International Journal of Data Science and Analysis
2026; 12(2): 17-24
http://www.sciencepublishinggroup.com/j/ijdsa
doi: 10.11648/j.ijdsa.20261202.11

Bayesian Geospatial Calibration of Reinforcement Learning for Malaria Transmission Control: Parameter Estimation

Kipngetich Gideon[1, *], Victor Muthama Musau[1], Margaret Wambui Kinyua[2]

[1]Pure and Applied Sciences, Kirinyaga University, Kerugoya, Kenya
[2]Mathematics, Statistics and Actuarial Sciences, Karatina University, Karatina, Kenya

Email address:

gideonlangat90@gmail.com (Kipngetich Gideon), vmusau@kyu.ac.ke (Victor Muthama Musau),
mwambui@karu.ac.ke (Margaret Wambui Kinyua)
*Corresponding author

To cite this article:

Kipngetich Gideon, Victor Muthama Musau, Margaret Wambui Kinyua. (2026). Bayesian Geospatial Calibration of Reinforcement Learning for Malaria Transmission Control: Parameter Estimation. *International Journal of Data Science and Analysis, 12*(2), 17-24. https://doi.org/10.11648/j.ijdsa.20261202.11

Received: 18 April 2026 ; **Accepted:** 3 May 2026 ; **Published:** 10 June 2026

Abstract: Malaria has been one of the major public health issues that has not been extensively addressed. Controlling the spread of infectious diseases in space and time requires robust adaptive policies that significantly account for heterogeneity, uncertainty, and optimal sequential decision-making. This study presents an innovative framework that integrates Bayesian spatiotemporal modeling with reinforcement learning (RL) with the 5D3 algorithm. The disease risk at location i and time t is modeled using a logistic regression with spatial random effects and Bayesian inference performed using the non-reversible Metropolis-Hastings algorithm, and the parameter estimates are used to calibrate a stochastic reinforcement learning environment via episodic parameter sampling. The study identified significant drivers of malaria risk: rainfall, temperature, secondary and tertiary levels of education, higher wealth index, female gender, treated nets, and spray repellents, while quantifying uncertainty via credible intervals. The spatial random effect captured unmeasured local heterogeneity, and the temporal effect accounted for seasonality, which is essential for reliable parameter estimation. Therefore, a reinforcement learning agent can learn optimal, spatially adaptive intervention policies under uncertainty, making the model suitable for public health decision-making where spatial heterogeneity and uncertainty are prominent. The proposed calibrated model within a policy-learning environment using posterior samples can be replicated to simulate realistic transmission scenarios for malaria and evaluate dynamic control strategies.

Keywords: Geospatial Calibration, Bayesian, Reinforcement Learning, Malaria, Transmission, Parameter Estimation

1. Introduction

Infectious diseases are human illnesses caused by bacteria, parasites, and viruses; these diseases include malaria, COVID-19, Chikungunya, Dengue, severe dengue, Yellow fever, Zika virus, Japanese encephalitis, Lymphatic filariasis, Leishmaniasis, Crimean-Congo hemorrhagic fever, Chagas disease (American Trypanosomiasis), Onchocerciasis, Trypanosomiasis, human African (sleeping sickness), Plague, and Schistosomiasis. These diseases continue to be a global health threat, as evidenced by the recent pandemic COVID-19 and the recurring Ebola outbreaks in West and Central Africa [1]. The burden of malaria, tuberculosis, and HIV has continued to persist despite significant interventions for control and prevention [2, 3]. Compartmental Susceptible-Infected-Recovered (SEIR models have been widely used to forecast outbreaks and evaluate intervention strategies. However, these models assume homogeneous mixing, fixed parameters, and deterministic dynamics [4, 5], thereby oversimplifying the complexities of real-world disease spread. Spatial heterogeneity is an important characteristic of disease spread. Characteristics such as population density, host and disease-carrying agent movement patterns, environmental factors such as temperature, rainfall, and healthcare access in different geographic regions contribute to a dynamic

disease environment. Public health interventions such as insecticide spraying and bed net distribution campaigns are implemented at the regional level. They must account for spatial spillovers, underscoring the importance of a model that incorporates spatial dependence. Another improvement in modelling the disease is the adoption of Bayesian methods to quantify parameter estimation uncertainty [6, 7]. This is important for selecting interventions without relying on point estimates, thereby allowing assessment of policy robustness. To enhance sequential decision-making under uncertainty, reinforcement learning is essential. Reinforcement learning agents learn optimal policies by taking actions, interacting with an environment, receiving rewards, and updating their policies [8]. Bayesian reinforcement learning, where the agent will maintain a belief over model parameters and updates it as new data is obtained, therefore addressing the challenge of the policy being overoptimistic due to fixed parameter assumptions.

Spatiotemporal disease mapping has a rich history in biostatistics with the BYM model combining spatially correlated and uncorrelated random effects [9]. Bayesian inference for spatiotemporal models has been performed via Markov chain Monte Carlo (MCMC) and integrated nested Laplace approximations [10]. Recent studies have demonstrated that the use of reinforcement learning in epidemiology for disease control has gained significant traction. Deep reinforcement learning determines the infectious disease (COVID-19) mitigation policies [11, 12]. Multi-agent deep reinforcement learning automatically learns the prevention strategies in the context of pandemic influenza [13]. In light of the adoption of reinforcement learning, studies have significantly assumed the common disease dynamics and simple compartmental models. Therefore, integration of Bayesian spatiotemporal models with reinforcement learning is rare; the study bridges this gap by using posterior samples from the non-reversible Metropolis-Hastings algorithm to calibrate a reinforcement learning dynamic environment. The proposed algorithm is an extension of the twin-delayed deep deterministic policy gradient (TD3) algorithm, which is applicable when states are continuous, and it mitigates overestimation bias by learning two critics [14]. The proposed algorithm uses five critics to stabilize further learning in a dynamic environment with high variance [15]. The study therefore makes the following contributions: (1) A geospatial calibration that transforms posterior samples into a stochastic environment for reinforcement learning, which enables the agent to experience a distribution of plausible disease dynamics, and (2) Estimation of parameters in an integrated five delayed deep deterministic policy gradient reinforcement learning algorithm adapted to the spatiotemporal logistic environment for reinforcement learning agent.

2. Methods

This section provide the development of proposed framework, model diagnostics and data description.

2.1. Spatiotemporal Malaria Model

The model is developed by letting Y_{it} as the number of disease cases observed in region i ($i = 1, \ldots, N$) at time t ($t = 1, \ldots, T$), n_{it} is the population that is at risk. Conditional on the true risk probability π_{it} given by

$$\pi_{it} = \left(1 + \exp\left(-(S_{it}\beta + u_i + \gamma_t + \epsilon_{it})\right)\right)^{-1} \tag{1}$$

$S_{it} = (S_{it1}, S_{it2}, \ldots, S_{itK})$ is a $1 \times K$ vector of covariates at location i and time t, which represent environmental and socio-economic factors, $\beta = (\beta_1, \ldots, \beta_K)^\top$ is a $K \times 1$ vector of regression coefficients. u_i is a spatial random effect attributed to region i that captures unmeasured spatial heterogeneity, inducing dependence among neighboring regions. γ_t is a temporal random effect for time t that accounts for global temporal trends, seasonality, and ϵ_{it} is an unstructured spatiotemporal random effect modeling independent noise.

The spatial random effect $u = (u_1, \ldots, u_N)^\top$ is modelled with intrinsic conditional autoregressive prior

$$u_i \mid u_{j \neq i} \sim \mathcal{N}\left(\frac{\sum_{j\sim i} w_{ij} u_j}{\sum_{j\sim i} w_{ij}}, \frac{\sigma_u^2}{\sum_{j\sim i} w_{ij}}\right) \tag{2}$$

where $j \sim i$ denotes that regions i and j are neighbors, and w_{ij} are spatial weights

$$w_{ij} = \begin{cases} 1 \text{ neighbors} \\ 0 \text{ Otherwise} \end{cases}$$

The joint distribution is therefore given by

$$p(u \mid \sigma_u^2) \propto \exp\left(-\frac{1}{2\sigma_u^2} \sum_{i\sim j} w_{ij}(u_i - u_j)^2\right) \tag{3}$$

Equation 3 is an improper with precision matrix $Q_u = \tau_u(D_w - W)$, rank $N - 1$, D_w is diagonal with $D_{ii} = \sum_j w_{ij}$. The model is made proper by imposing a sum-to-zero constraint $\sum_i u_i = 0$.

The temporal random effects γ_t is modelled as independent and identically distributed with a normal prior

$$\gamma_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2) \tag{4}$$

The unstructured spatiotemporal noise ϵ_{it} is assumed to be independent across both space and time $\epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, therefore accounting for measurement error not explained by the spatial and temporal components and overdispersion. The model is therefore represented by

$$\begin{aligned} &\pi_{it} = \left(1 + \exp\left(-(S_{it}\beta + u_i + \gamma_t + \epsilon_{it})\right)\right)^{-1} \text{ where} \\ &u \sim \text{ICAR}(\sigma_u^2), \gamma_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2), \epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2), \\ &\beta_k \sim \mathcal{N}(0, \tau_\beta^2), \sigma_u^2, \sigma_\gamma^2, \sigma_\epsilon^2 \sim \text{Inverse-Gamma}(a, b) \end{aligned} \tag{5}$$

2.2. Parameter Estimation

Model parameter estimations were performed using the Non-reversible Metropolis-Hastings algorithm, since the standard Metropolis-Hastings algorithm satisfies detailed balance, which can lead to random-walk behavior and slow mixing in high-dimensional, correlated spaces. The Non-reversible Metropolis-Hastings algorithm breaks detailed balance by introducing a "skew-symmetric" proposal distribution, resulting in a Markov chain that explores state space more efficiently [16].

Sampling from posterior $p(\Theta \mid Y)$ where $\Theta = (\beta, u, \gamma, \epsilon, \log\sigma_u^2, \log\sigma_\gamma^2, \log\sigma_\epsilon^2)$. The algorithm is given in the table below

Algorithm 1 Non-reversible Metropolis-Hastings

Require: Current state Θ, log-posterior gradient $\nabla L(\Theta)$, step size h, drift parameter δ, proposal covariance Σ.

1: Proposal step
2: Sample $\theta \sim \text{Uniform}\{+1, -1\}$ (direction).
3: Generate $\xi \sim \mathcal{N}(0, \Sigma)$.
4: Propose $\Theta' = \Theta + \theta \cdot \delta h + h^2 \nabla L(\Theta) + h\xi$.
5: Acceptance probability
6: Compute log-acceptance ratio

$$\log\alpha = \min\Big(0,\ L(\Theta') - L(\Theta) + \log q(\Theta \mid \Theta', \theta) - \log q(\Theta' \mid \Theta, \theta)\Big)$$

where

$$q(\Theta' \mid \Theta, \theta) = \mathcal{N}\left(\Theta' \mid \Theta + \theta\delta h + h^2\nabla L(\Theta), h^2\Sigma\right),$$

and the reverse proposal is

$$q(\Theta \mid \Theta', \theta) = \mathcal{N}\left(\Theta \mid \Theta' - \theta\delta h + h^2\nabla L(\Theta'), h^2\Sigma\right).$$

7: With probability $\min(1, \exp(\log\alpha))$, accept Θ'; otherwise stay at Θ.
8: Flip the direction for the next iteration, deterministic flip after each acceptance

δ-drift is a small constant vector chosen times the prior mode direction and h is a step size. The gradient of $L(\Theta) = \log p(\Theta \mid Y)$ is estimated as

$$\begin{aligned} L(\Theta) = &\sum_{i,t} \log \text{Bin}(Y_{it} \mid n_{it}, \pi_{it}) + \log p(\mathbf{u} \mid \sigma_u^2) + \\ &\sum_t \log\mathcal{N}(\gamma_t \mid 0, \sigma_\gamma^2) + \sum_{i,t} \log\mathcal{N}(\epsilon_{it} \mid 0, \sigma_\epsilon^2) \\ &+ \log p(\beta) + \log p(\sigma_u^2) + \log p(\sigma_\gamma^2) + \log p(\sigma_\epsilon^2) \end{aligned} \tag{6}$$

The diagnostic characteristics of the parameters are determined by running four independent Non-reversible Metropolis-Hastings chains, every 10000 iterations after a warm-up of 2000, and analyzing gelman-rubin with values $<$ 1.01 indicate convergence, effective sample size per iteration, and trace plots to assess mixing. Posterior sample thinning was also used to reduce autocorrelation.

2.3. Geospatial Calibration

Embedding the Bayesian posterior into the reinforcement learning, a stochastic calibration was applied, where at the beginning of each reinforcement learning episode, one set of parameters from the joint posterior is sampled

$$(\beta^{(m)}, u^{(m)}, \gamma^{(m)}, \sigma_\epsilon^{2(m)}) \sim p(\cdot \mid Y) \tag{7}$$

The estimated parameters are held constant throughout the reinforcement learning episode. The reinforcement learning agent will experience different plausible disease dynamics, therefore encouraging robustness.

2.4. Reinforcement Learning Integration

2.4.1. State Space

The state at time t capture all information needed for decision-making.

$$s_t = [\hat{\pi}_{1t}, \hat{\pi}_{2t}, \ldots, \hat{\pi}_{Nt}, S_{1t}, \ldots, S_{Nt}] \tag{8}$$

where $\hat{\pi}_{it}$ is the estimated risk probability/observed incidence malaria proportion. To address the dimensionality complexity infection counts Y_{it} are used

$$s_t = (Y_{1t}, Y_{2t}, \ldots, Y_{Nt}) \in \mathbb{R}^N \tag{9}$$

2.4.2. Action Space

The human agent select a vector of intervention intensities across all the regions

$$a_t = (a_t^{\text{spray}}, a_t^{\text{nets}}) \in [0, 1]^2 \tag{10}$$

The interventions affect the disease risk through reduction of linear predictor

$$\eta_{it}^{\text{interv}} = \mathbf{S}_{it}\beta + u_i + \gamma_t + \epsilon_{it} - \theta_{\text{spray}} a_t^{\text{spray}} - \theta_{\text{nets}} a_t^{\text{nets}} \tag{11}$$

$\theta > 0$ representing intervention efficacy. Therefore the risk probability is given by

$$\pi_{it} = \text{logit}^{-1}\left(\eta_{it}^{\text{interv}}\right) \tag{12}$$

2.4.3. Transition Function

For the state s_t, action a_t, and the sampled parameters $(\beta, u, \gamma, \sigma_\epsilon^2)$, next states are estimated as

1. For each region i linear predictor η_{it} is estimated as

$$\eta_{it} = \mathbf{S}_{it}\beta + u_i + \gamma_t + \epsilon_{it} - \theta^\top a_t. \tag{13}$$

$\epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ are drawn at each time step

2. Compute $\pi_{it} = (1 + e^{-\eta_{it}})^{-1}$

3. Draw the next number of infections

$$Y_{i,t+1} \sim \text{Binomial}(n_{i,t+1}, \pi_{it}) \tag{14}$$

The number of population at risk $n_{i,t+1}$ change due to interventions or natural demography, however in this study the assumption is that the population is constant

4. The covariates $S_{i,t+1}$ are updated according to exogenous factor seasonality and intervention effects. $S_{i,t+1}$ is assumed to be given by S_{it} and a_t.

The new state is therefore estimated as $s_{t+1} = (Y_{1,t+1}, \ldots, Y_{N,t+1})$

2.4.4. Reward Function

The reward for human agent is calculated as the negative total number of new infections plus a small penalty for high intervention levels reflecting the resource cost, this will encourage encourage the minimization

$$R_t = -\sum_{i=1}^{N} Y_{i,t+1} - \lambda \|a_t\|_2^2 \tag{15}$$

where $\lambda = 0.01$, discount factor is $\gamma_{\text{RL}} \in (0, 1)$

2.5. Reinforcement Learning Algorithm

The study proposed application of 5D3 algorithm which is an extension of TD3 [14], the algorithm learn five critic networks and use ensemling method to estimate Q-value [17]. The algorithm is described as follow

$$\begin{aligned} &\Delta\theta_i = \nabla\theta_i (y_t - Q_{\theta_i}(s_t, a_t))^2, \quad i = 1, \ldots, 5 \\ &\quad (\text{ Five critic networks } Q_{\theta_i}(s_t, a_t)) \\ &y_t = R_t + \gamma \max_{a' \in A} \min_{i \in N} Q_{\theta_i' \text{target}}(s_{t+1}, a_t') \\ &a_t^{'} = P_{\phi' \text{actor}}(s_{t+1}) + \text{clip}(\mathcal{N}(0, \sigma^2_{\text{smooth}}), -c, c) \end{aligned} \tag{16}$$

The soft updates on critic, actor network and target networks is given by

$$\begin{aligned} \mathcal{L}(\theta_k) =& \frac{1}{B} \sum_{i=1}^{B} \left(Q_{\theta_k}(s_i, a_i) - y_i\right)^2 \\ &(\text{ Adam optimizer with learning rate} 10^{-3}) \\ \Delta\phi =& \nabla_\phi Q_{\theta_i}\left(s_t, P_\phi\left(s_t\right)\right), \quad i = 1, \ldots, 5 \\ \theta_i' \leftarrow& \delta\theta_i + (1 - \delta)\theta_i' \\ \phi' \leftarrow& \delta\phi + (1 - \delta)\phi' \end{aligned} \tag{17}$$

Where δ is a small positive soft update attenuation coefficient ($\delta \leq 1$), the update ensures iterative target network updating, utilizing linear combination of network parameters ϕ, θ_i and target network parameters ϕ', θ_i'. The soft update ensures gradual and smooth changes of target network parameters, and the target values also change smoothly. Smaller value of soft update δ coefficient of attenuation leads more stable algorithm and slower variation target network, with more sluggish algorithm convergence rate.

2.6. Evaluation Metrics

This section provides an evaluation metric, the root mean squared error. This metric measures the accuracy of estimated model parameters, and the equation for the root mean squared error is given by

$$\text{RMSE}(\hat{\beta}) = \sqrt{\frac{1}{k} \sum_{i=1}^{k} \left(\text{E}[\beta_i \mid data] - \beta_i^{\text{true}}\right)^2} \tag{18}$$

Where k is the number of parameters, the measure is in the parameter's unit and a smaller Root Mean Squared Error (RMSE) indicates better parameter accuracy.

3. Results and Discussion

This section presents results of the study according to the study objective and discussion of the results.

3.1. Simulation

The simulation study was conducted to estimate the model's parameters. Parameter estimates, credible intervals and root mean square errors are presented in Table 1. The rainfall parameter -0.3993 indicates a negative association with malaria risk, where a unit increase in rainfall decreases the log-odds of malaria by 0.3993. CI $[-0.4078, -0.3752]$ does not contain zero, demonstrating the significance of the parameter and rainfall variable; the root mean square error 0.0121 implies a high precision of the parameter estimate. The temperature parameter estimate 0.6087 indicates a positive association with malaria risk, where a unit increase in temperature increases the log-odds of malaria by 0.6087; the CI $[0.5899, 0.6268]$ does not contain zero, demonstrating the significance of the parameter and temperature variable, and the root mean square error of 0.0127 implies a high precision of the parameter estimate. The age parameter 0.3010 indicate positive association with malaria risk where a unit increase in age increases the log-odds of malaria by 0.3010, indicating that older ages are associated with significantly higher risk of contracting the disease, the CI $[0.2819, 0.3213]$ does not contain zero demonstrating the significance of parameter and age variable, small root mean square error 0.0101 imply a high precision of the parameter estimate. Primary level education parameter -0.0761 indicate negative association with malaria risk, where a positive change in education (primary) level reduces the risk log-odds of malaria by 0.0761 compared to education (no education) level, the CI $[-0.330, 0.1384]$ contain zero demonstrating lack of parameter significance and education(primary) variable, small root mean square error 0.5808 imply a high precision of the parameter estimate. Education(secondary) parameter -0.4406 indicate negative association with malaria risk where a positive change in

education(secondary) level reduce the risk log-odds of malaria by 0.4406 compared to education(no education) level, the CI $[-0.6960, -0.2289]$ does not contain demonstrating the parameter significance of education(secondary) variable, root mean square error 0.5732 implies a higher precision with greater uncertainty. Education(tertiary) parameter -0.4227 indicate negative association with malaria risk where a positive change in education(tertiary) level reduce the risk log-odds of malaria by 0.4227 compared to education(no education) level, the CI $[-0.6763, -0.2092]$ does not contain demonstrating the parameter significance of education(tertiary) variable, root mean square error 0.5833 implies a higher precision with greater uncertainty. Wealth index (middle) level parameter 0.1688 indicate positive association with malaria risk where a positive change in wealth index (middle) level increase the risk log-odds of malaria by 0.1688 compared to wealth index(poorest) level, the CI $[-0.0205, 0.3668]$ contain zero demonstrating that the parameter is not significant as highlighted in Bernstein-von Mises theorem [18], root mean square error 0.1675 implies a higher precision with less uncertainty. Wealth index (richer) level parameter -0.2827 indicate negative association with malaria risk where a positive change in wealth index level (richer) decrease the risk log-odds of malaria by 0.2827 compared to wealth index(poorest) level, the CI $[-0.4897, -0.0822]$ does not contain zero demonstrating that the parameter is significant, root mean square error 0.1777 implies a higher precision with less uncertainty. Wealth index (richest) level parameter -0.3098 indicate negative association with malaria risk where a positive change in wealth index (richest) level decrease the risk log-odds of malaria by 0.3098 compared to wealth index(poorest) level, the CI $[-0.1205, -0.5057]$ does not contain zero demonstrating that the parameter is significant, root mean square error 0.4793 implies a higher precision with less uncertainty. Gender (female) level parameter -0.1841 indicate negatively association with malaria risk where a positive change in gender (female) level decrease the risk log-odds of malaria by 0.1841 compared to gender (male), this shows a lower risk for female, the CI $[-0.3438, -0.0547]$ does not contain zero demonstrating that the parameter is significant and hence the variable, root mean square error 0.1598 implying a higher precision with less uncertainty. The parameter estimates for treated mosquito nets and spray repellent have significant impacts on the odds of malaria, with reductions in the risk of malaria by log-odds of 0.1902 and 0.1841, respectively.

The trace plots in Figure 1 demonstrate strong evidence of convergence of parameters associated with temperature, rainfall, and age. Convergence was also observed for parameters associated with wealth index (middle), wealth index (richer), and wealth index (richest) after 15000 iterations, as shown by the MCMC chains' consistent fluctuations around the mean parameter value. The uniform spread and stable trajectories indicate that the chains have efficiently explored the posterior distributions and achieved stationarity. The overlapping traces suggest excellent chain mixing, thereby providing a reliable estimate of the parameter. Stability also enhances confidence in the accuracy of the posterior distribution and demonstrates that the MCMC process successfully captures the underlying relationships in the data.

Table 1. *Parameter Estimates, Root Mean Square Errors and Credible Intervals for the Estimated Parameters.*

Variable	Estimate	Credible Interval	RMSE
Rainfall	-0.3993	[-0.4078,-0.3752]	0.0121
Temperature	0.6087	[0.5899,0.6268]	0.0127
Age	0.3010	[0.2819,0.3213]	0.0101
Education$_{Primary}$	-0.0761	[-0.3300,0.1384]	0.5808
Education$_{Secondary}$	-0.4406	[-0.6960,-0.2289]	0.5732
Education$_{Tertiary}$	-0.4227	[-0.6763,-0.2092]	0.5833
Wealth Index$_{Middle}$	0.1688	[-0.0205,0.3668]	0.1675
Wealth Index$_{Richer}$	-0.2827	[-0.4897,-0.0822]	0.1777
Wealth Index$_{Richest}$	-0.3098	[-0.1205,-0.5057]	0.4793
Gender$_{Female}$	-0.1841	[-0.3438,-0.0547]	0.1598
Treated mosquito Net	-0.1902	[-0.2486,-0.0463]	0.1620
Spray repellent	-0.1841	[-0.3438,-0.0547]	0.1598

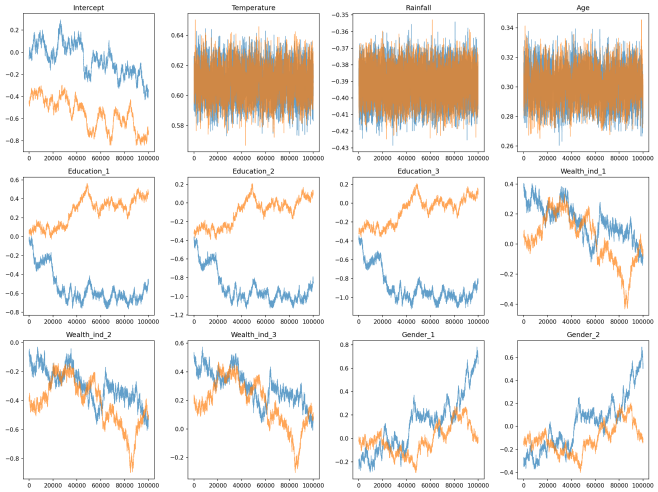

Figure 1. *Traceplots for Model parameters.*

3.2. Malaria Data

The section presents the estimation of model parameters using real-world malaria data; parameter estimates and their credible intervals were determined. The dataset represents malaria cases collected in Kenya for the 2015 and 2020 Malaria Indicator DHS surveys. The data was obtained from `https://dhsprogram.com/data/dataset/Kenya_MIS_2020.cfm?flag=1`.

The temperature parameter 0.1917 indicates a positive association with malaria risk, where a unit increase in temperature increases the log-odds of malaria by 0.1917; the CI $[0.0496, 0.9529]$ does not contain zero, demonstrating the significance of the parameter and temperature variable. The rainfall parameter -0.2311 indicates a negative association with malaria risk, where a unit increase in rainfall decreases

the log-odds of malaria by 0.2311. CI $[-0.4812, -0.0016]$ does not contain zero, demonstrating the significance of the parameter and the rainfall variable. The age parameter estimate 0.4666 indicates a positive association with malaria risk, where a unit increase in age increases the log-odds of malaria by 0.4666, indicating that older ages are associated with a significantly higher risk of the disease. CI $[0.0128, 0.9675, -0.0016]$ shows that the parameter is significant since it does not span over zero. Education(primary) parameter 0.1279 indicates a negative association with malaria risk, where a positive change in education (primary) level reduces the risk log-odds of malaria by 0.1279 compared to an individual with no education; the CI $[-1.0662, 0.8955]$ contains zero, demonstrating a lack of parameter significance. Education(secondary) parameter -0.1060 indicates a negative association with malaria risk, where a positive change in education(secondary) level reduces the risk log-odds of malaria by 0.1060 compared to an individual with no education; the CI $[-1.4397, -0.0168]$ does not contain, demonstrating the parameter significance of the education(secondary) variable. Education(tertiary) parameter -1.0108 indicates a negative association with malaria risk, where a positive change in tertiary education level reduces the risk log-odds of malaria by 1.0108 compared to an individual with no education; the CI $[-1.8976, -0.1329]$ does not contain, demonstrating the parameter significance of the education(tertiary) variable. Wealth index (middle) level parameter -0.0428 indicates a negative association with malaria risk, where a positive change in wealth index (middle) level decreases the risk log-odds of malaria by 0.0428 compared to wealth index(poorest), the CI $[-0.6107, 0.4438]$ contains zero, demonstrating that the parameter is not significant. Wealth index (richer) level parameter -0.5404 indicates a negative association with malaria risk, where a positive change in wealth index level (richer) decreases the risk log-odds of malaria by 0.5404 compared to wealth index(poorest); the CI $[-1.1169, -0.0042]$ does not contain zero, demonstrating that the parameter is significant. Wealth index (richest) level parameter -0.1607 indicates a negative association with malaria risk, where a positive change in wealth index (richest) level decreases the risk log-odds of malaria by 0.1607 compared to wealth index(poorest), the CI $[-0.6976, -0.1197]$ does not contain zero, demonstrating that the parameter is significant. Gender (female) level parameter -0.3824 indicates a negative association with malaria risk, where a positive change in gender (female) level decreases the risk log-odds of malaria by 0.3824 compared to gender (male). This shows a lower risk for females; the CI $[-1.3326, -0.1665]$ does not contain zero, demonstrating that the parameter is significant. The parameter estimates for treated mosquito nets and spray repellent indicate significant impacts on the odds of malaria, with reductions in the risk of malaria by log-odds of 0.5422 and 0.8839, respectively (Table 2).

Table 2. *Parameter Estimates and Credible Intervals for the Estimated parameters.*

Variable	Estimate	Credible Interval
Temperature	0.1917	[0.0496, 0.9529]
Rainfall	-0.2311	[-0.4812,-0.0016]
Age	0.4666	[0.0128,0.9675]
Education$_{Primary}$	0.1279	[-1.0662, 0.8955]
Education$_{Secondary}$	-0.1060	[-1.4397, -0.0168]
Education$_{Tertiary}$	-1.0108	[-1.8976, -0.1329]
Wealth index$_{Middle}$	-0.0428	[-0.6107,0.4438]
Wealth index$_{Richer}$	-0.5404	[-1.1169, -0.0042]
Wealth index$_{Richest}$	-0.1607	[-0.6976,-0.1197]
Gender$_{Female}$	-0.3824	[-1.3326,-0.1665]
Treated mosquito Net	-0.5422	[-1.3840, - 0.2450]
Spray repellent	-0.8839	[-2.4890, - 0.2570]

4. Discussion

The proposed model provides a robust approach for calibrating malaria transmission models. The simulation study demonstrates that the non-reversible Metropolis-Hastening algorithm successfully recovered the true parameter values with high precision, as evidenced by narrow credible intervals and small root mean square errors. The trace plots confirm convergence of parameters associated with temperature, rainfall, age, wealth index (middle), wealth index (richer), and wealth index (richest), indicating that the chains have efficiently explored the posterior distributions and achieved stationarity. This shows reliable inference given that the model includes spatially correlated random effects, which can lead to poor mixing.

Using Kenya's 2015 and 2020 Malaria Indicator Surveys reinforces several well-established epidemiological findings. Rainfall consistently exhibits a negative association with malaria risk, consistent with studies which demonstrated that excessive rainfall leads to flushing of mosquito breeding sites, therefore altering local hydrology [19]. Temperature shows a positive effect, which aligns with the temperature-dependent development of Plasmodium parasites and Anopheles vectors [20]. Education and wealth gradients reveal that only secondary or higher education and richer wealth quintiles contribute a significant protection, which is consistent with the findings that household assets and health literacy are threshold dependent [21]. Gender (female) indicates lower risk, highlighting a differential exposure and care?seeking behavior by females compared to males [22, 23]. The spatial random effect in the model captures unmeasured heterogeneity across locations; micro-climatic variation and local vector control effort that are critical for geospatial targeting of interventions, and the temporal effect accounts for seasonality and inter-annual trend. The credible intervals for rainfall, temperature, secondary and tertiary levels of education, richer and richest levels of wealth, and gender remain non-overlapping with zero, demonstrating their robustness. The convergence diagnostics validate that the Bayesian hierarchical model can be reliably used to inform

reinforcement learning policies. By providing posterior distributions of all parameters, including the spatial and temporal random effects, the calibration framework quantifies uncertainty in the transmission dynamics. Therefore, a reinforcement learning agent uses these distributions to learn optimal intervention allocation policies under uncertainty, adapting to the spatial heterogeneity and seasonal patterns.

5. Conclusion

The proposed model successfully identified significant drivers of malaria risk: rainfall, temperature, secondary and tertiary levels of education, higher wealth index, female gender, treated nets, and spray repellents, while quantifying uncertainty via credible intervals. The spatial random effect captured unmeasured local heterogeneity, and the temporal effect accounted for seasonality, which is essential for reliable parameter estimation. These posterior distributions provide a foundation for reinforcement learning; therefore, a reinforcement learning agent can now learn optimal, spatially adaptive intervention policies under uncertainty. The model is therefore suitable for public health decision-making where spatial heterogeneity and uncertainty are prominent. Future study can adopt the proposed calibrated model within a policy-learning environment using posterior samples to simulate realistic transmission scenarios for vector-borne infectious disease and evaluate dynamic control strategies.

ORCID

0000-0003-2874-4266 (Kipngetich Gideon)

Abbreviations

Cl	Credible Interval
RL	Reinforcement Learning
MCMC	Markov Chain Monte Carlo
5D3	5-Delayed Deep Deterministic Policy Gradient

Author Contributions

Kipngetich Gideon: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing-original draft, Writing-review & editing

Victor Muthama Musau: Supervision, Writing - review & editing

Margaret Wambui Kinyua: Supervision, Writing - review & editing

Acknowledgments

The author would like to thank the School of Pure and Applied Sciences at Kirinyaga University for providing conducive environment for this research. Victor Muthama Musau and Margaret Wambui Kinyua for their constructive suggestions and comments.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] P. Gholizadeh, M. Sanogo, A. Oumarou, M. N. Mohamed, Y. Cissoko, M. S. Sow, P. Pagliano, P. Akouda, S. Soufiane, A. A. Iknane, *et al.*, "Fighting covid-19 in the west africa after experiencing the ebola epidemic," *Health Promotion Perspectives*, vol. 11, no. 1, p. 5, 2021.

[2] W. Avis, "Malaria, hiv and tb in tanzania: Epidemiology, disease control challenges and interventions," 2022.

[3] L. Palombi and S. Moramarco, "Health in sub-saharan africa: Hiv, tb and malaria epidemiology," in *Multidisciplinary Teleconsultation in Developing Countries*, pp. 3–16, Springer, 2018.

[4] A. Yu, "Practical compartmental models for infectious disease dynamics in closed populations," *Intelligence Planet Journal of Mathematics and Its Applications*, vol. 2, no. 4, 2025.

[5] J. A. L. Marques, F. N. B. Gois, J. Xavier-Neto, and S. J. Fong, "Epidemiology compartmental models sir, seir, and seir with intervention," in *Predictive Models for Decision Support in the COVID-19 Crisis*, pp. 15–39, Springer, 2020.

[6] R. Van de Schoot, S. Depaoli, R. King, B. Kramer, K. Martens, M. G. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen, *et al.*, "Bayesian statistics and modelling," *Nature Reviews Methods Primers*, vol. 1, no. 1, p. 1, 2021.

[7] B. J. Reich and S. K. Ghosh, *Bayesian statistical methods*. Chapman and Hall/CRC, 2019.

[8] C. Szepesvari, *Algorithms for reinforcement learning*. Springer nature, 2022.

[9] M. Morris, K. Wheeler-Martin, D. Simpson, S. J. Mooney, A. Gelman, and C. DiMaggio, "Bayesian hierarchical spatial models: Implementing the besag york mollié model in stan," *Spatial and spatio-temporal epidemiology*, vol. 31, p. 100301, 2019.

[10] B. Schrodle and L. Held, "Spatio-temporal disease mapping using inla," *Environmetrics*, vol. 22, no. 6, pp. 725–734, 2011.

[11] M. L. Ozbilen, E. Eugriboz, R. Halepmollasi, I. Bilgen, and M. Haklidir, "Deep reinforcement learning for simulation-based determination of covid-19 pandemic mitigation policies," *Artificial Intelligence Theory and Applications*, vol. 1, no. 2, pp. 29–38, 2021.

[12] S. N. Khatami and C. Gopalappa, "Deep reinforcement learning framework for controlling infectious disease outbreaks in the context of multi-jurisdictions," *medRxiv*, pp. 2022–10, 2022.

[13] P. J. K. Libin, A. Moonens, T. Verstraeten, F. Perez-Sanjines, N. Hens, P. Lemey, and A. Nowe, "Deep reinforcement learning for large-scale epidemic control," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 155–170, Springer, 2020.

[14] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International conference on machine learning*, pp. 1587–1596, PMLR, 2018.

[15] Q. Lan, Y. Pan, A. Fyshe, and M. White, "Maxmin q-learning: Controlling the estimation bias of q-learning," *arXiv preprint arXiv: 2002.06487*, 2020.

[16] M. Vialaret and F. Maire, "On the convergence time of some non-reversible markov chain monte carlo methods," *Methodology and Computing in Applied Probability*, vol. 22, no. 3, pp. 1349–1387, 2020.

[17] Y. Song, P. N. Suganthan, W. Pedrycz, J. Ou, Y. He, Y. Chen, and Y. Wu, "Ensemble reinforcement learning: A survey," *Applied Soft Computing*, vol. 149, p. 110975, 2023.

[18] A. Katsevich, "Improved dimension dependence in the bernstein–von mises theorem via a new laplace approximation bound," *Information and Inference: A Journal of the IMA*, vol. 14, no. 3, p. iaaf020, 2025.

[19] 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.

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

[21] L. S. Tusting, B. Willey, H. Lucas, J. Thompson, H. T. Kafy, R. Smith, and S. W. Lindsay, "Socioeconomic development as an intervention against malaria: a systematic review and meta-analysis," *The Lancet*, vol. 382, no. 9896, pp. 963–972, 2013.

[22] M. Ngutu, D. O. Omia, T. O. Ngage, C. A. Oduor, N. O. Ouko, B. Oingo, I. Oluoch, S. Kariuki, J. Chikovore, W. Onyango-Ouma, *et al.*, "Gender-related factors affecting community malaria-related perceptions and practices in migori county, kenya," *Malaria journal*, vol. 24, no. 1, p. 196, 2025.

[23] G. M. Diiro, H. D. Affognon, B. W. Muriithi, S. K. Wanja, C. Mbogo, and C. Mutero, "The role of gender on malaria preventive behaviour among rural households in kenya," *Malaria Journal*, vol. 15, no. 1, p. 14, 2016.