Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2025, 2025:52
https://doi.org/10.28919/cmbn/9180
ISSN: 2052-2541

MODELING THE EFFECT OF CLIMATIC FACTORS ON MALARIA INCIDENCES USING AN AUTO-REGRESSIVE GENERALIZED LINEAR MODEL WITH LAGGED COVARIATES

SUZAN KUTEGEKA[1,*], CHARITY WAMWEA[2], ANTHONY WAITITU[2]

[1]Department of Mathematics, Pan African University Institute of Basic Science Technology and Innovation, Nairobi, Kenya

[2]Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

Abstract. Malaria remains a significant public health challenge globally, with its transmission intricately associated with climatic factors. Understanding the time lag effect of these variables on malaria incidence is crucial for developing effective control strategies. The research used monthly malaria incidence data obtained from the Uganda Ministry of Health, along with climate data from the Uganda National Meteorological Authority to investigate the impact of climatic factors, specifically rainfall, temperature, humidity and wind, on malaria incidence in Arua District, using monthly data from 2020 to 2024. To capture both immediate and delayed effects, the study employed an autoregressive generalized linear model (AR-GLM) with lagged covariates. The results indicate that increased rainfall and humidity with specific delays are positively associated with the incidence of malaria, while changes in temperature show a complex and unstable relationship. In addition, the AR-GLM model, which incorporates these time-dependent effects, outperforms a standard GLM, as evidenced by lower Akaike Information Criterion (AIC) and Mean Absolute Percentage Error (MAPE) values. Residual analysis confirms that AR-GLM adequately captures the temporal dependencies in the malaria data, showing no significant autocorrelation in residuals. The findings contribute to the development of effective malaria control and prevention strategies tailored to

*Corresponding author
E-mail address: suzankutegeka2020@gmail.com
Received February 07, 2025

the specific temporal dynamics of the climate-malaria interactions in Arua district, ultimately aiming to improve public health outcomes.

Keywords: malaria incidence; climatic factors; autoregressive model; Generalised Linear Models; lagged covariates; Uganda, AR-GLM.

2020 AMS Subject Classification: 92D30.

1. INTRODUCTION

Malaria is a disease caused by a parasitic infection transmitted by a mosquito (female Anopheles), and it is still a major health problem not only in Africa but worldwide. The world wide mortality rate for malaria spans from 0.3% to 2.2%; in tropical climate regions, the mortality rate is higher, ranging from 11% to 30%, particularly in cases involving severe forms of malaria [1]. According to the World Health Organization, children under 5 years of age account for nearly 80% of all malaria fatalities. According to [2], 249 million cases were estimated globally in 2022 and this surpassed the pre-pandemic level of 233 millions in 2019 by 16 million cases.

According to the World Health Organization, there was a decrease in the number of malaria cases from 238 million in 2000 to 229 million in 2019, but the disease remains widely spread in many countries, especially in Africa. Malaria cases increased steadily from 2019 to 2021, with a slower rise observed between 2020 and 2021, culminating in a global total of 247 million cases in 2021, up from 245 million in 2020 and 232 million in 2019 [3]. Climate change could result in a rise in annual fatalities from malnutrition, malaria, diarrhea, and heat-related illnesses from 2030 to 2050, according to the World Health Organization. It has been observed that the effects of climate change on the transmission of malaria are currently evident in many areas, some locations have shown positive advancements in combating malaria over the past decade [4]. According to [5], malaria transmission varies notably between urban and rural areas in Africa, with rural areas experiencing increased exposure because of the greater abundance and variety of mosquitoes that carry the disease. However, research indicates that malaria transmission has risen in many cities across Sub-Saharan Africa since 2003 [6]. Countries in East and Southern Africa experience elevated malaria transmissions, presenting a high risk of malaria, with many of them having consistently stable transmission patterns [7]. Uganda grappled with the

third-highest worldwide incidence of malaria cases, accounting for 5.1% of the global burden, and recorded the seventh-highest mortality rate at 3.2% [3].Uganda holds the highest global malaria incidence rate at 478 cases per 1,000 population annually, and the malaria-related death toll in the country is estimated to range between 70,000 and 100,000 deaths each year, surpassing the mortality rate associated with HIV/AIDS [8]. Climatic factors, especially temperature, rainfall, and humidity, are critical drivers of malaria transmission. These variables influence mosquito breeding, survival, and the development rate of malaria parasites. For instance, rainfall provides breeding sites for mosquitoes, while temperature affects mosquito and parasite life cycles, with warmer temperatures generally increasing transmission rates [9] [1]. In Uganda, malaria incidence remains high, particularly in regions like Acholi and West Nile, where recent data indicate a malaria prevalence rate nearly three times the national average [10]. In areas like Arua City, the prevalence reaches as high as 56%, over six times the national rate, emphasizing the urgent need for targeted malaria interventions in these hotspots [11].

Research has increasingly focused on understanding the temporal relationship between climate factors and malaria transmission to enhance prediction and control efforts. For example, studies in North Namibia and southwest China have highlighted the importance of incorporating time-lagged effects of rainfall and temperature to better model malaria transmission dynamics [12];[13]. Similarly, studies from the Colombian Amazon and Mozambique underscore the role of seasonal climate patterns, such as El Niño, in influencing malaria peaks, with optimal lag periods ranging from one to three months [14]; [15].

Despite the insights provided by these studies, there remains a significant research gap in explicitly modeling delayed climate effects and capturing temporal dependencies in malaria data. Most existing models, including Generalized Linear Models (GLMs) and Generalized Additive Models (GAMs), lack the flexibility to handle both complex time dependencies in malaria-climate relationships while considering lagged factors. Advanced models like the Generalized Autoregressive Moving Average (GARMA) model have been proposed to address these limitations, but their complexity often limits practical application [16].

This study seeks to bridge this gap by employing an Autoregressive Generalized Linear Model (AR-GLM) with lagged covariates to capture the immediate and delayed effects of climatic factors on malaria transmission. An auto-regressive model with sufficient lags can capture the same dependencies that a moving average component would due to the ability of a stationary AR process to be expressed as an infinite MA process [17]. By examining the climate-malaria relationship in Arua District Uganda from 2020 to 2024, this study aims to identify critical time lags for climate variables, thereby informing effective malaria control and prevention strategies. The AR-GLM approach offers a simpler yet robust framework for understanding these delayed associations, ultimately contributing to improved predictive modeling for malaria intervention planning.

2. METHODS

2.1. Study Area and Data. The study was conducted in Arua district, located approximately 475 kilometers (295 miles) northwest of Kampala, Uganda. Arua district has geographical coordinates of 03°02'07.0"N latitude and 30°54'39.0"E longitude. It is situated at an average elevation of 1,310 meters (4,298 feet) above sea level. As of 2020, the population of Arua was estimated to be 72,400 people.

In this study, secondary data on monthly malaria incidence in the Arua district, covering the period from January 2020 to July 2024, was obtained from the Ministry of Health, Uganda. Monthly climate data, including rainfall, temperature, humidity, and wind force, was obtained from the Uganda National Meteorological Authority. The malaria incidence rate serves as the response variable, representing the monthly count of reported malaria cases in the district. Climate variables act as explanatory variables, rainfall measured in millimeters, temperature measured in degrees Celsius, humidity indicating moisture level, and wind force measured in knots, monthly average wind speed can influence mosquito dispersion. These variables are included as potential drivers of malaria transmission patterns in the model.

2.2. Generalized linear model (GLM). Generalized linear models (GLMs) developed by [18] extend classical linear models by allowing the mean of a population to be linked to a linear predictor through a potentially nonlinear link function. This framework accommodates response

variables that follow any distribution within the exponential family, providing greater flexibility in modeling diverse types of data. Given a response y_i, $i = 1,2,...,n$ where n is the number of observations and set of covariates $\mathbf{X}_i = x_{i1},...,x_{im}$, the generalized linear model (GLM) is given as

$$g(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta}_i, \tag{1}$$

where $\boldsymbol{\beta}_i$ is a vector of coefficients. The GLM consists of three model components, the random component which specifies the probability distribution of the response variable y_i, systematic component which relates the linear predictor η_i to the explanatory variables (predictors) as

$$\eta_i = \mathbf{X}_i^T \boldsymbol{\beta}, i = 1,...,n, \tag{2}$$

for some vector of parameters $\boldsymbol{\beta} = (\beta_1,...,\beta_m)^T$ are m unknown regression parameters to be estimated and covariate $\mathbf{X}_i = (x_{i1},...,x_{im})^T$ associated with observation y_i [19]

and link components which transforms the mean of the response variable to the linear predictor scale in such away that

$$g(\mu_i) = \eta_i, \tag{3}$$

g is a monotone, differentiable function called the link function; that is, it is flat, or increasing or decreasing with mean μ_i , but it cannot be increasing for some values of μ_i and decreasing for other values [20].

In a GLM, the response variable y_i is assumed to come from an exponential family distribution such as normal, binomial, Poisson and others.

2.3. Count Data Distribution. Malaria incidence which is our response variable refers to the number of new cases of malaria observed within a specific population over a particular period like monthly or yearly and this can be referred to as count data. Poisson distribution and the negative binomial distribution are frequently employed to model disease case counts. Generalized Linear Models (GLMs) provide a framework for regression analysis of count data [21].

2.3.1. *The Poisson Distribution.* The Poisson distribution itself is well-suited for modeling count data and has a specific probability mass function that characterizes it. The probability mass function (PMF) for a response variable y_i with the mean μ is given by

$$P(Y_i = y_i) = \frac{\mu^{y_i} e^{-\mu}}{y_i!}, \quad y_i = 0, 1, 2, \ldots, \text{ and } \mu > 0, \tag{4}$$

where $P(Y_i = y_i)$ is the probability that the discrete random variable Y takes non-negative integer values.Under the Poisson distribution, the probability for negative values is 0, that is $p(y < 0) = 0$, and $\mathbb{E}(Y) = \mu = \text{Var}(Y)$, meaning the mean is equal to the variance, a property known as equi-dispersion of the Poisson distribution.

2.3.2. *Negative Binomial.* When analyzing disease counts, it's common to encounter over-dispersion where the variance exceeds the mean [22]. To address this, the negative binomial model serves as a refinement of the Poisson regression. This model is designed to handle the over-dispersion often seen in disease count data. It is an alternative to the Poisson model that introduces an additional parameter to account for the extra variability. The Negative Binomial probability mass function (PMF) is

$$P(Y = y) = \frac{\Gamma\left(\frac{1}{\alpha} + y\right)}{\Gamma\left(\frac{1}{\alpha}\right) y!} \left[\frac{\frac{1}{\alpha}}{\left(\frac{1}{\alpha}\right) + \mu}\right]^{\frac{1}{\alpha}} \left[\frac{\mu}{\left(\frac{1}{\alpha}\right) + \mu}\right]^{y}, \tag{5}$$

where $\Gamma(.)$ is a gamma function. The mean and the variance of Y are given by

$$\begin{aligned} E(Y) &= \mu \\ \text{Var}(Y) &= \mu + \alpha\mu^2 \end{aligned} \tag{6}$$

As α approaches zero, the variance of Y equals the mean μ implying that the negative binomial converges to the Poisson distribution.

2.4. The Auto-Regressive Model (AR). The auto-regressive model is a type of stochastic time series model used to describe certain time-varying processes.This model captures the influence of the past values of the variable on its current value. The auto-regressive process of order p or AR(p) is defined by the equation

$$Y_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + \omega_t, \tag{7}$$

where $\omega_t \sim N(0, \sigma^2)$ is the error term, $\phi = (\phi_1, \phi_2, \ldots, \phi_p)$ is the vector of model coefficients, Y_t is the value of the time series at time t, Y_{t-j} indicates lagged values of the series, p is the order of the auto-regressive model which indicates how many past values are included in the model. The AR model assumes stationarity and that the error term should be white noise, meaning that it has the following properties such as Zero Mean, Constant Variance over time, no Auto-correlation.

2.5. The Auto-regressive Generalized Linear model (AR-GLM). The AR-GLM is formulated by combining the generalized linear model with the AR model and based on the following assumptions.

- The response y_t depends not only on the current and lagged covariates $\mathbf{X}_t$ but also on the past observations y_{t-j} to capture temporal dependence.
- For count data models, zero values of the response variable are handled by replacing them with a small positive constant to avoid undefined logarithmic terms.
- Covariates can have both immediate and delayed effects on the response, capturing how past covariates influence the response variable.

2.5.1. *Model Formulation.* Consider a time series of count data $\{y_t\}_{t=1}^{n}$, where y_t represents the malaria incidences at time t, and the model aims to relate these counts to past values of the series and other covariates X_t.

The response variable y_t is assumed to follow a distribution from the exponential family (Poisson or Negative Binomial) conditional on the information set $D_t = \{y_{t-1}, y_{t-2}, \ldots, y_1, X_1, \ldots, X_t\}$ containing past observations y_{t-1} and covariate vectors $\mathbf{X}_t = (x_{t1}, \ldots, x_{tm})$ such that

$$f(y_t|\mathbf{D}_t) = \exp\left\{\frac{y_t\theta_t - b(\theta_t)}{\varphi} + c(y_t, \varphi)\right\}, \tag{8}$$

Similar to the standard GLM, the conditional mean $\mu_t = \mathbb{E}(y_t/\mathbf{D}_t) = b'(\theta_t)$is related to the variables by a twice-differentiable one-to-one monotonic function g, which is called the link function. $\text{var}(y_t \mid \mathbf{D}_t) = \varphi b''(\theta_t) = \varphi v(\mu_t)$ is the conditional variance, where $v(\mu_t)$ is called the

variance function. However, unlike the standard GLM, the formula here allows autoregressive terms to be included additively [16] such that

$$g(\mu_t) = \mathbf{X}_t^T \boldsymbol{\beta} + \sum_{j=1}^{p} \boldsymbol{\phi}_j \left\{ y_{t-j} \right\}, \tag{9}$$

where $\sum_{j=1}^{p} \boldsymbol{\phi}_j \left\{ y_{t-j} \right\}$ are auto-regressive terms.

For count data, the logarithmic link function is used to relate the linear predictor to the conditional mean $\mu_t = \mathbb{E}(y_t | D_t)$ as

$$g(\mu_t) = \ln(\mu_t),$$

where $g(\cdot)$ is the link function, ensuring that the mean μ_t is positive, as is appropriate for count data. equation(9) becomes

$$\ln(\mu_t) = X_t^T \beta + \sum_{j=1}^{p} \boldsymbol{\phi}_j \left\{ y_{t-j} \right\}, \tag{10}$$

since the count of y_t is always positive and as ϕ contributes positively to the past values, the influence of y_{t-j} increases the expected count largely. To reduce on this effect we use $\ln(y_{t-1})$ instead of y_{t-l} directly. Also to avoid undefined values in these counts we add a constant c to the past values so that equation (10) becomes

$$\ln(\mu_t) = X_t^T \beta + \sum_{j=1}^{p} \boldsymbol{\phi}_j \left\{ \ln\left(y_{t-j} + c \cdot \mathbb{I}(y_{t-j}) \right) \right\}, \tag{11}$$

where $c \in (0,1)$, and $\mathbb{I}(y_{t-j})$ is an indicator function defined as

$$\mathbb{I}(y_{t-j}) = \begin{cases} 1 & \text{if } y_{t-j} = 0 \\ 0 & \text{if } y_{t-j} > 0 \end{cases}$$

Consider the term $\sum_{i=1}^{m} \sum_{k=0}^{r} \beta_i x_{t-k,i}$ which captures the immediate effects (when $k = 0$) and delayed effects (when $k > 0$) of the covariates $x_{t-k,i}$ on the current response $\ln(\mu_t)$. This captures how current and past values of the covariates influence the current response, then (11) becomes

$$\ln(\mu_t) = \beta_0 + \sum_{i=1}^{m} \sum_{k=0}^{r} \beta_{i,k} x_{t-k,i} + \sum_{j=1}^{p} \phi_j \left(\ln(y_{t-j} + c \cdot \mathbb{I}(y_{t-j}) \right). \tag{12}$$

This includes the auto regressive term to account for the influence of past values y_{t-j} of the response variable adjusted by the effects of the covariates at the specific time lag. The coefficient $\boldsymbol{\phi}_j$ weights these influences.

2.5.2. *Parameter Estimation.* The autoregressive generalized linear model fitting procedure performs a maximum likelihood estimation (MLE). Let $\mathbf{Q}$ contain the model parameter to be estimated. The log-likelihood of the data $y_{r+1},....,y_n$ is conditional on the first r observations $y_1,.....,r$ and on $\eta_t = g(\mu_t)$ for $r >= max(p)$ is given by

$$
\begin{aligned}
l(\mathbf{Q}) &= \sum_{t=r+1}^{n} \log f(y_t/\mathbf{D}_t) \\
l(\mathbf{Q}) &= \sum_{t=r+1}^{n} \left\{ \frac{y_t\theta_t - b(\theta_t)}{\varphi} + c(y_t,\varphi) \right\} = \sum_{t=r+1}^{n} l_t
\end{aligned} \tag{13}
$$

Given that

$$
U(\mathbf{Q}) = \frac{\partial l}{\partial \mathbf{Q}} = \left(\frac{\partial l_t}{\partial \beta_1}, \ldots, \frac{\partial l_t}{\partial \beta_m}, \frac{\partial l_t}{\partial \phi_1}, \ldots, \frac{\partial l_t}{\partial \phi_p}, \ldots \right),
$$

is a score equation. At maximum the score function equates to zero,

$$
U(\mathbf{Q}) = \frac{\partial l}{\partial \mathbf{Q}} = 0. \tag{14}
$$

Maximum partial likelihood estimators are solved by fishers scoring algorithm.

2.6. Analysis of the Relationship between Climatic factors and Malaria Incidences. The study used cross correlation analysis to find the significant lags of different covariates, estimated the parameters using MLE and then used the z-values to test the hypothesis about coefficients that is whether they are statistically different from zero. These z- values are obtained as

$$
z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}, \tag{15}
$$

where $\hat{\beta}_j$ is the estimated coefficient for predictor j, $SE(\hat{\beta}_j)$ is the standard error of the estimated coefficient. The z-value is distributed normally with a mean of 0 and a standard deviation of 1 as

$$
z \sim N(0,1)
$$

The p-value associated with the z-value can be calculated using the cumulative distribution function (CDF) of the standard normal distribution.

$$
p\text{-value} = P(Z \geq |z|) = 2 \cdot (1 - \Phi(|z|))
$$

where $\Phi(z)$ is the CDF of the standard normal distribution. Decisions are made based on the significance level for the p-values under the statistical hypothesis that,

$$H_0 : \beta_i = 0 \quad H_a : \beta_i \neq 0$$

The p-value that is less than the chosen significance level 0.05 leads to rejection of the null hypothesis. This indicates that there is sufficient evidence to conclude that the predictor has a statistically significant effect on the response variable.

2.7. Model performance.

2.7.1. *Model Comparisions.* Model performance was perfomed using tools such as the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) to evaluate temporal dependencies, and statistical metrics like the Akaike Information Criterion (AIC) and Mean Absolute Percentage Error (MAPE) to measure model quality and predictive accuracy. The ACF examines correlations between a time series and its lagged values, with significant correlations at non-zero lags indicating potential autocorrelation in residuals. The PACF focuses on the direct effects of individual lags, helping to identify which lags contribute most to the observed autocorrelation.

The AIC is used to compare models by balancing goodness of fit and complexity, penalizing those with excessive parameters. The model with the lowest AIC value is preferred, as it achieves the best trade-off between accuracy and simplicity. MAPE evaluates prediction accuracy by calculating the average percentage error between observed and predicted values, with lower values indicating better performance. Combined, these tools ensure comprehensive evaluation of the model's ability to capture data patterns, handle temporal dependencies, and make accurate predictions.

2.7.2. *Residual Analysis of AR-GLM.* Residual analysis evaluates how well a model explains observed data by comparing predictions to actual values. For count data (e.g., number of events), traditional methods assuming normality are unreliable. Instead, this study uses the **DHARMa package** in R, which generates residuals through simulations tailored to count models. DHARMa avoids normality assumptions by standardizing residuals to a scale between 0 and 1, making patterns easier to interpret. The analysis relies on two key diagnostic tools.

First, the *residual vs. predicted plot* identifies systematic errors, such as overdispersion (excessive variance in the data) or autocorrelation (time-dependent patterns in residuals). Deviations from a smooth curve centered on zero suggest the model misses important patterns, such as non-linear relationships between predictors and the response variable. Second, the *Q-Q plot* compares residuals to the model's expected distribution (e.g Poisson or negative binomial). Deviations in the tails of the plot may indicate outliers or overdispersion, while alignment with the reference line supports adequate fit. Statistical tests further validate assumptions. The *dispersion test* measures whether the model's variance aligns with the data. Values greater than 1 signal overdispersion, while values less than 1 indicate underdispersion. The *outlier test* flags extreme residuals that could skew results. DHARMa's approach ensures robust diagnostics for count data models, addressing challenges like non-normality and overdispersion through simulation-based residuals. This method simplifies identifying issues like model misspecification or outliers.

3. RESULTS

3.1. Data exploration. Monthly malaria counts for the period 2020 January to 2024 July was used. The data was obtained from the Ministry of Health, Uganda. Table 1 shows results from analysis of Malaria incidences.

TABLE 1. Summary of Malaria Incidences

Metric	**Value**
Mean	2758
Standard Deviation	830.98
Range	1525–4371
Variance	690521
Overdispersion Statistic (Variance / Mean)	250.34

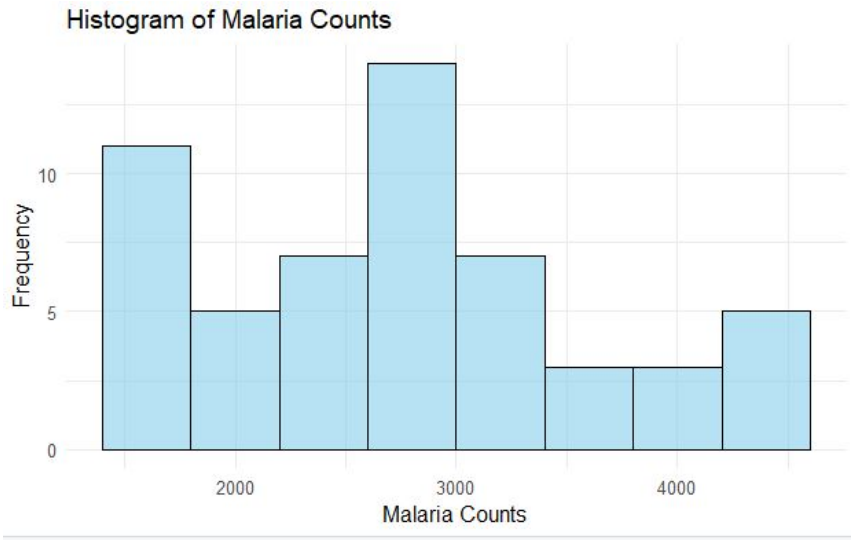

FIGURE 1. Histogram of malaria counts

From Table 1, the minimum value for malaria incidences is 1525, the maximum value is 4371 cases, the mean value for the whole period is 2758, the standard deviation is 830.98 and variance is 690521. the variance mean ratio is 250.34 which shows that malaria data is over-dispersed. From Figure 1, it is observed that the histogram is skewed to the right which is representation of count data. We conclude that malaria incidences is count data and we can model it using a distribution suited for count data which is the negative binomial distribution.

TABLE 2. Stationarity and Autocorrelation Tests

Variable/Test	**Statistic**	**p-value**
ADF Test Results for Stationarity		
Rainfall	-4.2407	0.0100
Mean Temperature	-4.0063	0.0158
Mean Wind	-4.9222	0.0100
Mean Humidity	-3.5146	0.0484
Malaria incidences	-3.7873	0.0256
Ljung-Box Test for Autocorrelation		
Malaria incidences	90.94	0.0000

From Table 2, Rainfall, Mean temperature, Mean wind, Mean humidity and malaria incidences have p-values $(0.0100, 0.0158, 0.0100, 0.0484, 0.0256)$ respectively which are below the the significant level (0.05) meaning the ADF test results indicate that each series is stationary that is these variables have relatively consistent statistical properties over time then we reject the null hypothesis of non-stationary. The p-value (0.0000) below 0.05 from Ljung-Box Test indicates significant autocorrelation in the malaria counts series. This means that an autoregressive component should be included in the model to capture this time-dependent structure.

3.2. Relationship between the different covariates and malaria incidence.

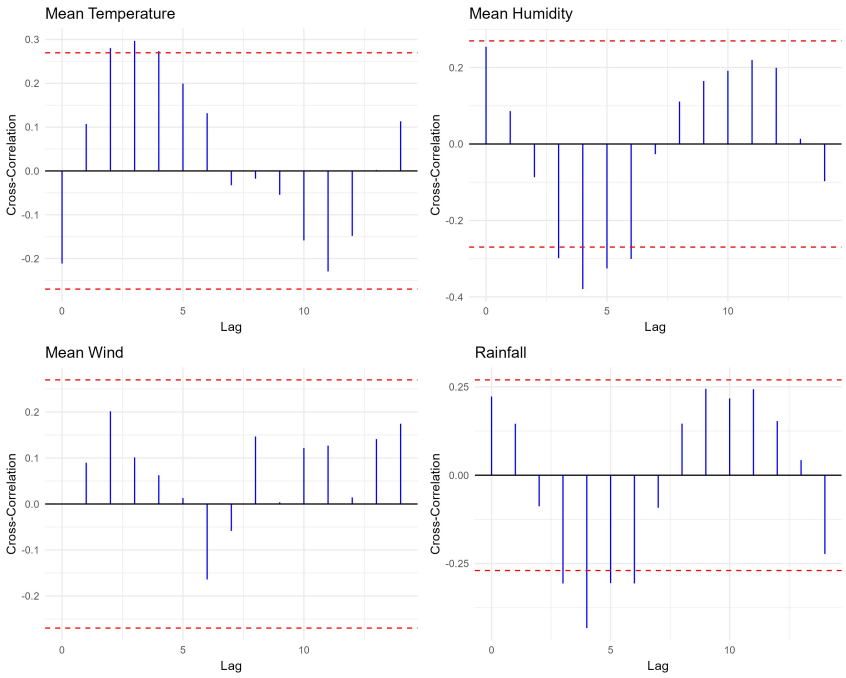

FIGURE 2. cross correlation plots

3.2.1. *Significant Time Lags.* Results from the Cross-Correlation Function (CCF) analysis Figure 2 show that certain climatic variables Mean Temperature, Mean Humidity, and Rainfall have significant lagged correlations with malaria incidences, suggesting that these factors may influence malaria incidence with a delayed effect. Each climate variable shows a distinct pattern of influence across different time lags for instance, Mean Temperature has significant positive correlations observed at lags 2, 3, and 4, Mean Humidity and Rainfall have significant positive

correlations at lags 3, 4, 5, and 6, while Mean Wind does not exhibit any significant lagged relationship.

The findings of previous studies suggest significant correlations that begin at lower lags, such as lag 0, lag 1, and lag 2 for most variables. For example, [23] observed that rainfall can influence mosquito breeding sites almost immediately, with significant impacts seen as early as lag 1. Similarly, [24] reported that rainfall at lag 1 and lag 2 can directly impact the dynamics of mosquito population in regions with favorable baseline conditions. Regarding temperature, [25] highlighted that even short-term temperature increases could accelerate mosquito and parasite development, with effects observable at lag 1 or lag 2. Furthermore, [26] demonstrated that short-term increases in humidity could sustain mosquito populations, influencing malaria transmission within a month (lag 1).

To fully capture both immediate and delayed effects, lags from 0 to 6 are included in the model to account for potential influences across a range of time lags.

3.2.2. *Regression Results for AR-GLM Model.*

TABLE 3. Regression Results for AR-GLM Model

Predictor	**Estimate**	**Std. Error**	**z-value**	**p-value**
Intercept	8.7811	2.2553	3.893	0.0001***
log(malaria_lag1)	0.5557	0.1008	5.512	0.0000***
Rainfall	0.0014	0.0006	2.382	0.0172*
Mean Temperature	-0.0750	0.0505	-1.485	0.1375
Mean Humidity	-0.0093	0.0059	-1.561	0.1184
Rainfall (Lag 1)	0.0015	0.0006	2.589	0.0096**
Temperature (Lag 4)	-0.1570	0.0594	-2.642	0.0082**
Fourier Cosine	0.4562	0.1095	4.167	0.0000***
Humidity (Lag 1)	-0.0100	0.0067	-1.499	0.1338
Humidity (Lag 2)	0.0136	0.0057	2.383	0.0172*

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

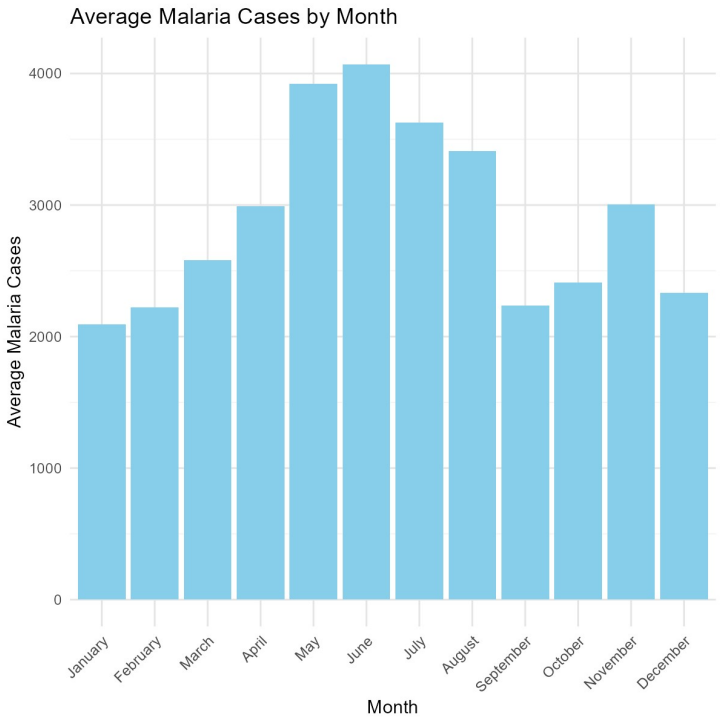

FIGURE 3. Average malaria cases by month.

From Table 3 the intercept (8.7811, p = 0.0001) represents the baseline log-transformed malaria incidence when all covariates are zero, A one-unit increase in log-transformed malaria cases from the previous month (lag 1) is associated with a 55.57% increase in current cases (0.5557, p=0.0000), highlighting the strong influence of past incidence on current trends. Rainfall shows a positive effect, where a one-millimeter increase leads to a 0.14% increase in malaria cases (0.0014, p = 0.0172), Additionally, rainfall from the previous month (lag 1) has a delayed effect, with a one-millimeter increase resulting in a 0.15% increase in cases (0.0015, p = 0.0096). Temperature from four months ago (lag 4) significantly reduces malaria cases by 15.70% for each one-degree Celsius increase (-0.1570, p = 0.0082), The Fourier cosine term (0.4562, p=0.0000) captures seasonal variability, reflecting cyclical environmental factors that significantly influence malaria transmission, while humidity from two months ago (lag 2) shows that a one-unit increase raises malaria cases by 1.36% (0.0136, p = 0.0172), indicating a delayed positive impact on mosquito development. In contrast, non-significant predictors include mean temperature and mean humidity.

3.3. Assessing the performance of the model. The best model for both the standard GLM and the AR-GLM were considered and compared. The models given below were obtained

$$\begin{aligned} E(y_t) &= \mu_t = \exp(\eta_t), \\ \eta_t &= \beta_0 + \log(\text{malaria\_lag1}) + \text{rainfall} + \text{mean\_temp} \\ &\quad + \text{mean\_humidity} + \text{rainfall\_lag1} + \text{temp\_lag4} \\ &\quad + \text{Fourier\_cos} + \text{humidity\_lag1} + \text{humidity\_lag2}. \end{aligned} \tag{16}$$

$$\begin{aligned} E(y_t) &= \mu_t = \exp(\eta_t), \\ \eta_t &= \beta_0 + \text{mean\_temp} + \text{rainfall\_lag1} + \text{Fourier\_cos}. \end{aligned} \tag{17}$$

where

$$\text{Fourier\_cos} = \cos\left(2\pi\frac{t}{12}\right)$$

Equations (16) and (17) are the best model for the AR-GLM with AIC value of 797.76 and standard GLM with AIC value of 839.01 obtained from the full model respectively using step function in R-studio and these were used to compare the performance.

3.3.1. *Comparision between the GLM and AR-GLM.*

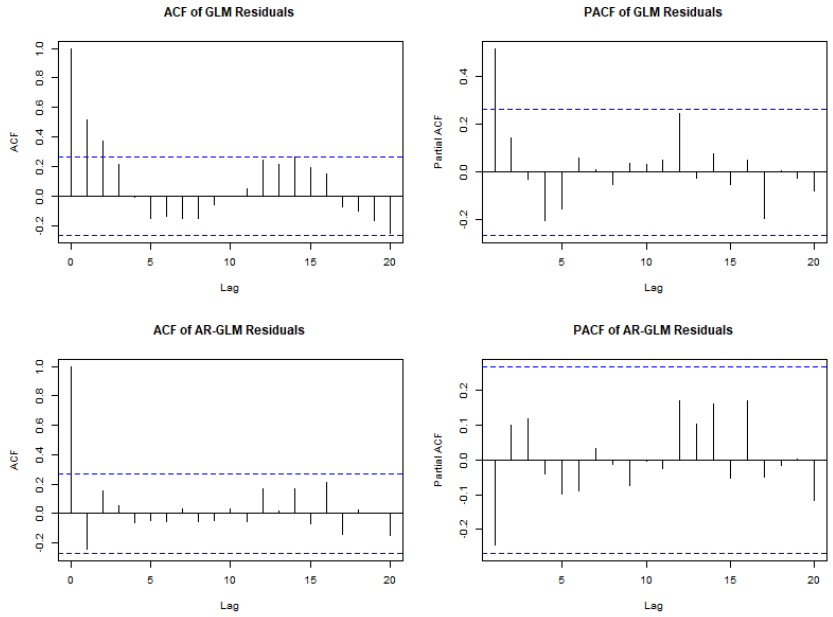

FIGURE 4. ACF and PACF Plots

From Figure 4, the ACF plot for the GLM residuals shows significant auto-correlation at multiple lags, with several spikes exceeding the confidence bounds. This indicates that the GLM model has not adequately captured the temporal dependencies in the malaria count data, leaving autocorrelation in the residuals. The PACF plot reveals significant partial autocorrelation at lag1 indicating the potential need for an autoregressive component model time-dependent patterns in the GLM residuals.

The ACF plot of the AR-GLM residuals shows a reduction in autocorrelation compared to the GLM residuals. Most spikes now fall within the confidence bounds, indicating that the inclusion of an autoregressive term in the model has addressed the temporal dependencies left by the GLM. The PACF plot also demonstrates a reduction in partial autocorrelation, with most values now within the confidence bounds. This result suggests that the autoregressive component in the AR-GLM effectively captures the autocorrelation structure in the malaria data hence the ability to produce residuals that approximate to white noise.

TABLE 4. Model Performance Comparison for GLM and AR-GLM

Model	**MAPE (%)**	**AIC**	**Ljung-Box p-value**
GLM	21.70	839.01	0.042
AR-GLM	14.37	797.76	0.994

Table 4 displays the AIC, MAPE, and Ljung-Box tests for the two models. The lower AIC value (793.23) for the AR-GLM as compared to the AIC value (808.42) for the GLM implies that the AR-GLM provides a better model fit as compared to the GLM. The AR-GLM has a MAPE of 14.37%, which is lower than the GLM's MAPE of 21.70% indicating that the AR-GLM provides more accurate predictions, with a smaller average prediction error compared to the GLM.

The Ljung-Box test shows a p-value of 0.994, which is greater than the significance level (0.05) implying that we fail to reject the null hypothesis of no auto-correlation in the residuals and conclude that there's no significant evidence of auto-correlation in the residuals.

3.3.2. *Analysis of the AR-GLM residuals.*

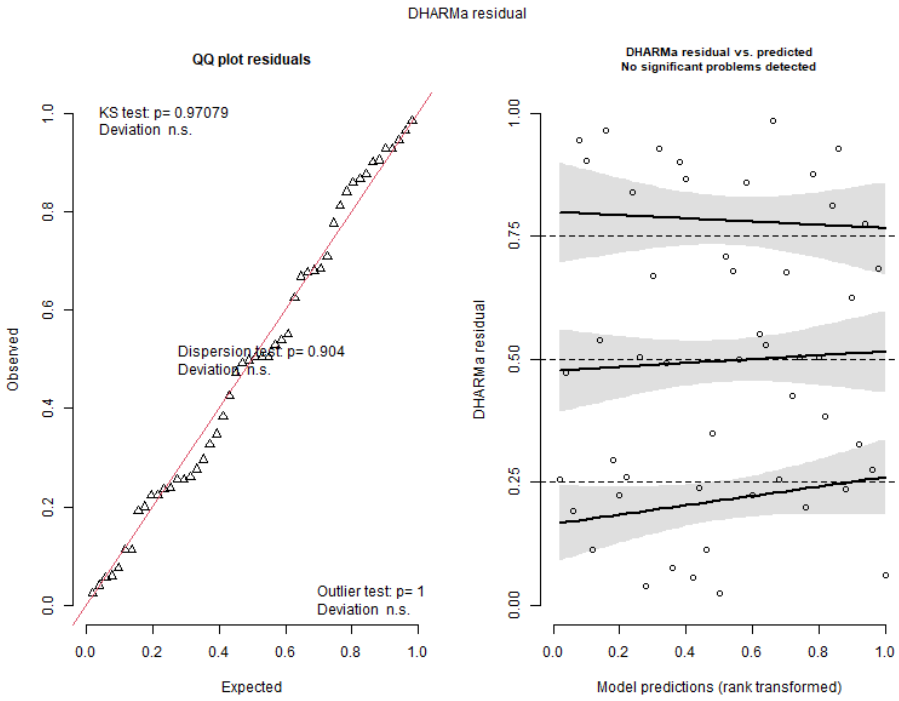

FIGURE 5. Residual Plots of AR-GLM Residuals.

From 5The points in the QQ plot generally follow the diagonal red line. This indicates that the distribution of the residuals is consistent with the theoretical distribution assumed by the model (negative binomial). The outlier test has a p-value of 1 and is labelled as ”Deviation n.s”, showing that no significant outliers detected in the data. The Kolmogorov-Smirnov test has a p-value of 0.97079, which is far greater than the significance level (typically 0.05). The ”Deviation n.s.” confirms that there is no reject the null hypothesis that the residuals come from the expected distribution. This means the residuals' distribution isn't statistically different from the theoretical distribution. The dispersion test has a p-value of 0.904, also much greater than 0.05. The ”Deviation n.s.” indicates that the model adequately accounts for the dispersion in the data, and there is no evidence of significant overdispersion or underdispersion. The QQ plot suggests that the model's distributional assumptions are reasonably well met. The residuals appear to be distributed as expected by the negative binomial distribution, and there are no strong indications of overdispersion or outliers. The residual plot indicates the residuals are generally evenly scattered around the line. The plot explicitly states, ”No significant problems detected.” This

suggests that the model residuals do not exhibit any obvious patterns heteroscedasticity, related to the predicted values. The dashed horizontal lines likely represent the mean residual value for different groups of predictions, and the gray shaded areas may represent pointwise 95% confidence intervals around those means. The fact that the points are relatively evenly distributed around these lines reinforces the conclusion that there are no major systematic patterns. This plot suggests that the model is adequately capturing the relationship between the predictors and the response variable. There are no apparent issues with non-linearity or heteroscedasticity.

4. DISCUSSIONS

From Table 3, the effect of temperature with a 4-month lag is significant and negative (Estimate = -0.1570, $p = 0.0082$), suggesting that mean temperature four months prior is associated with reduced malaria incidence. This result could reflect the detrimental effects of extreme temperatures on mosquito survival and malaria transmission. [27] discuss how temperature extremes, both high and low, can adversely affect mosquito development, reducing their population and thereby malaria transmission.

The lagged effects observed highlights the complexity of temperature's impact on malaria incidence, where both the timing and intensity of temperature variations are considered. These results are supported by the findings from [28] , who noted that temperature effects on malaria transmission are not always linear and can depend on the specific timing relative to mosquito and parasite development cycles. The effect of mean temperature (p =0.1035) was not statistically significant, suggesting that temperature at the current month does not have a direct, immediate effect on malaria incidence. The finding aligns with [28], who noted that immediate temperature changes might not strongly influence malaria if they do not persist long enough to affect mosquito and parasite development. Hence the need to consider lagged temperature other than immediate temperatures.

The results shows that rainfall with a 1-month lag has a significant positive effect on malaria incidence (estimate =0.0015,p=0.0096), indicating that rainfall from the previous month is associated with increased malaria cases. Rainfall provides essential breeding sites for mosquitoes, which subsequently increases malaria transmission in the following months. For instance, [29] found that rainfall significantly influences mosquito breeding by creating temporary water pools

that serve as habitats for mosquito larvae, thereby enhancing malaria transmission. This finding is also consistent with [29], who demonstrated that rainfall is a critical determinant of mosquito population dynamics Conversely,

There is a significant positive effect of humidity with a 2-month lag (Estimate = -0.0101, p = 0.0238), and not significant at lag 1 and current humidity. The 2-month lag suggests that higher humidity levels from the previous months could increase malaria incidence, perhaps by creating favorable conditions for mosquito activity if too high. However, These findings align with [30], who observed that varying humidity levels have complex impacts on mosquito dynamics, with both short-term and longer-term effects on malaria incidence.

The autoregressive term (Estimate = 0.7208, p= 0.001) is highly significant, indicating strong temporal dependence in malaria cases. This result suggests that previous malaria cases are a strong predictor of current malaria cases, which is a common finding in malaria modeling. Temporal dependency can occur due to sustained transmission cycles and ongoing mosquito breeding within a region.

The significant positive coefficient for Fourier cosine (estimate =0.4562,p=3.09e-05) confirms the presence of strong seasonal patterns in malaria transmission. This term captures the periodicity of malaria incidence, driven by climatic seasonality. Based on the cosine function and initial month of June, malaria incidence is expected to peak in May, June, and July. These months likely correspond to the optimal conditions for mosquito breeding and malaria transmission due to favorable climatic factors such as rainfall and temperature. In contrast, malaria incidence is expected to be lowest in December and January, when conditions are less favorable for transmission and these results are suported by figure 3. These findings corresponds with the results from [29], which show that post-rainy season months are associated with higher malaria cases due to increased mosquito breeding.

5. CONCLUSIONS

Malaria remains a serious public health problem in Arua District, Northern Uganda. The objective of this study was to model the effect of climate variables such as rainfall, temperature, humidity, and wind, for both current and lagged effects, on malaria incidences in Arua District. The significant lagged effects of mean temperature, rainfall, and humidity indicate

that these climatic factors influence malaria incidence with a delay. Specifically, higher temperatures four months prior are associated with a reduction in malaria incidence, suggesting that extreme temperatures might negatively impact mosquito survival, hence reducing malaria transmission. Similarly, the positive effect of rainfall with a one-month lag highlights the importance of rainfall in creating mosquito breeding habitats, contributing to higher malaria incidence in subsequent months.

Malaria incidence peaks in certain months, particularly in May, June, and July, likely due to the breeding conditions following the rainy season, while it decreases in September, October and December, which may correspond to drier or less favorable mosquito-breeding periods. This reflects the unique seasonality of malaria in Arua District, Uganda, which might experience peaks in malaria incidence primarily in the post-rainy season months rather than in the earlier or later months of the year. The AR term's high significance suggests that malaria incidence is strongly dependent on past values, indicating that previous cases are a strong predictor of current malaria cases. This temporal dependence might be due to sustained malaria transmission cycles within the population and the cyclical nature of mosquito breeding.

The AR-GLM outperformed the traditional GLM in capturing the autocorrelation structure in malaria incidence data, as evidenced by the AIC, MAPE, and R-squared values. The AR-GLM effectively reduces autocorrelation in residuals, suggesting it better captures the underlying data structure and provides more accurate predictions of malaria incidence. Given the seasonal peaks, interventions such as mosquito control efforts, awareness campaigns, and resource allocations (distribution of bed nets and insecticides) should be intensified in the months following the rainy season, particularly around May, to curb the anticipated rise in malaria cases. We encourage policymakers to integrate climate data into malaria control planning, enabling timely and targeted interventions based on climatic conditions to improve prediction accuracy, models for malaria should consider incorporating lagged climatic variables, especially for rainfall, temperature, and humidity, as these factors show delayed but significant effects on malaria incidence.

Despite its contributions, this study has limitations that need to be addressed in future research. Although this study primarily focused on climatic factors (temperature, rainfall, and

humidity) and temporal patterns, it did not incorporate socioenvironmental factors such as population density, land use, access to healthcare and socioeconomic status. These variables may have a critical influence on the dynamics of malaria transmission. Expanding the model to include these variables could provide a more comprehensive view of the determinants of malaria incidence. Furthermore, this study was conducted in a specific region (Arua District, Uganda) with unique climatic and environmental conditions, which may limit the generalizability of the findings to other regions with different climates or malaria transmission patterns. Expanding the study to a wider geographic area or comparing the findings between regions with various climatic conditions could provide a more robust understanding of the relationship between climate and malaria incidence.

Finally, the relatively small sample size used in this study limits the ability to detect some relationships, increases the risk of overfitting, and may prevent the capture of long-term trends and seasonal patterns. This can reduce the robustness, generalizability, and reliability of the model findings. Future research should consider collecting longer time series to increase statistical power and improve model stability.

ABBREVIATIONS

GLM Generalized Linear Model

AR-GLM Autoregressive Generalized Linear Model

NB-GLM Negative Binomial Generalized Linear Model

AIC Akaike Information Criterion

MLE Maximum Likelihood Estimation

PACF Partial Autocorrelation Function

ACF Autocorrelation Function

MAPE Mean Absolute Percentage Error

AR Autoregressive

ACKNOWLEDGEMENTS

Great thanks to the African Union through the Pan African University Institute for Basic Sciences, Technology and Innovation for the scholarship opportunity that made this work possible.

DATA AVAILABILITY

The data used in this study is included in this manuscript and any additional data may on reasonable grounds be requested from the corresponding author.

AUTHOR CONTRIBUTION

Suzan Kutegeka was responsible for conceptualization, methodology development, data analysis, manuscript drafting, and interpretation of results. Charity Wamwea and Anthony Waititu provided supervision, contributed to methodology refinement, statistical validation, and manuscript review. All authors reviewed and approved the final manuscript.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] J. Talapko, I. Škrlec, T. Alebić, M. Jukić, A. Včev, Malaria: The Past and the Present, Microorganisms 7 (2019), 179. https://doi.org/10.3390/microorganisms7060179.

[2] WHO, World Malaria Report 2023, (2023). https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023.

[3] WHO, World Malaria Report 2021, (2021). https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021.

[4] Y. Wu, Z. Qiao, N. Wang, et al. Describing Interaction Effect between Lagged Rainfalls on Malaria: An Epidemiological Study in South–West China, Malaria J. 16 (2017), 53. https://doi.org/10.1186/s12936-017-1706-2.

[5] Y.S. Kitawa, Z.G. Asfaw, Space-Time Modelling of Monthly Malaria Incidence for Seasonal Associated Drivers and Early Epidemic Detection in Southern Ethiopia, Malaria J. 22 (2023), 301. https://doi.org/10.1186/s12936-023-04742-9.

[6] P. Doumbe-Belisse, E. Kopya, C.S. Ngadjeu, et al. Urban Malaria in Sub-Saharan Africa: Dynamic of the Vectorial System and the Entomological Inoculation Rate, Malaria J. 20 (2021), 364. https://doi.org/10.1186/s12936-021-03891-z.

[7] WHO, World Malaria Report 2019, (2019). https://www.who.int/publications/i/item/9789241565721.

[8] WHO, World Malaria Report 2022, (2022). https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2022.

[9] S.J. Connor, M.C. Thomson, D.H. Molyneux, Forecasting and Prevention of Epidemic Malaria: New Perspectives on an Old Problem, Parassitologia 41 (1999), 439–448.

[10] M.G. Zalwango, L. Bulage, J.F. Zalwango, et al. Trends and Distribution of Severe Malaria Cases, Uganda, 2017–2021: Analysis of Health Management Information System Data, Quart. Epidemiol. Bull. 8 (2023), 2.

[11] Uganda Radio Network, Arua City Resorts to Using Drama As Malaria Cases Surge, (2023). https://ugandaradionetwork.net/story/arua-city-resorts-to-using-drama-as-malaria-cases-surge#google_vignette, Accessed: 2025-02-04.

[12] R.N. Katale, D.B. Gemechu, Spatio-Temporal Analysis of Malaria Incidence and Its Risk Factors in North Namibia, Malaria J. 22 (2023), 149. https://doi.org/10.1186/s12936-023-04577-4.

[13] X. Zhao, F. Chen, Z. Feng, et al. The Temporal Lagged Association between Meteorological Factors and Malaria in 30 Counties in South-West China: A Multilevel Distributed Lag Non-Linear Analysis, Malaria J. 13 (2014), 57. https://doi.org/10.1186/1475-2875-13-57.

[14] W. Gonzalez-Daza, R.J. Vivero-Gómez, M. Altamiranda-Saavedra, et al. Time Lag Effect on Malaria Transmission Dynamics in an Amazonian Colombian Municipality and Importance for Early Warning Systems, Sci. Rep. 13 (2023), 18636. https://doi.org/10.1038/s41598-023-44821-0.

[15] C.J. Armando, J. Rocklöv, M. Sidat, et al. Climate Variability, Socio-Economic Conditions and Vulnerability to Malaria Infections in Mozambique 2016–2018: A Spatial Temporal Analysis, Front. Public Health 11 (2023), 1162535. https://doi.org/10.3389/fpubh.2023.1162535.

[16] M.A. Benjamin, R.A. Rigby, D.M. Stasinopoulos, Generalized Autoregressive Moving Average Models, J. Amer. Stat. Assoc. 98 (2003), 214–223. https://doi.org/10.1198/016214503388619238.

[17] L. Yang, G. Qin, N. Zhao, et al. Using a Generalized Additive Model with Autoregressive Terms to Study the Effects of Daily Temperature on Mortality, BMC Med. Res. Methodol. 12 (2012), 165. https://doi.org/10.1186/1471-2288-12-165.

[18] J.A. Nelder, R.W.M. Wedderburn, Generalized Linear Models, J. R. Stat. Soc. Ser. A. 135 (1972), 370. https://doi.org/10.2307/2344614.

[19] A. Eunice, Statistical Modeling of Malaria Incidences in Apac District, Uganda, Ph.D. Thesis, Jomo Kenyatta University of Agriculture and Technology, (2018).

[20] A.J. Dobson, A.G. Barnett, An Introduction to Generalized Linear Models, Chapman & Hall/CRC, 2018.

[21] P. McCullagh, Generalized Linear Models, Routledge, 2019.

[22] T.Z. Nigussie, T.T. Zewotir, E.K. Muluneh, Seasonal and Spatial Variations of Malaria Transmissions in Northwest Ethiopia: Evaluating Climate and Environmental Effects Using Generalized Additive Model, Heliyon 9 (2023), e15252. https://doi.org/10.1016/j.heliyon.2023.e15252.

[23] G. Zhou, N. Minakawa, A.K. Githeko, G. Yan, Climate Variability and Malaria Epidemics in the Highlands of East Africa, Trends Parasitol. 21 (2005), 54–56. https://doi.org/10.1016/j.pt.2004.11.002.

[24] P. Bi, S. Tong, K. Donald, K.A. Parton, J. Ni, Climate Variability and Transmission of Japanese Encephalitis in Eastern China, Vector-Borne Zoonotic Dis. 3 (2003), 111–115. https://doi.org/10.1089/153036603768395807.

[25] M.C. Thomson, S.J. Mason, T. Phindela, S.J. Connor, Use of Rainfall and Sea Surface Temperature Monitoring for Malaria Early Warning in Botswana, Amer. J. Trop. Med. Hyg. 73 (2005), 214–221. https://doi.org/10.4269/ajtmh.2005.73.214.

[26] H.D. Teklehaimanot, M. Lipsitch, A. Teklehaimanot, J. Schwartz, Weather-Based Prediction of Plasmodium Falciparum Malaria in Epidemic-Prone Regions of Ethiopia I. Patterns of Lagged Weather Effects Reflect Biological Mechanisms, Malaria J. 3 (2004), 41. https://doi.org/10.1186/1475-2875-3-41.

[27] E.A. Mordecai, K.P. Paaijmans, L.R. Johnson, et al. Optimal Temperature for Malaria Transmission Is Dramatically Lower than Previously Predicted, Ecol. Lett. 16 (2013), 22–30. https://doi.org/10.1111/ele.12015.

[28] K.P. Paaijmans, A.F. Read, M.B. Thomas, Understanding the Link between Malaria Risk and Climate, Proc. Nat. Acad. Sci. 106 (2009), 13844–13849. https://doi.org/10.1073/pnas.0903423106.

[29] K.P. Paaijmans, S.S. Imbahale, M.B. Thomas, W. Takken, Relevant Microclimate for Determining the Development Rate of Malaria Mosquitoes and Possible Implications of Climate Change, Malaria J. 9 (2010), 196. https://doi.org/10.1186/1475-2875-9-196.

[30] S.J. Wang, C. Lengeler, D. Mtasiwa, et al. Rapid Urban Malaria Appraisal (RUMA) II: Epidemiology of Urban Malaria in Dar Es Salaam (Tanzania), Malaria J. 5 (2006), 28. https://doi.org/10.1186/1475-2875-5-28.