

https://dspace.org/

A Bayesian model for predicting monthly fire frequency in Kenya

L Orero; E Omondi; Bernard Omolo

PUBLIC LIBRARY SCIENCE
https://hdl.handle.net/10539/40994
WOS

Downloaded from a DSpace repository. https://dspace.org/

PLOS ONE

RESEARCH ARTICLE

A Bayesian model for predicting monthly fire frequency in Kenya

Levi Orero[1], **Evans Otieno Omondi**[1,2]*, **Bernard Oguna Omolo**[1,3,4]

1 Institute of Mathematical Sciences, Strathmore University, Nairobi, Kenya, **2** African Population and Health Research Center (APHRC), Nairobi, Kenya, **3** Division of Mathematics & Computer Science, University of South Carolina – Upstate, Spartanburg, SC, United States of America, **4** School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

* evansotieno@aims.ac.za

Check for updates

OPEN ACCESS

Citation: Orero L, Omondi EO, Omolo BO (2024) A Bayesian model for predicting monthly fire frequency in Kenya. PLoS ONE 19(1): e0291800. https://doi.org/10.1371/journal.pone.0291800

Editor: Sathishkumar Veerappampalayam Easwaramoorthy, Jeonbuk National University, KOREA, REPUBLIC OF

Received: January 6, 2023

Accepted: September 6, 2023

Published: January 25, 2024

Copyright: © 2024 Orero et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The datasets generated and/or analysed during the current study are available in the GitHub repository, from the following link: https://github.com/Komondi/Modelling-fire-frequency-in-Kenya.

Funding: The authors received no direct funding for this research.

Competing interests: The authors have declared that no competing interests exist.

Abstract

This study presents a comprehensive analysis of historical fire and climatic data to estimate the monthly frequency of vegetation fires in Kenya. This work introduces a statistical model that captures the behavior of fire count data, incorporating temporal explanatory factors and emphasizing the predictive significance of maximum temperature and rainfall. By employing Bayesian approaches, the paper integrates literature information, simulation studies, and real-world data to enhance model performance and generate more precise prediction intervals that encompass actual fire counts. To forecast monthly fire occurrences aggregated from the Moderate Resolution Imaging Spectroradiometer (MODIS) data in Kenya (2000-2018), the study utilizes maximum temperature and rainfall values derived from global GeoTiff (.tif) files sourced from the WorldClim database. The evaluation of the widely used Negative Binomial (NB) model and the proposed Bayesian Negative Binomial (BNB) model reveals the superiority of the latter in accounting for seasonal patterns and long-term trends. The simulation results demonstrate that the BNB model outperforms the NB model in terms of Root Mean Square Error (RMSE), and Mean Absolute Scaled Error (MASE) on both training and testing datasets. Furthermore, when applied to real data, the Bayesian Negative Binomial model exhibits better performance on the test dataset, showcasing lower RMSE (163.22 vs. 166.67), lower MASE (1.12 vs. 1.15), and reduced bias (-2.52% vs. -2.62%) compared to the NB model. The Bayesian model also offers prediction intervals that closely align with actual predictions, indicating its flexibility in forecasting the frequency of monthly fires. These findings underscore the importance of leveraging past data to forecast the future behavior of the fire regime, thus providing valuable insights for fire control strategies in Kenya. By integrating climatic factors and employing Bayesian modeling techniques, the study contributes to the understanding and prediction of vegetation fires, ultimately supporting proactive measures in mitigating their impact.

Introduction

Fires are a natural and integral part of many ecosystems in Kenya, particularly savannas and grasslands [1–3]. Historically, these ecosystems have relied on regular, low-intensity fires that

help maintain ecosystem health and biodiversity [1, 4–7]. However, fire dynamics have shifted in in recent decades, with an alarming increase in both the frequency and intensity of these fires. These changes can be attributed to various factors including changes in land use practices, human activities, and climate change [2, 3, 8–11]. It is therefore crucial to use tools and techniques that allow prediction of fire occurrence thus allowing prevention of substantial losses [8, 12–14].

Understanding the problem of accurately predicting fire frequency based on historical fire and weather data is critical in this context [14]. By analyzing the relationship between past fire occurrences and the corresponding weather conditions, we can develop tools and techniques that enhance our ability to forecast and manage future fires. This study delves into the complexities of this issue, investigating the underlying factors that contribute to changes in fire patterns and emphasizing the importance of predictive methodologies. By shedding light on the challenges posed by evolving fire dynamics, we can pave the way for more efficient fire prevention and mitigation strategies in order to safeguard ecosystems, human lives, and valuable assets [14, 15].

The National Aeronautics and Space Administration (NASA) detects thermal activity on Earth via its satellite system called the Earth Observing System Data and Information System (EOSDIS). This system is part of the Earth Science Data program [16]. NASA's Fire Information for Resource Management System (FIRMS) has gone through several iterations in order to improve its fire products, which are used by scientists in a variety of fields [17–19]. One of the products is the Moderate Resolution Imaging Spectroradiometer (MODIS). Researchers can analyze its data and cross-reference it with other climatic variables using specific and detailed country-level datasets [20–24]. MODIS fire data has been used in numerous studies on fire frequency and distribution all over the world [18, 19, 21, 23, 25, 26]. Rainfall and temperature are two important factors that influence the occurrence and frequency of fires [27, 28]. Data on these variables at the country level are available on several websites online, though at varying aggregation levels and formats [29, 30]. The interaction of rainfall and temperature has a significant impact on the fire regime and fire probability [31], which motivated the use of these variables in this study. By aggregating these data by month, it is possible to conduct research on the frequency of fires in a specific geographical region. Fire regimes vary due to weather changes and climate patterns, and the data show excessive dispersion, implying that it is critical to consider various techniques to deal with this feature. Tested models from simulation studies are applied to real-world fire frequency data in this study to evaluate performance on a larger scale.

In modeling fire occurrences, various techniques have been employed to predict fire frequency, particularly through discrete response regression models [32]. Among these models, count response models are widely regarded as the most suitable for datasets with discrete response variables in the form of counts. Commonly utilized count models include Poisson, Negative Binomial, Zero-Inflated, Zero-Truncated, Hurdle, and Random-effects models [33]. The foundational model in this category is the Poisson regression model, which serves as the basis for other related models. Count data often exhibit a right-skewed distribution, with the variance increasing proportionally to the mean of the underlying distribution [33]. To address situations where data display over-dispersion, an alternative type of count model is recommended [33]. Considering the well-established relationship between the mean and variance in Poisson and Binomial distributions, this study extends the Poisson generalized linear model (GLM) by incorporating an additional parameter to account for over-dispersion. By leveraging historical fire and weather data and employing Bayesian techniques, this work develops a robust predictive model for fire frequency, considering seasonal variations and addressing the potential over-dispersion in the data.

Several methods have been developed to deal with over-dispersion in count data in a variety of fields [34–37]. Unique features in the data, such as Markov patterns or the ability of a model to handle over-dispersion and under-dispersion have been explored [34, 37]. This study adds to the research that has been done on the statistical modeling of count response data with spatial and temporal features. Furthermore, it uses techniques that account for over-dispersion in count regression models [33, 38–40].

It is important to note that fires, whether caused by humans or natural causes, produce thermal anomalies that are detected by Earth's satellites [17]. Specifically, humans clearing farmland or disposing of waste frequently cause massive fires that are unpredictable and difficult to predict. However, the relationship between temperature, rainfall, and fires is evident as higher temperatures and lower rainfall are typically associated with increased fire occurrence. As a result, during certain months of the year characterized by seasonal variations in rainfall and temperature, a substantial surge in the number of fires is observed [11, 27, 28, 41]. Several researchers have published their findings on the relationship between climate change, forest fires, and the impact on forests by the turn of the millennium [42]. Since then, several methods for predicting fire occurrences in various parts of the world have been developed, though they are limited to specific areas [21, 27, 41, 43–45].

While fires pose a significant threat to the environment, predictive tools are limited, and some studies have used historical fire data to build a probability-based model that indicated relative fire frequency rather than intensity [31]. Instead, count data on monthly fire frequency were used in this study. It expanded on the spatial term by incorporating climatic data specific to the fire's location. Several authors have worked on different methods for predicting fires and their likelihood of occurrence. Several studies have found geo-statistical approaches to be useful in this prediction [21, 46, 47]. Lim et al. [21] compared the accuracy of satellite and field survey data on fire occurrence using a maximum entropy (MaxEnt) model in spatial modeling of fire probability. The satellite data used contained numerous detection errors thus reducing model accuracy. Although fire probability data is not used, this paper puts their recommendation to use spatial filtering to improve model accuracy into action.

Oliveira et al. [48] investigated fire occurrence in Mediterranean Europe and used random forest and multiple regression techniques to model its spatial patterns. The authors' dependent variable in the model was fire density (number of fires per km^2). The model data only covered the main fire season (June–September) thereby limiting the scope of the study and the impact of historical data on fire regimes. The present study uses modeling on data spanning 18 years to investigate and account for the potential effects of seasonality on prediction. Given that the number of fires in a given period is count data, some researchers decided to make predictions using Poisson processes [27, 49–52]. For the period 2007–2008, [49] used the number of forest fires by forest area as their response variable. This was a 24-month period, which introduced volatility because they admitted that their models underestimated the number of fires in months 2, 3, 8, and 9 of 2009, when there was an unusually high number of fires. This study acknowledges that predictions can only be accepted if the future behaves similarly to the past, and more data improves the model's ability to make predictions.

Traditional regression techniques, on the other hand, have been used to predict fires with varying degrees of success [43]. Su et al. [53] used a geographically weighted Negative Binomial regression model, which outperformed a standard Negative Binomial regression model but did not fully account for seasonality. When there is over-dispersion (i.e., when the variance is greater than or equal to the mean), a negative binomial model is preferable [54]. Several methods are used to deal with count data features such as over-dispersion and auto-correlation. Avanzi et al. [32] modeled count processes using a Markov-modulated non-homogeneous Poisson process framework, extending the standard Markov-modulated Poisson process

(MMNPP) model by using a more flexible approach that can capture both cyclical and non-recurring trends. By relaxing the intrinsic equi-dispersion assumption of Poisson regression, flexibility in modeling count data was introduced [33]. Given previous and current data on fires and weather data, a motivating problem for these Poisson-based models and the negative binomial model is forecasting the number of satellite-detected fires in a specific geographical region.

In this paper, a Bayesian Negative Binomial model for count data is developed and extended to include an informative time component. The selected count model is a good fit for the response variable, which is fire frequency. In addition, it allows incorporation of prior knowledge to improve predictions. Furthermore, it uses Bayesian inference, which provides a better intuitive framework to quantify uncertainty unlike other methods. Furthermore, the Bayesian Negative Binomial model is tested on real-world data to demonstrate its accuracy in predicting the number of future monthly and cumulative fires caused by these climatic variables. This allows for more accurate prediction intervals that can help in allocation of resources and planning of fire response strategies.

Materials and methods

Data sources

The data used in this study were extracted from different sources and were of two types; i) fire hot-spot data and ii) temperature and rainfall data.

Fire hotspot data. MODIS data on active fire pixels' timing, location, and radiative power are stored in different formats [55]. We used the Collection 6 Level 1 Fire Products (abbreviated MCD14DL) in this case because these data sets provide image and geographic coordinates, fire pixel and mean background brightness temperatures, and fire radiative power (FRP) for each individual 1 km active fire pixel detected by MODIS [56]. The fires included were those presumed vegetation fire, active volcano, other static land source and offshore. Text files (.csv) from November 2000 to December 2018, containing fire hotspot data were downloaded and then combined to form a single file. The number of fires recorded for each month was then calculated, taking into account mean values of features such as fire radiative power, brightness, and time of day. However, only fires that were presumed to be vegetation fires caused by human activity were included in subsequent analyses.

Monthly climate data. This study made use of three monthly climatic variables: precipitation, minimum temperature, and maximum temperature. The precipitation (rainfall) and temperature data were obtained from WorldClim, a database of high spatial resolution global weather and climate data developed by the Climatic Research Unit at the University of East Anglia [57, 58]. The database contains historical data aggregated over a target temporal range of 1970–2018, based on data from 9,000 to 60,000 weather stations [57]. For each month of the year from the year 2000 to 2018, global GeoTiff (.tif) files were downloaded. The spatial resolution of the rainfall and temperature data available on the database is defined at three levels: approximately 21 km^2 at the equator; around 85 km^2; and approximately 340 km^2 [57]. This study uses the level of 2.5 minutes, which corresponds to an area of approximately 21 km^2 at the equator. These were then imported into QGIS 3.28.1 software [59], which was used to crop them to country using a Kenya .shp shape file. To identify the rainfall or temperature values for a specific fire hotspot, GPS coordinates were used as intersection points as they were available for both sets of data files. This was computed in R software version 4.1.0 using the packages raster, stars, sf, sp and stringr [60–64]. The final datasets were then written into the local directory for use in subsequent modeling. Fig 1 shows the methodological flowchart used in data collection, preprocessing and modeling framework. The variables in the rainfall and

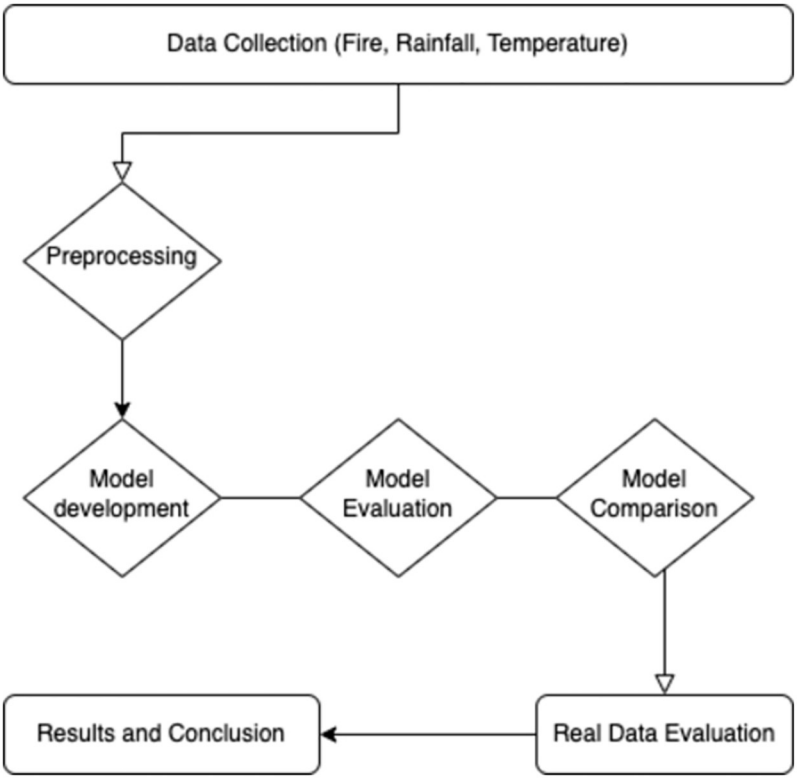


Fig 1. This methodological flowchart shows the approach used from data collection, preprocessing, modelling and presentation of results used.

https://doi.org/10.1371/journal.pone.0291800.g001

temperature datasets used in this study included month of year, year, rainfall amount in millimetres, minimum temperature and maximum temperature as variables.

While there was no explicit spatial variable in the model, it was ensured beforehand that, each fire hotspot was intersected with the corresponding temperature and rainfall value in that area at that time. This made the variables spatially valid.

Simulation study

This simulation study aimed to evaluate the performance of two statistical models using simulated data sets with defined and different sample sizes in the presence of varying degrees of over-dispersion. The simulation study consisted of a total of 16 scenarios. The variables were simulated on $n = 60, 120, 240, 360$, which represented 5 years, 10 years, 20 years, and 30 years worth of monthly data. This study simulated 1000 datasets for each combination of parameters using R software version 4.1.0 and set the seed at "76568".

To achieve the simulation of a dataset that was similar in characteristics to the real data, several steps were taken. First, the maximum temperature variable was assumed to follow a truncated normal distribution, given that the upper and lower bounds are known from the real data. Second, assuming a linear relationship between rainfall and temperature, a model was fitted on the two variables, and rainfall values were simulated to include random error terms. Lastly, the two simulated variables were fitted on to a standard negative binomial model with a specified initial θ to simulate a third response variable of the fire frequency, which is count data.

The mathematical formulation of a random number variable assumed to follow a truncated normal distribution. To generate random numbers from a truncated normal distribution, the lower and upper bounds to the range of possible values for x of the distribution, denoted by a and b, respectively are defined.

In this case, the lower bound a was 23.43 and the upper bound b was 34.82. The mean of the variable μ was 29.18 and its standard deviation σ was 2.29. This was implemented in the `TruncatedNormal` package in R.

To simulate the rainfall variable, initial mapping of the data showed a linear relationship between the two variables. To simulate rainfall data with a similar relationship with temperature, a simple linear regression model was fitted to the real data as as described in Eq (1).

$$\text{Rainfall} = \beta_0 + \beta_1 \text{Maximum temperature} + \varepsilon, \tag{1}$$

with the resulting model given by

$$\text{Rainfall} = 312.692 - 7.83 \times \text{Maximum temperature} + \varepsilon$$

To simulate rainfall values using the maximum temperature values, residuals of the model were calculated, then residual bounds were created with minimum and maximum residuals defined. Next, errors were simulated and assumed to follow a uniform distribution. Then, simulated values of rainfall were created by adding the simulated errors to the values of rainfall predicted by the regression model. Lastly, to simulate the fire counts, the two simulated variables were used as new data in predictions by a standard negative binomial fitted on the real data. The resulting model is given in (2).

$$y \sim \text{NegBinomial}(\mu = \phi, \alpha = \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)) \tag{2}$$

where y is the response variable and μ and α are the mean and dispersion parameters of the negative binomial model. Similar to the previous simulation, the residuals of the Negative Binomial model were calculated, residual bounds defined, errors simulated and then added to the values predicted by the model. These steps were implemented in a user-defined function in R that took four arguments; n = number of values to be simulated, theta = dispersion parameter in the NB model, mod = linear regression model object and data = real dataset. The output was a data frame. The synthetic covariate and response data were generated using a user-defined function, while varying sample size n and dispersion parameter θ. The subsequent analysis sought the best model for predicting the discrete count variable in the presence of an informative time component and over-dispersion.

This study evaluated two forms of the negative binomial models: a standard Negative Binomial model and a Bayesian Negative Binomial model. The simulation studies were used to evaluate the performance of the two statistical models [65]. Data were simulated, exported into Excel .csv files and then the two models were fitted and their performance evaluated. Because the data were time series, each dataset was split into a training and a testing dataset in a non-random and sequential 80:20 ratio. This means that in a dataset, the first 80 percent of the observations become the training set and the subsequent 20 percent are the testing set. Then, each model was fitted to the training dataset, and its predictive performance on both the training and test datasets was evaluated. Several metrics were used to assess model performance. Given the time series nature of the data, these included the Root Mean Square Error (RMSE), percent bias, and Mean Absolute Scaled Error (MASE). The model performance metrics were calculated using the `Metrics` package [66].

Statistical models and estimation

Standard negative binomial model. The gamma mixture of the Poisson distributions is called the *negative binomial distribution* for y. Its probability mass function is

$$p(y;\mu,\phi)=\frac{\Gamma(y+\phi)}{\Gamma(\phi)\Gamma(y+1)}\left(\frac{\mu}{\mu+\phi}\right)^{y}\left(\frac{\phi}{\mu+\phi}\right)^{\phi},\qquad y=0,1,2,\ldots \tag{3}$$

The fire hotspots counts Y_i were modelled using a negative binomial(NB) distribution in this case. The NB distribution has two parameters, the mean μ and a dispersion parameter ϕ:

$$Y_i \sim \mathrm{NB}(\mu_i,\phi) \tag{4}$$

Using a logarithmic link function, the mean μ_i is specified as:

$$\log(\mu_i)=\beta_0+\beta_1(\mathrm{Rainfall}_i)+\beta_2(\mathrm{Maximum\ \ temperature}_i)+\Theta_i+\varepsilon_i$$

Θ_i is an collection of two pseudo covariates that make up an informative component:

$$\Theta_i=\underbrace{\beta_3\cos\left(\frac{2p\pi}{\mathrm{time}_i}\right)+\beta_4\sin\left(\frac{2p\pi}{\mathrm{time}_i}\right)}_{\mathrm{sine-cosine\ \ function}},\ \text{where time}_i=1,2,\cdots,n\ \text{and p}\ =12$$

The raw real world data are often likely to be dominated by seasonal patterns and long term trends. To control for these patterns in the models, we add Θ_i, which is some function of time, into the models [67, 68]. The error terms are assumed to be normally distributed:

$$\varepsilon=(\varepsilon_1,\cdots,\varepsilon_N)\sim N(0,\Sigma)$$

The additional covariates Θ_i were included in the NB implementation on simulated data and were added during model fitting. To complete the description of the NB model, we specify the distribution of the dispersion parameter ϕ:

$$\phi \sim \mathit{Gamma}(r_j,d_j)$$

which in general means that when $X \sim \mathit{Gamma}(r,d)$, the probability density $f(x)$ is proportional to $x^{r-1}\exp(-dx)$. This model was implemented in R software using the glm.nb function in the MASS package [69]. The function performed a maximum likelihood estimation of the generalized Negative Binomial linear model.

Bayesian negative binomial model. In this type of model, at each loop stage of the estimation process, the respective prior distribution updates the posterior distribution for each predictor [70]. The posterior distribution in (3) is updated by multiplying the likelihood by the prior:

$$p(\theta|y)\propto L(\theta)\pi(\theta) \tag{5}$$

$p(\theta|y)$ is the posterior distribution that explains the predictors; $L(\theta)$ is the likelihood function and $\pi(\theta)$ is the prior distribution [70]. Assuming the Negative Binomial characteristics of the previous model in (3), the Bayesian Negative Binomial model takes the form described in Eq (6).

$$y_i \sim \mathrm{NB}(\mu_i,v_i). \tag{6}$$

The inverse link function can be expressed as

$$\mu_i=\ \exp(x_i'\alpha)\mu_i=\ \exp(\alpha_0+\alpha_1x_1+\alpha_2x_2+\alpha_3x_3+\alpha_4x_4), \tag{7}$$

where α_0, α_1, α_2 α_3, and α_3 are the model parameters and ϕ is the dispersion parameter. Using a logarithmic link function, the mean μ_i becomes

$$\log(\mu_i) = \alpha_0 + \alpha_1(\text{Rainfall}_i) + \alpha_2(\text{Maximum temperature}_i) + \Theta_i + u_i, \tag{8}$$

where Θ_i is an collection of two pseudo covariates that make up an informative time component, similar to the NB model in the previous section given by

$$\Theta_i = \underbrace{\beta_3 \cos\left(\frac{2p\pi}{\text{time}_i}\right) + \beta_4 \sin\left(\frac{2p\pi}{\text{time}_i}\right)}_{\text{sine-cosine function}}, \text{ where } \text{time}_i = 1, 2, \cdots, n \text{ and } \text{p} = 12 \tag{9}$$

The informative time component made up of the time term and the sine-cosine function is important to account for the seasonality and cyclic characteristics of the data [67, 68]. Then, the distribution of the dispersion parameter v_i:

$$v_i \sim \mathcal{G}amma(f_j, r_j) \tag{10}$$

which in general means that when $X \sim \mathcal{G}amma(f, r)$, the probability density $f(x)$ is proportional to $x^{f-1}\exp(-rx)$. The covariate coefficients α_0, α_1, α_2, α_3, α_4 and α_5 have the following prior distributions:

$$\alpha_m \sim N\left(0, \frac{1}{\tau\alpha_m}\right), \quad m = 0, 1, 2, 3, 4. \tag{11}$$

This represents a weak prior belief about the values of the model parameters. The prior for the dispersion parameter v_i is a Gamma distribution with small shape and rate parameters, representing a vague prior belief about the value of v_i. This model was implemented in R software using the stanglm.nb function in the `rstanarm` package [71]. The function performed a full Bayesian estimation of the generalized Negative Binomial linear model via Markov Chain Monte Carlo sampling algorithm. The outcome's posterior predictive distribution was used as the source of a single draw of the predictions used to calculate model performance.

Ethical approval

The data used did not have any personal identification data and the Strathmore University Ethics Review Committee approval was obtained.

Results and discussion

Simulation results

Simulations were carried out under different sample sizes (n) and varying degrees of the dispersion parameter (θ). The sample sizes used were 60, 120, 240 and 360 months while the θ values were 1.5, 5, 10 and 100. For each combination, the model performance of the Negative Binomial (NB) and the Bayesian Negative Binomial (BNB) are analysed and represented. The metrics used are bias on the testing datasets, Mean Absolute Scale Error (MASE) on the testing datasets, and Root Mean Squared Error (RMSE) on both the testing and training datasets. Table 1 shows the results obtained.

According to results in Table 1, the negative binomial mostly outperformed the BNB model in terms of the model percent bias, while the Bayesian negative binomial was a better model in terms of the MASE on the testing dataset, RMSE on the training datasets (rmse1), and RMSE on the testing datasets(rmse2). For instance, at $\theta = 1.5$ and $n = 60$, the bias was -4.56 for the

Table 1. Model performance metrics of the simulated number of datasets for each dispersion and sample size combination.

θ	n	BNB_{bias}	NB_{bias}	BNB_{mase}	NB_{mase}	BNB_{rmse1}	NB_{rmse1}	BNB_{rmse2}	NB_{rmse2}
1.5	60	-0.404	-4.555	0.037	0.051	4.083	29.142	4.373	5.988
1.5	120	0.899	-11.245	0.032	0.144	3.791	30.111	2.825	11.729
1.5	240	0.803	-10.492	0.029	0.171	3.889	35.434	2.070	10.969
1.5	360	0.713	-9.942	0.029	0.176	3.900	36.598	1.929	10.405
5.0	60	-0.405	-4.555	0.037	0.051	4.084	29.142	4.374	5.988
5.0	120	0.753	-11.498	0.030	0.147	3.751	30.260	2.675	11.978
5.0	240	0.817	-10.441	0.029	0.172	3.862	35.233	2.089	10.917
5.0	360	0.738	-9.901	0.029	0.175	3.863	36.464	1.930	10.369
10.0	60	-0.403	-4.554	0.037	0.051	4.083	29.142	4.376	5.988
10.0	120	0.899	-11.245	0.032	0.144	3.790	30.111	2.825	11.729
10.0	240	0.803	-10.492	0.029	0.171	3.888	35.435	2.071	10.969
10.0	360	0.714	-9.942	0.029	0.176	3.901	36.598	1.929	10.405
100.0	60	-0.405	-4.554	0.037	0.051	4.084	29.142	4.374	5.988
100.0	120	0.753	-11.498	0.030	0.147	3.752	30.260	2.674	11.979
100.0	240	0.004	-10.441	0.721	0.172	31.841	35.233	32.536	10.917
100.0	360	0.739	-9.819	0.029	0.174	3.840	36.161	1.940	10.287

https://doi.org/10.1371/journal.pone.0291800.t001

negative binomial model, which was lower compared to the bias of -0.40 for the Bayesian negative binomial model. The trend was similar at $\theta = 5$ and $n = 360$, where the negative binomial model had a lower bias of -9.90. Further, the same is the case when sample is of $\theta = 100$ and sample size n is 240 where the bias of the Bayesian negative binomial model (bias = 0.004) is higher than that of the negative binomial model under the same conditions (bias = -10.441). Looking at the MASE,Table 1 shows how the values are lower for the Bayesian negative binomial model. For instance, at $\theta = 1.5$ and $n = 60$, the MASE is 0.037 for the BNB model and 0.051 for the negative binomial model. When at $\theta = 10$ and $n = 360$, the BNB model outperforms the NB model with a lower value of MASE = 0.029. For the RMSE on training data, when $\theta = 1.5$ and $n = 60$, the value of the metric is lower for the BNB model (4.083) suggesting that it is a better model. This is similar across the board for the different sample sizes and values of θ. Finally, this is the same for the RMSE on testing data as seen in the 9th column of 1, where RMSE is 4.373 for the BNB model and 5.988 for the NB model when $\theta = 1.5$ and $n = 60$.

Overall, the BNB model becomes better (lower error) when sample size (n) increases, suggesting that when more training data is available, the predictive performance improves. This is is also visible in the distribution as shown by the violin plots in Fig 2. For all the metrics, it is expected that a lower metric value is representative of better predictive performance. The length of the violin plot represents the range of values, while the width indicates the density of the data at different levels along the y-axis. For the plot a in Fig 2, the BNB violin plots appear higher than the negative binomial model plots, suggesting that there is a greater concentration or frequency of bias data points in that region and thus, worse performance. This is different for the other three plots b, c and d, in which the BNB model shows lower error values and thus better performance.

Results on actual data

Descriptive statistics. Table 2 shows the summary statistics of the data and the selected variables where the sample size is 218. The overall minimum number of fires over the period (2000–2018) was 10 while the maximum number was 1661. In the period of 218 months, the

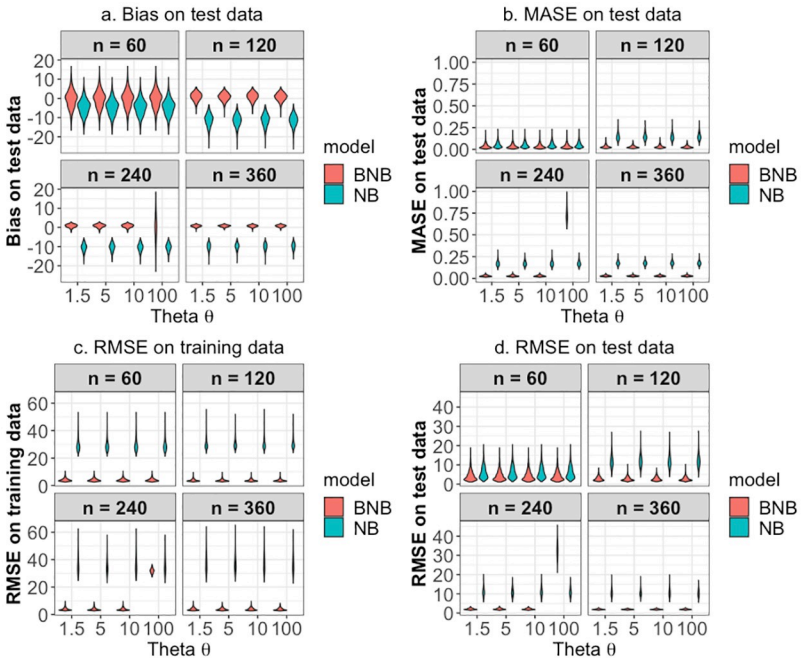


Fig 2. Violin plot of the four model performance metrics in the simulation study. Panel **a** shows the model results by bias on test data, Panel **b** shows the model results by MASE on test data, Panel **c** shows the model results by RMSE on training data and Panel **d** shows the model results by RMSE on test data.

https://doi.org/10.1371/journal.pone.0291800.g002

mean and median are 277.13 and 155.00, respectively, with a standard deviation of 304.91. On the other hand, the mean and median mean maximum temperatures were 29.19 and 29.29 respectively. For the mean minimum temperature, the mean was 18.17 degrees Celsius with a standard deviation of 2.14. Lastly, the mean amount of rainfall was 84.04mm with a standard deviation (SD) of 51.69.

Fig 3 shows the trend and patterns observed on the data and its four variables between the year 2000 and 2018. In Panel (a) of Fig 3, there are irregular patterns with the number of fires being high at certain times of the years and low in others. The highest number of fires was above 1500 in a month. In In Panel (b) of Fig 3, there appears to be a similar irregularity and there average monthly maximum temperatures have been rising over time. In Panel (c) of Fig 3, there is a similar trend as described in Panel (b) of Fig 3, which suggests that on average, minimum temperatures have been rising in the period. Lastly, the results suggest that there seem to be an increase in the average rainfall from 2015 as shown in Panel (d) of Fig 3. Furthermore, the results suggest that there are visible fluctuations that can be expected to happen within a calendar year. The irregular patterns in each graph show the presence of seasonality

Table 2. Summary of real world dataset variables (n = 218).

Variable	Minimum	Maximum	Median	Mean	SD
Fire frequency	10.00	1661.00	155.00	277.13	304.91
Maximum temperature	23.44	34.82	29.27	29.19	2.30
Minimum temperature	11.26	22.34	18.36	18.17	2.14
Rainfall	8.43	297.86	70.66	84.04	51.69

https://doi.org/10.1371/journal.pone.0291800.t002

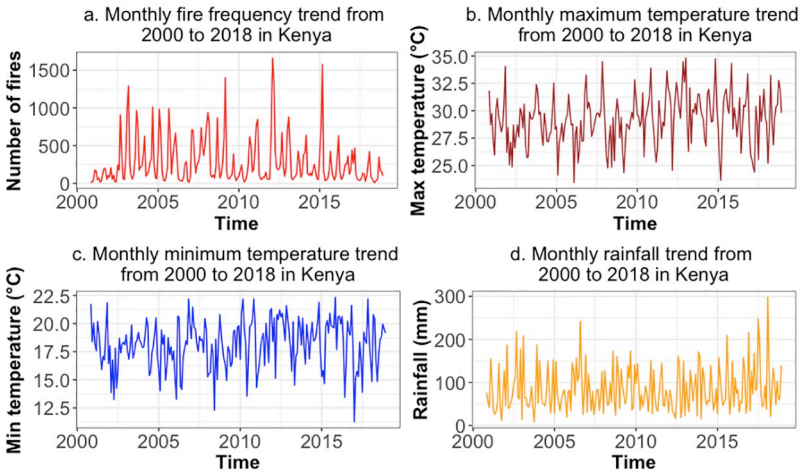


Fig 3. Time series of the four variables in the study from the real world data. Panel **a** shows the time series of monthly fire frequency, Panel **b** shows the monthly maximum temperature, Panel **c** shows the monthly minimum temperature and Panel **d** shows the monthly rainfall amounts.

https://doi.org/10.1371/journal.pone.0291800.g003

in the data, which is to be expected as fires, rainfall and temperature change with season of the year.

The relationship between maximum temperature and fire frequency was irregular (See S1 Fig). In March, the two variables show a strong inverse relationship and this suggests that an increase in maximum temperature is associated with fewer fire incidences. On the other hand, this is an opposite behaviour in January, February, May, June, October and December. In these months, an increase in maximum temperature is associated with more fire incidences. The other months also have specific trends though suggest weak relationships. The relationship between rainfall and fire frequency was also irregular by month (See S2 Fig). Here, there seems to be a pattern suggesting an inverse relationship between the two variables. This means that for all months except September and December, more rainfall is associated with fewer vegetation fires. To get more information from the data on the relationship between these variables, we introduced a dummy variable called era, which took two values: 00's if data were from the 2000–2009 period and 10's if the data were from the 2010–2018 period. The relationship between maximum temperature and fire frequency is shown in Fig 4. Here, there seems to be a pattern suggesting an inverse relationship between the two variables in the 00's and a direct relationship between the two variables in the 10's period. This suggests that since the 10's higher maximum temperature is associated with more vegetation fires, contrary to the previous period. The relationship between rainfall and fire frequency is shown in Fig 5. Here, there seems to be a pattern suggesting a direct relationship between the two variables in the 00's and an inverse relationship between the two variables in the 10's period. This suggests that since the 10's more rainfall is associated with fewer vegetation fires, contrary to the previous period.

Model performance on actual data. Both models were fitted on real world data and the results evaluated as shown in Table 3. The results show that the BNB outperformed the NB on the RMSE on training set, RMSE on the testing set, MASE on the testing set and the percentage bias on the testing set as shown in Table 3.

For the training set, the Bayesian Negative Binomial model achieved a slightly better performance (RMSE = 307.58) as compared to the traditional Negative Binomial regression

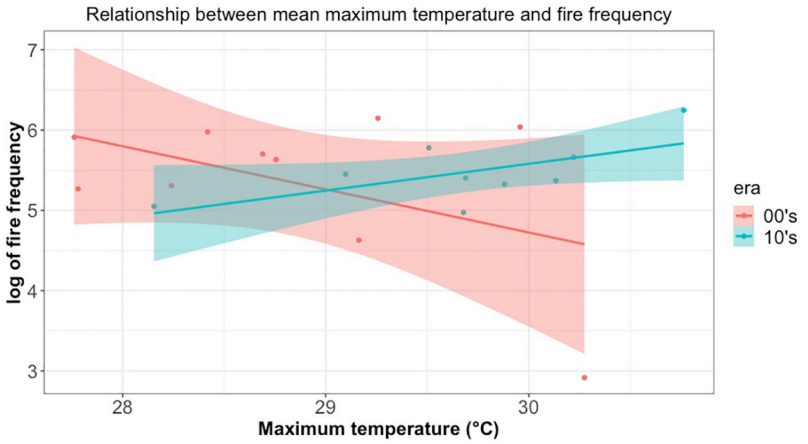


Fig 4. Scatterplot of relationships between mean annual maximum temperature and fire frequency. Graphs show different relationship for each era suggesting the difference in time periods and effect of other variables.

https://doi.org/10.1371/journal.pone.0291800.g004

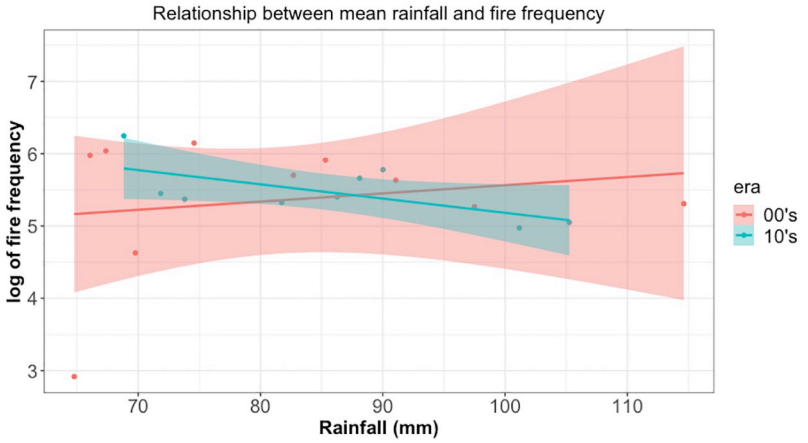


Fig 5. Scatterplot of relationships between mean annual rainfall and fire frequency. Graphs show different relationship for each era suggesting the difference in time periods and effect of other variables.

https://doi.org/10.1371/journal.pone.0291800.g005

(RMSE = 308.88). This superiority of the Bayesian approach was also evident in the testing set, where the Bayesian Negative Binomial model resulted in an RMSE of 163.22, whereas the Negative Binomial model exhibited an RMSE of 166.67. On the testing set, both models exhibited closely comparable performance, with MASE values of 1.15 and 1.12 for the Negative Binomial and Bayesian Negative Binomial models, respectively. The bias in predictions on the testing set

Table 3. Model performance metrics on data with n = 218.

Metric	Negative Binomial	Bayesian Negative Binomial
RMSE on training set	308.88	307.58
RMSE on testing set	166.67	163.22
MASE on testing set	1.15	1.12
Bias on testing set	-2.62%	-2.52%

https://doi.org/10.1371/journal.pone.0291800.t003

Table 4. BNB model beta estimates and credible intervals on data.

Term	Estimate	Standard error	CI-2.5%	CI-97.5%
(Intercept)	7.69	1.55	4.00	10.19
Max temperature	-0.09	0.05	-0.18	-0.00
Mean rainfall	0.002	0.002	-0.002	0.006
Sine function	0.83	0.19	0.43	1.19
Cosine function	-0.05	0.17	-0.40	0.27

https://doi.org/10.1371/journal.pone.0291800.t004

was found to be slightly less for the Bayesian Negative Binomial model (-2.52%) in comparison to the traditional Negative Binomial model (-2.62%).

Table 4 summarizes the beta estimates of the Bayesian Negative Binomial Model and associated 95% credible intervals for the predictors applied to the data. The estimated baseline frequency, when all other predictors are zero, is 7.69, with the associated 95% credible intervals extending from 4 to 10.19. In summary, the BNB model indicates significant effects of the max temperature and sine function on the response variable. The influences of mean rainfall and the cosine function, however, are statistically uncertain suggesting possibility of interaction terms and unexplained variance from other variables.

To better understand the significance of prediction intervals, the NB and BNB models were used to calculate prediction intervals on the test dataset. Fig 6 depicts these outcomes. It can be seen that the actual values are mostly within the prediction interval with the deviation from the upper prediction intervals being slightly larger for the NB model. While both models seem to have a similar pattern and trend, the predicted values of the BNB seems to be a better fit as they have a larger variance. It can be said that if sampling process was repeated indefinitely, 95 percent of the predicted intervals would include the new observation. The correlation between the BNB's observed fire counts in the test data and predicted counts was not statistically significant, suggesting a weak non-linear relationship, $r = -0.07$, $t(42) = -0.45$. On the other hand, the correlation between the NB model's observed fire count in the test data and and the predicted values was statistically significant, indicating a moderate positive relationship, $r = 0.32$, $t(42) = 2.19$. This indicates that the relationship between the observed and predicted values is not well captured by a linear association.

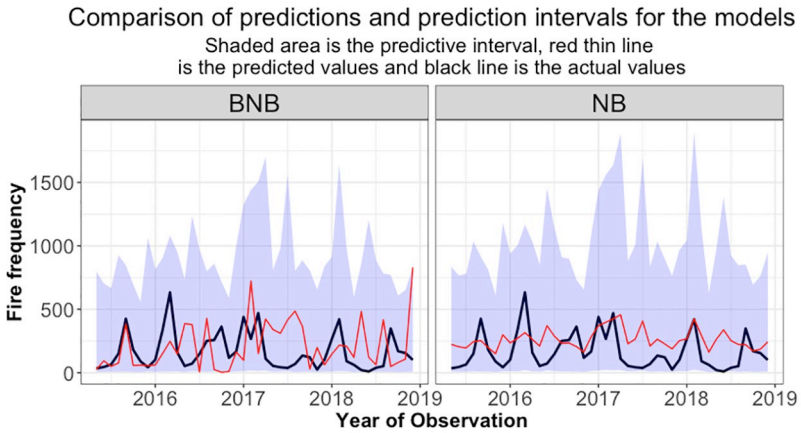


Fig 6. The actual fire counts, the predicted values and the prediction intervals of the NB and BNB models on testing data. Desired probability is 0.9.

https://doi.org/10.1371/journal.pone.0291800.g006

The seasonal cycle is visible in the time series plots of the four variables of fire frequency, minimum temperature, maximum temperature, and rainfall. The simulation results show that both models perform well on overdispersed data, with the BNB model outperforming the Negative Binomial (NB) in terms of Mean Absolute Scaled Error (MASE), Root Mean Squared Error (RMSE) on testing sets, and RMSE on training sets. Furthermore, when applied to real data, the BNB model outperformed the NB model for all the four metrics. While the simulations were designed to handle some of the complexities found in real-world data, they clearly did not cover all potential complications.

However, it should be acknowledged that while the simulations addressed some complexities in real-world data, they did not encompass all potential complications. Real-world environmental variables may possess diverse autocorrelation structures, which the BNB model seems to handle better than the NB model. The empirical findings are limited to the specific case studied, and future research should explore and address data complexity while confirming the results through validation with real data, employing more appropriate simulation methods.

The results show how different periods in a time series can have different patterns, making any underlying relationships difficult to establish. For instance, the relationships between fire frequency and rainfall or temperature vary by month, potentially influenced by climatic factors, farming seasons, or human activities. Moreover, variations are observed when analyzing the data by decade or time period. Notably, there has been a shift in the relationship from 2010 to 2018, suggesting an increasing impact of climate change, particularly in higher temperatures, which are associated with more fires. The BNB model also provided prediction intervals that were close to the actual predicted values. This feature is valuable as large prediction intervals without nearby real data would lack practical significance. In the real data, MASE values greater than 1 for both models indicate that the actual forecasts performed worse than a superior time series model in terms of mean absolute error. Achieving MASE values below 1 may have been challenging due to structural characteristics of the data.

In the context of count data, a discrete probability distribution like Poisson or Negative Binomial is the most appropriate representation to avoid bias and misspecification in the model [72]. Given the time-related nature of the data, time-trend analysis plays a crucial role in understanding system dynamics and changes. The authors acknowledge the complexity of predicting the number of fires due to various influencing factors [73]. However, due to the randomness of the number of fire occurrences each month, application of the Bayesian model was more appropriate [74, 75]. The value offered by the Bayesian technique was the assumption that the parameter of interest is random, and that it lies within the prediction intervals [74]. In many instances, the BNB model generated results similar to that of the frequentist NB model [74, 75]. The intention was to allow for more precise predictions through this method as reported by [76], which the results in this study have achieved.

A problem this study sought to solve was on the time dimensions of data used to predict fire frequency. Unlike the shorter time spans used in [48–50], this study simulated data under different scenarios and thus had sample sizes of 60, 120, 240 or 360 months. As is utilised in literature, the application of the Negative Binomial distribution as in [43, 53, 54] allowed better modeling of over-dispersed count data. To tackle the issue of under-predictions as in [49], the implementation of the Bayesian Negative Binomial model allowed for the calculation of better prediction intervals. This is coming up as a useful technique as presented by [77, 78] who acknowledge the importance of Bayesian inference in fire prediction. This study contributes to understanding fire frequency prediction, highlighting the superiority of the Bayesian Negative Binomial model in handling overdispersed data and providing reliable prediction intervals. However, further research is necessary to explore and account for data complexity, validate the findings with real data, and refine simulation methods.

Conclusion

This study utilized the assumption that the occurrence of fire incidents followed a Poisson process. Although the results are intriguing, they do not provide practical insight into predicting the timing of future events. Additionally, the absence of outliers in the simulation runs is inconsistent with real-world systems, where outliers would be expected. Nevertheless, the fitted models offer valuable guidance on the behavior of real-world variables. It is important to note that the fire regime system is complex, influenced by various environmental and human factors, and cannot be entirely explained by the variables used in this study. It is also noted that although historical data shows variation in fire behaviour over two time periods analysed (2000–2009 (00) and 2010–2018 (10)), these have not been included in the simulations.

To enhance the comprehensiveness of this study, several avenues for expansion can be pursued. First, additional variables that may impact the fire regime could be incorporated. Furthermore, the inclusion of outliers in the simulated datasets would better reflect real-world scenarios. Additionally, the proposed modeling framework can be extended to include predictions of monthly fire frequency for shorter time horizons (for example, less than 6 months ahead). Refining the proposed models to optimize performance metrics such as MASE and RMSE on the testing datasets would also be beneficial. Incorporating polynomials into the model specification might offer a way to improve the models.

In conclusion, this study highlights the potential for further statistical modeling of count response data with spatial and temporal features. The authors recommend the use of the Bayesian Negative Binomial model, which accounts for and addresses overdispersion, as the preferred model for real data. Although not exhaustive, this work provides valuable insights into the behavior of models under different conditions, leading to more comprehensive predictions. Ultimately, these predictions can contribute to effective fire planning and mitigation efforts in Kenya.

Supporting information

S1 Fig. The relationship between maximum temperature and fire frequency. Scatter-plots of monthly relationships between maximum temperature and fire frequency. Graphs show different relationship for each month suggesting the seasonality and effect of unknown variables.
(TIF)

S2 Fig. The relationship between rainfall and fire frequency. Scatter-plots of monthly relationships between mean rainfall and fire frequency. Graphs show different relationship for each month suggesting the seasonality and effect of unknown variables.
(TIF)

Acknowledgments

The authors acknowledges, with thanks, the support of Strathmore University, Institute of Mathematical Sciences for the production of this manuscript.

Author Contributions

Conceptualization: Levi Orero, Bernard Oguna Omolo.

Data curation: Levi Orero.

Formal analysis: Levi Orero, Evans Otieno Omondi.

Methodology: Evans Otieno Omondi, Bernard Oguna Omolo.

Project administration: Levi Orero.

Supervision: Evans Otieno Omondi, Bernard Oguna Omolo.

Validation: Levi Orero.

Visualization: Levi Orero.

Writing – original draft: Levi Orero, Bernard Oguna Omolo.

Writing – review & editing: Levi Orero, Evans Otieno Omondi, Bernard Oguna Omolo.

References

1. Poletti Claudio, Dioszegi Gergo, Nyongesa Kevin Wafula, Vacik Harald, Barbujani Marco, and Kigomo John Ngugi. Characterization of forest fires to support monitoring and management of mount kenya forest. Mountain Research and Development. 2019; 39 (3): R1–R12. https://doi.org/10.1659/MRD-JOURNAL-D-18-00104.1
2. Faith N Wangari. The effects of fires on plants and wildlife species diversity and soil physical and chemical properties at Aberdare Ranges, Kenya. PhD thesis, University Of Nairobi. 2016.
3. Wafuta Patrick Wasike, Sirmah Peter Kipkosgei, and Mbira George Kere. Forest fires incidences, causes, and associated losses: A case study of londiani forest in kenya. East African Journal of Forestry and Agroforestry. 2019; 1(2):32–41.
4. Palumbo Ilaria. The importance of fire ecology in protected areas management. Developments in Earth Surface Processes, Elsevier. 2013; 1(16):181–191. https://doi.org/10.1016/B978-0-444-59559-1.00014-1
5. Kamau Peter Ngugi and Medley Kimberly E. Anthropogenic fires and local livelihoods at Chyulu hills, Kenya. Landscape and Urban Planning. 2014; 124:76–84. https://doi.org/10.1016/j.landurbplan.2014.01.010
6. Kinyanjui James Mwangi, Karachi Moses, and Ondimu Kennedy Nyambuti. Natural regeneration and ecological recovery in mau forest complex, Kenya. Open Journal of Ecology. 2013; 3(6):417. https://doi.org/10.4236/oje.2013.36047
7. Vehrs Hauke-Peter and Heller Gereon Romanus. Fauna, fire, and farming: landscape formation over the past 200 years in pastoral east Pokot, Kenya. Human Ecology. 2017; 45:613–625. https://doi.org/10.1007/s10745-017-9926-1
8. Henry Mary C, Maingi John K, and Jessica McCarty. Fire on the water towers: mapping burn scars on mount Kenya using satellite data to reconstruct recent fire history. Remote Sensing. 2019; 11(2):104. https://doi.org/10.3390/rs11020104
9. Sari Fatih. Identifying anthropogenic and natural causes of wildfires by maximum entropy method-based ignition susceptibility distribution models. Journal of Forestry Research, 2023; 34(2):355–371. Springer. https://doi.org/10.1007/s11676-022-01502-4
10. Grünig Marc, Seidl Rupert, and Senf Cornelius. Increasing aridity causes larger and more severe forest fires across Europe. Global Change Biology, 2023; 29(6):1648–1659. Wiley Online Library. https://doi.org/10.1111/gcb.16547 PMID: 36517954
11. Jones Matthew W, Abatzoglou John T, Veraverbeke Sander, Andela Niels, Lasslop Gitta, Forkel Matthias, et al. Global and regional trends and drivers of fire under climate change. *Reviews of Geophysics*, 2022; 60(3):e2020RG000726. Wiley Online Library. https://doi.org/10.1029/2020RG000726
12. Japheth Omondi Aseko. Application of forest fire detection index in forest fires: a case study of Aberdare forest. PhD thesis, University of Nairobi. 2018.
13. Oladapo Muyiwa Oladoja, Adesola Grace Folorunso, Taiwo Mobolaji Adegoke, Sule Omeiza Bashiru, Kingley Chinedu Arum, and Aliyu Abba Mustapha. On the Application of Modelling Forest Fire in the Environment: A Bayesian Model Averaging Approach. In *2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG)*, 2023 (1): 1–5. IEEE.
14. Aaishwarya Gaikwad, Nishi Bhuta, Tejas Jadhav, Param Jangale, and Swati Shinde. A Review On Forest Fire Prediction Techniques. In *2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, 2022; 1–5. IEEE.
15. Bar Somnath, Parida Bikash Ranjan, Pandey Arvind Chandra, Shankar B Uma, Kumar Pankaj, Panda Santosh K, et al. Modeling and prediction of fire occurrences along an elevational gradient in Western Himalayas. *Applied Geography*, 2023; 151:102867. Elsevier. https://doi.org/10.1016/j.apgeog.2022.102867

16. Ramapriyan Hampapuram K., Behnke Jeanne, Sofinowski Edwin, Lowe Dawn, and Esfandiari Mary Ann. Evolution of the Earth Observing System (EOS) Data and Information System (EOSDIS). Springer Berlin Heidelberg, Berlin, Heidelberg. 2010; 63–92.

17. NASA. Fire information for resource management system (firms). Retrieved from https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms, on 05th March 2022.

18. Fornacca Davide, Ren Guopeng, and Xiao Wen. Performance of three modis fire products (mcd45a1, mcd64a1, mcd14ml), and esa fire_cci in a mountainous area of Northwest yunnan, China, characterized by frequent small fires. Remote Sensing. 2017; 9(11):1131. https://doi.org/10.3390/rs9111131

19. Giglio Louis, Schroeder Wilfrid, and Justice Christopher O. The collection 6 modis active fire detection algorithm and fire products. Remote Sensing of Environment. 2016; 178:31–41. https://doi.org/10.1016/j.rse.2016.02.054 PMID: 30158718

20. Wooster Martin J, Roberts G, Perry GLW, and Kaufman YJ. Retrieval of biomass combustion rates and totals from fire radiative power observations: Frp derivation and calibration relationships between biomass consumption and fire radiative energy release. Journal of Geophysical Research: Atmospheres. 2005; 110 (D24). https://doi.org/10.1029/2005JD006318

21. Lim Chul-Hee, Kim You Seung, Won Myungsoo, Kim Sea Jin, and Lee Woo-Kyun. Can satellite-based data substitute for surveyed data to predict the spatial probability of forest fire? A geostatistical approach to forest fire in the Republic of Korea. Geomatics, Natural Hazards and Risk. 2019; 10 (1):719–739. https://doi.org/10.1080/19475705.2018.1543210

22. Schroeder Wilfrid, Oliva Patricia, Giglio Louis, and Csiszar Ivan A. The new viirs 375 m active fire detection data product: Algorithm description and initial assessment. Remote Sensing of Environment. 2014; 143:85–96. https://doi.org/10.1016/j.rse.2013.12.008

23. Cai Zhanzhang, Jönsson Per, Jin Hongxiao, and Eklundh Lars. Performance of smoothing methods for reconstructing ndvi time-series and estimating vegetation phenology from modis data. Remote Sensing. 2017; 9(12):1271. https://doi.org/10.3390/rs9121271

24. Tong Qi and Gernay Thomas. A hierarchical Bayesian model for predicting fire ignitions after an earthquake with application to California. Natural Hazards, 2022; 1–24. Springer.

25. Storey Michael Anthony and Price Owen F. Prediction of air quality in Sydney, Australia as a function of forest fire load and weather using Bayesian statistics. *Plos one*, 2022; 17(8):e0272774. Public Library of Science San Francisco, CA USA. https://doi.org/10.1371/journal.pone.0272774 PMID: 36001532

26. Kadir Evizal Abdul, Kung Hsiang Tsung, Nasution Arbi Haza, Daud Hanita, AlMansour Amal Abdullah, Othman Mahmod, et al. Fires Hotspot Forecasting in Indonesia Using Long Short-Term Memory Algorithm and MODIS Datasets. In *Vegetation Fires and Pollution in Asia*, 2023; 589–602. Springer.

27. Lima Carlos HR, AghaKouchak Amir, and Randerson James T. Unraveling the role of temperature and rainfall on active fires in the Brazilian Amazon using a nonlinear poisson model. Journal of Geophysical Research: Biogeosciences. 2018; 123(1):117–128. https://doi.org/10.1002/2017JG003836

28. Liu Dingli, Xu Zhisheng, and Fan Chuangang. Predictive analysis of fire frequency based on daily temperatures. Natural Hazards. 2019; 97(3):1175–1189. https://doi.org/10.1007/s11069-019-03694-1

29. Group World Bank. World bank climate change knowledge portal: Kenya. Retrieved from https://climateknowledgeportal.worldbank.org/download-data, 2021.

30. Verdin Andrew, Funk Chris, Peterson Pete, Landsfeld Martin, Tuholske Cascade, and Grace Kathryn. Development and validation of the CHIRTS-daily quasi-global high-resolution daily temperature data set. ISSN 2052-4463. 2020; 7(1):303. https://doi.org/10.1038/s41597-020-00643-7 PMID: 32929097

31. Trauernicht Clay. Vegetation-rainfall interactions reveal how climate variability and climate change alter spatial patterns of wildland fire probability on big island, Hawaii. Science of The Total Environment. 2019; 650:459–469. https://doi.org/10.1016/j.scitotenv.2018.08.347 PMID: 30199690

32. Avanzi Benjamin, Taylor Greg, Wong Bernard, and Xian Alan. Modelling and understanding count processes through a markov-modulated non-homogeneous poisson process framework. *European Journal of Operational Research*, 2021; 290(1):177–195. https://doi.org/10.1016/j.ejor.2020.07.022

33. Hilbe Joseph M. *Negative Binomial Regression*. Cambridge University Press, 2011.

34. Trocóniz Iñaki F, Plan Elodie L, Miller Raymond, and Karlsson Mats O. Modelling overdispersion and markovian features in count data. *Journal of pharmacokinetics and pharmacodynamics*, 2009; 36 (5):461–477. https://doi.org/10.1007/s10928-009-9131-y PMID: 19798550

35. Lindén Andreas and Mäntyniemi Samu. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 2011; 92(7):1414–1421. https://doi.org/10.1890/10-1831.1 PMID: 21870615

36. Payne Elizabeth H, Hardin James W, Egede Leonard E, Ramakrishnan Viswanathan, Selassie Anbesaw, and Gebregziabher Mulugeta. Approaches for dealing with various sources of overdispersion in

modeling count data: Scale adjustment versus modeling. *Statistical methods in medical research*, 2017; 26(4):1802–1823. https://doi.org/10.1177/0962280215588569 PMID: 26031359

37. Huang A and Kim ASI. Bayesian conway–maxwell–poisson regression models for overdispersed and underdispersed counts. *Communications in Statistics-Theory and Methods*, 2021; 50(13):3094–3105. https://doi.org/10.1080/03610926.2019.1682162

38. Hardin James W, Hardin James William, Hilbe Joseph M, and Hilbe Joseph. *Generalized linear models and extensions*. Stata press, 2007.

39. Hayat Matthew J and Higgins Melinda. Understanding poisson regression. *Journal of Nursing Education*, 2014; 53(4):207–215. https://doi.org/10.3928/01484834-20140325-04 PMID: 24654593

40. Xia Yinglin, Morrison-Beedy Dianne, Ma Jingming, Feng Changyong, Cross Wendi, and Tu Xin. Modeling count outcomes from hiv risk reduction interventions: A comparison of competing statistical models for count responses. *AIDS research and treatment*, 2012; 2012. https://doi.org/10.1155/2012/593569 PMID: 22536496

41. Charizanos Georgios and Demirhan Haydar. Bayesian prediction of wildfire event probability using normalized difference vegetation index data from an Australian forest. *Ecological Informatics*, 2023; 73:101899. Elsevier. https://doi.org/10.1016/j.ecoinf.2022.101899

42. Flannigan Michael D, Stocks Brian J, and Wotton B Mike. Climate change and forest fires. *Science of the total environment*, 262(3):2000; 221–229. PMID: 11087028

43. Kwak Hanbin, Lee Woo-Kyun, Saborowski Joachim, Lee Si-Young, Won Myoung-Soo, Koo Kyo-Sang, et al. Estimating the spatial pattern of human-caused forest fires using a generalized linear mixed model with spatial autocorrelation in south korea. *International Journal of Geographical Information Science*, 2012; 26(9):1589–1602. https://doi.org/10.1080/13658816.2011.642799

44. Majid Saima, Alenezi Fayadh, Masood Sarfaraz, Ahmad Musheer, Gündüz Emine Selda, and Polat Kemal. Attention based CNN model for fire detection and localization in real-world images. *Expert Systems with Applications*, 2022; 189:116114. Elsevier. https://doi.org/10.1016/j.eswa.2021.116114

45. Gao Chao, Lin Honglei, and Hu Haiqing. Forest-Fire-Risk Prediction Based on Random Forest and Backpropagation Neural Network of Heihe Area in Heilongjiang Province, China. *Forests*, 2023; 14 (2):170. MDPI. https://doi.org/10.3390/f14020170

46. Cracknell Matthew J and Reading Anya M. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, 2014; 63:22–33. https://doi.org/10.1016/j.cageo.2013.10.008

47. Pereira JMC and Turkman KF. Statistical models of vegetation fires: Spatial and temporal patterns. In *Handbook of Environmental and Ecological Statistics*, Chapman and Hall/CRC, 2019; 401–420.

48. Oliveira Sandra, Oehler Friderike, San-Miguel-Ayanz Jesús, Camia Andrea, and Pereira José MC. Modeling spatial patterns of fire occurrence in mediterranean europe using multiple regression and random forest. *Forest Ecology and Management*, 2012; 275:117–129. https://doi.org/10.1016/j.foreco.2012.03.003

49. Boubeta Miguel, Lombardía María José, Marey-Pérez Manuel, and Morales Domingo. Poisson mixed models for predicting number of fires. *International journal of wildland fire*, 2019; 28(3):237–253. https://doi.org/10.1071/WF17037

50. Marchal Jean, Cumming Steve G, and McIntire Eliot JB. Exploiting poisson additivity to predict fire frequency from maps of fire weather and land cover in boreal forests of québec, canada. *Ecography*, 2017; 40(1):200–209. https://doi.org/10.1111/ecog.01849

51. Kim Taeho, Lieberman Benjamin, Luta George, and Peña Edsel A. Prediction intervals for poisson-based regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2021; page e1568.

52. Miguel Boubeta, MJ Lombardía, Francisco Marey-Pérez, and Domingo Morales. Area-level spatio-temporal poisson mixed models for predicting domain counts and proportions. *arXiv preprint arXiv 2020; 2012.00069*.

53. Su Zhangwen, Hu Haiqing, Tigabu Mulualem, Wang Guangyu, Zeng Aicong, and Guo Futao. Geographically weighted negative binomial regression model predicts wildfire occurrence in the great xing'an mountains better than negative binomial model. *Forests*, 2019; 10(5):377. https://doi.org/10.3390/f10050377

54. Dong Kai, Zhao Hongyu, Tong Tiejun, and Wan Xiang. Nblda: negative binomial linear discriminant analysis for rna-seq data. *BMC bioinformatics*, 2016; 17(1):1–10. https://doi.org/10.1186/s12859-016-1208-1 PMID: 27623864

55. Giglio Louis, Descloitres Jacques, Justice Christopher O, and Kaufman Yoram J. An enhanced contextual fire detection algorithm for modis. *Remote sensing of environment*, 2003; 87(2-3):273–282. https://doi.org/10.1016/S0034-4257(03)00184-6

56. NASA. Firms, 2021. URL https://firms.modaps.eosdis.nasa.gov/download/. Accessed on 05th March 2022.

57. Fick Stephen E and Hijmans Robert J. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 2017; 37(12):4302–4315. https://doi.org/10.1002/joc.5086

58. Harris IPDJ, Jones Philip D, Osborn Timothy J, and Lister David H. Updated high-resolution grids of monthly climatic observations–the cru ts3. 10 dataset. *International journal of climatology*, 2014; 34 (3):623–642. https://doi.org/10.1002/joc.3711

59. QGISDevelopmentTeam. *QGIS Geographic Information System*, 2021. URL http://qgis.org. Open Source Geospatial Foundation.

60. Robert J. Hijmans. *raster: Geographic Data Analysis and Modeling*, 2022. URL https://CRAN.R-project.org/package=raster. R package version 3.6-11.

61. Edzer Pebesma. *stars: Spatiotemporal Arrays*, *Raster and Vector Data Cubes*, 2022. URL https://CRAN.R-project.org/package=stars. R package version 0.6-0.

62. Pebesma Edzer. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal, 10(1):439–446, 2018. URL https://doi.org/10.32614/RJ-2018-009.

63. Bivand Roger S., Pebesma Edzer, and Gomez-Rubio Virgilio. *Applied spatial data analysis with R*, Second edition. Springer, NY, 2013. URL https://asdar-book.org/.

64. Hadley Wickham. *stringr: Simple*, *Consistent Wrappers for Common String Operations*, 2022. URL https://CRAN.R-project.org/package=stringr. R package version 1.4.1.

65. Morris Tim P, White Ian R, and Crowther Michael J. Using simulation studies to evaluate statistical methods. *Statistics in medicine*, 38(11):2074–2102, 2019. https://doi.org/10.1002/sim.8086 PMID: 30652356

66. Ben Hamner and Michael Frasco. *Metrics: Evaluation Metrics for Machine Learning*, 2018. URL https://CRAN.R-project.org/package=Metrics. R package version 0.1.4.

67. Bhaskaran Krishnan, Gasparrini Antonio, Hajat Shakoor, Smeeth Liam, and Armstrong Ben. Time series regression studies in environmental epidemiology. *International journal of epidemiology*, 2013; 42(4):1187–1195. https://doi.org/10.1093/ije/dyt092 PMID: 23760528

68. Stolwijk AM, Straatman HMPM, and Zielhuis GA. Studying seasonality by using sine and cosine functions in regression analysis. *Journal of Epidemiology & Community Health*, 1999; 53(4):235–238. https://doi.org/10.1136/jech.53.4.235 PMID: 10396550

69. Venables W. N. and Ripley B. D.. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. URL https://www.stats.ox.ac.uk/pub/MASS4/. ISBN 0-387-95457-0.

70. Hilbe Joseph M. *Modeling Count Data*. Cambridge University Press, 2014.

71. Goodrich Ben, Gabry Jonah, Ali Imad, and Brilleman Sam. rstanarm: Bayesian applied regression modeling via stan. *R package version*, 2020; 2(1).

72. De Souza RS, Hilbe JM, Buelens Bart, JD Riggs, Cameron Ewan, de Oliveira Ishida Emille Eugenia, et al. The overlooked potential of generalized linear models in astronomy–iii. bayesian negative binomial regression and globular cluster populations. *Monthly Notices of the Royal Astronomical Society*, 2015; 453(2):1928–1940. https://doi.org/10.1093/mnras/stv1825

73. Halim S Zohra, Janardanan Sunder, Flechas Tatiana, and Mannan M Sam. In search of causes behind offshore incidents: Fire in offshore oil and gas facilities. *Journal of Loss Prevention in the Process Industries*, 2018; 54:254–265. https://doi.org/10.1016/j.jlp.2018.04.006

74. Grzenda Wioletta. The advantages of bayesian methods over classical methods in the context of credible intervals. *Information systems in management*, 4, 2015.

75. de Oliveira Maristela Dias, Colosimo Enrico A, and Gilardoni Gustavo L. Bayesian inference for power law processes with applications in repairable systems. *Journal of Statistical Planning and Inference*, 2012; 142(5):1151–1160. https://doi.org/10.1016/j.jspi.2011.11.016

76. BahooToroody Ahmad, Abaei Mohammad Mahdi, Arzaghi Ehsan, Song Guozheng, De Carlo Filippo, Paltrinieri Nicola, et al. On reliability challenges of repairable systems using hierarchical bayesian inference and maximum likelihood estimation. *Process Safety and Environmental Protection*, 2020; 135:157–165. https://doi.org/10.1016/j.psep.2019.11.039

77. Pimont François, Fargeon Hélène, Opitz Thomas, Ruffault Julien, Barbero Renaud, Martin-StPaul Nicolas, et al. Prediction of regional wildfire activity in the probabilistic bayesian framework of firelihood. *Ecological applications*, 2021; e02316. https://doi.org/10.1002/eap.2316 PMID: 33636026

78. Halim Syeda Zohra, Quddus Noor, and Pasman Hans. Time-trend analysis of offshore fire incidents using nonhomogeneous poisson process through bayesian inference. *Process Safety and Environmental Protection*, 2021; 147:421–429. https://doi.org/10.1016/j.psep.2020.09.049