and Kennedy Nyongesa¹

¹Department of Mathematics, Masinde Muliro University of Science and Technology, P.O. BOX 190-5000, Kakamega, Kenya.

Authors' contributions

This work was carried out in collaboration between both authors. Author DAM designed the study, performed the statistical analysis, wrote the protocol, the literature and the first draft of the manuscript. Author KN managed the analyses of the study and approved the final manuscript. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2020/v10i430253

Editor(s):

(1) Dr. Belkacem Chaouchi, Khemis Miliana University, Algeria.

Reviewers:

(1) Abdulrahman AL-Sanea, Kocaeli University, Turkey.

(2) Augustine Adu Frimpong, Southern University and A&M College, USA.

Complete Peer review History: <http://www.sdiarticle4.com/review-history/64479>

Received: 04 November 2020

Accepted: 09 January 2021

Published: 02 February 2021

Original Research Article

Abstract

Aims/ objectives: To develop a state-transition model for malaria symptoms.

Study design: Longitudinal study.

Place and Duration of Study: Department of Mathematics Masinde Muliro University of Science and Technology between January 2015 and December 2015.

Methodology: We included 300 students (patients) with liver malaria disease, with or without the medical history of malaria disease, physical examination for signs and symptoms for both specific and non-specific symptom, investigation of the disease through laboratory test (BS test) and diagnostic test results. the focus of this study was to develop state-transition model for malaria symptoms. Bayesian method using Markov Chain Monte Carlo via Gibbs sampling algorithm was implemented for obtaining the parameter estimates.

Results: The results of the study showed a significant association between malaria disease and observed symptoms

Conclusion: The study findings provides a useful information that can be used for predicting malaria disease in areas where Blood slide test and rapid diagnostic test for malaria disease is not possible.

*Corresponding author: E-mail: drinoldmbete123@gmail.com;

Keywords: Bayesian; Posterior; Malaria; Symptoms.

1 Introduction

Malaria is an ancient disease that has been affecting people since the beginning of recorded time. It poses serious economic, social and health burdens in tropical and subtropical countries where it is predominantly found [1]. Malaria still remains a huge public health issue regardless of how many years of research has been conducted on how to combat this disease. According to WHO [2], the latest world malaria report released in November 2017 shows that the number of malaria cases reported in the year 2016 was 216 million up from 211 million cases reported in 2015. The report also shows that malaria death estimates in 2016 stood at 445,000 compared to 446,000 deaths in 2015. The high burden of malaria cases in 2016 was in Africa at 90% with 91% cases of deaths reported in children. According to WHO report on malaria cases in Kenya, malaria is one of the leading causes of morbidity and fatality with about 3.5 million children at risk of developing severe malaria, out of which an estimated 34,000 children under five years die every year. The disease is also responsible for 30% of out-patient visits at health centres, economically, it is estimated that 170 million working hours are lost each year because of malaria illness [3].

Symptoms are experienced deviations from an individual's perception of his or her normal healthy state of being, yet not necessarily an indicator of illness. A symptom can emerge from sensitivity to certain combinations of biological, social and environmental processes and vary in magnitude, severity, persistence and character. Symptoms can be subjectively reported or objectively observed. Depending on the disease, the scope and intensity, the duration of symptoms can vary over time.

The malaria symptoms can be grouped into two; symptoms for uncomplicated malaria (suspected malaria) and symptoms for complicated malaria (severe malaria). Malaria is considered uncomplicated when symptoms are present but there are no clinical or laboratory signs to indicate severity or vital organ dysfunction. The symptoms of uncomplicated malaria are non-specific i.e. they are self-reported symptoms that do not indicate a specific disease process, they are initial symptoms and include fever (temperature), chills, headache, pains (joint, muscle, abdominal), muscle aches, loose stool, tiredness, nausea, high pressure, vomiting and diarrhoea. Infection with *Plasmodium falciparum* if not promptly treated can quickly progress to complicated malaria (severe malaria). The main symptoms of severe malaria include coma, severe breathing difficulties, low blood sugar, hallucination, prostration, immobility, confusion and incoherent speech, seizure, loss of consciousness, hyperparasitaemia, black quarter urine and low blood haemoglobin [4]

2 Literature Review

Martens *et al* [5] examined the relationship between malaria and environmental and socio-economic variables in Sudan using health production modified model. They used regression analysis method to analyse their results, the regression results showed significant relationships between malaria, rainfall and water bodies while other variables such as Human Development Index, temperature, population density and percent of cultivated areas were not significant. Teklehaimanot *et al* [6] used robust Poisson regression model to model the daily average number of cases in 10 districts of Ethiopia that was associated with rainfall, minimum temperature and maximum temperature as explanatory variable in a polynomial distributed lag model. To improve reliability and generalizability within similar climatic conditions, the districts were grouped into two climatic zones, hot and cold. The results showed that malaria was associated with rainfall and minimum temperature in Ethiopia. In cold districts, rainfall was associated with a delayed increase in malaria cases while the association in the hot districts occurred at relatively shorter lags. The results also showed that in cold districts,

minimum temperature was associated with malaria cases with a delayed effect while in hot districts, the effect of minimum temperature was non-significant at most lags, and much of its contribution was relatively immediate.

In many studies of medical treatment, symptoms are measured repeatedly over time in observation called longitudinal observation. Though we cannot observe directly latent variables, we learn about it by measuring symptom. For the longitudinal models, two latent variables govern disease, one for the probability of experiencing a particular symptom and another for the severity of the experienced symptom. Thus the probability of a symptom and the severity of it depends on both latent variables and observed variables [7]. Latent variable link observable data in the real world to symbolic data in the model. Bayesian statistics is often used for inferring latent variables, the common method used inferring latent variables in Bayesian statistics are; Hidden Markov Model (HMM), factor analysis, principal component analysis and Expectation Maximization (EM) algorithm [7].

Zammit *et al* [8] developed an intra-individual consistency model using a logistic-type latent variable model. The latent variable in the model was used to represent the propensity of symptoms and intensity of episodes as these could not be observed directly and needed to be estimated through observation of symptoms episodes in hypoglycaemia. The model results showed that there was individual difference in symptom reporting and that adults exhibit distinct intra-individual variability in symptom reporting. Hans *et al* extended on the model developed by Zammit *et al* by allowing for different forms of symptom experiencing thresholds between groups variability when symptoms are classified in groups and performing variable selection to determine a predictive model for the effect of patient characteristics and their interactions on symptom consistency. The study was conducted in several health centers in the United Kingdom and data collected from 381 participants aged between 17-75 years. Bayesian estimation was performed for all coefficients in the developed model without grouped symptoms and with grouped symptoms. The analysis shows that a multiplicative form of symptom propensity and episode intensity provides the most suitable symptom experiencing threshold and groups of symptoms show distinct propensity and that gender subjects had significant impact on the consistency of symptom reporting.

3 Materials and Methods

3.1 Observed symptoms

From a population point of view, one has the disease if he/she exhibits malaria symptoms. Following an infective bite by mosquito, symptoms appear in about 9 – 14 days. The initial symptoms of malaria are non-specific and mimic a flu-like syndrome [5]. In a study by Martins *et al* [5], there are 19 common symptoms associated with malaria disease which were confirmed and assessed by blood smear microscopy. The symptoms are; fever, chills, sweating, headache, myalgia, arthralgia, abdominal pain, nausea, vomiting, dizziness, cough, diarrhoea, weakness, inappetence, bitter mouth, pallor, coryza, sneezing and sore throat. Some of these symptoms are observable symptoms in patients for instance fever (body temperature), chills, sweating, vomiting, cough, diarrhoea, pallor and sneezing. A healthy person infected with malaria, develops mild to moderate and finally severe symptoms depending on treatment measures taken.

In this study, the observations are data of observed symptoms recorded by the health officer for each student who visited the health facility with malaria related symptoms. The observed symptoms used in this study are; fever (body temperature), rigors, sweating, vomiting, diarrhoea, pallor, cough, convulsion and prostration.

4 Model Assumption

The following assumption were used in the development of the model;

- (i) At a given time, individual i is in a particular state i.e either in an infective state (I) or susceptible state (S).
- (ii) The probability of transiting from state S to state I is time dependant.
- (iii) We do not explicitly model the distinction between state E and I as these states are not distinguishable with the data considered because it is only symptoms that allow for distinction of states and therefore state S is characterised by no symptoms or minimal symptoms and state I is one in which symptoms are observed.

5 The State-Transition Model

State-transition model is developed from a Markov chain, in this model, only the pathways or transitions are considered and the probability of every transition is calculated, furthermore, the model can be used on population of any size. To aid our discussion, we first provide an overall modelling framework.

Fig. (1) provides a transition diagram on how an i^{th} individual susceptible to malaria evolves.

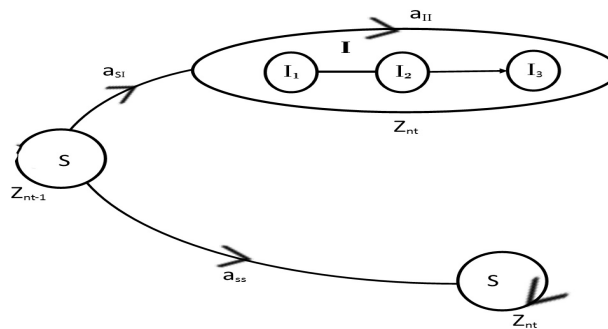


Fig. 1. Malaria transition diagram

An i^{th} student can become infected at time t given that he/she was not infected at time $t - 1$. The process of individual i becoming infected is denoted by Z_i , thus when in state S (susceptible state) at time $t - 1$, the process is Z_{it-1} . When a student transit to the next state for instance from S to I (infected state), the process becomes Z_{it} . When in state I , the process evolves with time and the three possible categories within state I are mild (I_1), moderate (I_2) and severe (I_3).

6 Computation of Transition Probability for the i^{th} Student

In this section, we are interested in the computation of transition probabilities in Fig. (1) which are vital to our model. The transition probabilities are:

$$\begin{aligned} a_{iSS}^{(t)} &= P(Z_{i(t)} = S | Z_{it-1} = S), \\ a_{iSI}^{(t)} &= P(Z_{it} = I | Z_{it-1} = S), \end{aligned} \quad (1)$$

and

$$a_{iII}^{(t)} = P(Z_{i(t)} = I | Z_{it-1} = I)$$

where $a_{iSI}^{(t)}$ is the probability that i^{th} individual transits from state S at time $t-1$ to state I at time t , $a_{iSS}^{(t)}$ is the probability that i^{th} individual remains at state S at time t and $a_{iII}^{(t)}$ is the probability that i^{th} individual evolves within state I at time t .

To calculate the probabilities of $a_{iSI}^{(t)}$ or $a_{iSS}^{(t)}$, we require probit models. This model simply use the cumulative Gaussian normal distribution to calculate the probability of being in one state/category or not. For instance the probability of $a_{iSI}^{(t)}$ is given by

$$a_{iSI}^{(t)} = P(Z_{it} = I | Z_{it-1} = S) = \int_{-\infty}^{\psi_{it}} \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}t^2 dt = \phi(\psi_{it}) \quad (2)$$

where ϕ is the cumulative standard normal distribution and the upper bound parameter ψ_{it} is a transition parameter ($\psi_{it} \in \mathfrak{R}$) which defines the transition of the i^{th} student from a state of susceptibility at time $t-1$ to a state of illness at time t . In the next Section (3.6), we shall show how to compute ψ_{it} .

Similarly $a_{iSS}^{(t)}$ is given by

$$a_{iSS}^{(t)} = P(Z_{it} = S | Z_{it-1} = S) = \int_{-\infty}^{\psi_{it}^c} \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}t^2 dt \quad (3)$$

where ψ_{it}^c is the complement of ψ_{it} . We rewrite the compliment in Equation (3) as

$$a_{iSS}^{(t)} = P(Z_{it} = S | Z_{it-1} = S) = \int_{-\infty}^{\mathfrak{R} - \psi_{it}} \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}t^2 dt \quad (4)$$

where \mathfrak{R} is in the scale of time t given by the interval $\mathfrak{R} = [-\infty, +\infty]$. Equation (4) can further be simplified to

$$a_{iSS}^{(t)} = P(Z_{it} = S | Z_{it-1} = S) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}t^2 dt - \int_{-\infty}^{\psi_{it}} \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}t^2 dt \quad (5)$$

upon simplification of Equation (5), we get

$$a_{iSS}^{(t)} = P(Z_{it} = S | Z_{it-1} = S) = 1 - \int_{-\infty}^{\psi_{it}} \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}t^2 dt \quad (6)$$

Therefore Equation (3) can be written as

$$a_{iSS}^{(t)} = P(Z_{it} = S | Z_{it-1} = S) = 1 - a_{iSI}^{(t)} \quad (7)$$

7 Computation of Transition Parameter

The probability of student i transiting from a state of susceptibility at time $t - 1$ to a state of illness at time t is modelled as

$$P(Z_{it} = I | Z_{it-1} = S) = \phi(\psi_{it}) \quad (8)$$

where ϕ is a cumulative distribution function, with $N(\varphi; 0, 1)$ a normal distribution function for variable φ characterized by zero mean and unit variance. Therefore, the transition parameter $\psi_{i(t)}$ in concert with a probit link function, defines the probability with which one transits from a susceptible state to infective state. We model this transition parameter via two terms:

$$\psi_{it} = I_t + \gamma_{it}^* \quad (9)$$

where the term $I_t \in \mathfrak{R}$ is the malaria incidence rate of an area as influenced by environmental/climatic factors at time t and γ_{it}^* is a latent regression term.

To determine the value of transition parameter ψ_{it} , we compute the values of I_t and γ_{it}^* . In the next Section, we show how to compute I_t .

7.1 Computation of expected malaria incidence rate

In this section, we first determine the association between environmental/climatic factors with malaria disease then afterwards determine the influence of factors on malaria incidence rate using the model proposed by Alberto *et al* [9]. According Alberto *et al*, the influence of environmental/climatic factors on the malaria incidence rate of an area can be explained by the relation

$$\sum \kappa I_t * \sum \nu \sin\left(\frac{2\pi}{\rho}\right) [R_t * T_t * V_t] \rightarrow I_{t+d} \quad (10)$$

where

I_t is the malaria incidence rate at time t , R_t, T_t, V_t , are the environmental/climatic factors at time t , which are rainfall, the average temperature and the mean vegetation density respectively. ρ is the seasonal oscillation at time t , I_{t+d} is malaria incidence at time $t + d$ i.e d days from t , κ is the linear regression coefficient for the incidence rate, ν is the parameter that determines the amplitude of seasonal oscillations estimated by regression and $*$ is the operator that link the components and expresses the lack of a priori knowledge of how the variables are related.

Equation (10) expresses cumulative combination of previous values of the incidence rate as an estimator of population reservoir and the cumulative combination of past levels of rain, temperature and vegetation density as estimators of vector capacity that combine to influence future values of the incidence rate. The term that includes rainfall, temperature and vegetation density implies that the malaria incidence oscillates with a period that is proportional to their common seasonality. The model in Equation (10) combines all those terms having significant autocorrelation and cross-correlation coefficients with the malaria incidence rate in their corresponding lags at a significance level of $p \leq 0.05$. To this end we need to estimate parameters of the model in Equation (10) by using historical data.

8 Computation of Regression Term in the Transition Model

To determine the value of γ_{it}^* , we assume the transition probabilities to vary based on various observed symptoms. We summarize these symptoms for student i in period t in vector form as

$X_{it} = \{X_1 - \text{fever (body temperature)}, X_2 - \text{rigors}, X_3 - \text{convulsion}, X_4 - \text{sweating}, X_5 - \text{vomiting}, X_6 - \text{diarrhoea}, X_7 - \text{pallor}, X_8 - \text{cough and } X_9 - \text{prostration} \}$

The transition of an individual from one state to another can only occur as a result of the observed symptoms. Thus, we assume that there exist a linear relationship defined by

$$\gamma_{it}^* = X_{it}\theta + \epsilon_{it} \quad (11)$$

where the random error ϵ_{it} is assumed to be randomly normally distributed with mean 0 and unity variance i.e $\epsilon \sim N(0, 1)$. θ is a $p \times 1$ vector of unknown regression parameters. X_{it} is a $1 \times p$ observed symptoms.

To obtain γ_{it}^* , we need to estimate the parameter θ which we show how to estimate in the next section.

8.1 Estimation of regression coefficients

The density function for the i^{th} observation is given by

$$f(\epsilon) = (2\pi)^{-\frac{1}{2}} \exp -\frac{1}{2}\epsilon'\epsilon \quad (12)$$

From Equation (11), $\epsilon = \gamma_{it}^* - X\theta$, replacing ϵ_{it} in Equation (12) becomes

$$f(\epsilon) = (2\pi)^{-\frac{1}{2}} \exp -\frac{1}{2}(\gamma_{it}^* - X\theta)'(\gamma_{it}^* - X\theta) \quad (13)$$

when we use the proportionality sign (\propto), we do not write the term that does not involve θ . Therefore Equation (13) becomes

$$f(\epsilon) \propto \exp -\frac{1}{2}(\gamma_{it}^* - X\theta)'(\gamma_{it}^* - X\theta) \quad (14)$$

distributing the transpose in the first term of the exponential, we get

$$f(\epsilon) \propto \exp\{-\frac{1}{2}(\gamma_{it}' - \theta'X')(\gamma_{it}^* - X\theta)\} \quad (15)$$

we expand the exponent to get

$$f(\epsilon) \propto \exp\{-\frac{1}{2}(\gamma_{it}'\gamma_{it}^* - \gamma_{it}'X\theta - \theta'X'\gamma_{it}^* + \theta'X'X\theta)\} \quad (16)$$

Since the first term is a constant with respect to θ and so it can be removed as a multiplicative proportionality constants so as to get

$$f(\epsilon) \propto \exp\{-\frac{1}{2}(-\gamma_{it}'X\theta - \theta'X'\gamma_{it}^* + \theta'X'X\theta)\} \quad (17)$$

Suppose we assume that $g(\theta)$ to be the conjugate prior of Equation (17) i.e $g(\theta) \sim N(0, \sigma^2)$. Then the distribution of $g(\theta)$ is given as

$$g(\theta) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp -\frac{1}{2\sigma^2}\theta'\theta \quad (18)$$

when we use the proportionality sign (\propto), we do not write the term that does not involve θ

$$g(\theta) \propto \exp -\frac{1}{2\sigma^2}\theta'\theta \quad (19)$$

Next we find the posterior distribution ($P(\theta|\gamma_{it}^*, X)$) by multiplying Equation (17) and Equation (19) to get

$$P(\theta|\gamma_{it}^*, X) \propto \exp -\frac{1}{2}(\gamma_{it}' - X\theta)'(\gamma_{it}^* - X\theta) \exp -\frac{1}{2\sigma^2}\theta'\theta \quad (20)$$

we expand the term in the exponent to get

$$P(\theta|\gamma_{it}^*, \sigma_2, X) \propto \exp -\frac{1}{2}\{(\gamma_{it}^{*'} \gamma_{it}^* - \gamma_{it}^{*'} X\theta - \theta' X' \gamma_{it}^* + \theta' X' X\theta + \frac{1}{\sigma^2} \theta' \theta)\} \quad (21)$$

Since the first term is a constant with respect to θ , it can be removed as a multiplicative proportionality constant, thus we have

$$P(\theta|\gamma_{it}^*, \sigma_2, X) \propto \exp -\frac{1}{2}\{(-\gamma_{it}^{*'} X\theta - \theta' X' \gamma_{it}^* + \theta' X' X\theta + \frac{1}{\sigma^2} \theta' \theta)\} \quad (22)$$

rewrite Equation (22) as

$$P(\theta|\gamma_{it}^*, \sigma_2, X) \propto \exp -\frac{1}{2}\{(-\gamma_{it}^{*'} X\theta - \theta' X' \gamma_{it}^* + \theta' (X' X + \frac{1}{\sigma^2})\theta)\} \quad (23)$$

In Equation (23), the first two terms are identical to one another i.e one is just a transposed version of the other, therefore the two terms can be added and Equation (23) rewritten as

$$P(\theta|\gamma_{it}^*, \sigma_2, X) \propto -\frac{1}{2}\{\exp(\theta' (X' X + \frac{1}{\sigma^2})\theta) - 2\gamma_{it}^{*'} X\theta\} \quad (24)$$

To simply the expression in Equation (24), we let

$$M = (X' X + \frac{1}{\sigma^2})$$

and

$$b = X' \gamma_{it}^*$$

Then Equation (24) becomes

$$P(\theta|\gamma_{it}^*, \sigma_2, X) \propto \exp -\frac{1}{2}\{\theta' M\theta - 2b'\theta\} \quad (25)$$

completing the square (quadratic form) in Equation (25) and applying the following identity

$$\theta' M\theta - 2b'\theta = (\theta - M^{-1}b)' M(\theta - M^{-1}b) - b' M^{-1}b$$

Equation (25) becomes

$$P(\theta|\gamma_{it}^*, \sigma_2, X) \propto \exp -\frac{1}{2}\{(\theta - M^{-1}b)' M(\theta - M^{-1}b) - b' M^{-1}b\} \quad (26)$$

which can be written as a product of likelihood and prior as follow;

$$P(\theta|\gamma_{it}^*, \sigma_2, X) \propto \exp -\frac{1}{2}\{(\theta - M^{-1}b)' M(\theta - M^{-1}b)\} \exp -\frac{1}{2}[b' M^{-1}b] \quad (27)$$

The Gaussian density of Equation (27) is given by

$$P(\theta|\gamma_{it}^*, \sigma_2, X) \propto \exp -\{[(\theta - (X' X + \frac{1}{\sigma^2})^{-1} X' \gamma_{it}^*) (X' X + \frac{1}{\sigma^2})] [(\theta - (X' X + \frac{1}{\sigma^2})^{-1} X' \gamma_{it}^*)]\}$$

but γ_{it}^* is a constant value and not a density in θ so it can be removed as a multiplicative proportionality constants, therefore we rewrite it in Gaussian form as as

$$P(\theta|\gamma_{it}^*, \sigma_2, X) \propto \exp -(\theta - (X' X + \frac{1}{\sigma^2})^{-1} X' \gamma_{it}^*) (X' X + \frac{1}{\sigma^2}) X' (\theta - (X' X + \frac{1}{\sigma^2})^{-1} X' \gamma_{it}^*)$$

which implies that a posterior mean is given by

$$E[\theta] = (X' X + \frac{1}{\sigma^2})^{-1} X' \gamma_{it}^* \quad (28)$$

thus, the Bayes estimate of $\hat{\theta}$ is

$$\hat{\theta} = (X' X + \frac{1}{\sigma^2})^{-1} X' \gamma_{it}^* \quad (29)$$

In Equation (29), γ_{it}^* and X are observed from an experiment.

Next we obtain the estimate of $\hat{\sigma}$ which we show how to estimate in the next section.

8.2 Estimation of variance

To estimate σ^2 in Equation (21), we assume that ϵ is randomly distributed with mean 0 and variance σ^2 i.e $\epsilon \sim N(0, \sigma^2)$. Therefore the density function for the i^{th} observation is given as

$$f(\epsilon|\sigma^2) = \prod_{i=1}^n \left(\frac{1}{\pi}\right)^{\frac{1}{2}} \left(\frac{1}{2\sigma^2}\right)^{\frac{1}{2}} \exp -\frac{1}{2\sigma^2} (\gamma_{it}^* - X\theta)^2 \quad (30)$$

when we use the proportionality sign (\propto), we do not write the term that does not involve σ^2

$$f(\epsilon|\sigma^2) \propto \left(\frac{1}{2\sigma^2}\right)^{\frac{n}{2}} \exp -\frac{1}{2\sigma^2} \sum_{i=1}^n (\gamma_{it}^* - X\theta)^2 \quad (31)$$

To estimate σ^2 , we need a family of prior distribution that has a support on $(0, \infty)$ and one such family of distribution is the gamma family. When using such a prior distribution, we say that σ^2 has inverse-gamma distribution. Therefore, it is convenient to work with the precision parameter τ .

Let $\tau = (\frac{1}{2\sigma^2})$. Then Equation (31) becomes

$$f(\epsilon|\sigma^2) \propto \tau^{\frac{1}{2}} \exp -\tau \sum_{i=1}^n (\gamma_{it}^* - X\theta)^2 \quad (32)$$

Suppose the conjugate prior of τ is the gamma density $G(\alpha, \beta)$ i.e

$$p(\tau) = \frac{\beta}{\Gamma\alpha} \tau^{\alpha-1} \exp -\tau\beta \quad (33)$$

then the posterior distribution $P(\tau|\gamma_{it}^*, X)$ of τ is given by

$$p(\tau|\gamma_{it}^*, X) \propto \tau^{\frac{1}{2}} \exp -\tau \sum_{i=1}^n (\gamma_{it}^* - X\theta)^2 \times \frac{\beta}{\Gamma\alpha} \tau^{\alpha-1} \exp -\tau\beta \quad (34)$$

dropping the term that does not involve a function of τ , we have

$$p(\tau|\gamma_{it}^*, X) \propto \tau^{\frac{1}{2}} \exp -\tau(\sum_{i=1}^n (\gamma_{it}^* - X\theta)^2) \times \tau^{\alpha-1} \exp -\tau\beta \quad (35)$$

Let $r \triangleq \frac{1}{2}$, then Equation (35) simplifies to

$$p(\tau|\gamma_{it}^*, X) \propto \tau^{r+\alpha-1} \exp -\tau \sum_{i=1}^n ((\gamma_{it}^* - X\theta)^2 - X\theta)^2 + \beta) \quad (36)$$

thus

$$p(\tau|\gamma_{it}^*, X) \propto G(r + \alpha, \sum_{i=1}^n ((\gamma_{it}^* - X\theta)^2) + \beta) \quad (37)$$

therefore

$$E[\tau] = \frac{r + \alpha - 1}{\sum_{i=1}^n (\gamma_{it}^* - X\theta)^2 + \beta} \quad (38)$$

and

$$var[\tau] = \frac{r + \alpha - 1}{\sum_{i=1}^n (\gamma_{it}^* - X\theta)^2 + \beta} \quad (39)$$

Next, we compute the Bayes estimator of σ^2 . We know that $2\sigma^2 = \frac{1}{\tau}$

then

$$\sigma^2 = \frac{1}{2\tau}$$

Therefore, the Bayes estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{2} \left[\frac{1}{\hat{\tau}} \right] = \frac{1}{2} \left[\frac{\sum_{i=1}^n (\gamma_{it}^* - X\theta)^2 + \beta}{r + \alpha - 1} \right]$$

we simplify to get

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (\gamma_{it}^* - X\theta)^2 + \beta}{2r + 2\alpha - 2}$$

on further simplification, we get

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (\gamma_{it}^* - X\theta)^2 + \beta}{n + 2\alpha - 2} \quad \text{provided } n + 2\alpha > 2 \quad (40)$$

Since the Fisher information for σ is proportional to $\frac{1}{\sigma^2}$ as it will be seen Jeffreys prior density in the present case is proportional to the improper density $\frac{1}{\sigma^2}$ which induces for τ the density $\frac{1}{\tau} d\tau$. This corresponds to the limiting case $\alpha = 0$ and $\beta = 0$. Hence the Bayes estimator is

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (\gamma_{it}^* - X\hat{\theta})^2}{n - 2}, \quad n > 2 \quad (41)$$

Now we need to obtain the empirical parameter estimates of the model (29) and the model (41) using collected data. The data was collected from MMUST health facility for the time period of 1st January, 2015 to 31st December, 2015.

9 Results

9.1 Environmental data

In this study a case of malaria was defined as a patient (student) seeking medical attention. Microscopically confirmed malaria cases was collected from MMUST health facility over an average of 5 years. The data was extracted from records of students who visited the health facility for consultations with medical officer for the years 2011 through 2015. The average monthly rainfall in mm, normalized difference vegetation index (NDVI), minimum and maximum temperature in ($^{\circ}C$) data was obtained from Kakamega Central meteorological station for the time period between 2011-2015.

9.2 Analysis of data to determine malaria incidence rate

The association between climatic/environmental variables and malaria incidence was carried out using Pearson's correlation analysis and outcome of the association shown in Table 1.

Pearsons correlation analysis showed that malaria incidences were highly correlated with climatic variables at 5% significance level. All correlations were positive and the highest correlation was found to exist between maximum temperature and NDVI (coefficient= .512).

Exploration of malaria incidence, temperature, rainfall and vegetation (NDVI) was first carried out to identify regularities in the data. The results in Fig. 2 showed that exploration of malaria incidence, rainfall, temperature and vegetation for the time period 2011-2015 exhibits numerous peaks, many of which appear to be equally spaced with no clear upward and downward trend. The equally spaced peaks suggests the presence of a periodic component with seasonal dependency in the series of data with 12-month period. From the figure we can also see that there peaks that do

not appear to be part of the seasonal pattern and which represent significant deviations from the neighbouring data points.

Periodogram of the malaria incidence rate with Fast Fourier and Tukey Transformation was carried out to identify the periodic oscillations to be modelled. Fig. 3 shows the periodogram of malaria incidence.

Table 1. Association between Climatic Variables and Monthly Malaria Incidence

Variable	Test	Mal incidence	Rainfall	MinTemp	MaxTemp	NDVI
Mal incidence	Pearson Cor	1	0.409	0.326	0.301	0.407
	Sig. (2-tailed)		0.001	0.011	0.02	0.001
Rainfall	Pearson Cor	0.409	1	0.486	0.546	0.481
	Sig. (2-tailed)	0.001	0.000	0.000	0.000	0.000
Min Temp	Pearson Cor	0.326	0.486	1	0.152	0.347
	Sig. (2-tailed)	0.011	0.000		0.246	0.007
Max Temp	Pearson Cor	0.301	0.546	0.152	1	0.512
	Sig. (2-tailed)	0.020	0.000	0.246		0.000.
NDVI	Pearson Cor	0.407	0.481	0.347	0.512	1
	Sig. (2-tailed)	0.001	0.000	0.007	0.000	

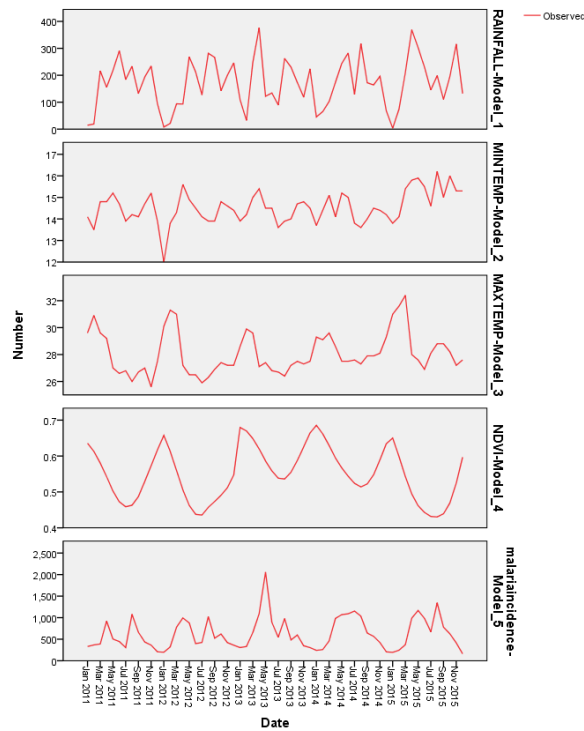


Fig. 2. Observed monthly data for malaria incidence, rainfall, temperature and Vegetation

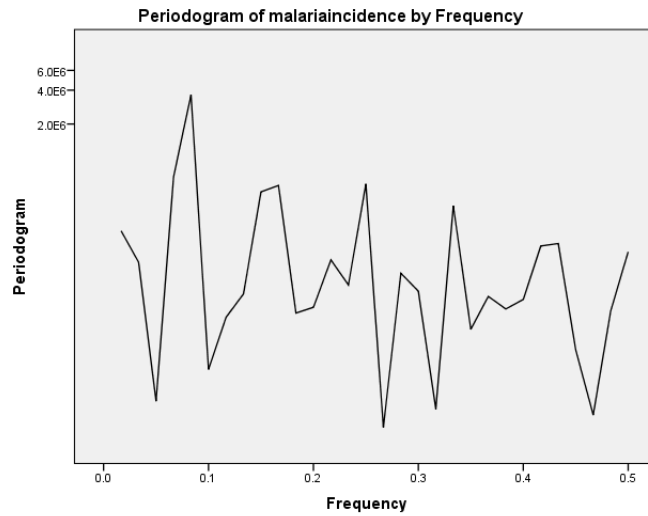


Fig. 3. Periodogram of malaria incidence

Table 2. Univariate Statistics for malaria incidence

Frequency	Period	Sine Transform	Cosine Transform	Periodogram
.00000	.000	627.100	.000	1519980.364
.01667	-60.112	-62.157	224308.821	1274539.658
.03333	-58.492	22.543	117885.095	2269020.477
.05000	-3.205	-14.763	6846.454	10300179.316
.06667	-149.465	14.129	676186.498	9940058.033
.08333	142.869	-317.734	3640998.571	9792371.683
.10000	14.539	15.040	13127.341	9861048.189
.11667	30.484	18.444	38083.043	9368580.069
.13333	-44.158	10.248	61648.306	2615516.270
.15000	112.032	63.637	498027.285	2697150.533

The plot of the periodogram shows a sequence of peaks that stand out from the background noise, with the lowest frequency peak at a frequency of just less than 0.1. Each of the data points in the time series represents a month, so an annual periodicity corresponds to a period of 12 in the current data set. Because period and frequency are reciprocals of each other, a period of 12 corresponds to a frequency of 0.08333 shown in Table 2. Therefore, an annual component implies a peak in the periodogram at 0.08333, which seems not to be the case with the peak just below a frequency of 0.1. The univariate statistics in Table 2 contains the data points that were used to plot the periodogram. From the table, we can see that for frequencies that less than 0.1, the largest value in the Periodogram column occurs at a frequency of 0.01667 and not 0.08333 as it is expected to be for the annual periodic component. Periodic components that have the shape of a sine or cosine function (sinusoidal) show up in the periodogram as single peaks. Periodic components that are not sinusoidal show up as a series of equally spaced peaks of different heights, with the lowest frequency peak in the series occurring at the frequency of the periodic component. Therefore, from the results in Table 2, we conclude that the data contains no periodic oscillation.

To remove seasonality from the data we carried out trend analysis using seasonal decomposition in time series. Fig. 4-8 shows seasonal adjustment for the series of data.

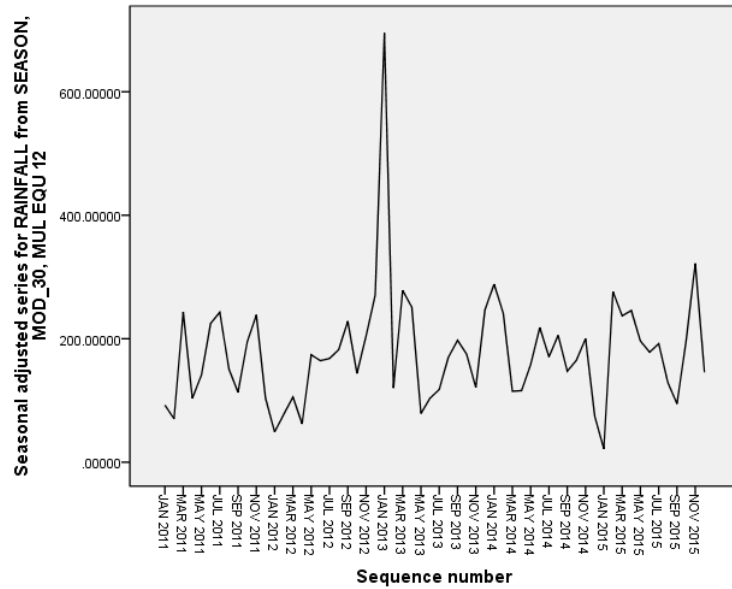


Fig. 4. Seasonal Adjustment for Rainfall data

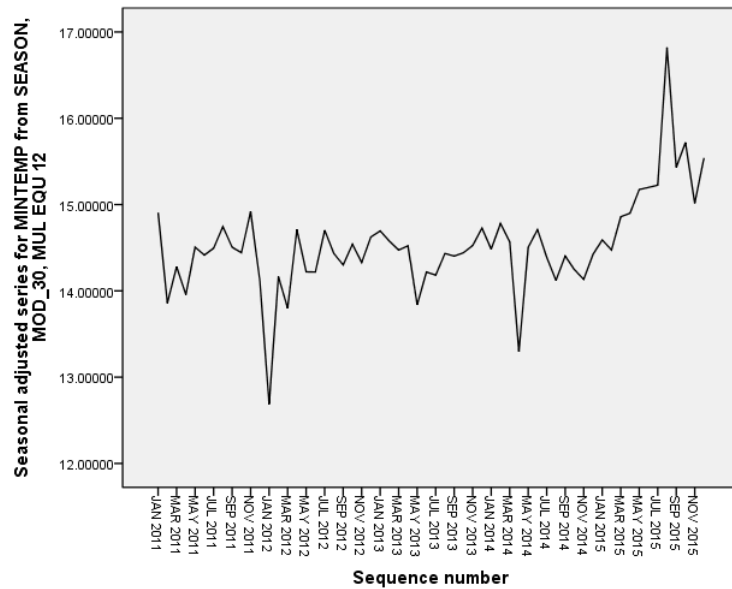


Fig. 5. Seasonal Adjustment for minimum temperature data

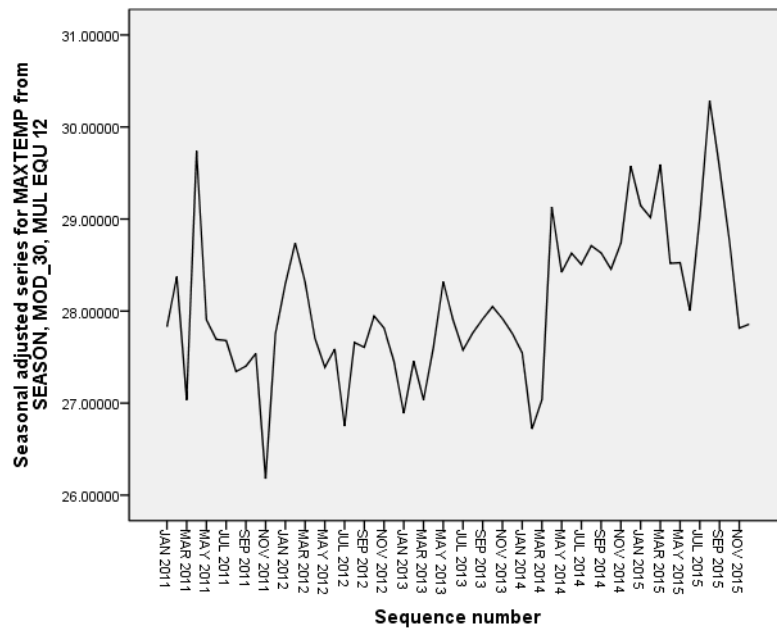


Fig. 6. Seasonal Adjustment for maximum temperature data

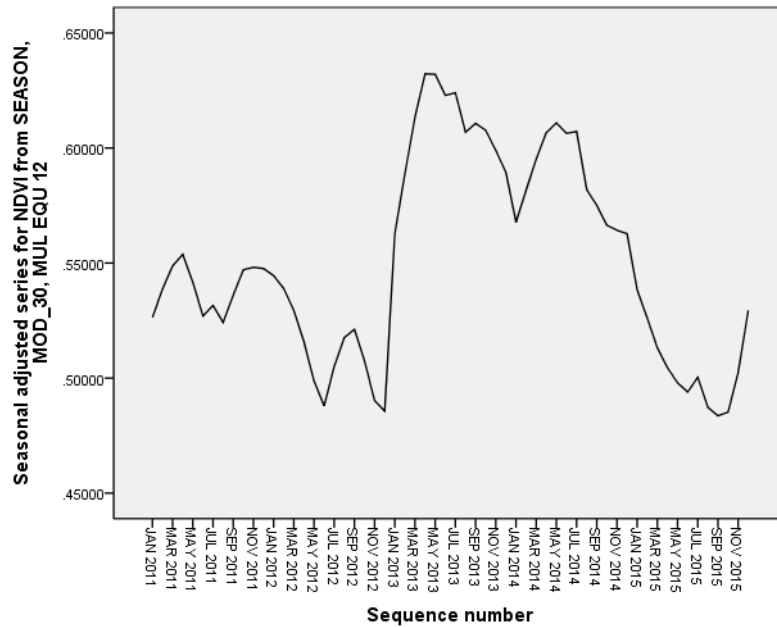


Fig. 7. Seasonal Adjustment for NDVI data

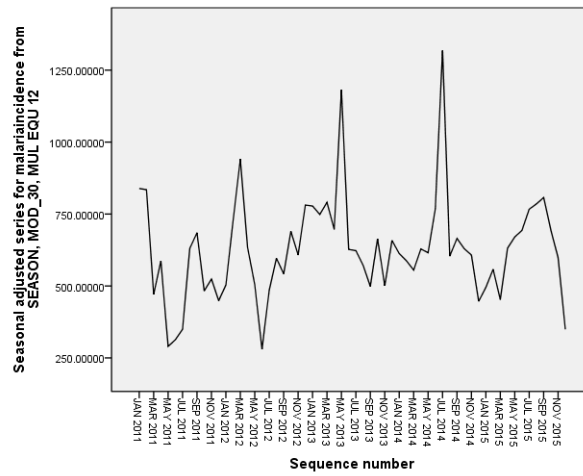


Fig. 8. Seasonal Adjustment for malaria incidence data

The seasonally adjusted series shows a clear upward trend. A number of peaks are evident, but they appear at random intervals, showing no evidence of an annual pattern.

Correlograms was carried out for the simple autocorrelation function (ACF) for the malaria incidence, temperature, rainfall and NVDI variables with lags equal to their period of oscillation. From Figs. 9 - 13, the estimated correlations in the ACF decay to zero, this suggests that the time series is stationary and should not be differenced. The results further shows how the ACF looks like for the malaria incidence, rainfall, temperature and NDVI variable. The autocorrelation function shows a significant peak at a lag of 1 with a long exponential tail i.e typical pattern for time series. We can clearly see that their is moderately large negative/positive spike at the first lag followed by correlations that bounce around between being positive and negative of which they are either not statistically significant or just barely cross the threshold of statistical significance.

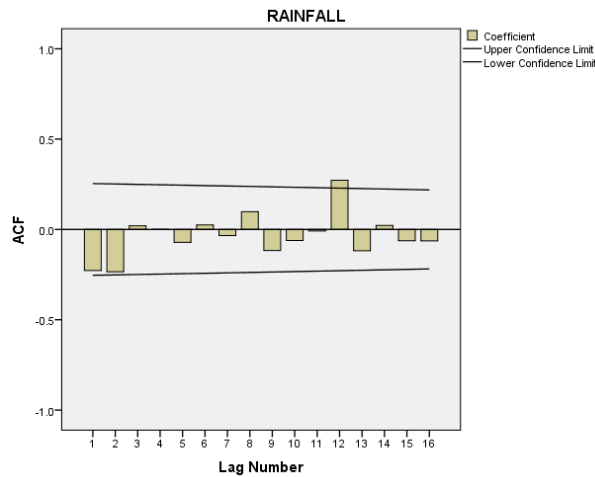


Fig. 9. Observed Autocorrelation plot for rainfall data

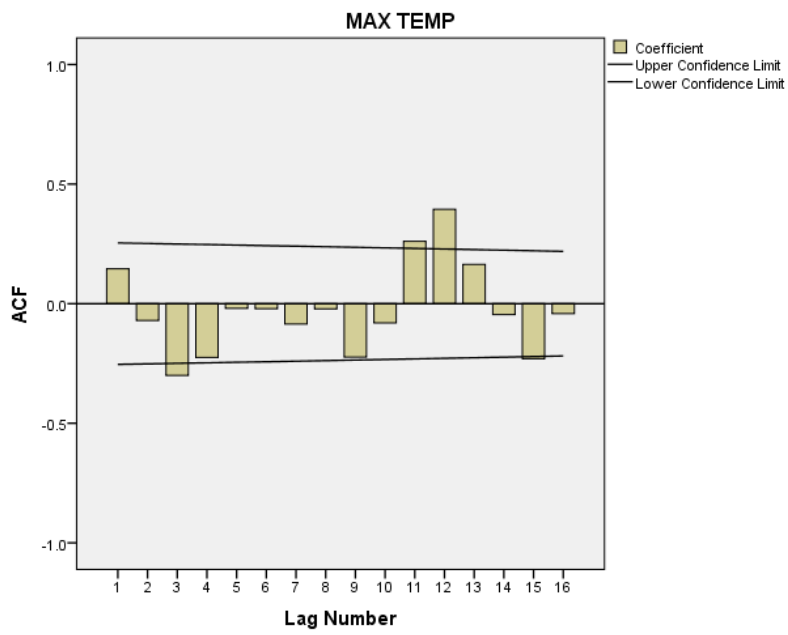


Fig. 10. Autocorrelation plot for maximum temperature data

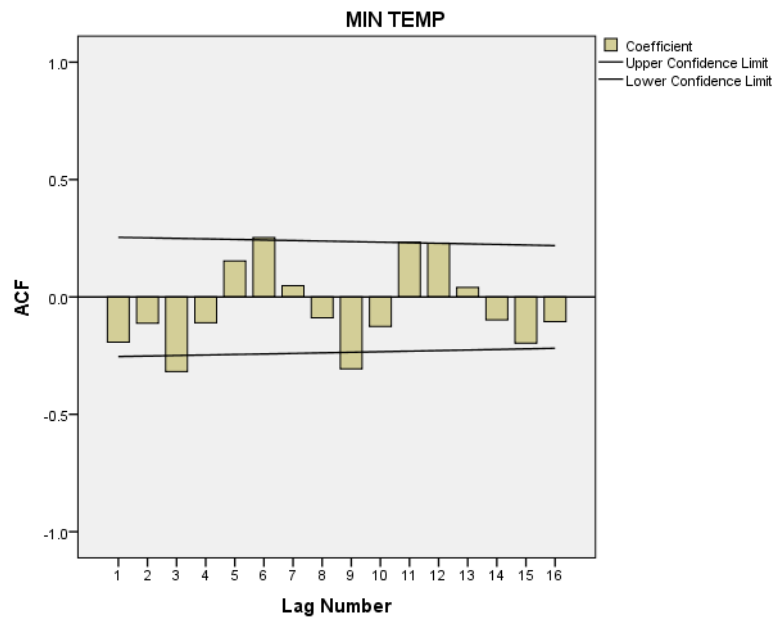


Fig. 11. Autocorrelation plot for minimum temperature data

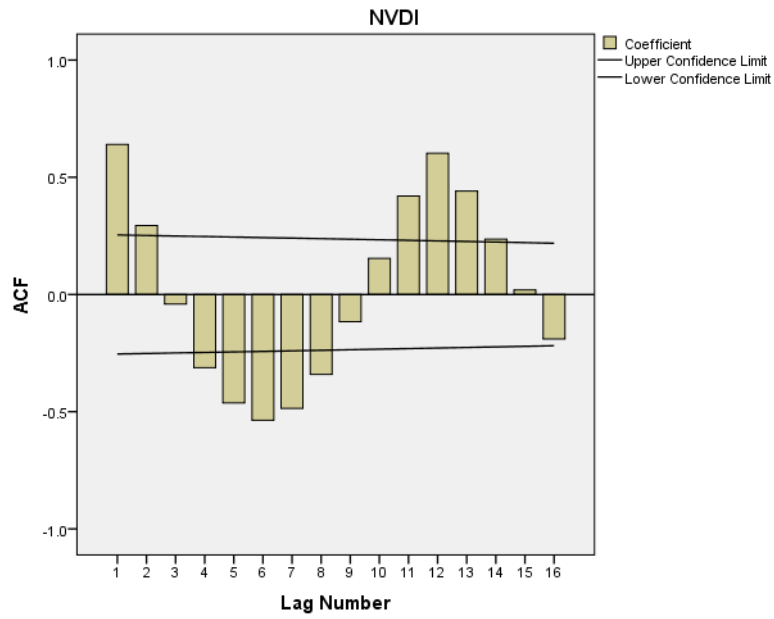


Fig. 12. Autocorrelation plot for NDVI data

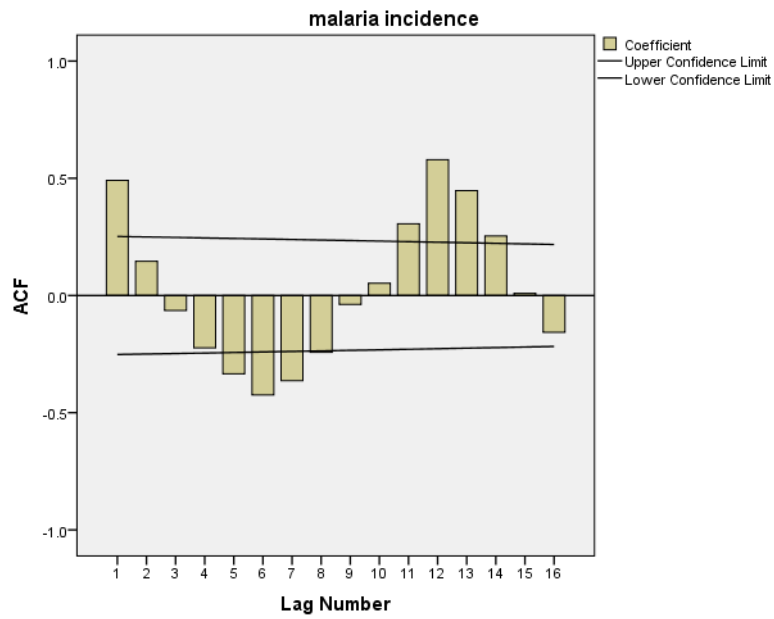


Fig. 13. Autocorrelation plot for malaria incidence data

Correlograms was also carried out for the partial autocorrelation function (PACF) for malaria incidence, rainfall, temperature and NDVI variables as shown in Figs. 14 - 18.

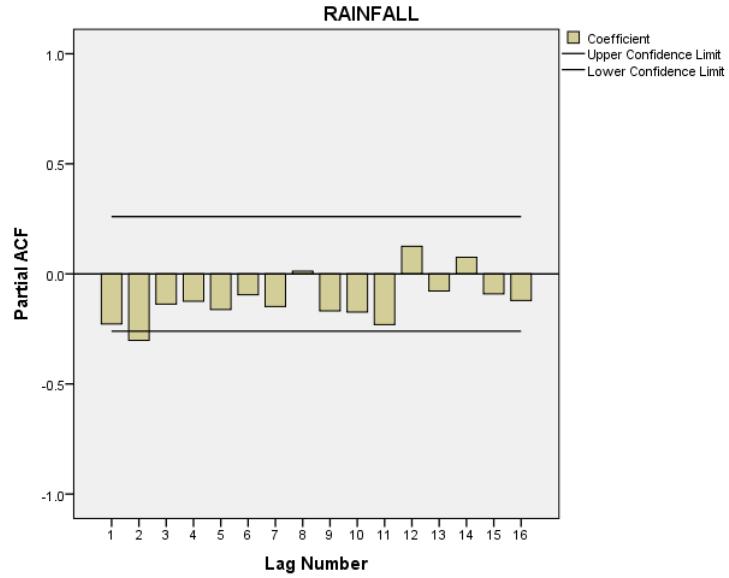


Fig. 14. Partial autocorrelation plot for rainfall data

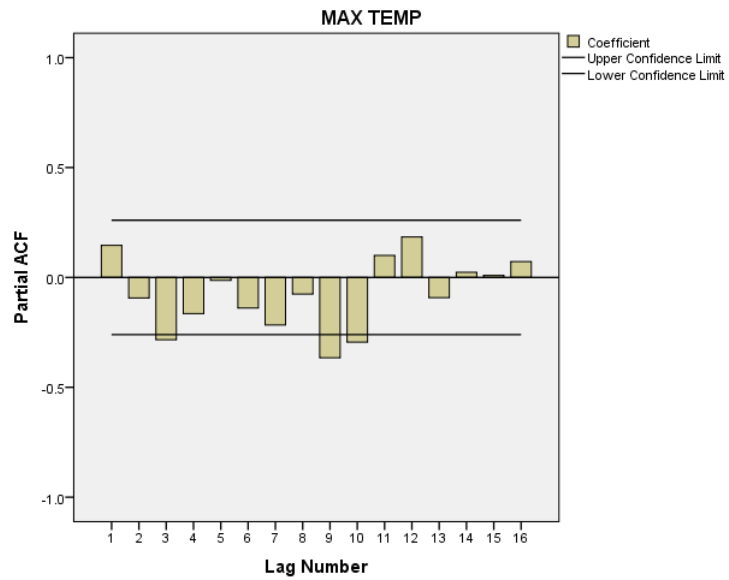


Fig. 15. Partial autocorrelation plot for maximum temperature data

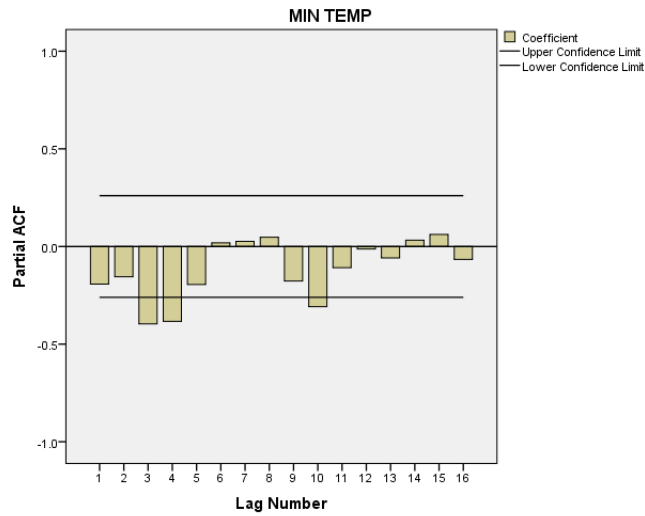


Fig. 16. Partial autocorrelation plot for minimum temperature data

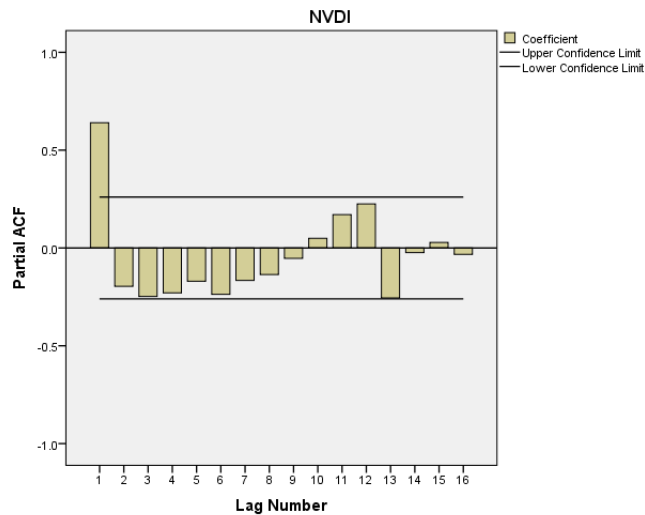


Fig. 17. Partial autocorrelation plot for NDVI data

The results in Figure 14 - 18 shows a decay in the partial correlations toward zero. From the figures, we can conclude that we estimate the model using the first difference i.e ARIMA(0,1,0)

Next we carried out estimation to determine the values that influences malaria incidence.

In Table 3, results showed that malaria incidence is best described by a seasonal ARIMA model with one order of differencing. The seasonal nature of the model accounted for the seasonal peaks in the series of plots and the single order of differencing reflects the upward trend that was evident in the data.

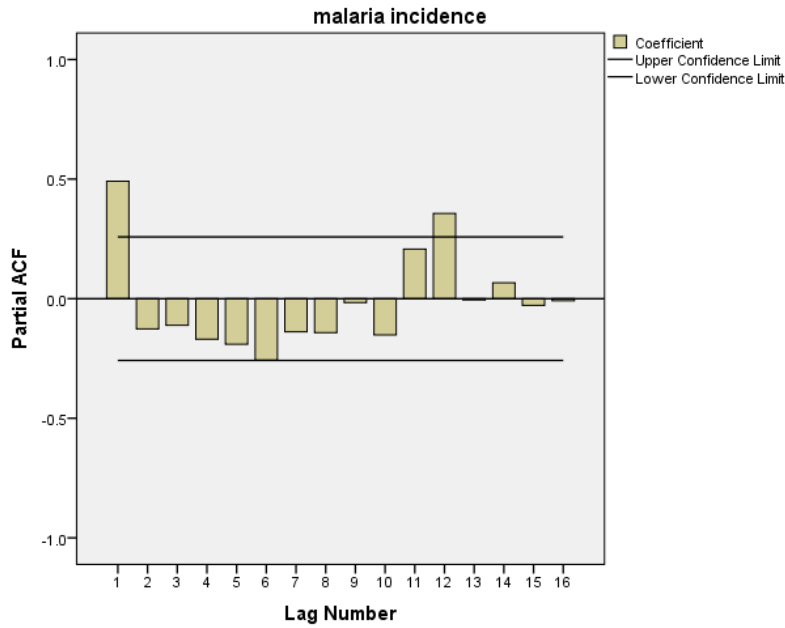


Fig. 18. Partial autocorrelation plot for malaria incidence data

Table 3. Model description

	Model Type
malaria incidence $Model_1$	ARIMA(0,0,0)(0,1,0)

Table 4. Model statistics

Model	Predictors	Model Fit statistics Stationary R-squared	Ljung-Box Q(18)		
			Statistics	DF	Sig.
malaria incidence ₁	3	.732	28.886	15	.571

The model statistics in Table 4 provides summary information and goodness-of-fit statistics for each estimated model. From the table, the model contains three predictors out of the four candidate predictors that were originally specified. Therefore, the Expert Modeller identified three independent variables used in for forecasting. The stationary R-squared value provides an estimate of the proportion of the total variation in the series that is explained by the model and is preferable to ordinary R-squared when there is a trend or seasonal pattern, as is the case with the model. Larger values of stationary R-squared indicates better fit. For instance a value of 0.732 means that the model does an excellent job of explaining 73% of the observed variation in the series. The Ljung-Box statistic provides an indication of whether the model is correctly specified. A significance value less than 0.05 implies that there is structure in the observed series which is not accounted for by the model. The value of 0.571 shown in Table 4 is not significant, therefore we conclude that the model is correctly specified.

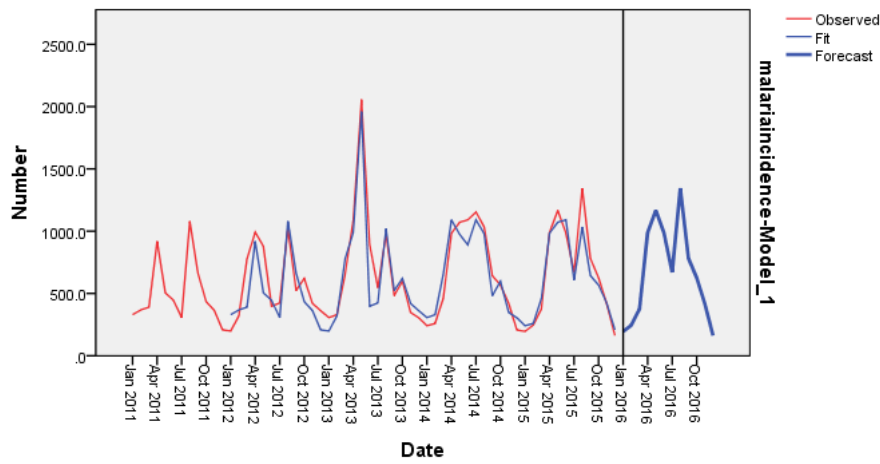


Fig. 19. Predicted and observed values of the data

Figure 19 shows the predicted value and the observed data. From the figure, the predicted values shows good agreement with the observed values indicating that the model had satisfactory predictive ability. For instance, the figure shows how the model predicts the seasonal peaks by capturing the upward and downward trend of the data.

Table 5. Regression Modelling results

Variable	coefficient	SE	95% CL	p-value
I_{t-1}	.67	.088	0.55 -0.75	.0001
$\sin(0.52R_{t-1}T_{t-1}V_{t-1})$.83	.185	0.45-0.95	.0001

In Table 5, the results showed that the model explained a substantial percentage of the observed variability in the malaria incidence rate ($R^2=73\%$) with a coefficient of 0.67 for the autoregressive term and 0.83 for the environmental/climatic factors. The model leaves a base rate residual in the form of white noise which is normally distributed with a mean 0 and standard deviation of 0.088. Therefore, after adding the autoregressive term and the term representing the influence of environmental/climatic variables, we get the parameter estimates of in model (10) as

$$I_t = 0.67I_{t-1} + 0.83\sin(0.52R_{t-1}T_{t-1}V_{t-1}) \tag{42}$$

where I_t is the malaria incidence rate for any month t , I_{t-1} is the observed rate in the preceding month, R_{t-1} is the cumulative rainfall in the preceding month, T_{t-1} is the mean maximum temperature of the preceding month and V_{t-1} is the mean vegetation density in the preceding month.

9.3 Symptomatic data

The data for malaria symptoms was extracted from records of students who visited the health facility for consultations with medical officer for the year 2015. The data was then divided into two; structured test data and unstructured test data. The structured test data consisted of laboratory results (based on doctor’s recommendation for blood slide (BS) test), age, gender and registration number of the students while the unstructured test data consisted of students narration of their

illness and doctors' interrogation and diagnosis. For this study, unstructured test data was used for analysis. The 9 observed symptoms used for the the study were; fever (body temperature), rigors, convulsion, sweating, vomiting, diarrhoea, pallor, cough and prostration. The number of symptoms for each student was then recorded on an ordinal scale from 0 to 3 with 0 being symptoms that cannot cause malaria, 1 being symptoms that can cause mild malaria, 2 being symptoms that can cause moderate malaria and 3 being symptoms that cause severe malaria (Appendix VIII). These information is summarised in Table 3.6 using frequency distribution.

Table 6. Frequency distribution of Malaria severity

malaria severity	Frequency	Percent	ordinal scale
No malaria symptoms	20	6.7	0
mild symptoms	46	15.3	1
moderate symptoms	52	17.3	2
severe symptoms	182	60.7	3
Total	300	100.0	

In Table 6, results showed that 20(6.7%) of the students who visited the health facility with malaria-like symptoms had no malaria infection despite the fact that they had displayed initially displayed malaria-like symptoms. The absence of malaria infection may be due to symptom(s) overlap or misdiagnosis of the disease. The results also showed that 46 (15.3%) had mild symptoms (mild malaria infection), 52(17.3%) had moderate symptoms (moderate malaria infection) while 182(60.7%) had severe symptoms (severe malaria infection). The presence of malaria infection among the students can highly be attributed to the influence of environmental factors on the malaria incidence and the endemicity of the area of study which is Hyperendemic i.e the prevalence of malaria infection is between 51% and 75%.

Let

X_1 =fever, X_2 =rigors, X_3 =convulsion, X_4 =sweating, X_5 =vomiting, X_6 =diarrhoea
 X_7 =pallor, X_8 =cough, X_9 =prostration and y =status of an individual (ordinal scale) as a result of transition due to observed symptoms.

Table 7 gives the descriptive statistics of the collected data.

Table 7. Descriptive statistics

Variable	%	Mean	Std. Dev.	S.E	skewness	kurtosis
X_1	73.7	1.806667	0.901108	0.0520255	-0.0022965	1.909835
X_2	50.7	1.716667	1.194912	0.0689883	-0.1478512	1.445367
X_3	53	1.743333	1.250066	0.0721726	-0.2266415	1.381524
X_4	57.3	0.79	0.407988	0.0235552	-0.423983	3.022728
X_5	72.3	1.903333	1.047687	0.0604882	-0.5580152	2.101434
X_6	44.3	1.23	1.077483	0.0622085	0.3369825	1.832919
X_7	58.0	1.8	1.075728	0.0621072	-0.435974	1.92676
X_8	0.7	0.34	1.214002	0.0700905	-0.0160609	1.43978
X_9	80.3	1.91333	0.4951943	0.0285901	1.087602	3.991302
y	78.0	1.923333	1.217537	0.0702945	0.0009702	2.68562

From Table 7, results showed that prostration (80.3%), fever (73.7%) and vomiting (72.3%) were the highly observed symptoms among students with cough (0.7%) being the least observed. The

small standard error (SE) in the results shows that the sample mean is a reflection of the actual population mean. For instance, fever (X_1) had a mean of 1.8067 with a S.E of 0.0521 while cough (X_8) had a mean of 0.34 with S.E of -0.01606. The standard deviation (SD) indicates a variation between set of scores with the mean, for instance, fever X_1 with a mean of 1.8067 had SD of 0.9011 implies that 73.7% on average of fever X_1 was over 0.9011 point away from the mean.

9.4 Simulation of regression parameters

To estimate the parameters of the model, we used Normal distribution as a prior for θ parameter and the Normal-gamma distribution as the prior for σ^2 . The regression model in Equation (11) was fitted to the data to obtain the parameter estimates of the model using Bayesian method where random samples were generated based on the posterior distribution of each parameter using Gibbs sampler via Markov Chain Monte Carlo (MCMC) method which produced an iteration of about 12500 with Burn-in of 1000 and thinning of 10 resulting to effective sample size of 10000 as shown in Table 3.8

Table 8. Bayesian simulation results

Name	value
Number of chains Per MCMC	= 3
Iterations	= 12500
Burn-in	= 1000
Sample size	= 10,000
Number of observation	= 300
Average acceptance rate	= 0.3394
Average efficiency:minimum	= .9927
Average efficiency:maximum	= .9993
Maximum Gelman-Rubin Rc	= 1.176
Average marginal likelihood	= -231.54658

From Table 8, the results showed that Gibbs algorithm achieved an overall acceptance rate (AR) of 34% and an average efficiency of about 99%. The acceptance rate specifies the proportion of proposed parameter values that was accepted by the algorithm. For the Gibbs algorithm, this number rarely exceeds 50% and efficiencies of greater than 10% is considered good [10]. An acceptance rate of 0.3394 in the results means that 34% out of 10000 proposal parameter values was accepted by the algorithm with efficiencies of 99%.

The summary of Bayesian simulation results is shown in Table 9.

Table 9. Bayesian Normal Regression with MH sampling

parameter	Mean	Std. Dev.	MCSE	Median	HPD[95%	Cred.Interval]
θ_1	0.1502703	.0317274	.000106	.0503697	.0110428	.1130387
θ_2	0.0260683	.0339402	.000113	.0459884	.0193016	.1131818
θ_3	0.1460505	.0437595	.000146	.4605534	.3747775	.5460086
θ_4	-0.0096916	.069681	.000232	-.0094585	-.1461358	.1256172
θ_5	0.1381898	.0331668	.000111	.0820614	.0190056	.1488435
θ_6	0.1452622	.0271493	.00009	-.1134825	-.1663481	-.0601344
θ_7	0.1344508	.0432297	.000144	.4525424	.3671538	.5370451
θ_8	-0.1135517	.0292297	.000097	.0643512	.0052016	.1200136
θ_9	0.1678915	.0584017	.000195	.0078181	-.1069035	.1222403
σ^2	0.2395643	.0198361	.000066	.2384632	.2019022	.278897

From the results in Table 9, the median column provides estimates of median of the posterior distribution which is used to assess the symmetries of the posterior distribution. From the results, the estimates of means and medians were very close implying that the posterior distribution is symmetric. The results also shows credible intervals which have probabilistic interpretation, for instance, the probability of observing fever (X_1) has a credible interval that lies between 0.0110428 and 0.1130387. The lower bound of the interval is greater than 0, which implies that fever has an effect on the transition of an individual from one disease state to the next disease state.

Therefore the Bayes estimates of the parameter θ in Equation (29) is

$$\hat{\theta} = \begin{bmatrix} 0.1502703 \\ 0.0260683 \\ 0.1460505 \\ -0.0096912 \\ 0.1381898 \\ 0.1452622 \\ 0.1344508 \\ -0.1135517 \\ 0.1678915 \end{bmatrix} \quad (43)$$

Therefore, the regression components in Equation (??) can now be written as

$$\hat{\gamma}_{it}^* = 0.1502703X_1 + 0.0260683X_2 + 0.1460505X_3 - 0.0096912X_4 + 0.1381898X_5 + 0.1452622X_6 + 0.1344508X_7 - 0.1135517X_8 + 0.1678915X_9$$

Based on Equation(9), the value of transition parameter is now given as

$$\begin{aligned} \psi_{it} &= 0.67I_{t-1} + 0.83\sin(0.52R_{t-1}T_{t-1}V_{t-1}) + 0.1502703X_1 + 0.0260683X_2 + 0.1460505X_3 \\ &- 0.00969142X_4 + 0.1381898X_5 + 0.1452622X_6 + 0.1344508X_7 - 0.1135517X_8 \\ &+ 0.1678915X_9 \end{aligned}$$

Therefore the transition probability model (1) is now given as

$$\begin{aligned} a_{iSI}^{(t)} &= \phi(0.83 + 0.5027028X_1 + 0.0460683X_2 + 0.4605055X_3 - 0.0093X_4 \\ &+ 0.3818986X_5 + 0.4526221X_6 + 0.1344508X_7 - 0.1135517X_8 + 0.7778915X_9) \end{aligned}$$

We illustrate the above computation by using an example.

Example 3.1: Consider a student displaying the following symptoms; fever, diarrhoea, cough and rigors. The likelihood that the student will transit from susceptible state to infective state is given by

$$\begin{aligned} a_{iSI}^{(t)} &= \phi(0.83 + 0.5027028X_1 + 0.0460683X_2 + 0.4526221X_6 - 0.1135517X_8) \\ &= \phi(1.6717732) \\ &= 0.9525 \end{aligned}$$

Therefore, from the results, the student has a higher chance (95.25%) of transiting from susceptible state to infective state.

10 Conclusions

Bayesian method using Markov Chain Monte Carlo via Gibbs sampling algorithm was implemented for obtaining the parameter estimates of the model. The results of the study showed a significant association between malaria disease and observed symptoms. the estimate also indicates that a one

unit change in observed symptom is has a significant effect on the transition of an individual from one disease state to the other disease state.

Acknowledgment

I happily express my thanks to Prof. Kennedy for providing valuable comments that helped to improve this paper. Special gratitude to Mathematics Department, Masinde Muliro University of Science and Technology, I can't forget to thank my lovely family; my wife Stellah Makokha for their endless motivation and support throughout my study, may God bless you.

Competing Interests

Authors have declared that no competing interests exist.

References

- [1] Mandal S, Sarkar RR, Sinha S. Mathematical models of malaria-a review. *Malaria Journal*. 2011;10:202.
- [2] WHO. Malaria Fact Sheet 2017 Report; 2017.
- [3] WHO. World Malaria Report; 2012.
- [4] Dondorp AM, Day NP. The treatment of severe malaria. *Trans R Soc Trop Med Hyg*. 2007;101:633-634.
- [5] Martins, et al. Clustering symptoms of non-severe malaria in semi-immune Amazonian patients. *PeerJ*; 2015.
- [6] Teklehaimanot HD, Lipsitch M, Teklehaimanot A, Schwartz J. Weather-based prediction of plasmodium falciparum malaria in epidemic-prone regions of Ethiopia I. patterns of lagged weather effects reflect biological mechanisms. *Malaria Journal*. 2004;3:41.
- [7] Xu J, Zeger S. Joint analysis of longitudinal data comprising repeated measures and time to events. *Journal of the Royal Statistical Society series C. Applied Statistics*. 2001;50:375-87.
- [8] Zammit, Nicola N,George Streftaris, Gavin J. Gibson, Ian J. Deary, Brian M. Frier. Modelling the consistency of hypoglycaemic symptoms: High variability in diabetes. *Diabetes Technology and Therapeutics*. 2011;13(5):571-578.
- [9] Alberto Gomez-Elipe, Angel Otero, Michel Van Herp Armando. Aguirre-Jaime Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997-2003. *Malaria Journal*. 2007;6:129.
- [10] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. Chapman and Hall. London; 1995.

© 2020 Mbete and Nyongesa; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

<http://www.sdiarticle4.com/review-history/64479>