STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **14**, November 2025, pp 2930–2956.
Published online in International Academic Press (www.IAPress.org)

Integrating Climate and Environmental Data with Bayesian Models for Malaria Prediction

Makwelantle Asnath Sehlabana [1,*], Daniel Maposa [1], Alexander Boateng [2], Sonali Das [3]

[1] *Department of Statistics and Operations Research, University of Limpopo, South Africa*
[2] *Department of Mathematics and Computer Science, Modern College of Business and Science, Sultanate of Oman*
[3] *Department of Business Management, University of Pretoria, South Africa*

Abstract Malaria remains a notable public health challenge in endemic regions, with an estimated 263 million cases and 579,000 malaria-related deaths globally in 2023. Climate and environmental factors, such as temperature, rainfall, and the Normalised Difference Vegetation Index (NDVI), play a crucial role in malaria transmission. While statistical models aid in malaria prediction, Bayesian methods remain underutilised despite their ability to incorporate prior knowledge into predictive models. The major contribution of this study is to develop a Bayesian malaria prediction model incorporating climate and environmental data. Both objective and subjective prior distributions are evaluated to determine their effectiveness in improving model performance. The results indicate that a subjective prior outperforms other priors. Additionally, Ehlanzeni (Mpumalanga), Vhembe and Mopani districts (Limpopo) are identified as high-risk malaria regions. The findings suggest that malaria transmission peaks in summer and autumn, particularly in areas where temperatures during the night range from 12°C-20°C, rainfall is moderate (100–200 mm), and NDVI exceeds 0.6. Malaria risk intensifies following months of accumulated rainfall, creating optimal mosquito breeding conditions. These insights may assist malaria control programmes in developing targeted interventions, such as early warning systems and vector management strategies. Future research will explore Bayesian machine learning for malaria prediction.

Keywords Bayesian Methods, Malaria Prediction, Prediction Model, Objective Prior, Subjective Prior, Climate Factors, Environmental Factors

AMS 2010 subject classifications 46N30, 62C10, 62F15

DOI: 10.19139/soic-2310-5070-2514

1. Introduction

Climate change is expected to increase the dangers and risks associated with malaria, particularly in vulnerable regions. Notably, rising temperatures are projected to have the most harmful effects on malaria prevalence in Africa during the period from 2021 to 2040 [1]. In 2023, the global number of malaria cases was estimated at 263 million, representing an increase of 11 million cases compared to 2022. During the same year, malaria-related deaths worldwide were estimated at 597 000, with the African region accounting for an overwhelming 94% of global cases and 95% of deaths [2]. Between September 2022 and August 2023, 5 813 malaria case notifications were recorded in South Africa through the Notifiable Medical Conditions Surveillance System. Of these, the endemic provinces of Limpopo, Mpumalanga, and KwaZulu-Natal accounted for 49%, 15%, and 7% of the cases respectively, while the non-endemic province of Gauteng contributed 19% [3]. Vulnerable populations, such as children under five years of age and pregnant women, are severely affected by malaria, which further deepens

*Correspondence to: Makwelantle Asnath Sehlabana (Email: asnath.sehlabana@ul.ac.za). Department of Statistics and Operations Research, University of Limpopo, Limpopo Province, Polokwane, South Africa(0727).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

socio-economic inequalities and hinders development efforts in endemic regions [4].

The Intergovernmental Panel on Climate Change estimates that by the 2030s, 51-62 million people in Eastern and Southern Africa will be at increased risk of malaria due to global warming [5]. Key climate and environmental factors influencing malaria transmission include temperature, which affects larval development, mosquito survival, parasite development, and biting rates, as well as rainfall, which creates breeding sites, supports the reproductive cycle, and sustains mosquito survival and habitat. Additionally, the Normalised Difference Vegetation Index (NDVI) shapes the nature of breeding sites, habitat types, and human-mosquito interactions [6, 7]. In addition, factors such as conflict and humanitarian crises, resource constraints, and biological challenges, including drug and insecticide resistance, continue to hinder global progress in the fight against malaria [8].

Malaria remains a crucial public health challenge in endemic countries, necessitating the implementation of diverse control interventions. Common strategies include insecticide-treated nets, advancements in genomic surveillance, rapid diagnostic tests, innovative treatment options, and vector control methods [4]. Among these interventions, the RTS,S/AS01 malaria vaccine, endorsed by the World Health Organization (WHO), has shown promise. Pilot studies in Ghana, Kenya, and Malawi, involving over two million children, demonstrated a 13% reduction in child mortality due to its use [9]. Despite these achievements, challenges persist in scaling vaccine distribution due to limited production capacity. While malaria vaccines offer varying degrees of protection and minimal severe side effects, public perception of the vaccine in many African communities is high. However, awareness and understanding of the vaccine remain relatively low [10]. This underscores the importance of exploring every possible solution to mitigate malaria transmission and associated mortality.

Research emphasised the critical role of effective malaria outbreak early warning systems in predicting incidence before outbreaks occur[11]. Additionally, studies highlighted the pressing need to address the impacts of climate change on malaria pathogenesis. These studies suggests that predictive modelling and artificial intelligence applications can play a pivotal role in addressing these challenges [12]. Malaria is one of the major vector-borne diseases, highly sensitive to climate and environmental factors [13]. Recent advancements in malaria predictive modelling have increasingly incorporated climate data. For instance, Kim et al. [14] and Landman et al. [15] developed malaria prediction models using time-series forecasting techniques, integrating climate variables. Similarly, studies by Adamu et al. [16], Lee et al. [17], and Nkiruka et al. [18] employed climate data within Machine Learning algorithms to predict malaria trends. However, Lu et al. [19] conducted a review of ten studies focused on developing prediction models for malaria re-introduction. Their findings revealed that while diverse approaches, including mathematical, Machine Learning, Delphi, and statistical methods, were utilised, most models were designed for European regions. Additionally, these prediction models exhibited a high risk of bias due to poor reporting, suboptimal methodological rigor, and a lack of proper validation. These findings highlight the urgent need for the development and thorough evaluation of robust malaria prediction models, particularly tailored to endemic regions.

Bayesian methods offer distinct advantages in statistical modelling, particularly through the incorporation of prior knowledge via both objective and subjective prior distributions. Despite these strengths, they remain underutilised in scientific research [20]. As one study alone often cannot solve the entire problem, existing research tends to address only specific components of the broader issue. For instance, Ibeji et al. [21] integrated only socio-economic variables in their malaria prediction model, while Semakula [22] relied solely on demographic and health surveys data. In contrast, the study by Rotejanaprasert [23] demonstrated progress by incorporating both malaria vector species and climate variables. Nonetheless, their analysis was limited to provincial-level data, potentially obscuring critical local patterns of malaria transmission. Motivated by these limitations, this study makes a major contribution by developing a Bayesian predictive model that integrates climate and environmental variables, specifically tailored to malaria-endemic regions in South Africa. Through the incorporation of prior knowledge, Bayesian frameworks can yield more informed and robust predictions. Furthermore, this research will evaluate and compare the performance of objective and subjective prior distributions within the Bayesian

framework to assess their respective predictive capabilities.

A scoping review by Nkiruka et al. [18] on malaria prediction models using climate variability and machine learning highlighted a range of methodologies, such as Seasonal Autoregressive Integrated Moving Average (SARIMA), Autoregressive Integrated Moving Average (ARIMA), binomial regression, the Vector-Borne Disease Community Model of Epidemics, sparse learning with elastic net regularisation, the Scale Interaction Experiment–Frontier Version 2 (SINTEX-F2), support vector machines (SVM), artificial neural networks (ANN), and ensemble learning techniques. However, the absence of Bayesian models in the review further highlights their continued underutilisation in malaria prediction research. By addressing this gap, the present study contributes to the limited body of literature on Bayesian methods in malaria prediction. The findings are expected to support enhanced regional malaria surveillance and provide evidence-based insights for policy development and intervention strategies.

2. Methodology

2.1. Study Area and Data Collection

This study focuses on malaria cases in three endemic South African provinces: KwaZulu-Natal, Limpopo, and Mpumalanga. Each province offers unique climatic, geographical, and demographic characteristics that influence malaria transmission dynamics. KwaZulu-Natal has three malaria-endemic districts namely, King Cetshwayo, uMkhanyakude, and Zululand. These districts are susceptible to malaria transmissions, owing to their favourable conditions for the malaria carrying mosquitoes, which thrive in regions with moderate rainfall and temperatures conducive to the mosquito lifecycle. Limpopo, known for its predominantly warm climate, features two malaria-endemic districts: Mopani and Vhembe. Mpumalanga contains one malaria-endemic district, Ehlanzeni, which is distinguished by its warm temperatures and plateau grasslands. These ecological factors contribute to suitable conditions for malaria transmission, particularly during rainy seasons when mosquito breeding sites are abundant.

For this study, secondary data sources have been utilised. Malaria case data is provided by the South African National Department of Health, comprising confirmed cases of malaria, which include both locally transmitted and imported instances. Climate data, obtained from the South African Weather Service (SAWS), includes daily rainfall, maximum, and minimum temperature readings. These data are aggregated to generate monthly averages, enabling a detailed analysis of the relationship between climatic factors and malaria incidence. In addition to climate data, the study incorporates the Normalised Difference Vegetation Index (NDVI), a satellite-derived measure that reflects the health and density of vegetation, which influences the abundance and distribution of malaria vectors. NDVI data is accessed through EarthExplorer (EE), an online platform hosted by the United States Geological Survey (USGS). The analysis is based on data spanning from 2018 to 2022, providing a robust temporal dataset to explore the interactions between environmental factors and malaria incidence in the selected provinces.

2.2. Model Specifications

2.2.1. Prior Distributions: The choice of prior distributions is crucial in Bayesian statistics. Different rules suggested for choosing prior distributions are objective (non-informative or weakly informative) and subjective (informative) [24]. This study tests and compares objective and subjective prior distributions with the goal of developing malaria prediction models.

i. Objective prior distributions

Objective prior distributions are derived from the sample data embedded in the likelihood function [25]. These priors, often referred to as weakly informative or non-informative priors, are used when prior knowledge is either limited or ambiguous, making it difficult to specify a fully subjective prior distribution.

In this study, we examine both non-informative and weakly informative priors. For the non-informative case, we test Jeffreys prior due to its strong theoretical foundation in information theory and, most importantly, its invariance to parameterisation. For the weakly informative case, we consider the conjugate prior corresponding to the sample data distribution. The dataset used in this study is best described by a negative binomial model, as demonstrated in the study by Sehlabana et al. (Inpress). Jeffreys prior for this model is defined as follows:

$$\pi(\theta) \propto |I(\theta)|^{1/2},$$

where $|\cdot|$ represents the determinant and $I(\theta)$ denotes the expected Fisher information matrix. The dataset used in this study is best described by a negative binomial model, with Jeffreys prior approximated by $beta(a, b)$, where $a = \epsilon$ and $b = 0.5$ as $\epsilon \to 0$, as demonstrated in the study by Sehlabana et al. (Inpress). Conjugate priors are widely used in Bayesian analysis, particularly for exponential family distributions such as the normal, Poisson, gamma, binomial, and negative binomial models. In the case of the negative binomial distribution, the conjugate prior for the success probability P follows a beta distribution, $P \sim beta(\alpha, \beta)$. The parameters α and β determine the shape of the beta distribution, with the mean and variance given by:

$$E[P] = \frac{\alpha}{\alpha + \beta},$$

and

$$\text{Var}(P) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

The total prior strength, S, measures how informative the prior is. It is derived from the mean μ_P and variance σ_P^2 of the beta distribution and is given by:

$$S = \frac{\mu_P(1 - \mu_P)}{\sigma_P^2} - 1.$$

From the properties of the Beta distribution, we find that:

$$\alpha + \beta = S \implies \alpha = S\mu_P, \quad \text{and} \quad \beta = S(1 - \mu_P).$$

Thus, the Beta parameters α and β are determined by μ_P and S, based on the dataset.

ii. Informative prior distributions

Subjective prior distributions are based on experts' beliefs or knowledge, which are quantified into probabilities and probability distributions through an elicitation process. For the dataset used in this study, Sehlabana et al. (In progress) elicited prior distributions, and statistical assessments revealed that beta, gamma, and normal distributions best captured expert perspectives on factors affecting malaria transmission. These prior distributions are tested in this study, and their specifications are provided in Table 1.

Table 1. Specifications of elicited prior distributions

Probability Distribution	**Estimated Parameters**
Beta	(27.948; 83.957)
Gamma	(37.585; 150)
Normal	(0.250; 0.041)

2.2.2. Posterior distributions: The posterior distribution in a Bayesian framework is derived using Bayes' theorem. It is defined as:

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)},$$

where:

$P(\theta|Y)$ is the posterior distribution, representing the updated belief about the parameter θ given the observed data Y,
$P(Y|\theta)$ is the likelihood function, representing the probability distribution of the observed data given θ,
$P(\theta)$ is the prior distribution, reflecting the initial belief about θ before observing the data, and
$P(Y)$ is the marginal likelihood, which serves as a normalising constant, given by:

$$P(Y) = \int P(Y|\theta)P(\theta)\, d\theta.$$

Since $P(Y)$ is independent of θ, the posterior is often written as:

$$P(\theta|Y) \propto P(Y|\theta)P(\theta).$$

Here, "$\propto$" means "proportional to," indicating that the posterior is obtained by multiplying the likelihood and prior, followed by normalisation. Consider that malaria counts Y follow a negative binomial distribution. The probability mass function of Y is given by:

$$P(Y|r, P) = \frac{\Gamma(Y+r)}{\Gamma(r)\Gamma(Y+1)} P^r(1-P)^Y,$$

and its likelihood function is given by:

$$L(r, P|Y) = \prod_{i=1}^{n} \frac{\Gamma(Y_i+r)}{\Gamma(r)\Gamma(Y_i+1)} P^r(1-P)^{Y_i}. \tag{1}$$

where $P_i = \frac{e^{X_i\theta}}{1+e^{X_i\theta}}$. Consider that an informative normal prior specified in Table 1 is used in the model. That is, the prior $\theta \sim N(0.25, 0.001681)$. Then, the probability density function of the normal prior is given by:

$$P(\theta) = \frac{1}{\sqrt{2\pi \times 0.001681}} \exp\left(\frac{-(\theta-0.25)^2}{2 \times 0.001681}\right). \tag{2}$$

Combining the likelihood function in equation (1) and the prior distribution in equation (2), the posterior is given by:

$$P(\theta|Y) = \prod_{i=1}^{n} \left[\frac{\Gamma(Y_i+r)}{\Gamma(r)\Gamma(Y_i+1)} P_i^r(1-P_i)^{Y_i} \times \frac{1}{\sqrt{2\pi \times 0.001681}} \exp\left(\frac{-(\theta-0.25)^2}{2 \times 0.001681}\right)\right]. \tag{3}$$

The log posterior is given by:

$$\log P(\theta|Y) = \sum_{i=1}^{n} [\log\Gamma(Y_i+r) - \log\Gamma(r) - \log\Gamma(Y_i+1) + r\log P_i + Y_i\log(1-P_i)] - \frac{(\theta-0.25)^2}{2 \times 0.001681} + C,$$

where C is a constant that does not depend on θ. The posterior does not have a closed-form solution and is approximated in this study using the Markov Chain Monte Carlo (MCMC) method.

2.2.3. Posterior predictive distribution: The posterior predictive distribution represents the distribution of future malaria case counts, denoted by Y^*, given the observed malaria case counts Y. It accounts for both parameter uncertainty (through the posterior distribution) and the stochastic nature of new observations (through the likelihood). Taking into account the likelihood function of Y in equation (1) and the posterior in equation (3), the posterior predictive distribution is given by:

$$P(Y^*|Y) = \int P(Y^*|\theta)P(\theta|Y)d\theta.$$

Considering that $Y^*|\theta \sim \text{NegBin}(r, P^*)$, where P^* is computed using predictor X^*, then:

$$P(Y^*|Y) = \int \prod_{i=1}^{n} \left[\frac{\Gamma(Y_i^* + r)}{\Gamma(r)\Gamma(Y_i^* + 1)} P_i^r (1 - P_i)^{Y_i^*} \right] \times \prod_{i=1}^{n} \left[\frac{\Gamma(Y_i + r)}{\Gamma(r)\Gamma(Y_i + 1)} P_i^r (1 - P_i)^{Y_i} \times \frac{1}{\sqrt{2\pi \times 0.001681}} \right] \times \exp\left(\frac{-(\theta - 0.25)^2}{2 \times 0.001681} \right) d\theta. \quad (4)$$

Since the integral in equation (4) does not have a closed-form solution, sampling methods like Monte Carlo integration are used to approximate it.

2.2.4. Model validation: In this study we use the Leave-One-Out (LOO) cross-validation method to estimate the predictive accuracy of the fitted Bayesian model. The key metrics of LOO used for model validation in this study include:

i. The expected log pointwise predictive density (elpd)
A primary measure of model predictive accuracy is the expected log pointwise predictive density (elpd), defined as:

$$\text{elpd} = \sum_{i=1}^{n} \int \log P(Y_i^*|y) P_t(Y_i^*) dY_i^*,$$

where $P_t(Y_i^*)$ is the true data-generating process. Since $P_t(Y_i^*)$ is unknown, we approximate elpd using leave-one-out cross-validation (LOO-CV), where elpd is expressed as:

$$\text{elpd}_{\text{loo}} = \sum_{i=1}^{n} \log P(Y_i|Y_{(-i)}),$$

where $P(Y_i|Y_{(-i)})$ is the predictive density computed by leaving out the i^{th} observation, given by:

$$P(Y_i|Y_{(-i)}) = \int P(Y_i|\theta) P(\theta|Y_{(-i)}) d\theta.$$

The predictive density $P(Y_i|Y_{(-i)})$ is estimated using posterior draws from the full model.

ii. The effective number of parameters (P_{loo})
The effective number of parameters, denoted by P_{loo}, quantifies the model complexity. It reflects how much flexibility the model has in predicting new data. The P_{loo} is given by:

$$P_{\text{loo}} = \sum_{i=1}^{n} \left[\log P(Y_i|Y) - \log P(Y_i|Y_{(-i)}) \right],$$

where $\log P(Y_i|Y)$ is the log-likelihood of the observed data, and $\log P(Y_i|Y_{(-i)})$ is the log-likelihood for the i^{th} observation computed using the model without the i^{th} observation. Larger values of P_{loo} indicate more complex models with higher flexibility.

iii. The LOO information criterion (Looic)
Looic is a model selection criterion based on LOO. It is calculated by adding a penalty term for the effective number of parameters to the LOO estimate of predictive accuracy. Looic is given by:

$$\text{Looic} = -2 \times \text{elpd}_{\text{loo}} + 2 \times P_{\text{loo}}.$$

Looic helps in comparing the models by balancing the fit of the model to the data with the complexity of the model. Lower Looic values are preferred, indicating better predictive accuracy relative to their complexity.

In addition to the three LOO metrics, we compute the difference in expected log predictive density ($\text{elpd}_{\text{diff}}$) between each model and a reference model. Given M models ($M_1, M_2, \ldots, M_M$), we select a reference model M_{ref}, typically the one with the higher elpd_{loo} value. For each model M_j, we compute $\text{elpd}_{\text{diff}}$ given by:

$$\text{elpd}_{\text{diff}} = \text{elpd}_{(100,j)} - \text{elpd}_{(100,\text{ref})},$$

where $\text{elpd}_{(100,j)}$ is the LOO elpd for model j, and $\text{elpd}_{(100,\text{ref})}$ is the LOO elpd of the reference model. A positive $\text{elpd}_{\text{diff}}$ value indicates that the model has better predictive accuracy than the reference model, while a negative $\text{elpd}_{\text{diff}}$ value indicates that the reference model has better predictive accuracy.

3. Results and Discussion

This section presents and discusses the results of the study, derived from the data exploration and statistical modelling techniques outlined in Section 2. Table 2 summarises the variables included in the malaria count models, providing their descriptions and the corresponding coding in the dataset.

Table 2. Description and Coding of Variables .

Variable	**Description**	**Data Set Code**	**Data Type**
District	Districts in the three Provinces of interest, namely: Amajuba, Capricorn, Ehlanzeni district, Gert Semenya, Harry Gwala, iLembe, King Cetshwayo, Mopani, Nkangala, Sekhukhune, Ugu, uMgungundlovu, uMkhanyakude, uMziny-athi, uThukela,Vhembe, Waterberg, and Zulu-land.	District with levels: Amaj, Capr, Ehlan, Gert, Harry G, iLembe, KingC, Mop, Nkang, Sekh, Ugu, uMgun, Umkha, Umzin, Uthu, Vhem, Water, and Zulu respectively	Character
Malaria counts	Malaria cases reported.	Mal	Numeric
Maximum temperatures	This variable represents the average tempera-tures during the day (in $^\circ$C).	MaxT	Numeric
Minimum temperatures	This variable captures the average temperatures during the night (in $^\circ$C).	MinT	Numeric
Month	Months of the year, namely:January, February, March, April, May, June, July, August, Septem-ber, October, November, and December.	Month, with levels: Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec	Character
Normalised differenced vegetation index (NDVI)	It is a widely used metric for quantifying health and density of vegetation using sensor data.	NDVI	Numeric
Province	The three malaria endemic provinces of South Africa, namely: KwaZulu-Natal, Limpopo, and Mpumalanga.	Province, with lev-els: KZN, LP, MP	Character
Rainfall	This variable measures the amount of precipita-tion in each period (in mm).	Rain	Numeric
Year	The period in which the data was collected, in years from 2028 to 2022.	Year, levels: 2018, 2019, 2020, 2021, 2022	Character

Table 3. Summary Statistics of Predictors Across Provinces.

Province	Predictor	Summary Statistics					
		Mean	SD	Min	Max	Median	N
KZN	Malaria Cases	9.99	32.15	0	339	1.0	580
	Max Temperature (°C)	26.01	3.14	0.00	37.60	26.10	449
	Min Temperature (°C)	12.94	5.13	-0.20	22.30	13.90	448
	Rainfall (mm)	67.52	64.69	0	465.40	56.95	460
	NDVI	0.59	0.12	0.28	0.79	0.62	580
LP	Malaria Cases	101.26	238.62	0	3005	19.5	290
	Max Temperature (°C)	27.98	3.10	20.10	34.13	28.19	232
	Min Temperature (°C)	13.85	4.49	3.30	20.90	14.46	232
	Rainfall (mm)	50.79	75.32	0	570.50	21.83	290
	NDVI	0.44	0.12	0.24	0.72	0.45	290
MP	Malaria Cases	148.03	301.61	0	2451	9.5	174
	Max Temperature (°C)	24.22	3.23	16.65	30.76	24.65	174
	Min Temperature (°C)	10.37	4.67	0.05	18.98	10.82	174
	Rainfall (mm)	55.22	56.66	0	262.24	38.52	174
	NDVI	0.49	0.14	0.28	0.73	0.48	174

Table 3 presents a summary of key predictors across three South African provinces: KwaZulu-Natal (KZN), Limpopo (LP), and Mpumalanga (MP). The mean malaria cases (Mal_mean) vary significantly across provinces, with KZN reporting the lowest average at approximately 9.99 cases, while LP and MP exhibit much higher means at 101.26 and 148.03, respectively. The standard deviation of malaria cases is also highest in LP (238.62) and MP (301.61), indicating greater variability in malaria incidence compared to KZN (32.15). Minimum malaria cases recorded in all provinces are zero, while maximum cases are highest in LP (3005) and MP (2451), with KZN reaching only 339 cases.

Regarding temperature variables, the mean maximum temperature (MaxTemp_mean) is highest in LP (27.98°C), followed by KZN (26.01°C) and MP (24.22°C). However, the range of maximum temperatures varies, with MP having the lowest recorded MaxTemp_min (16.65°C) and LP the highest (20.10°C). The minimum temperature (MinTemp_mean) follows a similar pattern, with LP having the warmest nights (13.85°C) and MP the coolest (10.37°C). The temperature variability (MaxTemp_sd and MinTemp_sd) is fairly consistent across provinces, with standard deviations ranging between 3°C and 5°C.

Rainfall patterns show that mean rainfall (Rain_mean) is highest in KZN (67.52 mm), followed by MP (55.22 mm) and LP (50.79 mm). The standard deviation of rainfall is highest in LP (75.32 mm), suggesting more fluctuations in rainfall compared to MP (56.66 mm) and KZN (64.69 mm). The maximum recorded rainfall (Rain_max) is also highest in LP (570.5 mm), while MP has the lowest maximum rainfall (262.24 mm).

The Normalised Difference Vegetation Index (NDVI), shows that KZN has the highest mean NDVI (0.59), followed by MP (0.49) and LP (0.44). This suggests that KZN has denser or healthier vegetation compared to the other two provinces. The NDVI variability (NDVI_sd) is slightly higher in MP (0.14) compared to LP (0.12) and KZN (0.12), indicating that vegetation cover is more heterogeneous in MP.

Overall, the data suggest significant differences in malaria incidence, temperature, rainfall, and vegetation among the three provinces. Malaria cases are more frequent and variable in LP and MP than in KZN, which also has the highest vegetation index and the most rainfall. Temperature differences may contribute to variations in malaria incidence, with LP generally being the warmest province and MP the coolest.

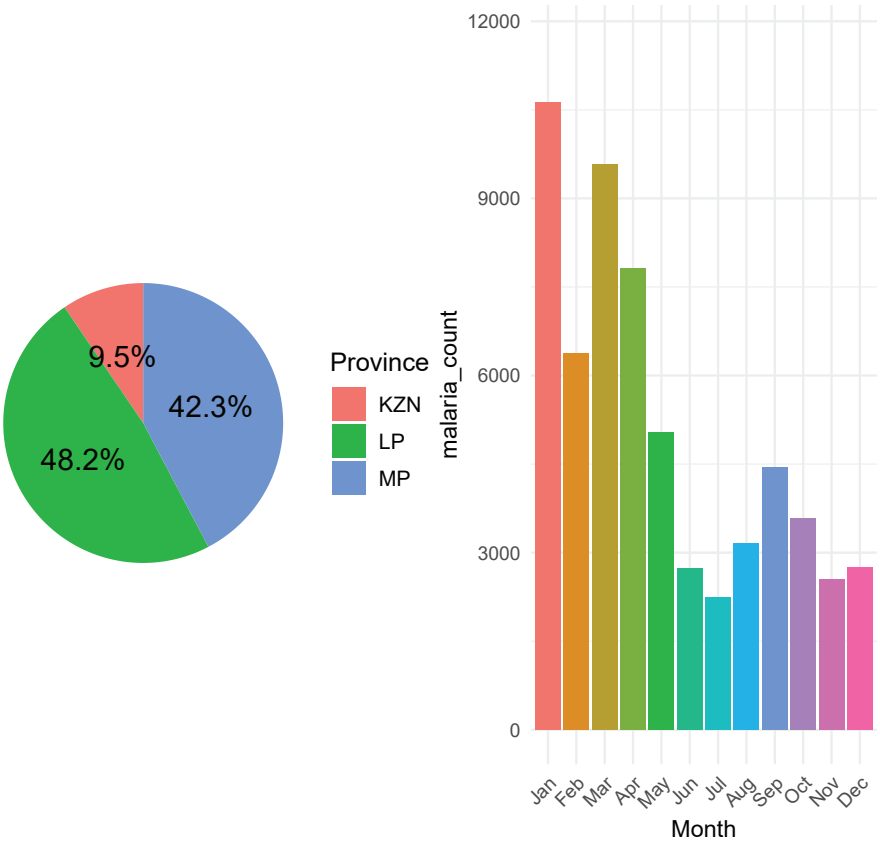

Figure 1. Malaria Distribution Across Provinces and Monthly Variations.

The composite Figure 1 provides an overview of malaria cases across three provinces, namely: KwaZulu-Natal (KZN), Limpopo (LP), and Mpumalanga (MP) and their monthly distribution. The pie chart illustrates the proportion of malaria cases reported in each province over the years of the study. Limpopo accounted for the highest share, contributing 48.2% of total cases, followed by Mpumalanga with 42.3%, and KwaZulu-Natal (KZN) with the lowest proportion at 9.5%. This suggests that malaria burden is heaviest in Limpopo and Mpumalanga, possibly due to environmental conditions that favour transmission.

The bar chart further breaks down malaria case patterns by month, revealing distinct seasonal trends. The highest transmission is observed from January to April, with January recording the peak number of cases, exceeding 10,000. Cases gradually decline from May onwards, reaching the lowest levels between June and August. A slight increase is observed from September to December, indicating the start of the malaria season. This trend aligns with known malaria transmission cycles, where cases peak during and after the rainy season due to increased mosquito breeding and favourable conditions for parasite development.

Figure 2 presents multiple scatterplots, each showing the relationship between malaria counts and a numeric predictor variable: maximum temperature (MaxTemp), minimum temperature (MinTemp), normalised difference vegetation index (NDVI), and rainfall (Rain). These visualisations help assess potential associations between malaria cases, climate, and environmental factors.

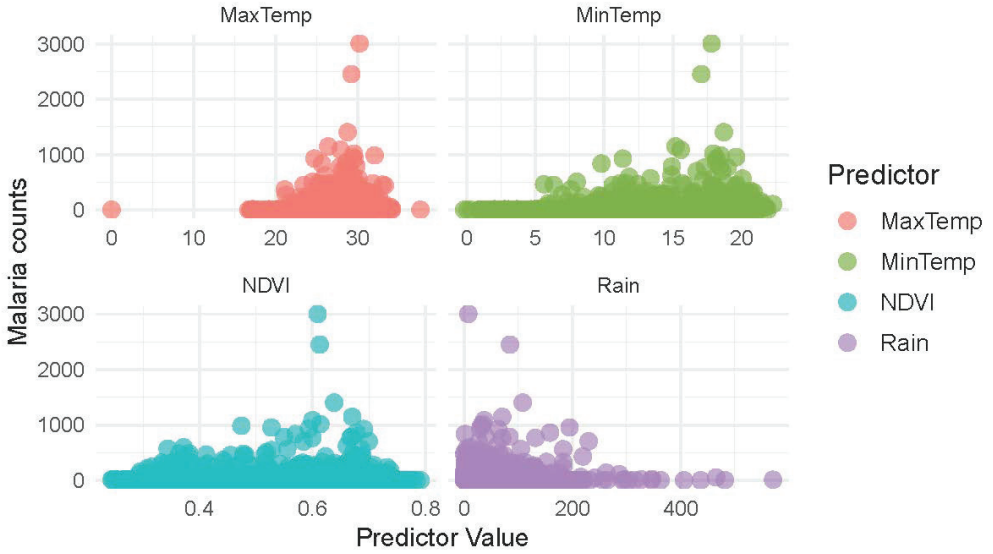

Figure 2. Scatterplots Showing Malaria Incidence Across Key Predictors

In the MaxTemp plot, malaria cases appear to increase with temperature around 25–30°C, after which they decline. This suggests that malaria transmission may be highest within this optimal temperature range, beyond which extreme heat might reduce mosquito survival. The MinTemp plot shows a somewhat similar pattern, with malaria cases increasing as minimum temperature rises, peaking around 15–20°C, and then tapering off. This suggests that malaria transmission is more favourable in regions where temperatures during the night remain moderate, supporting mosquito survival.

The NDVI plot suggests a weak but slightly positive relationship, where higher vegetation index values (around 0.5–0.7) correspond to an increased malaria count. This aligns with the idea that dense vegetation provides breeding grounds for mosquitoes, although the relationship does not appear strongly linear. The Rain plot shows that malaria cases are more frequent at low to moderate rainfall levels (approximately 0–200 mm), with fewer cases at very high rainfall levels (around 250 mm and more).

The line graph in Figure 3 display malaria case trends across different districts from 2018 to 2022. Each line represents a district, and the y-axis indicates the malaria count over the years. The legend at the bottom provides abbreviated district names. From the trends, two districts stand out with the highest malaria cases: one represented by a brownish-yellow line (Ehlanzeni district) and another by a pink line (Vhembe district). The Ehlanzeni district showed a sharp decline in malaria cases from 2018 to 2021 before stabilising in 2022. Conversely, the Vhembe district experienced fluctuations, with cases dropping initially but rising sharply in 2020 before declining again. Mopani and Umkhanyakude districts had moderate malaria counts, showing either a steady trend or a slight increase in cases over the years. Other districts, represented by thin lines near the bottom, had consistently low malaria counts, suggesting lower transmission levels.

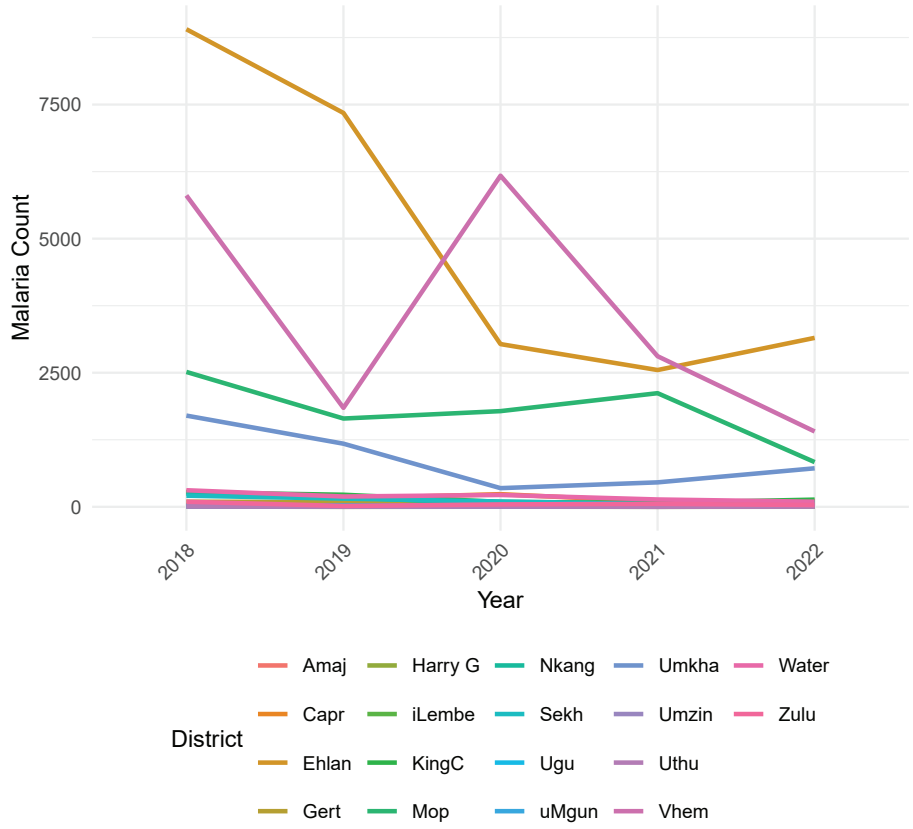

Figure 3. Malaria Case Trends Over the Years Across Districts.

Table 4. Model Comparison for Bayesian Negative Binomial Regression Using Different Prior Distributions.

Bayesian Negative Binomial Model Prior	**Elpd_loo Estimate**	**Elpd_loo Std Error**	**P_loo Estimate**	**P_loo Std Error**	**Looic Estimate**	**Looic Std Error**
Normal (0.25, 0.00168)	-2334.0	63.9	8.5	2.7	4669.7	127.8
Jeffreys beta(0.000009, 0.5)	-2388.3	63.7	4.9	1.4	4776.7	127.4
Betawi (0.090, 65.430)	-2368.0	63.3	6.0	1.2	4736.1	126.7
Gamma(37.585, 150)	-2370.5	64.2	5.2	1.1	4741.0	128.3
Betai (27.948, 83.957)	-2374.2	64.4	5.3	1.2	4748.5	128.8

Table 4 presents the results of a Bayesian negative binomial regression model using different prior distributions, with key metrics including the Expected Log Predictive Density using Leave-One-Out Cross-Validation ($elpd_loo$), the estimated effective number of parameters (P_loo), and the Leave-One-Out Information Criterion (Looic). The key metrics estimates for each prior distribution tested provide a means to evaluate the sensitivity of the model to prior specifications. Among the priors tested, the Normal(0.25, 0.00168) prior demonstrates the best predictive performance, having the highest $elpd_loo$ (-2334) and the lowest Looic (4669.7). However, this model is also the most complex, as indicated by its highest P_loo estimate (8.5). In contrast, the Jeffreys [beta(0.000009, 0.5)] prior performs the worst, with the lowest $elpd_loo$ (-2388.3) and the highest Looic (4776.7), suggesting weaker predictive accuracy despite its lower complexity ($P_loo = 4.9$). In particular, the model using Jeffreys prior was simplified

due to convergence issues with a more complex version (see Appendix 2). The beta (0.090, 65.430) and gamma (37.585, 150) priors offer a balance between model complexity and predictive performance, with moderately high $elpd_loo$ values and lower P_loo estimates. While the Normal prior provides the best overall predictive fit, simpler models like the beta or gamma priors remain viable alternatives depending on the trade-off between complexity and performance. However, given that this study focuses on developing a predictive model, the normal prior will be used for malaria count predictions. Additionally, there is no extreme gap in predictive performance across models, suggesting that there are no sensitivity issues.

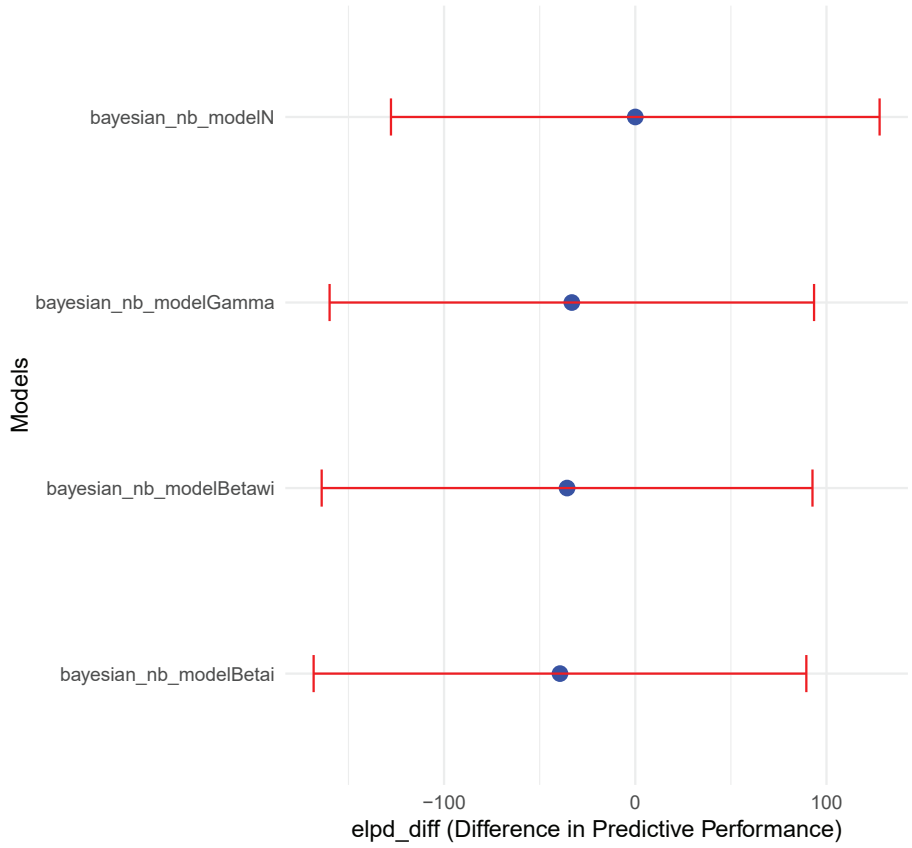

Figure 4. LOO-CV Model Comparisons.

The plot in Figure 4 presents a Leave-One-Out Cross-Validation (LOO-CV) comparison of four Bayesian negative binomial models: the model with informative normal prior (bayesian_nb_modelN), informative gamma prior (bayesian_nb_modelGamma), weakly informative beta prior (bayesian_nb_modelBetawi), and informative beta prior (bayesian_nb_modelBetai). In this LOO-CV comparison, the best model serves as the reference, meaning its expected log predictive density difference (elpd_diff) is set to zero. Based on the plot, bayesian_nb_modelN is the best-performing model. The second-best model is bayesian_nb_modelGamma, as its elpd_diff is close to zero with overlapping uncertainty intervals. Following that, bayesian_nb_modelBetawi ranks third, showing a slightly lower elpd_diff than the second-best model. The least-performing model is bayesian_nb_modelBetai, which has the lowest elpd_diff. However, due to the large uncertainty intervals, the differences between the models are not major, suggesting that their predictive performances may be quite similar.

Table 5. Bayesian Negative Binomial Regression Model Summary (with Informative Normal Prior).

Coefficients	**Estimate**	**Est. Error**	**l-95%CI**	**u-95%CI**	**Sig**
Intercept	-3.724	0.599	-4.896	-2.538	***
factorProvinceKZN (Ref)	———	———	———	———	———
factorProvinceLP	0.253	0.040	0.175	0.332	***
factorProvinceMP	0.343	0.041	0.264	0.422	***
factorDistrictCapr (Ref)	———	———	———	———	———
factorDistrictEhlan	0.358	0.040	0.280	0.437	***
factorDistrictGert	0.239	0.041	0.160	0.318	**
factorDistrictHarryG	0.233	0.041	0.152	0.314	**
factorDistrictiLembe	0.240	0.041	0.159	0.319	**
factorDistrictMop	0.251	0.042	0.168	0.332	**
factorDistrictNkang	0.247	0.039	0.169	0.324	**
factorDistrictUgu	0.236	0.041	0.156	0.314	**
factorDistrictuMgun	0.236	0.041	0.157	0.315	**
factorDistrictUmkha	0.252	0.041	0.173	0.331	**
factorDistrictUmzin	0.234	0.041	0.154	0.315	**
factorDistrictUthu	0.237	0.041	0.158	0.317	**
factorDistrictVhem	0.272	0.040	0.193	0.350	***
factorDistrictWater	0.239	0.040	0.162	0.316	**
factorDistrictZulu	0.237	0.041	0.157	0.317	**
factorMonthJan	0.255	0.040	0.177	0.333	***
factorMonthFeb	0.241	0.040	0.164	0.319	**
factorMonthMar	0.245	0.040	0.164	0.324	**
factorMonthApr (Ref)	———	———	———	———	———
factorMonthMay	0.260	0.040	0.183	0.337	***
factorMonthJun	0.272	0.040	0.193	0.352	***
factorMonthJul	0.257	0.041	0.177	0.336	**
factorMonthAug	0.251	0.040	0.173	0.329	**
factorMonthSep	0.246	0.041	0.164	0.326	**
factorMonthNov	0.240	0.040	0.162	0.317	**
factorMonthOct	0.238	0.041	0.160	0.318	**
factorMonthDec	0.239	0.040	0.159	0.318	**
factorYear2018 (Ref)	———	———	———	———	———
factorYear2019	0.247	0.041	0.169	0.326	**
factorYear2020	0.227	0.040	0.149	0.306	**
factorYear2021	0.238	0.040	0.159	0.317	**
factorYear2022	0.257	0.041	0.178	0.337	***
MaxTemp	0.186	0.027	0.133	0.239	***
MinTemp	0.156	0.023	0.111	0.201	***
Rain	-0.007	0.001	-0.010	-0.005	***
NDVI	0.248	0.040	0.170	0.326	**

Table 5 presents the results of a Bayesian negative binomial model developed using an informative normal prior distribution. It shows the estimated coefficients (Estimate), their estimation errors (Est. Error), and the 95% credible intervals (l-95% CI and u-95% CI). Statistical significance (Sig) is indicated with asterisks, three asterisks indicate very strong significance (the credible intervals are far from zero) while two asterisks indicates moderate significance (the credible intervals are still not including zero but are closer). This model produces reliable and stable parameter estimates, enabling valid inferences and accurate predictions. Model convergence confirms this. See Appendix 1 for model diagnostics and evidence of convergence.

The intercept has an estimate of -3.724, suggesting a baseline log-count of the response variable in the reference categories. The province effects show that compared to the reference province (KZN), both Limpopo (LP) and Mpumalanga (MP) have positive and significant estimates (0.253 and 0.343, respectively), indicating higher expected counts. Similarly, the district effects suggest that all included districts have positive associations with the response variable relative to the reference district (Capr), with Ehlanzeni showing the highest estimate (0.358). Seasonal patterns are evident, with all months except April (the reference category) showing positive and significant effects. Notably, June (0.272) has the highest estimate, indicating an unexpected peak in activity during mid-year. The year effects indicate that cases were higher in all years relative to 2018, with 2022 showing the strongest increase (0.257).

Regarding climatic variables, maximum and minimum temperatures have positive and significant associations, meaning that higher temperatures correspond to increased cases. Rainfall, however, has a small but significant negative effect (-0.007), suggesting that increased rainfall is linked to fewer cases. Lastly, NDVI is positively associated (0.248), indicating that greener vegetation correlates with higher case counts.

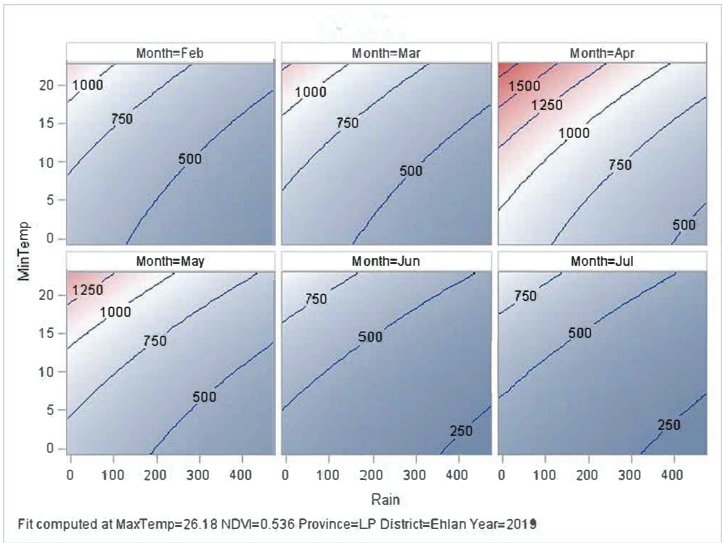

Figure 5. Malaria risk contours for different months based on rainfall and mimimum temperature.

The contour plot in figure 5 illustrates the relationship between minimum temperature, rainfall, and predicted malaria cases across different months (February–July) with reference categories, Ehlanzeni district, Limpopo Province, and 2019. The contour lines represent malaria predictions, with blue shades indicating lower values and red shades representing higher values. The results suggest that malaria risk is highest in March, April, and May, particularly in warm and wet conditions, with April showing the highest predicted cases exceeding 1,500. By June and July, malaria predictions drop remarkably, aligning with cooler and drier conditions. The plot also highlights the nonlinear effect of rainfall, moderate rainfall combined with higher temperatures leads to increased malaria risk, while excessive rainfall (>300 mm) does not always correspond to higher cases, possibly due to habitat flushing. These findings suggest that peak malaria transmission occurs in early autumn, warranting intensified control efforts during this period.

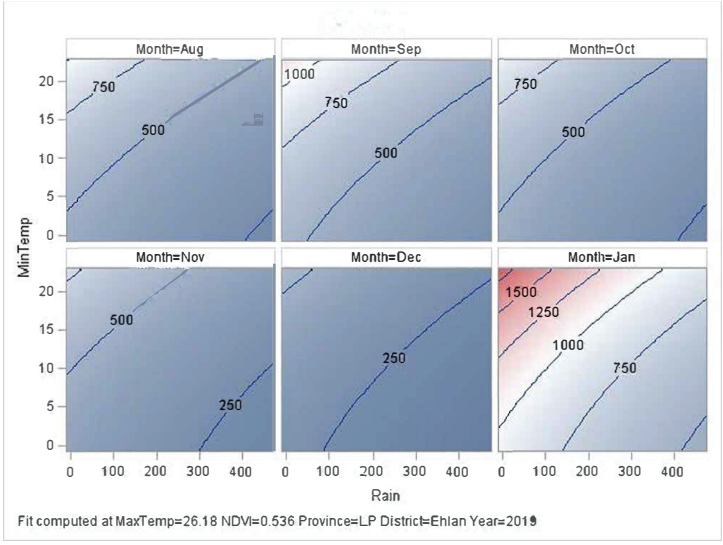

Figure 6. Malaria risk contours for different months based on rainfall and mimimum temperature.

The contour plot in Figure 6 depicts the relationship between minimum temperature, rainfall, and predicted malaria cases across different months (August–January). The highest malaria predictions occur in January, where warm temperatures and moderate rainfall contribute to cases exceeding 1,500. In contrast, from August to December, malaria predictions remain relatively low, with contour values mostly below 1,000 and an expected substantial decline in November and December. This seasonal pattern suggests that malaria transmission is highest in the summer months, particularly in January, when conditions are most favourable for mosquito breeding and parasite development. The gradual increase in predicted cases from September onward highlights the transition from low-risk winter months to peak transmission season.

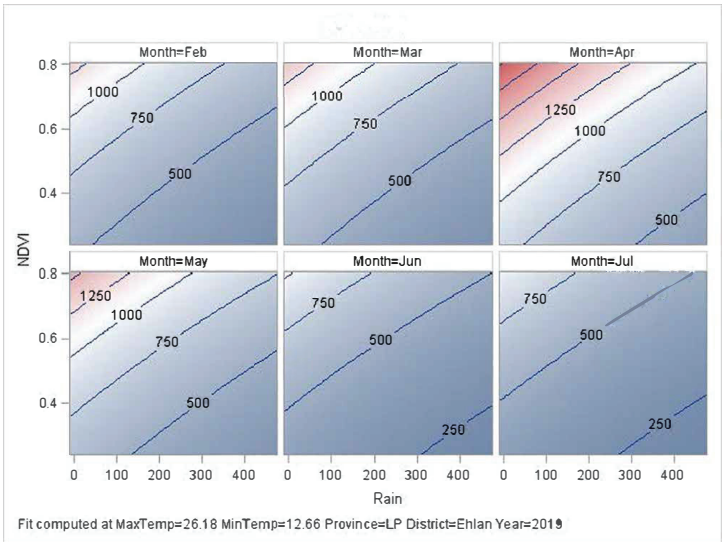

Figure 7. Malaria risk contours for different months based on NDVI and rainfall.

The contour plot in Figure 7 depicts the fitted relationship between rainfall, NDVI, and predicted malaria cases across different months from February to July. The results suggest that malaria cases are expected to peak between March and May, particularly when NDVI is high (above 0.6) and rainfall is moderate (around 100–200 mm). In contrast, malaria predictions decline in June and July, with lower values dominating the plot. The model also indicates that excessive rainfall beyond 300 mm may reduce malaria cases, possibly due to unfavourable breeding conditions for mosquitoes. The fit was computed at an average maximum temperature of 26.18°C and an average minimum temperature of 12.66°C relative to Ehlanzeni district, Limpopo province, and the year 2019. It highlights the role of climatic factors in malaria transmission.

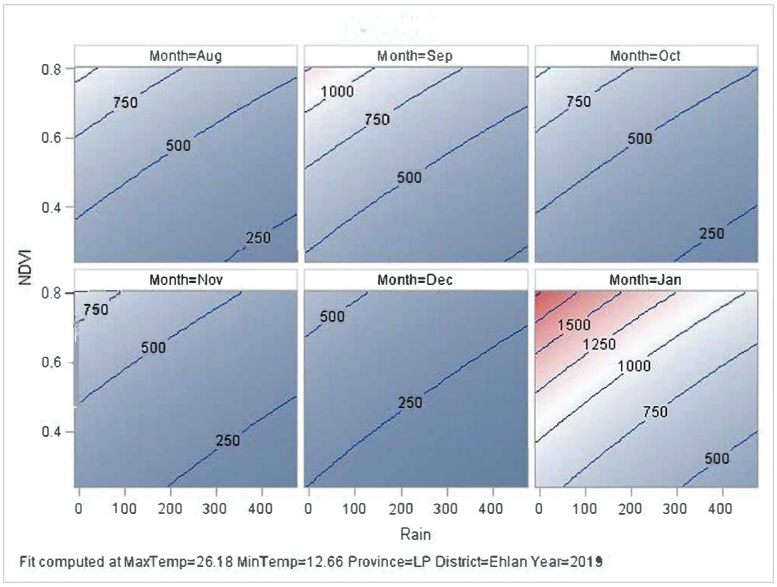

Figure 8. Malaria risk contours for different months based on NDVI and rainfall.

The contour plot in Figure 8 illustrates the relationship between rainfall, NDVI, and predicted malaria cases across different months from August to January. The results suggest a seasonal pattern in malaria risk, with relatively lower case counts predictions from August to December and a noticeable increase in January. This surge in January, indicated by the red-shaded region, corresponds to higher malaria cases, reaching values up to 1500. Rainfall appears to be a key driver of malaria incidence, as moderate rainfall levels generally coincide with an increase in malaria cases, particularly in January. In contrast, during months like August and November, the malaria burden is expected to be relatively low despite varying rainfall levels. NDVI, which represents vegetation greenness, also influences malaria cases, with higher NDVI values sometimes associated with an increased malaria burden. However, its effect appears less pronounced compared to rainfall.

Conclusion and Recommendations

3.1. Conclusion

This study integrates climate and environmental data to develop a predictive malaria model for three malaria-endemic provinces of South Africa. The climate factors considered include maximum and minimum average daily temperatures and rainfall, while the environmental factor is NDVI. In addition to exploratory data analysis, the study employs Bayesian methods, testing and comparing both objective and subjective prior distributions. The objective priors tested include Jeffreys' non-informative prior and a beta distribution, which serves as a weakly informative conjugate prior. For subjective informative priors, beta, gamma, and normal distributions are evaluated.

The findings reveal that among the three malaria-endemic provinces, Limpopo accounted for nearly half of all reported malaria cases between 2018 and 2022, followed by Mpumalanga with just over 40%. Ehlanzeni district in Mpumalanga emerges as the most susceptible to malaria transmission, followed by Vhembe and Mopani districts in Limpopo. The results indicate that location, seasonality, and climate factors influence malaria transmission remarkably. Warmer temperatures and greener environments contribute to higher malaria counts, while excessive rainfall has a suppressive effect due to habitat flushing.

Malaria cases are predicted based on climate and environmental conditions, with the highest transmission expected between January and May in areas where temperatures during the night range between 12°C-20°C and rainfall is moderate (about 100–200 mm). Similarly, malaria peaks in regions where NDVI exceeds 0.6, coinciding with moderate rainfall levels.

Overall, malaria transmission is expected to peak in summer and autumn while remaining low in winter and spring. The study highlights rainfall as a key risk factor, with heavy rainfall reducing malaria transmission by flushing breeding habitats. These findings reinforce the ongoing challenge climate change poses to malaria control and eradication efforts worldwide.

3.2. Recommendations

The predictions of the model can be integrated into early warning systems to anticipate malaria outbreaks, particularly in areas where transmission peaks are expected between January and May. Furthermore, the model could be integrated into decision-support tools that allow public health officials to monitor climatic conditions and predict malaria risks on a weekly or monthly basis. The tool could guide resource allocation, such as the distribution of insecticide-treated bed nets, spraying campaigns, and health outreach in high-risk districts.

We recommend that malaria control programme of South Africa place greater emphasis on control and eradication efforts in Ehlanzeni district (Mpumalanga), Vhembe and Mopani districts (Limpopo), as these areas experience the highest transmission rates. We also recommend closer collaboration between malaria control programmes and researchers to continuously update the model with new data, improving its accuracy and responsiveness over time. By utilising the model in this way, it could support data-driven, proactive interventions to reduce the burden of malaria in the most affected regions of South Africa.

3.3. Limitations of the Study and Future Research

One limitation of this study is that the current model may not fully capture potential non-linear relationships between climatic variables, particularly rainfall and malaria incidence. While the model demonstrated good predictive performance overall, the assumption of linear effects may overlook complex or threshold-dependent associations. Future research should explore more flexible modelling approaches, such as Gaussian processes, to better represent these dynamics and improve model interpretability and accuracy.

While this study successfully develops a Bayesian malaria prediction model incorporating climate and environmental factors, several areas warrant further investigation. Future research should explore Bayesian machine learning techniques, such as Bayesian neural networks and Gaussian processes, to enhance prediction accuracy while maintaining the ability to integrate prior knowledge. These methods could provide more flexible and robust models for malaria prediction. Additionally, expanding the model to incorporate additional climate and environmental variables, such as relative humidity, forest cover, land surface temperature, and elevation, may improve predictive performance. These factors could offer deeper insights into malaria transmission dynamics, particularly in regions with complex ecological interactions.

Appendices

Appendix 1

The inferences and malaria predictions in this study are based on a Bayesian Negative Binomial model developed using an informative normal prior distribution. The following figures present the model diagnostics, demonstrating that the model has successfully converged.

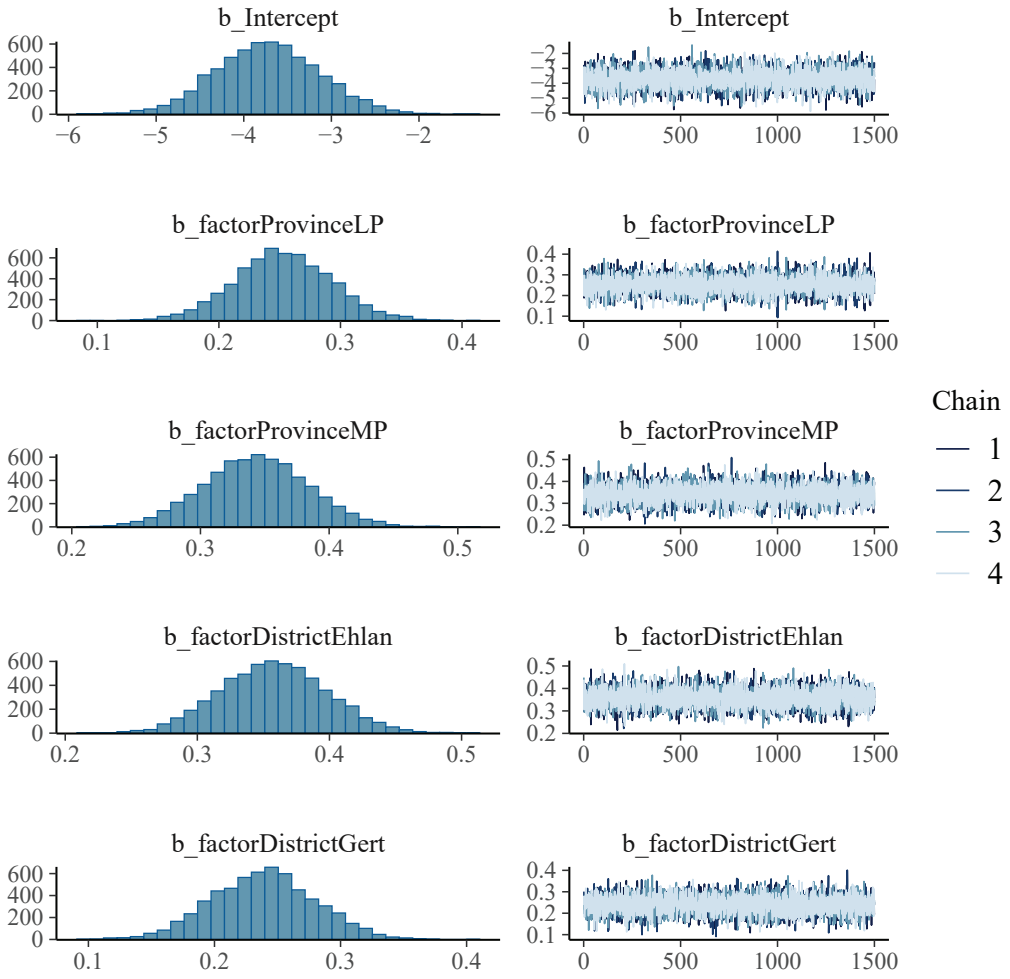

Figure 9. MCMC convergence diagnostics for the model intercept and the predictors (province and districts).

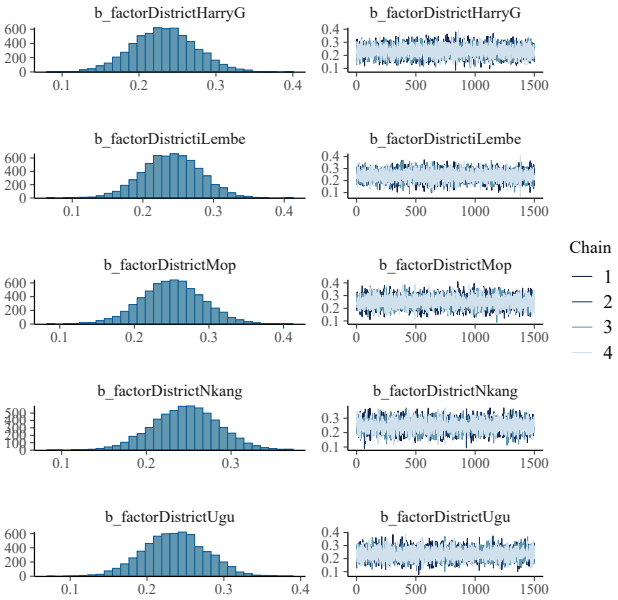

Figure 10. MCMC convergence diagnostics for the model predictors (districts).

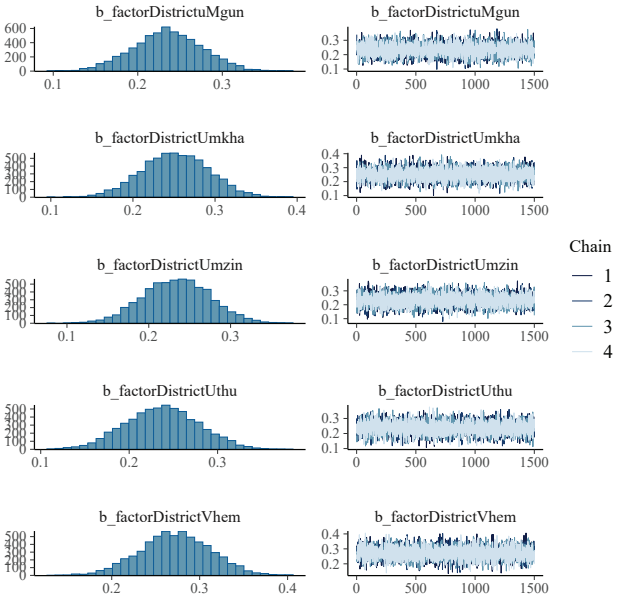

Figure 11. MCMC convergence diagnostics for the model predictors (districts).

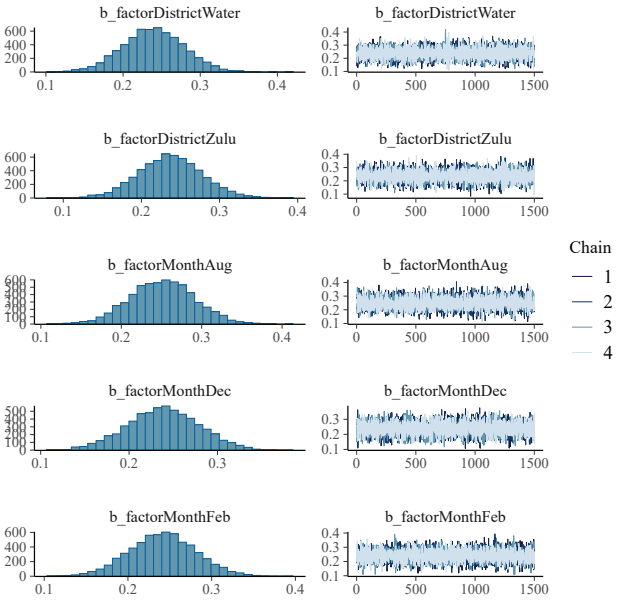

Figure 12. MCMC convergence diagnostics for the model predictors (districts and months).

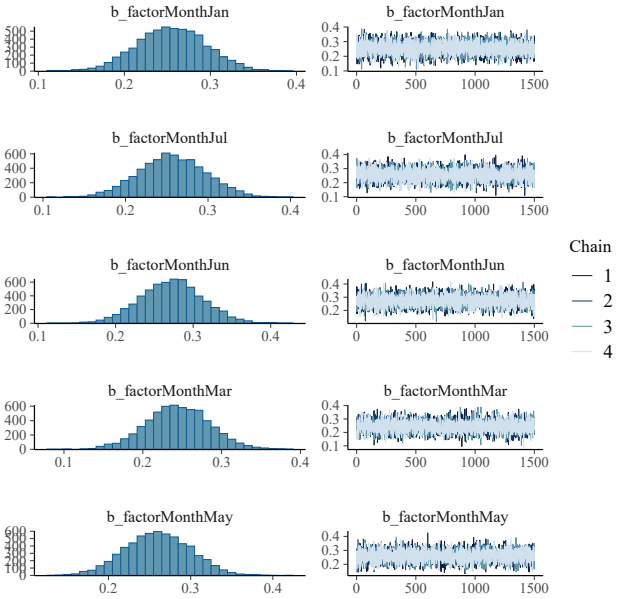

Figure 13. MCMC convergence diagnostics for the model predictors (months).

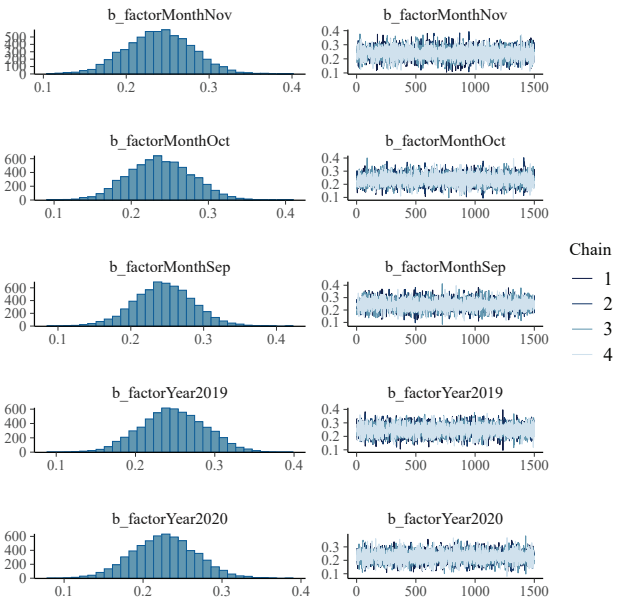

Figure 14. MCMC convergence diagnostics for the model predictors (months and year).

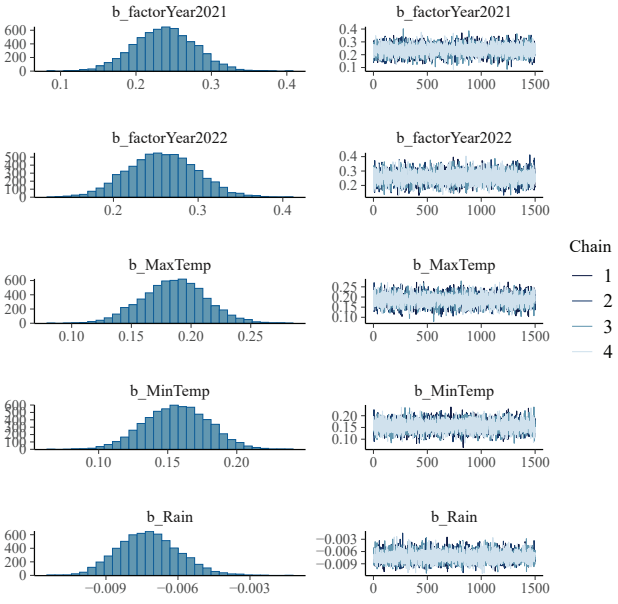

Figure 15. MCMC convergence diagnostics for the model predictors (year, maximum temperature, minimum temperature, and rain).

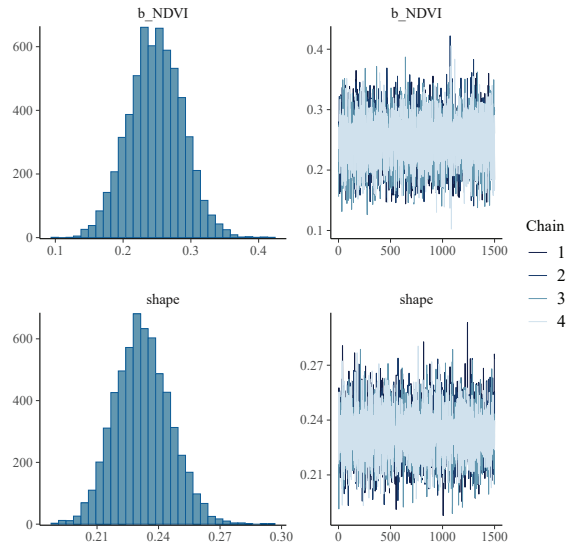

Figure 16. MCMC convergence diagnostics for the model predictor (NDVI).

Appendix 2

The figures in this appendix illustrate the model diagnostics for a fully Bayesian Negative Binomial model developed using Jeffreys prior distribution. Despite increasing the number of iterations and chains, convergence issues persisted. As a result, a simplified model was adopted for comparison purposes.

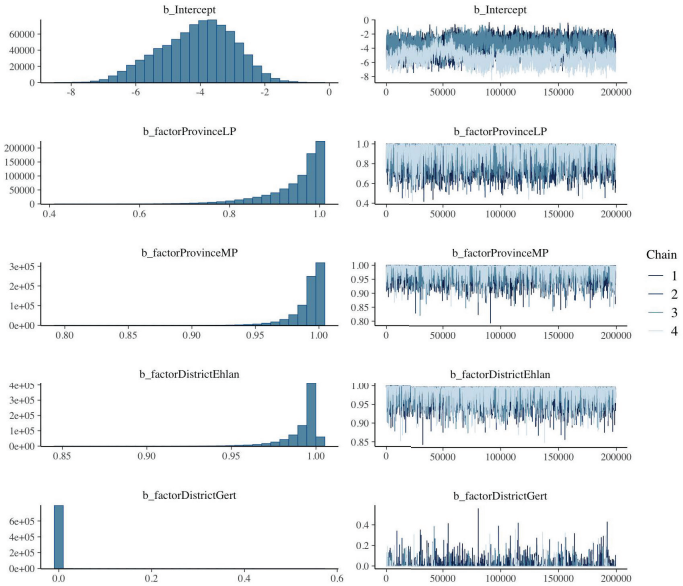

Figure 17. MCMC convergence diagnostics for the model intercept and the predictors (province and districts).

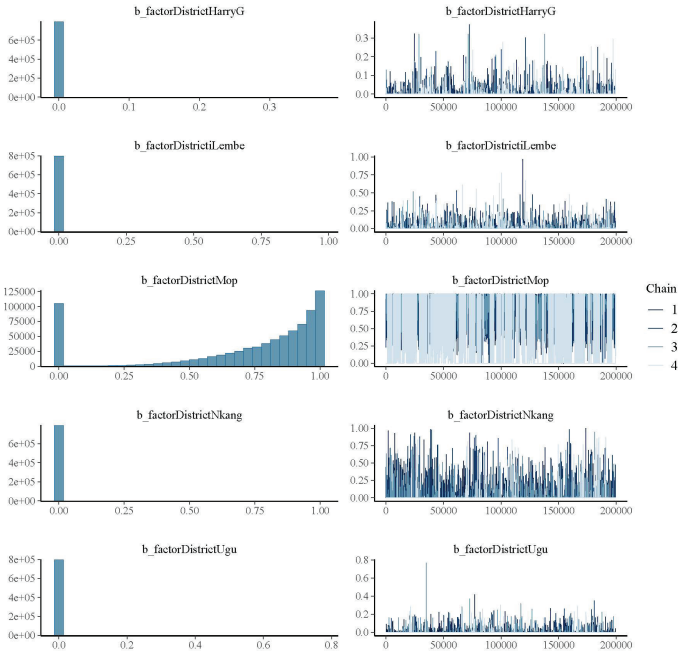

Figure 18. MCMC convergence diagnostics for the model intercept and the predictors (districts).

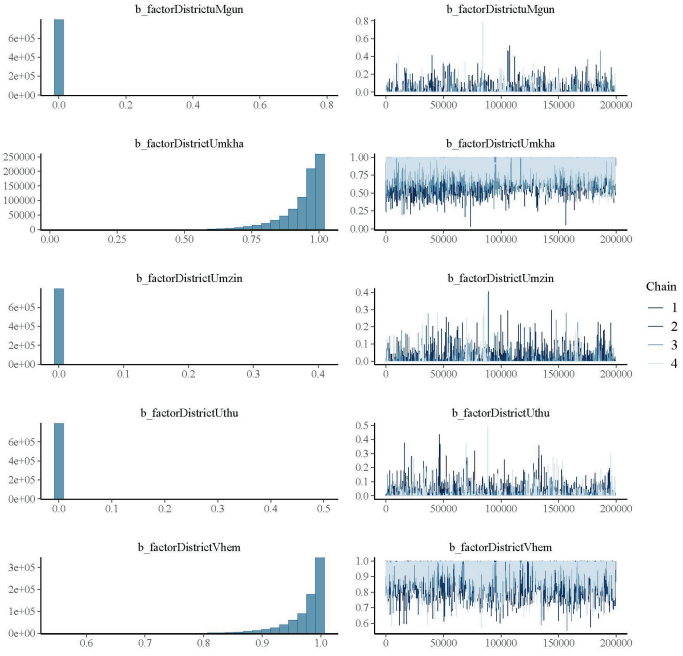

Figure 19. MCMC convergence diagnostics for the model intercept and the predictors (districts).

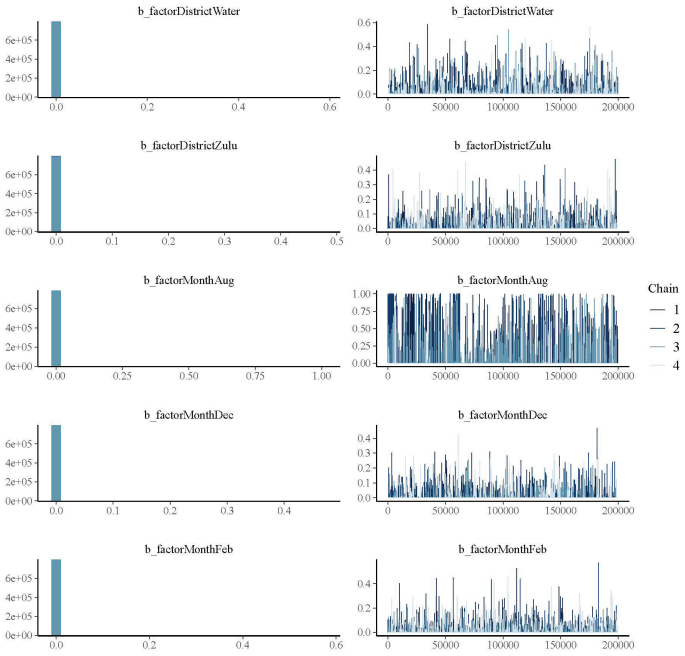

Figure 20. MCMC convergence diagnostics for the model predictors (districts and months).

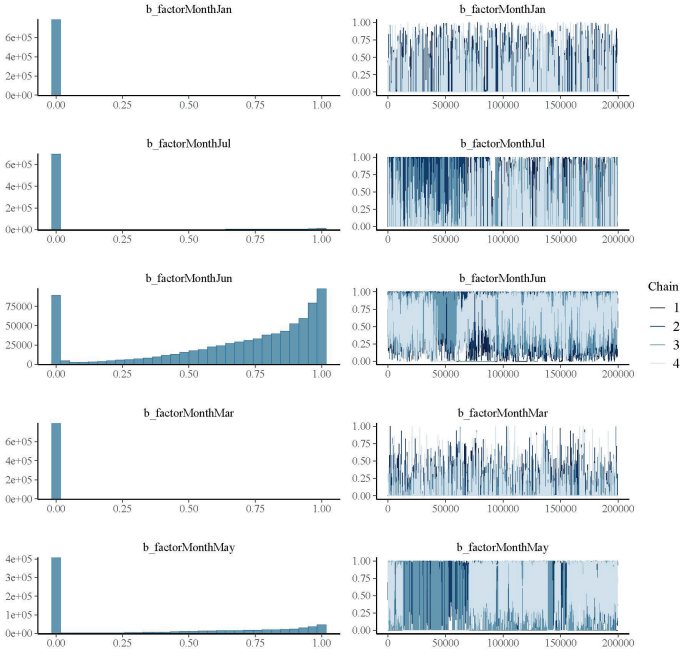

Figure 21. MCMC convergence diagnostics for the model predictors (months).

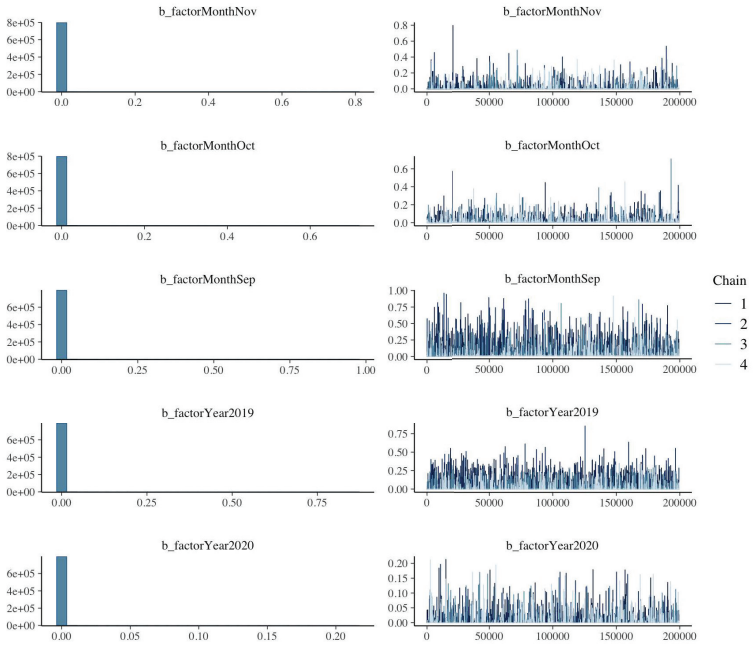

Figure 22. MCMC convergence diagnostics for the model predictors (months and year).

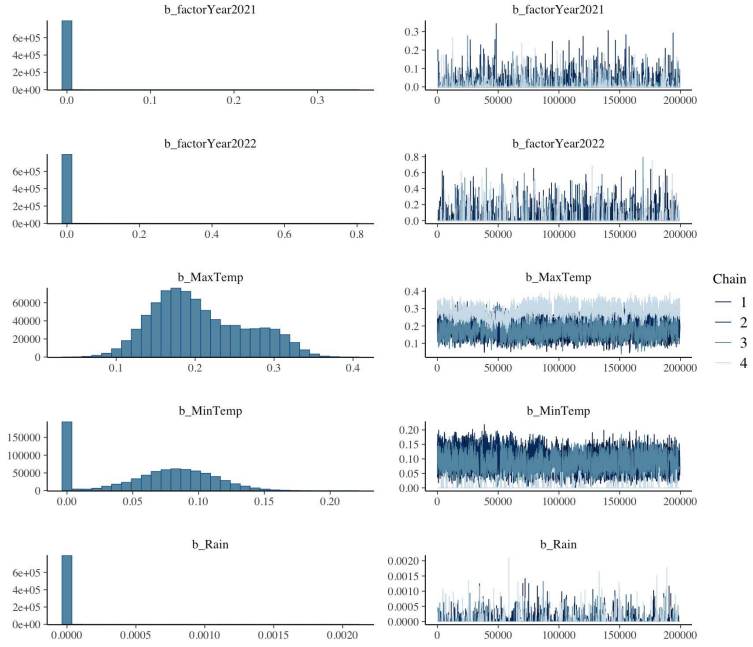

Figure 23. MCMC convergence diagnostics for the model predictors (year, maximum temperature, minimum temperature, and rain).

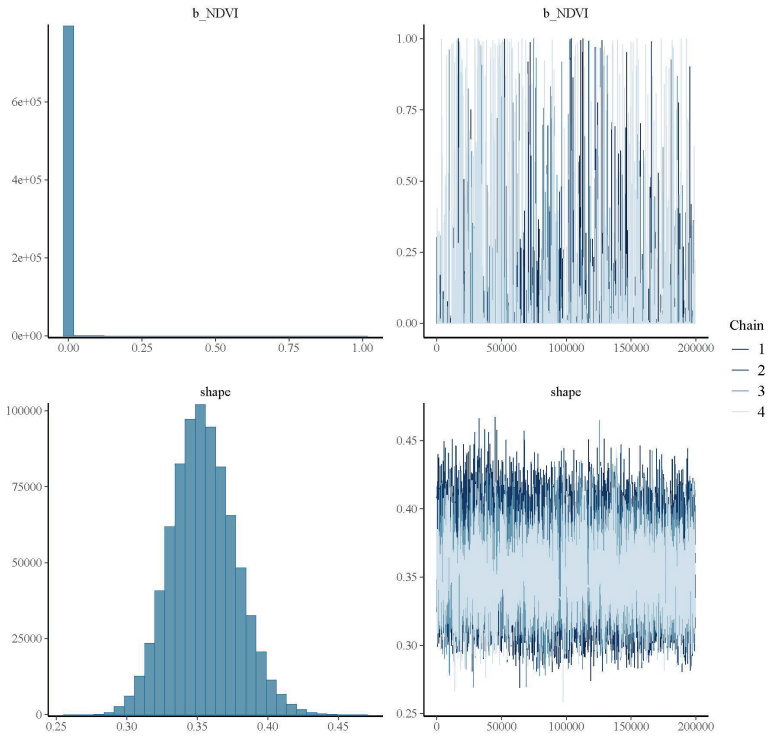

Figure 24. MCMC convergence diagnostics for the model predictor (NDVI).

Acknowledgement

The authors extend their sincere thanks to the anonymous referees, the Associate Editor, and the Editor for their valuable and constructive feedback, which significantly enhanced the quality of this paper.

REFERENCES

1. C. Li and S. Managi, *Global malaria infection risk from climate change*, Environmental Research, vol. 214, p. 114028, 2022.
2. World Health Organization, *World malaria report 2024: addressing inequity in the global malaria response*, World Health Organization, 2024.
3. M. Mabona, T. Zwane, J. Raman, L. Kuenza, B. Mhlongo, and P. Phafane, *Evaluation of the malaria case surveillance system in KwaZulu-Natal Province, South Africa, 2022: a focus on DHIS2*, Malaria Journal, vol. 23, no. 1, p. 47, 2024.
4. E. I. Obeagu and G. U. Obeagu, *Emerging public health strategies in malaria control: innovations and implications*, Annals of Medicine and Surgery, vol. 86, no. 11, pp. 6576–6584, 2024.
5. U. Samarasekera, *Climate change and malaria: predictions becoming reality*, The Lancet, vol. 402, no. 10399, pp. 361–362, 2023.
6. S. Bhatt, D. J. Weiss, E. Cameron, D. Bisanzio, B. Mappin, U. Dalrymple, K. E. Battle, C. L. Moyes, A. Henry, P. A. Eckhoff, et al., *The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015*, Nature, vol. 526, no. 7572, pp. 207–211, 2015.
7. S. Brugueras, B. Fernández-Martínez, J. Martínez-de la Puente, J. Figuerola, T. M. Porro, C. Rius, A. Larrauri, and D. Gomez-Barroso, *Environmental drivers, climate change and emergent diseases transmitted by mosquitoes and their vectors in southern Europe: A systematic review*, Environmental Research, vol. 191, p. 110038, 2020.
8. World Health Organization, *World malaria report 2023*, World Health Organization, 2023.
9. The Lancet, *Malaria vaccines: a test for global health*, Lancet (London, England), vol. 403, no. 10426, p. 503, 2024.
10. M. Chutiyami, P. Saravanakumar, U. M. Bello, D. Salihu, K. Adeleye, M. A. Kolo, K. K. Dawa, D. Hamina, P. Bhandari, S. K. Sulaiman, et al., *Malaria vaccine efficacy, safety, and community perception in Africa: a scoping review of recent empirical studies*, Infection, pp. 1–22, 2024.

11. P. Phoobane, M. Masinde, and T. Mabhaudhi, *Predicting infectious diseases: a bibliometric review on Africa*, International Journal of Environmental Research and Public Health, vol. 19, no. 3, p. 1893, 2022.
12. W. Nkenguye, *Climate change impacts on malaria: a call to action*, IJS Global Health, vol. 7, no. 2, p. e0428, 2024.
13. T. F. Tokponnon, R. Ossè, B. Yovogan, E. Guidi, C. J. Adoha, A. Sominanhouin, J. Ahouandjinou, A. Sidick, and M. C. Akogbeto, *Presence of *Plasmodium vivax* in *Anopheles gambiae* and absence in other malaria vectors in Cove-Zagnanando-Ouinhi health zone in southern Benin, West Africa*, Malaria Journal, vol. 23, no. 1, p. 20, 2024.
14. Y. Kim, J. V. Ratnam, T. Doi, Y. Morioka, S. Behera, A. Tsuzuki, N. Minakawa, N. Sweijd, P. Kruger, R. Maharaj, et al., *Malaria predictions based on seasonal climate forecasts in South Africa: A time series distributed lag nonlinear model*, Scientific Reports, vol. 9, no. 1, p. 17882, 2019.
15. W. A. Landman, N. Sweijd, N. Masedi, and N. Minakawa, *The development and prudent application of climate-based forecasts of seasonal malaria in the Limpopo province in South Africa*, Environmental Development, vol. 35, p. 100522, 2020.
16. Y. A. Adamu et al., *Malaria prediction model using machine learning algorithms*, Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 10, pp. 7488–7496, 2021.
17. Y. W. Lee, J. W. Choi, and E. H. Shin, *Machine learning model for predicting malaria using clinical information*, Computers in Biology and Medicine, vol. 129, p. 104151, 2021.
18. O. Nkiruka, R. Prasad, and O. Clement, *Prediction of malaria incidence using climate variability and machine learning*, Informatics in Medicine Unlocked, vol. 22, p. 100508, 2021.
19. G. Lu, D. Zhang, J. Chen, Y. Cao, L. Chai, K. Liu, Z. Chong, Y. Zhang, Y. Lu, A. K. Heuschen, et al., *Predicting the risk of malaria re-introduction in countries certified malaria-free: a systematic review*, Malaria Journal, vol. 22, no. 1, p. 175, 2023.
20. L. Held, *Bayesian methods in epidemiology*, in *Handbook of Epidemiology*, pp. 1–35, Springer, 2024.
21. J. U. Ibeji, H. Mwambi, and A.-K. Iddrisu, *Bayesian spatio-temporal modelling and mapping of malaria and anaemia among children between 0 and 59 months in Nigeria*, Malaria Journal, vol. 21, no. 1, pp. 311, 2022.
22. M. Semakula, F. Niragire, and C. Faes, *Spatio-temporal Bayesian models for malaria risk using survey and health facility routine data in Rwanda*, International Journal of Environmental Research and Public Health, vol. 20, no. 5, pp. 4283, 2023.
23. C. Rotejanaprasert, V. Malaphone, M. Mayfong, K. Chindavongsa, V. Banouvong, B. Khamlome, P. Vilay, V. Vanisavaeth, and R. J. Maude, *Spatiotemporal patterns and association with climate for malaria elimination in Lao PDR: a hierarchical modelling analysis with two-step Bayesian model selection*, Malaria Journal, vol. 23, no. 1, pp. 231, 2024.
24. E. Torsen, *Objective versus subjective Bayesian inference: a comparative study*, Int J, vol. 3, pp. 56–65, 2015.
25. J. O. Berger, J. M. Bernardo, and D. Sun, *Objective priors for discrete parameter spaces*, Journal of the American Statistical Association, vol. 107, no. 498, pp. 636–648, 2012.