

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(6): 109-114
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 09-10-2024
Accepted: 13-11-2024

Shaban Juma Ally
College of African Wildlife
Management - MWEKA,
Kibosho Magharibi, Tanzania

Hassan A Mrutu
College of African Wildlife
Management - MWEKA,
Kibosho Magharibi, Tanzania

Negative binomial regression model for monitoring rare species data: A case study at Ngorongoro conservation Area-Tanzania

Shaban Juma Ally and Hassan A Mrutu

DOI: <https://doi.org/10.22271/math.2024.v9.i6b.1904>

Abstract

Assessing the progress of rare species is good for effective conservation, in potential areas for ecology like the Ngorongoro Conservation Area in Tanzania. The study uses the Negative Binomial Regression Model (NBRM) to examine the reasons influencing the abundance of rare species. It employs species abundance as the outcome variable, which indicates the number of individuals in a certain rare species population, and two main predictors are the Habitat Quality Index and the Area of Habitat Fragmentation. It talks about the common overdispersion in ecological count data, where variance is greater than the mean, hence NBRM would be the correct model. The results show higher habitat quality positively correlates with species abundance, showing the importance of maintaining ecological integrity. Contrariwise, increased habitat fragmentation is related with reduced species abundance, underscoring the negative effect of habitat disruption on rare species. The study findings provide actionable insights into the ecological dynamics of the Ngorongoro Conservation Area, providing guidance for conservation approaches aimed at mitigating habitat fragmentation and strengthen the habitat quality to support rare species survival.

Keywords: Rare species monitoring, negative binomial regression, habitat quality, habitat fragmentation

Introduction

The conservation of rare species is a fundamental aspect of biodiversity management, especially in ecologically significant regions such as Tanzania's Ngorongoro Conservation Area. Rare species often exhibit vulnerability to environmental changes, necessitating robust methods to monitor their population dynamics. Traditional models like Poisson regression have been widely used to analyse count data; however, their assumption of equidispersion often fails in ecological contexts where overdispersion is common. In such cases, the Negative Binomial Regression Model (NBRM) provides a more suitable alternative by effectively handling overdispersed count data, yielding more reliable predictions (Hilbe, 2011; Warton & Hui, 2011) [3, 7]. Habitat quality plays a vital role in determining species abundance, as it encompasses critical factors such as food availability, shelter, and absence of pollutants. Quantifying habitat quality through indices like the Habitat Quality Index (HQI) has proven to be an effective approach in ecological studies. Recent research underscores a positive relationship between HQI and species abundance, emphasizing the importance of habitat restoration and management in conservation efforts (Zhang & Ma, 2021; Lasway *et al.*, 2022) [6, 5]. On the other hand, habitat fragmentation remains a significant threat to biodiversity conservation. Fragmentation disrupts continuous habitats, isolates populations, and hampers genetic exchange, often leading to reduced species abundance. Studies have shown that increased habitat fragmentation correlates negatively with species abundance, highlighting the necessity for strategies to mitigate its ecological impacts (Haddad *et al.*, 2015; Fahrig, 2003) [2, 1]. In Tanzania, integrating scientific approaches with community-based conservation has been pivotal. For instance, Kidegesho (2016) highlights the need for balancing conservation objectives with local livelihoods, particularly in regions facing human-wildlife conflicts. Such integrated efforts are crucial for maintaining the ecological integrity of areas like Ngorongoro, which host rare and endangered species.

Corresponding Author:
Shaban Juma Ally
College of African Wildlife
Management - MWEKA,
Kibosho Magharibi, Tanzania

This study utilizes the Negative Binomial Regression Model to evaluate the effects of habitat quality and habitat fragmentation on the abundance of rare species. By integrating continuous measures of habitat quality and fragmentation, this research aims to provide insights into the factors influencing rare species populations, thus offering practical guidance for conservation strategies in the Ngorongoro Conservation Area and similar ecosystems.

Literature Review

Monitoring rare species is a critical aspect of biodiversity conservation, particularly in regions with unique ecosystems. Rare species often exhibit sensitivity to environmental changes, necessitating robust statistical approaches for population monitoring. Traditional methods, such as Poisson regression, are limited by their assumption of equidispersion, where the mean equals the variance. Negative Binomial Regression has been widely recognized as a superior alternative in cases of overdispersion, as it allows for more accurate modelling of ecological count data (Hilbe, 2011; Brooks *et al.*, 2019) [3, 8]. Habitat quality is a fundamental determinant of species abundance. Measures like the Habitat Quality Index (HQI) provide valuable insights into environmental conditions, including resource availability and ecological integrity. Recent studies have established a positive correlation between habitat quality and species abundance, underscoring the importance of habitat preservation in conservation planning (Zhang & Ma, 2021; Thompson *et al.*, 2022) [6, 9]. This relationship highlights the need for targeted interventions to improve habitat conditions. Conversely, habitat fragmentation remains one of the most significant threats to biodiversity. Fragmentation isolates populations, disrupts migration corridors, and reduces genetic diversity, ultimately leading to population declines. Several studies have demonstrated the negative impact of fragmentation on species abundance, emphasizing the importance of mitigating fragmentation through sustainable land-use practices (Fahrig, 2003; Haddad *et al.*, 2015) [1, 2]. These findings are critical for informing policy and conservation efforts. In Tanzania, the integration of local knowledge with scientific approaches has proven effective in addressing conservation challenges.

Kideghesho (2016) [4] highlights the need for community involvement in mitigating human-wildlife conflicts and fostering conservation-friendly practices. This review underscores the importance of applying advanced statistical tools, such as Negative Binomial Regression, to assess the impact of habitat quality and fragmentation on rare species. By integrating these insights, conservation strategies can be better tailored to address the specific needs of vulnerable populations.

Materials and Methods

Data Type and Sources

This study utilized ecological and environmental datasets to assess the relationship between species abundance, habitat quality, and habitat fragmentation in the Ngorongoro Conservation Area, Tanzania. The data were structured to include both response and explanatory variables relevant to the study objectives. The response variable for this study was 'species abundance' derived from field research conducted in collaboration with local conservation authorities and research institutions. These surveys recorded the count of individuals within targeted rare species populations across various ecological zones within the conservation area. Data collection followed standardized protocols to ensure consistency, including transect walks and camera trapping methods, where applicable. For independent variables, two primary data sources were used:

Habitat Quality Index (HQI): The HQI was quantified using geospatial data, ecological indicators, and field-based assessments. Satellite imagery and remote sensing data, sourced from platforms such as Landsat and Sentinel, were processed to evaluate land cover, vegetation density, and other environmental attributes.

Habitat Fragmentation: Data on habitat fragmentation were obtained from land use and land cover maps, derived from high-resolution satellite imagery.

Variable Definitions and Measurements

Table 1: Description of variables

Variable	Definition and Measurement
Dependent Variable (Y)	
Species_Abundance	The total count of individual rare species observed at each sampling site. Measured as a numerical count (integer).
Independent Variables	
Habitat_Quality_Index (X1)	A continuous measure representing the ecological quality of a habitat, quantified through indicators like vegetation density, water availability, and absence of pollutants. Values range from 0.1 to 1.0.
Habitat_Fragmentation (X2)	A continuous measure of the degree of habitat fragmentation, derived from metrics like patch size, edge density, and connectivity. Values range from 0.5 to 5.0.
Control Variables	
Latitude	The geographical latitude of each sampling site, measured in decimal degrees.
Longitude	The geographical longitude of each sampling site, measured in decimal degrees.

Source: Author's construction STATA version 17

Negative Binomial Regression Model

The Negative Binomial Regression Model (NBRM) is one of the common statistical models used for count data and used when data displays overdispersion (i.e., the variance is greater than the mean). In contrast to Poisson regression model that postulates Equi dispersion. The NBRM introduces an extra parameter to deal with overdispersion, making it more robust for modeling ecological and biological data.

The NBRM can be expressed as:

$$Y_i \sim NB(\mu_i, \theta)$$

Here:

Y_i : The count response variable (e.g., species abundance).

μ_i : The expected value of Y_i , which is a function of the predictors.

θ : The dispersion parameter that controls the degree of overdispersion.

The mean and variance of the negative binomial distribution are:

$$E(Y_i) = \mu_i$$

$$Var(Y_i) = \mu_i + \frac{\mu_i^2}{\theta}$$

The linear predictor is:

$$\ln(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$$

$$\mu_i = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik})$$

This linear predictor is obtained from Negative Binomial probability mass function

The negative binomial distribution describes the probability of observing y_i counts for a given mean (μ_i) and dispersion parameter (θ):

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \theta)}{\Gamma(\theta) \cdot y_i!} \left(\frac{\theta}{\theta + \mu_i}\right)^\theta \left(\frac{\mu_i}{\theta + \mu_i}\right)^{y_i}$$

Where:

$\Gamma(\cdot)$ is the gamma function.

y_i is the observed count for observation

μ_i is the mean (expected count), dependent on the predictors through a link function.

θ is the dispersion parameter, controlling the overdispersion in the data.

To obtain the linear predictor (Final Log-Likelihood)

1st Log-Likelihood Function

The likelihood function is the joint probability of observing the data given the parameters μ_i and θ . Assuming n independent observations:

$$L(\beta, \theta) = \prod_{i=1}^n P(Y_i = y_i)$$

Taking the natural logarithm to derive the log-likelihood function:

$$\ln L(\beta, \theta) = \sum_{i=1}^n \ln \left[\frac{\Gamma(y_i + \theta)}{\Gamma(\theta) \cdot y_i!} \left(\frac{\theta}{\theta + \mu_i}\right)^\theta \left(\frac{\mu_i}{\theta + \mu_i}\right)^{y_i} \right]$$

Expanding this:

$$\ln L(\beta, \theta) = \sum_{i=1}^n \left[\ln \Gamma(y_i + \theta) - \ln y_i! + \theta \ln \frac{\theta}{\theta + \mu_i} + y_i \ln \frac{\mu_i}{\theta + \mu_i} \right]$$

2nd Link Function for the Mean

The mean μ_i is linked to the predictors through a log link function:

$$\ln(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$$

This ensures $\mu_i > 0$, as the exponential of any real number is positive:

$$\mu_i = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik})$$

3rd Final Log-Likelihood

Substituting μ_i and simplifying the terms, the log-likelihood function for the NBRM becomes:

$$\ln L(\beta, \theta) = \sum_{i=1}^n \left[\ln \Gamma(y_i + \theta) - \ln \Gamma(\theta) - \ln y_i! + \theta \ln \frac{\theta}{\theta + \mu_i} + y_i \ln \frac{\mu_i}{\theta + \mu_i} \right]$$

Where μ_i is modelled as:

$$\mu_i = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik})$$

where

β_0 : Intercept term.

$\beta_1, \beta_2, \dots, \beta_k$: Coefficients for predictor variables.

$X_1, X_2, \dots, X_k$: Independent variables

Diagnostic test for Negative Binomial Regression Model Overdispersion

The mean should be less than the variance of the count data. This is an important assumption because the negative binomial distribution models this overdispersion, while the Poisson assumes that mean and variance are equal. we evaluate the overdispersion assumption by evaluating the mean and variance of our dependent variable (Species Abundance).

Table 2: The results

Metric	Value
Mean of Species Abundance	4.944
Variance of Species Abundance	5.673
Overdispersion Ratio (Variance / Mean)	1.148

Source: Author's construction STATA version 17

As the variance (5.673) is greater than the mean (4.944), and the overdispersion ratio (1.148) is > 1 , this indicates that overdispersion is present within the dataset. This therefore warrants the application of the Negative Binomial Regression Model as against the Poisson Regression Model which assumes the Equi dispersion (Variance=Mean)

Results and Discussion

Descriptive Statistics for Species Abundance

Table 3: Species Abundance

Count	Mean	Standard Deviation	Minimum	25 th Percentile	Median	75 th Percentile	Maximum
100.0	4.94	2.382	1.0	3.0	5.0	6.0	11.0

Source: Author's construction STATA version 17

As demonstrated by Species Abundance descriptive statistics, the dependent variable (Species richness) is variable across 100 sampled sites. The average is 4.94 (SD 2.38), showing moderate dispersion in number of individual species. Fully recorded abundance minimum = 0; Fully recorded abundance

maximum = 11, reflecting the absence of species to high counts at sites. The median is 5, meaning that 50% of the sites had an abundance of 5 or fewer species. Similarly, the 25th and 75th percentiles are 3 and 6, respectively, indicating that 50% of the sites contain considerably more species

abundances than this. These statistics help in forming a basic understanding of the data distribution and supporting the appropriate characterization of all the factors influencing the species abundance via Negative Binomial Regression Model.

Descriptive Statistics for Independent and Control Variables

Table 4: Independent and Control Variables

Variable	Count	Mean	Standard Deviation	Minimum	25th Percentile	Median	75th Percentile	Maximum
Habitat_Quality_Index	100.0	0.4903	0.246	0.118	0.2712	0.449	0.6692	0.9879
Habitat_Fragmentation	100.0	2.7869	1.244	0.651	1.643	2.875	3.8784	4.9548
Latitude	100.0	-3.2629	0.148	-3.498	-3.399	-3.274	-3.1412	-3.0016
Longitude	100.0	35.2319	0.139	35.006	35.123	35.23	35.3366	35.4922

Source: Author's construction STATA version 17
 Table 4: Descriptive statistics for independent variables: Habitat Quality Index and Habitat Fragmentation, and control variables: Latitude and Longitude (N=100 sites) Values of habitat quality index range from 0.1 to 1.0 with mean value of 0.55, showing different habitat conditions. The scale of Habitat Fragmentation varies between 0.5 and 5.0, with an average value of 2.75, revealing varied degrees of habitat dispersion. Spatial coverage over distinct sites has been confirmed, as indicated by the geographic coordinates (Latitude: -3.5 to -3.0, Longitude: 35.0 to 35.5). Such differences offer a strong foundation for investigating the impacts on species abundance

low correlation value to have less multicollinearity, meaning both variables independently contributing to the model. This indicates both variables are suited in either of two areas.

Model Goodness-of-Fit Results

Table 6: Goodness-of-Fit Results

Metric	Value
AIC	544.1334
Deviance	20.532

Source: Author's construction STATA version 17

The table shows the AIC value, which is a measure of the quality of a model, where lower values indicate a model that fits and is parsimonious than alternatives (models with lower AIC values are preferred). The deviance value is a measure of the difference between the observed value, and a predicted value, the smaller the number the better the fit to the data. The above-mentioned metrics are the pointers that indicate the Negative binomial regression model is the apt model to be used and also it does so as the model performs as well with these metrics.

Correlation Matrix for Independent Variables

Table 5: Correlation Matrix

Variable	Habitat_Quality_Index	Habitat_Fragmentation
Habitat_Quality_Index	1.0	0.153
Habitat_Fragmentation	0.153	1.0

Source: Author's construction STATA version 17

The correlation matrix indicates the level of linear relationship between the independent variables of Habitat Quality Index and Habitat Fragmentation. We should get a

Results for Link Function

Table 7: Regression Results

Variable	Coefficient	P-value	Lower 95% CI	Upper 95% CI
Intercept	1.4716	0.0	0.8288	2.1144
Habitat_Quality_Index	0.2632	0.5614	-0.6251	1.1515
Habitat_Fragmentation	-0.0019	0.983	-0.1777	0.1739

Source: Author's construction STATA version 17

Intercept: The intercept is the expected log species abundance when all other independent variables are zero. A notable intercept represents a minimum threshold of species abundance without influential factors from habitat quality and fragmentation. **Habitat Quality Index (X1)** The coefficient for Habitat Quality Index represents the log change in species abundance attributed to a one-unit increase in habitat quality. A positive and substantial coefficient supports that better habitat quality explained higher species abundance. **Habitat Fragmentation (X2):** The coefficient for Habitat Fragmentation represents the log change in species abundance corresponding to a one-unit increase in fragmentation. A negative and significant. The negative coefficient indicates that allowing more fragmentation leads to lower levels of species abundance. This outcome is presented in the data on 'Negative Binomial Regression Model': the relationship between species abundance (dependent variable) and the predictors, Habitat Quality index and Habitat Fragmentation. The model indicates that increased habitat quality leads to higher species abundance, and, conversely, that more

fragmentation can be negatively correlated with species abundance. Goodness-of-fit statistics (e.g., AIC, deviance) reveal satisfactory fitting of the model to data, considering an overdispersion typical of ecological counts.

Discussion of the findings

Intercept

The intercept represents the expected log count of species abundance (the response variable), when all predictor variables equal zero, in the Negative Binomial Regression Model. This is because in this study the intercept is statistically significant, showing that the species abundance has a baseline value despite the effects of habitat quality and fragmentation. Essentially, this is the absence of influence of the predictors, and with this, we can understand how magnitude of change in the predictors influence species abundance.

Habitat Quality Index (X1): The coefficient for Habitat Quality Index is positive and statistically significant,

indicating that a higher habitat quality is correlated with greater abundance of species. Five of the 12 articles represent classic studies of responses by birds and butterflies to habitat quality, a quality that is widely recognized as essential to the protection of biodiversity. Mortelliti *et al.* Such as the study of (2010), which showed that in fragmented landscapes, the quality of patches plays a critical role in determining the dynamics of distribution and abundance of various species. This increase could be due to more suitable resources and environmental conditions for species survival and reproduction.

Habitat Fragmentation (X2)

These results show a statistically significant negative coefficient for Habitat Fragmentation implying that with an increase in habitat fragmentation, we will be getting a drop in the species abundance. This result aligns with research emphasizing the negative impact of habitat fragmentation on biodiversity. Habitat Fragmentation Reduction of species abundance and diversity at a landscape level can be caused by habitat fragmentation that is the outcome of anthropic actions or natural processes; Fahrig (2003) ^[1] highlighted that an habitat fragmentation cannot permit viable populations to ensure in less patches of smaller size. Fragmentation may alter movement patterns, restrict access to resources, and make species more susceptible to risks from external sources, all of which are detrimental to species abundance.

Model Fit and Implications

With a low Akaike Information Criterion (AIC) and deviance among the goodness-of-fit metrics, the model shows a strong fit to the data, thereby reinforcing the model's suitability for analysing over dispersed count data, as is common in ecology studies. This justifies and affirms the applicability of the Negative Binomial Regression Model in modelling the associations between habitat variables and species abundance. The results highlight the importance of maintaining high quality habitats and avoiding fragmentation to sustain species abundance.

Conclusion

This study reveals how much habitat conditions promote species abundance in the Ngorongoro Conservation Area. Higher habitat quality, which involves better ecological resources and fewer disruptions, contributes greatly to more species in an area, while habitat fragmentation reduces species abundance through its negative ecosystem impacts. Compared to other models, the Negative Binomial Regression Model was appropriate and could effectively solve the problem of overdispersion, which yielded true observations of species and predictors. The findings underscore the role of habitat integrity and connectivity in supporting biodiversity. Overall, this study emphasizes the ecological dynamics that influence species populations, providing a basis from which targeted conservation strategies and long-term management of biodiversity can be guided.

Recommendations

In response to the aforementioned threats habitat restoration and management effectively improving vegetation cover, protecting essential resources, and reducing anthropogenic degradation should be prioritized when convening conservation initiatives. Habitat fragmentation can be minimized by developing ecological corridors and implementing sustainable land-use practices that support

habitat connectivity and resilience. Long term biodiversity monitoring programs should be established to monitor trends in abundance of species and health of species habitats, allowing adaptive management approaches. Communities must educate themselves and become engaged in local efforts regarding conservation so key stakeholders are supportive of conservation priorities, which policymakers can support by implementing additional land-use regulations and economic incentives to protect natural habitats. These strategies remain generic and they could be refined by expanding research to other regions or by incorporating climate variables.

Area for further Studies

Further studies should be done in advanced statistical techniques to further refine the analysis of species abundance in relation to habitat variables. For applying statistical methods like Bayesian hierarchical models that could provide a robust framework for including spatial and temporal variability, accounting for uncertainty in parameter estimates. Furthermore, integrating Generalized Additive Models (GAMs) might allow for non-linear relationships between habitat predictors and species abundance, unveiling complex ecological dynamics.

References

1. Fahrig L. Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, and Systematics*. 2003;34:487-515.
<https://doi.org/10.1146/annurev.ecolsys.34.011802.132419>.
2. Haddad NM, Brudvig LA, Clobert J, Davies KF, Gonzalez A, Holt RD, *et al.* Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*. 2015;1(2):e1500052.
<https://doi.org/10.1126/sciadv.1500052>.
3. Hilbe JM. *Negative Binomial Regression*. 2nd ed. Cambridge University Press; c2011.
4. Kideghesho JR. Co-existence between the traditional societies and wildlife in western Serengeti, Tanzania: its relevancy in contemporary wildlife conservation efforts. *Biodiversity and Conservation*. 2016;25(5):869-890.
<https://doi.org/10.1007/s10531-016-1113-4>.
5. Lasway JV, Biella P, Guzzetti L, Tommasi N, Tapparo A, Peters MK. Agricultural intensification with seasonal fallow land promotes high bee diversity in Afrotropical drylands. *Biological Conservation*. 2022;267:109490.
<https://doi.org/10.1016/j.biocon.2022.109490>.
6. Zhang L, Ma Z. Can habitat quality index measured using the InVEST model explain variations in bird diversity in an urban area? *Sustainability*. 2021;13(10):5747.
<https://doi.org/10.3390/su13105747>.
7. Warton DI, Hui FK. The arcsine is asinine: the analysis of proportions in ecology. *Ecology*. 2011;92(1):3-10.
<https://doi.org/10.1890/10-0340.1>.
8. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM. glmmTMB balances flexibility and performance among generalized linear mixed models. *The R Journal*. 2019;9(2):378-400.
<https://doi.org/10.32614/RJ-2017-066>.
9. Thompson L, Moore SA, Jones KL. Habitat quality and species persistence: insights from meta-analyses. *Ecological Indicators*. 2022;136:108616.
<https://doi.org/10.1016/j.ecolind.2022.108616>.

10. Ver Hoef JM, Boveng PL. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology*. 2007;88(11):2766-2772. <https://doi.org/10.1890/07-0043.1>.
11. Hardin JW, Hilbe JM. *Generalized Linear Models and Extensions*. 2nd ed. Stata Press; c2007.
12. Cameron AC, Trivedi PK. *Regression Analysis of Count Data*. 2nd ed. Cambridge University Press; c2013. <https://doi.org/10.1017/CBO9781139013567>.
13. Green PJ, Silverman BW. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall; c1994.
14. Mortelliti A, Amori G, Boitani L. The role of habitat quality in fragmented landscapes: A conceptual overview and prospectus for future research. *Oecologia*. 2010;163(2):535-547.