



Available online at <http://scik.org>

J. Math. Comput. Sci. 9 (2019), No. 1, 73-86

<https://doi.org/10.28919/jmcs/3881>

ISSN: 1927-5307

BAYESIAN APPROACH IN MODELLING CHOLERA OUTBREAK IN ILALA MUNICIPAL COUNCIL, TANZANIA

YOHANA MAIGA MARWA^{1,2,*}, SAMUEL MWALILI³, ISAMBI SAILON MBALAWATA⁴

¹Mathematics and Statistics Department, Pan African University,

Institute for Basic Sciences, Technology and Innovation, P.O. Box 62000 – 00200, Nairobi, Kenya

²Tanzania Public Service College, P.O. Box 1534, Singida, Tanzania

³Jomo Kenyatta University of Agriculture and Technology, P.O. Box 62000 – 00200, Nairobi, Kenya

⁴African Institute for Mathematical Sciences, P.O. Box 106077, Dar es Salaam, Tanzania

Copyright © 2019 the authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract. In this paper, we consider a deterministic susceptible-infected-removed-bacteria (SIR-B) cholera epidemic model with water treatment as a control strategy. The infected compartments are divided into two classes namely; symptomatic infected and asymptomatic infected individuals in order to observe the contribution of vibrio cholerae to the environment from each compartment. The main objective of this paper is to validate the model using Ilala[†] municipal council data. We estimate the model parameters using least square and Bayesian approach via Markov chain Monte Carlo (MCMC) methods. The sample of 10,000 number of simulations was carried out in MCMC run to study the behaviour of each parameter in the proposed model. The results show that the parameters have good convergence, the dynamic behaviour of prediction fits the observed data.

Keywords: cholera epidemic; Bayesian; least square method; Markov chain Monte Carlo; adaptive Markov chain Monte Carlo.

2010 AMS Subject Classification: 92D30, 91G70.

[†]Ilalal municipal council is found in Dar es Salaam, Tanzania.

*Corresponding author

E-mail address: yohanamaiga2016@gmail.com

Received August 31, 2018

1. Introduction

Cholera is a water borne infectious disease caused by the ingestion of vibrio cholerae bacteria because of insufficient access to safe water and adequate sanitation [4]. It is considered as a serious infectious disease and can lead to high morbidity and mortality when remain untreated. Cholera is capable to spread rapidly, basing on the contract of exposure and the population exposed. It proved that pathogenic vibrio cholerae survive refrigeration and freezing in food supplies [3].

Measuring cholera cases in a population is a bit challenge in most of the developing countries due to the fact that disease incidences and deaths occur outside of the formal health care, particularly at home [8, 9]. The data from hospital provide a proxy for measuring the incidence of mild, severe cholera cases and estimating the morbidity rates [10]. The analysis of this data allows to assess and ultimately improve the health cares, assist in monitoring and planning of resources and provide appropriate interventions. There are several measures suggested by World Health Organization (WHO) and other scholars [1], to prevent cholera. These measures are environmental sanitation, water treatment, provision of clean water, provision of education on the effect of cholera [5].

The mathematical model for modeling cholera dynamics with a control strategy in Ghana is proposed in [2]. In their study, the results showed that the disease free equilibrium is locally asymptotically stable. But in their paper, there are some limitations such as, the model considers the infected individuals as a single compartment (I) and limited numerical analysis. Until then [12] developed a mathematical model that split the infected compartment (I) into two classes namely; symptomatic infected $I_s(t)$ and asymptomatic infected $I_a(t)$ individuals in order to observe the contribution of vibrio cholerae to the environment from each compartment and incorporate water treatment as a control strategy. The model was then synthetically tested, which in some cases it is hard to prove its validity in real life. Thus, in this paper we use Ilala municipal council data to test and validate the model developed by [12]. We use classical least square and Bayesian approach via Markov chain Monte Carlo simulation techniques to study uncertainties of model parameters.

2. Cholera Model

The mathematical model for cholera dynamics with two subpopulation; bacteria (pathogen) and individuals developed in [12]. Individuals were subdivided into four compartments namely susceptible individuals $S(t)$, symptomatic infected individuals $I_s(t)$, asymptomatic infected individuals $I_a(t)$ and recovered individuals $R(t)$. Susceptible individuals $S(t)$ contract disease at rate β and the influx of the susceptible comes from a constant recruitment rate b . The symptomatic infected individuals $I_s(t)$ become infected at a probability p and contribute vibrio cholerae through excretion to the environment at a rate α_1 , dies due to natural death and due to disease at rates μ and d respectively. The asymptomatic infected individuals $I_a(t)$ become infected at a probability q and contribute vibrio cholerae through excretion to the environment at a rate α_2 and recovery class $R(t)$ at rates r_1 and r_2 for $I_s(t)$ and $I_a(t)$ respectively. The pathogen concentration of vibrio cholerae as denoted by $B(t)$ decrease due to mortality rate ϕ and due to water treatment at a rate δ . Thus, the total human population is given by

$$N = S(t) + I_s(t) + I_a(t) + R(t),$$

at any given time t .

The formulated model equations by [12] is as follows;

$$\begin{aligned} \frac{dS(t)}{dt} &= bN - \frac{\beta B(t)S(t)}{\kappa + B(t)} - \mu S(t), \\ \frac{dI_s(t)}{dt} &= \frac{p\beta B(t)S(t)}{\kappa + B(t)} - (\mu + r_1 + d)I_s(t), \\ \frac{dI_a(t)}{dt} &= \frac{q\beta B(t)S(t)}{\kappa + B(t)} - (\mu + r_2)I_a(t), \\ \frac{dR(t)}{dt} &= r_1 I_s(t) + r_2 I_a(t) - \mu R(t), \\ \frac{dB(t)}{dt} &= \alpha_1 I_s(t) + \alpha_2 I_a(t) - (\delta + \phi)B(t), \end{aligned} \tag{2.1}$$

with initial conditions $S(0) > 0$, $I_s(0) \geq 0$, $I_a(0) \geq 0$, $R(0) \geq 0$, $B(0) \geq 0$ and $p + q = 1$.

Theoretically, the stability of equilibrium points were proved and the value of R_0 was computed using literature values, estimated values and MCMC mean and was found to be greater than 1. Numerically, the results showed that the parameters were identifiable[12]. However,

this model was not tested using the real data. Therefore, in this paper we validate the model using real data from ilala municipal council.

3. Materials and Methods

Bayesian approach via Markov chain Monte Carlo (MCMC), least square methods and model fitting are considered in this section.

3.1. Bayesian estimation and inference

Bayesian inference is the method of analysis that combines information collected from experimental data with the knowledge one has prior to performing the experiment. In this approach to statistics, the uncertainties are expressed in terms of probabilities. The classical approach to statistical inference based on the random sample. That is, if a probability distribution depends on a set of parameters θ , the classical approach makes inferences about θ solely on the basis of a sample x_1, x_2, \dots, x_n . This approach to inference is based on the concept of sampling distribution.

For parameter estimation approach, parameters θ are estimated based on some measurements y_i , ideally using e.g., least square method. For Bayesian estimation, parameters θ are interpreted as a random variable and the main goal always is to find the posterior distribution $P(\theta|\text{data})$ of the parameters. The posterior distribution is needed because it gives the probability density for the values of θ , given measurements y_i . Using Bayes' formula, the posterior density can be written as

$$P(\theta|\text{data}) = \frac{l(\text{data}|\theta)P(\theta)}{\int l(\text{data}|\theta)P(\theta)d\theta}$$

The likelihood $l(\text{data}|\theta)$ contains the measurement error model and it gives the probability density of observing measurement data given the parameter values of θ . Consider the model

$$y_i = f(x_i, \theta) + e_i,$$

and employing a Gaussian i.i.d error model, i.e., $e \sim N(0, \sigma^2 I)$ and noting that $e_i = y_i - f(x_i, \theta)$, it gives the likelihood of

$$l(y|\theta) \propto \prod_{i=0}^n l(y_i|\theta) \propto \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^n (y_i - f(x_i, \theta))^2\right),$$

where σ^2 is an estimate for the measurements error which is obtained using the mean square error (MSE)

$$\sigma^2 \approx MSE = \frac{RSS}{n - p},$$

where RSS (residual sum of squares) is the fitted value of the least square function, n is the number of measurements and p is the number of parameters.

Hence, in our typical applications, the likelihood distribution reads as

$$l(y|\theta) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - f(x_i, \theta))^2\right),$$

and the prior distribution $P(\theta)$ contains all existing information about parameters. However, the proposal distribution should be chosen so that it becomes easy to sample from and should be close to the target distribution (posterior) as possible. It should be noted that unsuitable proposal distribution has a high chance of giving inefficient implementations.

3.2. Parameter Estimation using Least square method

The least-square fit of the model provides estimates for the cholera epidemic parameters, these estimated parameters will be further analyzed using Markov chain Monte Carlo (MCMC) methods as initial values. For least square method, the idea is to minimize the sum of squares of residuals defined as:

$$l(\theta) = \sum_{i=1}^n (y_i - f(x_i, \theta))^2, \quad (2.2)$$

where y_i is the observed data, x_i is the model state, θ is the vector of parameter and $f(x_i, \theta)$ is the measurement model function at a given parameter value.

3.3. Markov chain Monte methods

Markov chain Monte Carlo methods are based on simulating high dimensionality and complex patterns of dependence in statistical model, where the proposal distribution is used which should result to the good mixing of chain. To determine a suitable proposal distribution for the required target distribution is a difficult task and time consuming as it involves trials and error. The most preferred proposal distribution is the Gaussian distribution, but we do not know how to obtain a suitable covariance matrix, therefore to overcome this challenge we use adaptive Markov chain Monte Carlo, where the proposal distribution is adapted during the MCMC run. Below is the adaptive MCMC algorithm developed by [7].

Algorithm . *Adaptive Markov chain Monte Carlo algorithm.*

- (i) Start with the initial value θ_0 and initial covariance Σ_0 and then select λ , ε , where ε is a small positive value whose role is to make sure that Σ_n is not singular matrix, and λ is a co-variance scaling factor which optimizes the mixing property of metropolis algorithm and an initial non-adapting period n_0 . For $n_0 = 0$ means the adaptation start as the algorithm start. If the target density is Gaussian then

$$\lambda = \frac{2.4^2}{m},$$

where m is the number of parameters.

- (ii) For $n = 1, 2, \dots, w$, where w is the total number of samples
- (a) Propose a new θ_* , from the Gaussian distribution $N(\theta_{n-1}, \Sigma_n)$
- (b) Accept/reject θ_* , according to the Markov chain Monte Carlo accepting probability
- (c) For $n \geq n_0$ adapt the proposal co-variance matrix using

$$\Sigma_n = \lambda(\text{cov}(\theta_0, \theta_1, \dots, \theta_{n-1}) + \varepsilon I_m),$$

where I_m is the $m \times m$ identity matrix.

4. Numerical analysis

Data were obtained from the daily cases records of all patients in hospital admissions at Ilala municipal council, between August to December, 2015. A total of 13 deaths were recorded

among 2010 infected individuals. Basing on our cholera model we divided the group of infected individuals into two groups i.e., people with mild symptoms (asymptomatic infected individuals) and people with severe symptoms (symptomatic infected individuals). Due to this criteria the total number of asymptomatic infected individuals are 1472 people and symptomatic infected individuals are 538 people. According to census report of 2012 the total population of ilala municipal council was 1,220,611 people [11]. Then using the population growth formula, the total population considered in this study is approximately 1,323,595 people in the year 2015. The data collected for asymptomatic infected individuals from Ilala municipal council are seen in Figure 1. From Figure 1, we see that, as the time goes $I_a(t)$ are increasing and later, decrease after a period of time, this is due to the immediate various measures taken by the Government of Tanzania to stop the spread of the disease. These measures were; providing good conditions relating to public health, especially the provision of clean drinking water, adequate sewage disposal and providing antibiotics to infected individuals e.g., using first line Doxycycline, second line Chloramphenicol for pregnant women and children, third line erythromycin.

The initial values Q_L used in this study are shown in Table 3.1 and the fitting of model parameters is carried using least square method. The estimated parameters Q_{LSQ} are shown in Table 3.1. However, the estimated parameters were further used in solving the ODE (2.1) and the solution of the cholera ordinary differential equation model is seen in Figure 2 by blue colour. From Figure 2, we see that the solution of ordinary differential equation cholera model fits true data.

The parameters were further analysed using Markov chain Monte Carlo method, we generated 10000 MCMC samples using the model parameters. The results are shown in Table 1 and in Figure 4 to Figure 8.

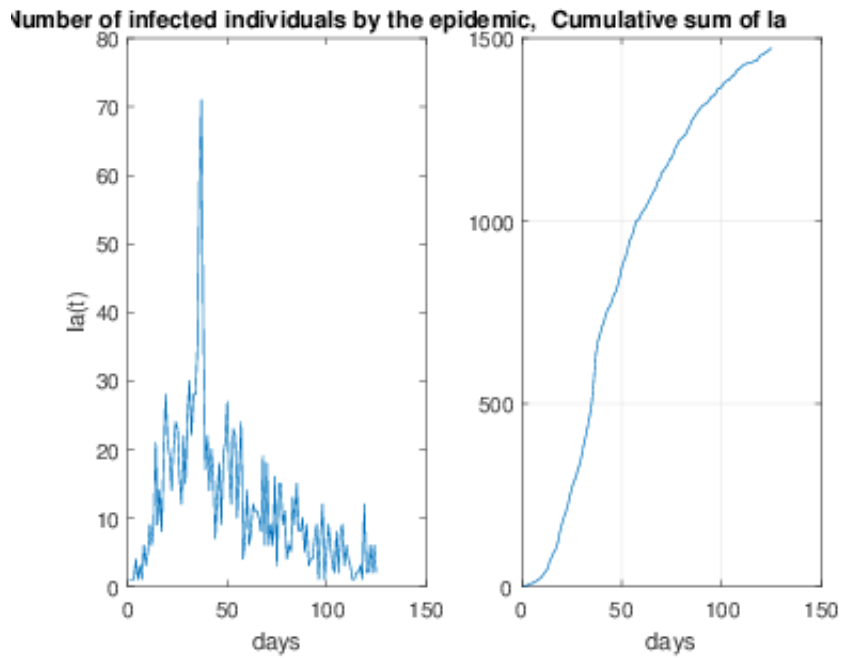


FIGURE 1. The infected individuals left and cumulative frequency of infected individuals right.

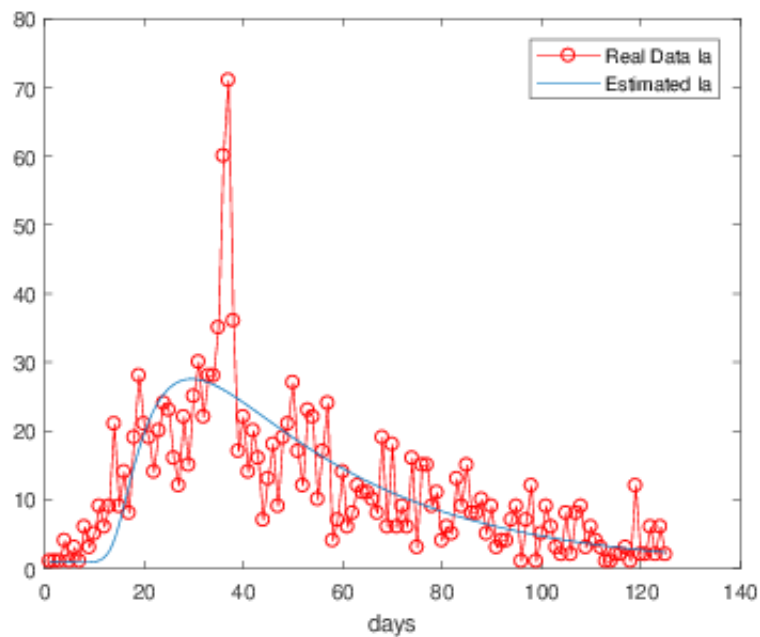


FIGURE 2. The dynamic behaviour of the solution of ODE cholera model vs real data for asymptomatic infected individuals.

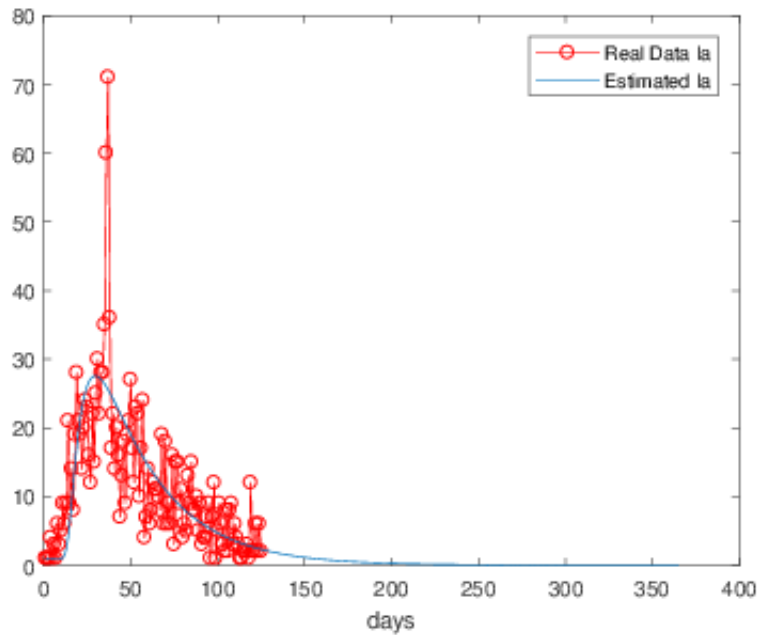


FIGURE 3. The dynamic behaviour of the solution of ordinary differential equation cholera model vs real data for asymptomatic infected individuals, when predicted annually.

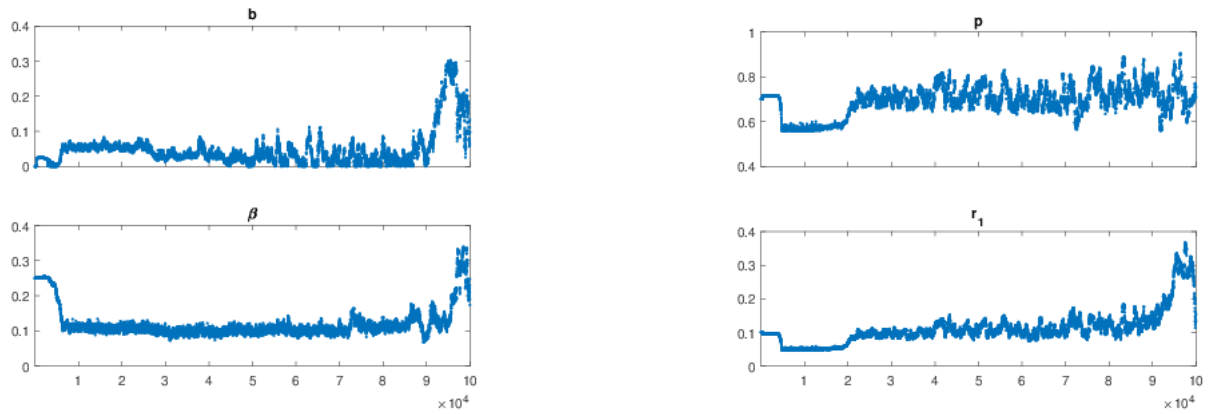
In Figure 3, we see that the model predicted well the observed data for cholera cases in Ilala municipal council. This implies that, if measures were not taken to stop cholera outbreak, the epidemic could continue and affects an inordinately large number of people within a very short period of time.

We test the convergence of the chain by examining the MCMC plots, parameter estimation by LSQ and prediction by both LSQ and MCMC estimates using cholera outbreak cases at Ilala municipal council in Tanzania.

TABLE 1. MCMC statistics for 10,000 numbers of simulations.

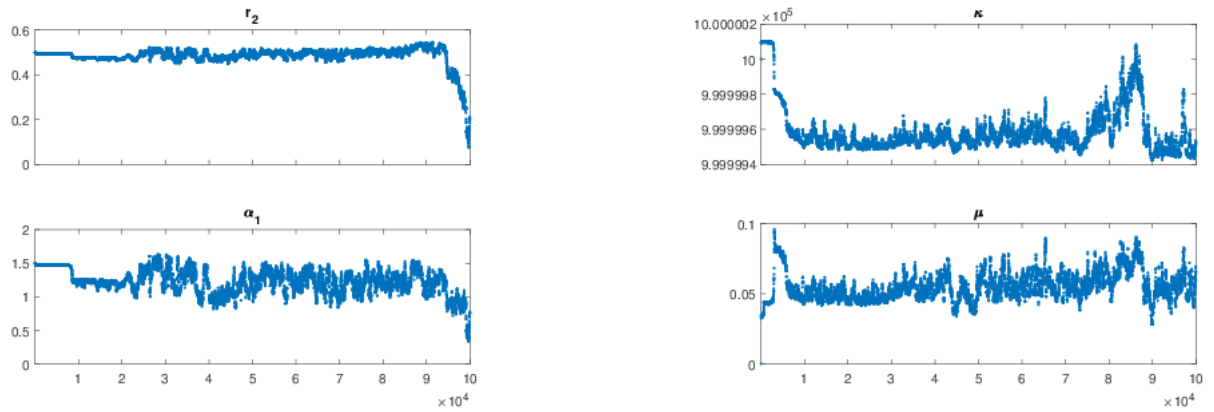
Symbol	θ_L	θ_{Lsq}	θ_{mean}	θ_{std}	Convergence	Kurtosis	Skewness
b	0.000072	0.000073947	0.002851	0.000645	0.2249	4.8134	1.7833
β	0.35	0.3516	0.33331	0.018586	0.86247	3.5072	-1.3376
κ	1000000	1051296.99	1000000	0.0738	1	3.0815	1.2372
μ	0.000044	0.00004075	0.1944	0.0756	0.0340	1.2549	-0.3656
p	0.7	0.6333	0.7139	0.0035	0.9809	1.3335	-0.0044
r_1	0.1	0.0859	0.1119	0.0021	0.9561	2.8123	1.1116
d	0.015	0.014945	0.014638	0.014228	0.8157	2.8917	1.2498
r_2	0.5	0.4886	0.5784	0.0135	0.9389	2.9136	-1.2692
α_1	1.5	1.4646	1.502	0.0583	0.9148	2.3783	-0.5904
α_2	0.1	0.0893	0.10603	0.0606	0.6107	2.5767	0.9368
δ	0.05	0.0471	0.0551	0.0213	0.55905	1.9206	-0.0846
ϕ	0.025	0.0238	0.0137	0.0268	0.0284	3.0178	1.1438
q	0.3	0.2987	0.2781	0.0183	0.7559	1.4354	-0.05

From Table 3.1, we see that the estimates of the parameters are indeed close to the true values. However, most of the parameters have the kurtosis values close to 3 that approximate to the kurtosis for a standard normal distribution.



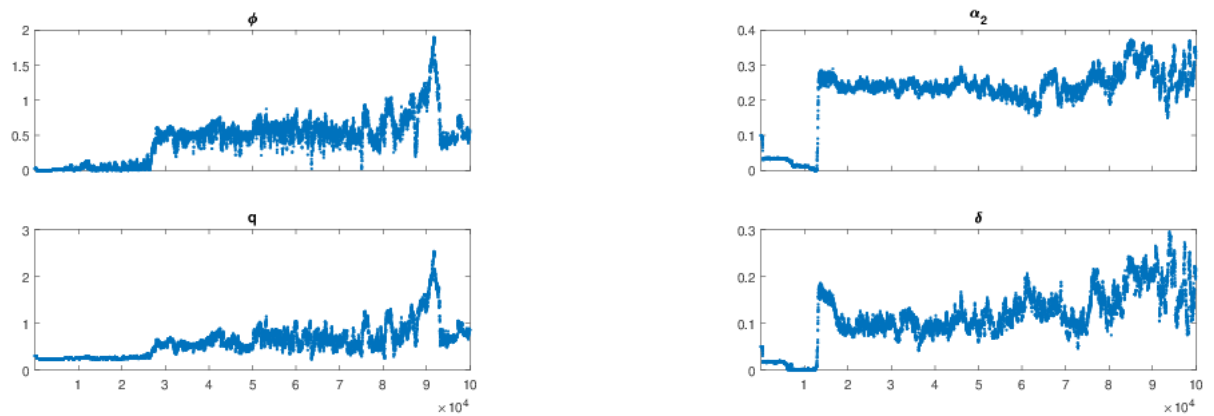
(A) Trace plots of estimated unknown parameters b and β using MCMC method. (B) Trace plots of estimated unknown parameter p and r_1 using MCMC method.

FIGURE 4. Trace plots of estimated unknown parameter b , β , p and r_1 using MCMC method.



(A) Trace plots of estimated parameters r_2 and α_1 using MCMC method. (B) Trace plots of estimated parameters κ and μ using MCMC method.

FIGURE 5. Trace plots of estimated parameters r_2 , α_1 , κ and μ using MCMC method.



(A) Trace plots of estimated parameters ϕ and q using MCMC method. (B) Trace plots of estimated parameters α_2 and δ using MCMC method.

FIGURE 6. Trace plots of estimated parameters α_2 , q , δ and ϕ using MCMC method.

From Figures 4 to 6, the results shows that the parameters have good mixing of chain.

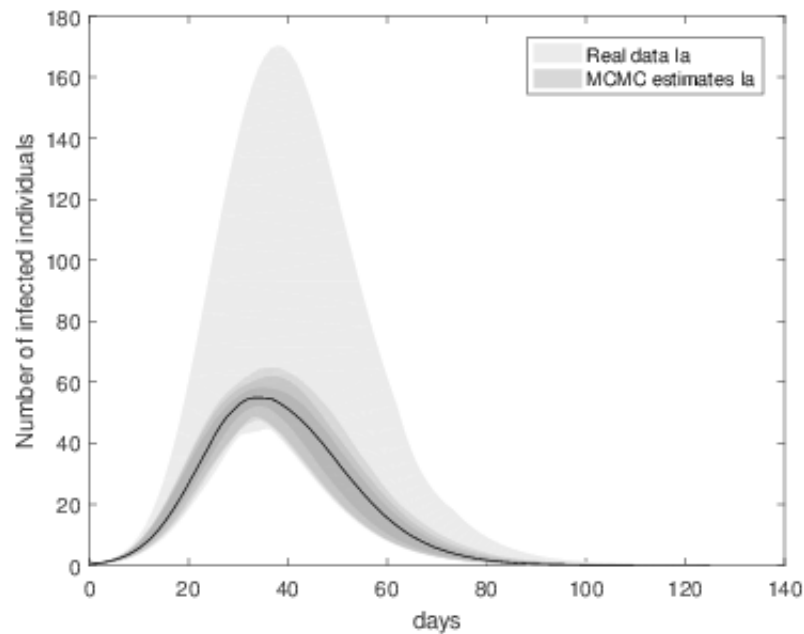


FIGURE 7. Prediction of infected individuals using MCMC estimates.

Also, we check the accuracy of the model through MCMC predictive plot. From Figure 7, the model predicted well the real data at 95 % posterior limits which is seen with the gray colour.

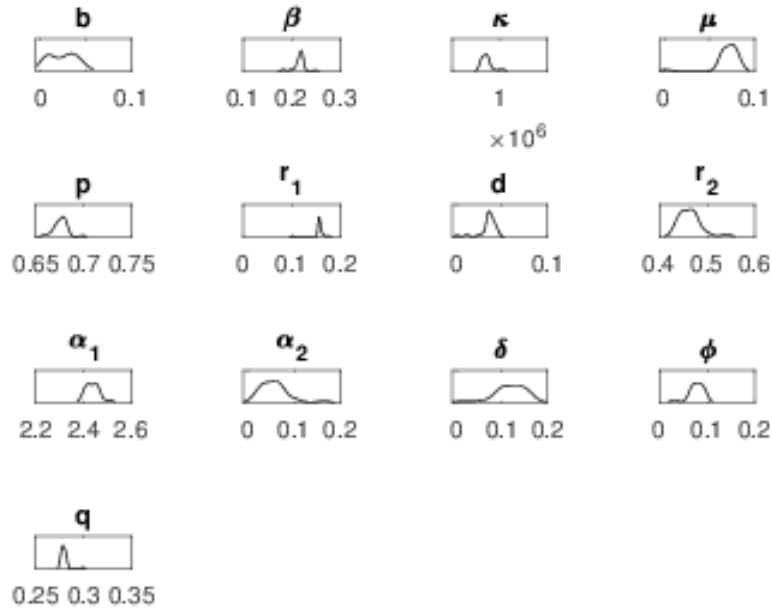


FIGURE 8. Posterior distribution of parameters with 10000 iteration.

Figure 8 represents the posterior distribution of 13 samples that have the distributions nearly to normal distribution.

4. Conclusion

The least square and MCMC methods are used to fit the proposed model in [12] using 2015 real data collected from Ilala municipal council in Tanzania. It was observed that both least square and MCMC methods performed well as the estimates are close to the true values. The parameters show good mixing of the chain with majority of posterior distributions of parameters taking the shape nearly to normal distribution. Similarly, it was observed that the dynamic behaviour of prediction fits the true data from Ilala municipal council.

Conflict of Interests

The authors declare that there is no conflict of interests.

Acknowledgment

The authors wish to thank Pan African University, AIMS-Tanzania and Tanzania Public Service College for their financial support, Ilala municipal council for data and the anonymous reviewers for the insightful comments that made this manuscript better.

REFERENCES

- [1] S. Edward, N. Nyerere, A mathematical model for the dynamics of cholera with control measures, *Appl. Comput. Math.*, 4 (2) (2015), 53–63.
- [2] W. Obeng-Denteh, E. A. Andam, L. Obiri-Apraku and W. Agyeil, Modeling cholera dynamics with a control strategy in Ghana, *Br. J. Res.*, 2 (1) (2015), 30–41.
- [3] W. J. Ivan, A. I. Scott, Y. K. Anna, M. S. Corinne, D. B. Anthony, M. Kevin, A. C. Thomas, P. Bindu, E. James and T. G. William, Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation, *Proc. Nat. Acad. Sci.*, 94 (1997), 5395-5400.
- [4] World Health Organization and others, Cholera vaccines: WHO position paper, *Weekly Epidemiological Record*, 85 (13) (2010), 117-128.
- [5] D. R. Hill, L. Ford and D. G. Lalloo, Oral cholera vaccines: use in clinical practice, *Lancet Infect. Diseases*, 6 (6) (2006), 361-373.
- [6] World Health Organization, World health statistics 2016: monitoring health for the SDGs sustainable development goals, *World Health Org.*, 2016
- [7] H. Haario, L. Laine, L. Marko, A. Mira and E. Saksman, DRAM: efficient adaptive MCMC, *Stat. Comput.*, 16 (4) (2006), 339-354.
- [8] J. G. Breman, The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Amer. J. Tropical Med. Hygiene*, 64 (1) (2001), 1-11.
- [9] D. D. Savigny, C. Mayombana, E. Mwangeni, H. Masanja, Care-seeking patterns for fatal malaria in Tanzania, *Malaria J.*, 3 (1) (2004), 27.
- [10] United Nations Millennium Project, Coming to grips with malaria in the new millennium Task force on HIV/AIDS, Malaria, TB, and Access to Essential Medicines, Working Group on Malaria, 2005.
- [11] National Bureau of Statistics, 2012 Population and Housing Census, *Tanzania National Website*, 2012.
- [12] Y. M. Marwa, S. Mwalili, I. S. Mbalawata, Markov chain Monte Carlo Analysis of Cholera Epidemic, *J. Math. Comput. Sci.*, 8 (5) (2018), 584-610.