

OPEN ACCESS



African Journal of
**Mathematics and
Computer Science Research**

January 2018
ISSN 2006-9731
DOI: 10.5897/AJMCSR
www.academicjournals.org

academicJournals



Academic
Journals

ABOUT AJMCSR

The African Journal of Mathematics and Computer Science Research (ISSN 2006-9731) is published bi-monthly (one volume per year) by Academic Journals.

The African Journal of Mathematics and Computer Science Research (AJMCSR) (ISSN:2006-9731) is an open access journal that publishes high-quality solicited and unsolicited articles, in all areas of the subject such as Mathematical model of ecological disturbances, properties of upper fuzzy order, Monte Carlo forecast of production using nonlinear econometric models, Mathematical model of homogenous tumour growth, Asymptotic behavior of solutions of nonlinear delay differential equations with impulse etc. All articles published in AJMCSR are peer-reviewed.

Contact Us

Editorial Office: ajmcsr@academicjournals.org

Help Desk: helpdesk@academicjournals.org

Website: <http://www.academicjournals.org/journal/AJMCSR>

Submit manuscript online <http://ms.academicjournals.me/>

Editors

Prof. Mohamed Ali Toumi

*Département de Mahtématiques
Faculté des Sciences de Bizerte
7021, Zarzouna, Bizerte
Tunisia.*

Associate Professor Kai-Long Hsiao,

*Department of Digital Entertainment,
and Game Design,
Taiwan Shoufu University,
Taiwan,
R. O. C.*

Dr. Marek Galewski

*Faculty of Mathematics and Computer Science,
Lodz University
Poland.*

Prof. Xianyi Li

*College of Mathematics and Computational Science
Shenzhen University
Shenzhen City
Guangdong Province
P. R. China.*

Editorial Board

Dr. Rauf, Kamilu

*Department of mathematics,
University of Ilorin,
Ilorin, Nigeria.*

Dr. Adewara, Adedayo Amos

*Department of Statistics,
University of Ilorin.
Ilorin.
Kwara State.
Nigeria.*

Dr. Johnson Oladele Fatokun,

*Department of Mathematical Sciences
Nasarawa State University, Keffi.
P. M. B. 1022, Keffi. Nigeria.*

Dr. János Toth

*Department of Mathematical Analysis,
Budapest University of Technology and
Economics.*

Professor Aparajita Ojha,

*Computer Science and Engineering,
PDPM Indian Institute of Information
Technology,
Design and Manufacturing, IT Building,
JEC Campus, Ranjhi, Jabalpur 482 011 (India).*

Dr. Elsayed Elrifai,

*Mathematics Department,
Faculty of Science,
Mansoura University,
Mansoura, 35516, Egypt.*

Prof. Reuben O. Ayeni,

*Department of Mathematics,
Ladoke Akintola University,
Ogbomoso,
Nigeria.*

Dr. B. O. Osu,

*Department of Mathematics,
Abia State University,
P. M. B. 2000,
Uturu,
Nigeria.*

Dr. Nwabueze Joy Chioma,

*Abia State University,
Department of Statistics,
Uturu,
Abia State,
Nigeria.*

Dr. Marchin Papzhytski,

*Systems Research Institute,
Polish Academy of Science,
ul. Newelska 6 0-60-66-121-66,
03-815 Warszawa,
POLAND.*

Amjad D. Al-Nasser,

*Department of Statistics,
Faculty of Science, Yarmouk University,
21163 Irbid,
Jordan.*

Prof. Mohammed A. Qazi,

*Department of Mathematics,
Tuskegee University,
Tuskegee,
Alabama 36088,
USA.*

Professor Gui Lu Long

*Dept. of Physics,
Tsinghua University,
Beijing 100084,
P. R. China.*

Prof. A. A. Mamun, Ph. D.

*Ruhr-Universitaet Bochum,
Institut fuer Theoretische Physik IV,
Fakultaet fuer Physik und Astronomie,
Bochum-44780,
Germany.*

Prof. A. A. Mamun, Ph. D.

*Ruhr-Universitaet Bochum,
Institut fuer Theoretische Physik IV,
Fakultaet fuer Physik und Astronomie,
Bochum-44780,
Germany.*

African Journal of Mathematics and Computer

Table of Content: Volume 11 Number 1 January 2018

ARTICLE

**Distributional analysis with line transect methodology
of the distance sampling techniques: Case of large mammals
of the Mole National Park (MNP) of Ghana**

Katara Salifu, S.K Amponsah and Bashiru I.I. Saeed

1

Full Length Research Paper

Distributional analysis with line transect methodology of the distance sampling techniques: Case of large mammals of the Mole National Park (MNP) of Ghana

Katara Salifu^{1*}, S.K Amponsah² and Bashiru I.I. Saeed³

¹University for Development Studies, Department of Statistics, Navrongo Campus, Navrongo, Ghana.

²Kwame Nkrumah University of Science and Technology, Department of Mathematics, Ghana.

³Kumasi Technical University, Department of Mathematics and Statistics, Kumasi, Ghana.

Received 5 June, 2017; Accepted 18 August, 2017

Distance sampling with line transect method has been applied by many researchers to monitor and observe varied animals and plants with the aim to estimating population density and/or abundance. The detection and detectabilities of plants and animals with line transect methodology appear to be described as seen in the application software DISTANCE by some four specific models or mathematical functions. This study has established the need for an expansion of the functional space of the application software of the distance sampling with line transect application to cater for differences in species with different biological characteristics, and revealed that the detection of plants and animals of all species of different kinds can be described differently with distinct functions, as differences in biological characteristics of different species can be manifested in their detections and detectabilities.

Key words: DISTANCE, distance sampling, detectability, line transect method, Mole National Park (MNP).

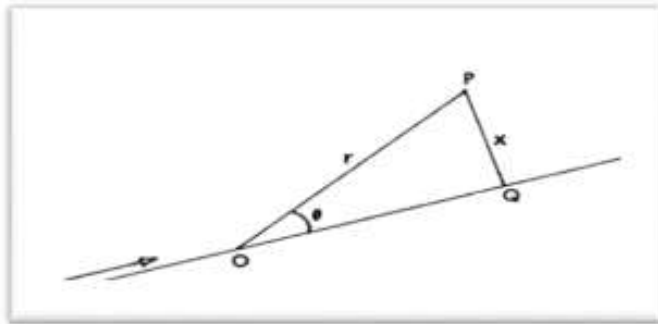
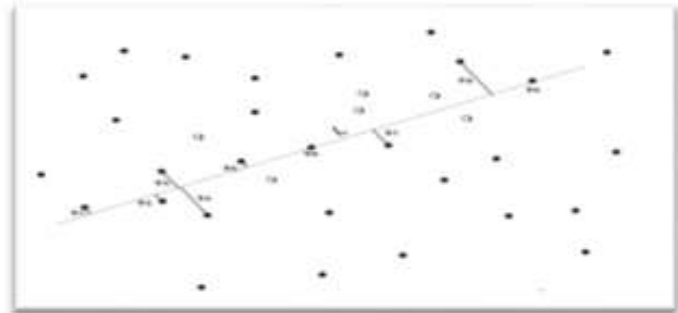
INTRODUCTION

Line transect primarily has had a relatively short history. The earliest attempts to use line transect in detecting animals and estimating their abundance date back to the 1930's (Buckland et al., 2001), and the first line transect estimator with a rigorous mathematical basis was due to Hayne (1949). Nearly 20 years later, Gates (1968) and Eberhardt (1968) made important contributions to the development of line transect sampling methodology. Eberhardt's (1968) work was more conceptual and

attempted to provide a class of models that were robust to the differing detection process.

Line transect methodology involves a process wherein an observer traverses through a line to observe units of interest on both left and right of the line whilst taking recordings of the distances and number of units of interests observed. Observation of objects of interest can either be perpendicular or through an angle θ as Figure 1 illustrates. Burnham et al. (1976) developed a

*Corresponding author. E-mail: katadeous@gmail.com.

**Angular detections.****Perpendicular detections****Figure 1.** Angular and Perpendicular observations along a single transect line.

general theory needed for reliable estimation. Anderson and Pospahala (1970) also used polynomials to fit the distance data, but did not provide an underlying theory. The field experiments of Robinette et al. (1974) were important in providing data sets with known abundance, on which estimation methods could be tested. Sen et al. (1974) gave an erroneous formulation. Burnham and Anderson (1976) corrected this formulation, provided a general framework for both parametric and nonparametric method and applied to data that were either grouped or ungrouped and truncated or untruncated.

In conventional line transect sampling, all animals on the line are assumed to have a certain detection probability which falls off steadily as distances increase (Buckland et al., 2015). An established assumption of this method is that the probability of detection of an object depends solely on its perpendicular distance from the transect line. However, many variables such as visibility, time of visits, weather, and availability of units of interest in the study region, among others may affect the detection probability. These variables however can be factored into the model wherein detections can appear to be conditioned on them with distances of detections as suggested by Marques et al. (2007). The inclusion of these may appear to influence either the shape or the scale or both of the detection function identified.

Buckland et al. (1993) provided an extensive review of distance sampling design and requirements with line transect methodology while Laake et al. (1993) developed the software DISTANCE allowing the computation of the density and abundance parameters based on some specified mathematical functions. The key elements presented by Buckland et al. (1993) relied on the assumption that animals will remain undetected in distance sampling and that the detectability decreases with increasing the distance from the line. The fundamental concept central to the theory is the detection function $g(x)$ where $g(x=0) = 1$. That is, at distance zero, all the animals on the line are detected with

a probability of 1. To derive robust estimates, a minimum of 60 to 80 observations are required across the study area (Buckland et al., 1993).

Line transect sampling has been developed along rigorous statistical inference principles. Parametric approaches to modelling $g(x)$ were predominant, with the notable exception of Anderson and Pospahala (1970), who rather inadvertently introduced some of the basic ideas that underlie a non-parametric or semi-parametric approach to the analysis of line transect data. Emlen (1971) proposed an ad hoc method that found use in Avian studies. A general model structure for line transects on perpendicular distances was presented by Seber (1986). For an arbitrary detection function, Seber gave the probability distribution of the distances x_1, x_2, \dots, x_n and the general form of the estimator of animal density D . This development was left at the conceptual stage and not pursued to the final step of a workable general approach for deriving line transect estimators.

Burnham and Anderson (1980) pursued the general formulation of line transect sampling and gave a basis for the general construction of line transect estimators. They developed the general result $\hat{D} = n \cdot \hat{f}(0) / 2L$, wherein the parameter $f(0)$ is a well-defined function of the distance data. The key problem of line transect data analysis was seen to be the modeling of $g(x)$ or $f(x)$ and the subsequent estimation of $f(0)$.

PROBLEM STATEMENT

The detection and detectabilities of plants and animals of all species of different kinds are seen generally to be described by the Half Normal, the Uniform, the Negative Exponential, or the Hazard Rate distributions as specified in the DISTANCE software. However differences in biological characteristics of different species of plants and animals including mammals of all types cannot be underestimated. This study is thus aimed at exploring

other potential functions that can equally play similar or better roles as those generally employed in describing and modeling detectabilities of units of interest under investigation.

Research questions

- i) Are all animals being described by same distributional forms as those specified in DISTANCE software?
- ii) Do different species of animals and plants possess different characteristics that can be described by different functions in detections, monitoring and conservation?

Objectives

- i) To identify an appropriate statistical distribution that best describes detections of units of interest under study.
- ii) To enhance detectability estimation process of rare and elusive events in the distance sampling realm.
- iii) To perform a Bayesian analysis of detectability of units of interest conditional on perpendicular distances and other identified covariates in the study area.
- iv) To provide appropriate detections periods of units under investigation to visitors or tourists and researchers who intend to visit the Mole National Park for the observation of units under study.

METHODS

Study area

The study area is located in the West Gonja District of Northern Ghana and about 184 km from the Northern Regional capital of Tamale. It covers a total area of 4,755 km² of the West Gonja District which inhabits several plants and animal species whose densities are of concern to wildlife management and conservation.

Briggs (2007) revealed that the MNP represents Ghana's largest wildlife refuge which is located northwest Ghana on grassland savannah and riparian ecosystems at an elevation of 150 m, with sharp escarpment forming the southern boundary of the park. The park's entrance is reached through the nearby town of Larabanga. The Lovi and Mole Rivers are ephemeral rivers flowing through the park, leaving behind only drinking holes during the long dry season (Bowell et al., 1993). This area of Ghana as seen in Figure 1 receives over 1000 mm per year of rainfall.

Survey design and transect selection

The sampling design in a line transect study is the procedure by which the transect locations are selected. Desired properties of unbiasedness of estimators will be based as much as possible on the design rather than on assumption about the population (Thompson, 2002).

The study employs a direct method of observation along transect lines placed in the study area. Transect lines are identified by employing simple and systematic sampling techniques. Direct methods are based on actual observation of the species in question (large mammals). Visibility and detectability can often pose a

problem when surveying terrestrial species; thus, surveys relying on signs are ideal for estimating mammal abundance and habitat use (Hayward et al., 2005; Sadler et al., 2004; Sullivan et al., 2002), calling for indirect methods of detection as a means to establishing the signs of animal presence.

To select the appropriate transects for the study, a combination of both probability and non-probability sampling designs were employed in order to improve data quality and analysis. These designs included stratification, convenient, and systematic, respectively. Due to the timing of the research and the nature of the study area, it was decided to stratify study area into two major strata (a stratum with and without water bodies) within which separate designs were employed in transect placement in order to optimize detections and improve precision.

Conveniently identifying a random start line in each stratum, a combination of both continuous systematic design and discrete parallel transect lines were placed, respectively in the strata so as to avoid discontinuation in detection from one transect to the other whilst ensuring an even spatial distribution of lines in surveyed region. Thompson (2002) indicates that estimates from *n* transects are more preferred to those based on single transects. In view of the nature of the study area coupled with resource constraints, a total effort of 4 km was estimated for coverage per day in the study area, for a period of 20 days, spanning 10 months, for at least 2 days per every visit, to traverse through a sample of ten (10) transects of equal lengths in the identified strata using a major route that cuts across the study area as a base.

A route perpendicular to the baseline was selected conveniently in each stratum and represented as the first transects and considered as a random start samples from which the remaining eight lines of non-overlapping and widely placed apart were systematically positioned to cover and include specific locations with prior knowledge of having high concentrations of the species of interest in the study area as shown in Figures 2a and b.

Data type

Here, both primary and secondary data were considered. Primary data was gathered within the region of study involving investigator and some supporting staff. Data gathered included both perpendicular distances and the number of Elephants, Hartebeests, Waterbucks and Warthogs detected at various locations of the study region using 10 conveniently and systematically placed transect lines within two stratified locations. Existing information from management of the MNP about the units of interest within the study area constituted the secondary data.

A principle of data augmentation was adopted to cater for units undetected but present in the study region and statistical packages such as the Distance, Minitab, R, Easy fit, Open Bugs to mention just a few were extensively used in both data exploration and modelling process involved in this study.

ANALYSIS AND DISCUSSION

The probability of detection of a unit of interest assumes a major component of the analysis of this thesis. The distance data measured are empirically categorized into different bins of varied sizes to determine the best data grouping in the exploration stage. This also included histogram plots of distance data to check for evasiveness, random movement, heaping, outliers and or truncation distance as shown in Figure 3(a, b, c and d).

Figure 3(a, b, c, and d) represent histogram plots of observed species in the MNP at different perpendicular

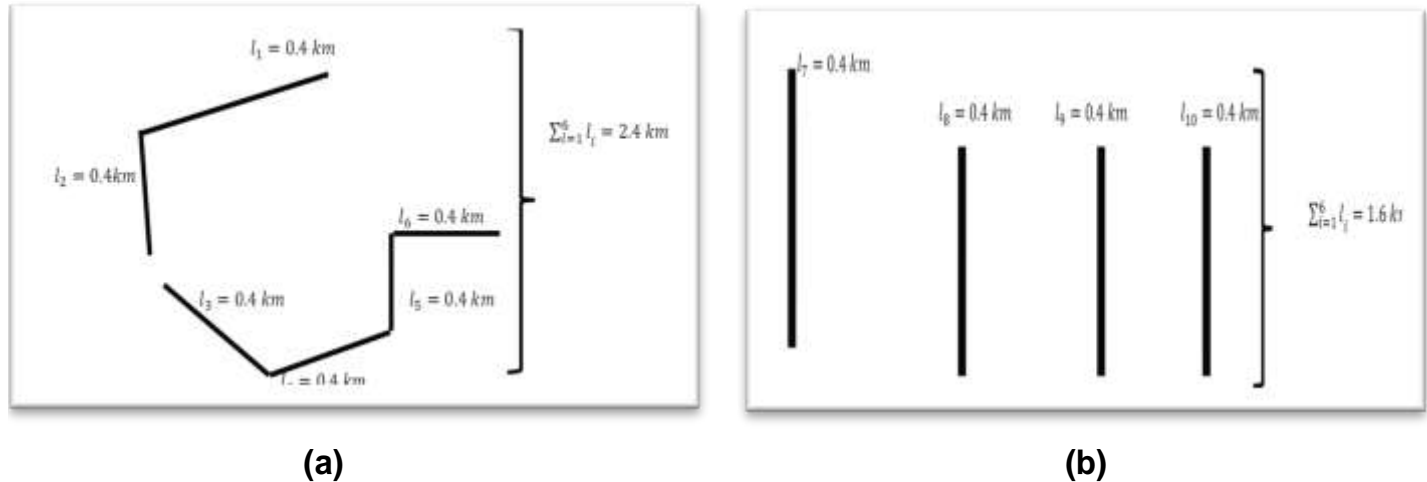


Figure 2. (a) Continuous zigzag lines placed in Stratum 1 (b) Systematically parallel lines placed in Stratum 2.

distances relative to the transect lines. In these plots, it is evident that ability to observe units of interest decreases with increasing distances. The plots also reflect differences in detection abilities among the four different species as a reflection of differences in biological characteristics.

Possible key functions and series expansions with their formulations in the DISTANCE software

Key function	Formula	Series Expansion	Formula
Uniform key	$\frac{1}{w}$	Cosine series	$\sum_{j=2}^m a_j \cos(j\pi x_s)$
Half Normal key	$e^{-(x^2/2\sigma^2)}$	Simple polynomial	$\sum_{j=2}^m a_j x_s^{2j}$
Hazard Rate key	$1 - e^{-(x/\sigma)^{-b}}$	Hermite polynomial	$\sum_{j=2}^m a_j H_{2j}(x_s)$
Negative Exponential	$e^{-(ax)}$		

With the application of these, where in detections are conditioned on perpendicular distances only, it is observed that the detection of elephants in the MNP is described by the hazard function based on the lowest AIC value. The use of the AIC is due to the fact that it contains a penalty term to penalize functions with more parameters so that over parameterization will be checked. The detectability function appears in the form:

$$g(x) \propto key(x)[1 + series(x_s)] \tag{1}$$

Equation 1 is found to be represented mathematically as:

$$1 - e^{[-(x/\sigma)^{-b}] } \tag{2}$$

with the scale parameter $\sigma = A(1) = 6.3894963 \approx 6.4$ and the shape parameter $b = A(2) = 1.4545765 \approx 1.5$ respectively. Thus the general form of the model is mathematically expressed as:

$$g(x) = 1 - e^{[-(x/6.4)^{-1.5}] } \tag{3}$$

Conditioning on both distance and other identified covariates also suggest a hazard rate key function with no adjustment term as the best model with minimum AIC = 299.5287. The detectability function assumes the form: $g(x) \propto key = 1 - \text{Exp}(-(x/s) ** -A(2))$ (4)

where the shape parameter = $A(2) = 2.507$ and the scale parameter = $s = A(1) * \text{Exp}((A(3)) + (A(4)) + (A(5)))$ with estimated values $A(1) = 13.13, A(3) = 0.6966, A(4) = -0.4176$ and $A(5) = -0.06031$ where $s=16.34$ and

$$g(x) = 1 - \exp(-(x/16.33957)^{-2.507}) \tag{5}$$

Comparing the two conditional cases, it can be observed that the covariates included in the Hazard rate model have increased both the scale and shape parameters from 6.4 to about 16.34 and 1.5 to about 2.507, respectively.

Distribution identification for distances of detected elephants in the MNP

To identify the distribution, observed data is assumed to follow or best fit the study and explores several probability plots to obtain a fit that satisfies the following as reflected in Figure 3:

- i) The plotted points roughly forming or following a straight line and appears close to the fitted distribution line and
- ii) Produces a minimum Anderson-Darling statistic with

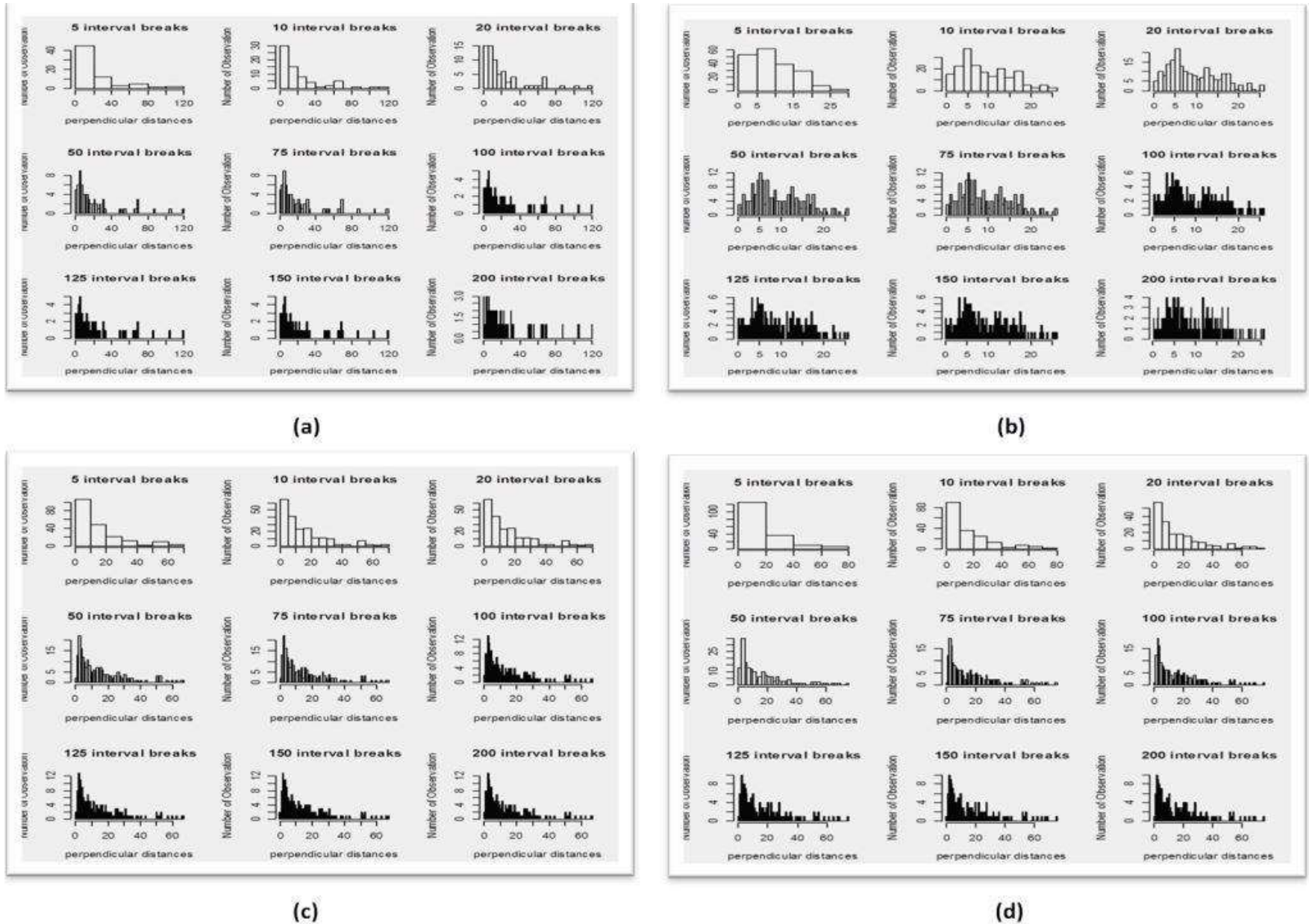


Figure 3. Histogram plots of observed perpendicular distances of (a) detected elephants (b) Warthogs (c) detected hartebeests (d) waterbucks.

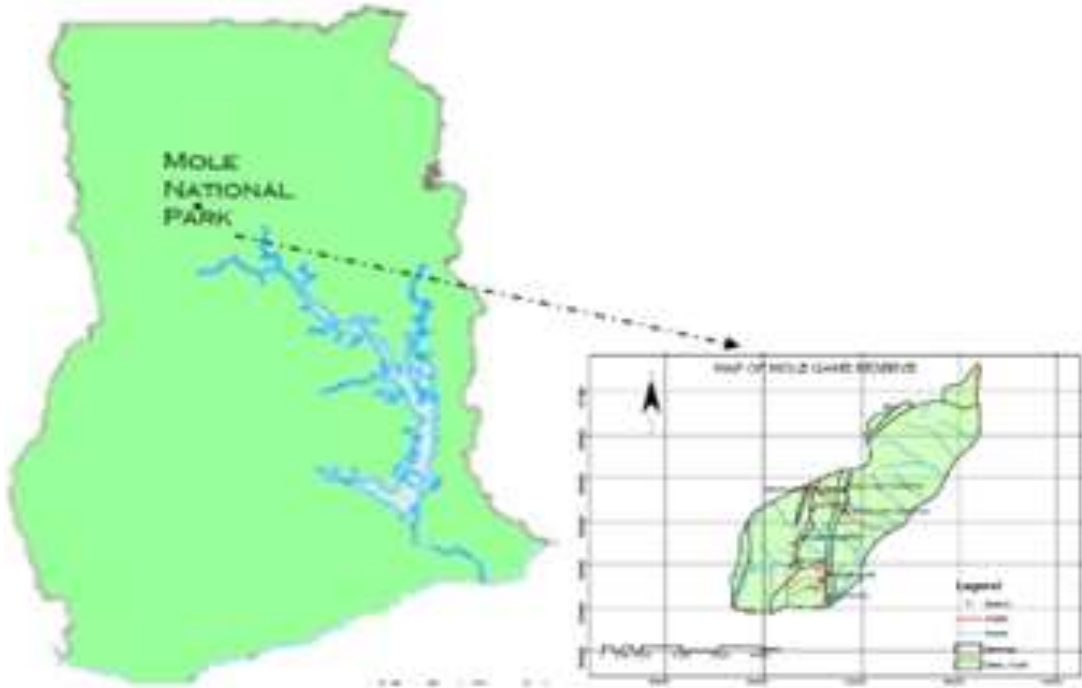


Figure 4. Map of Ghana with location of the Mole National Park.

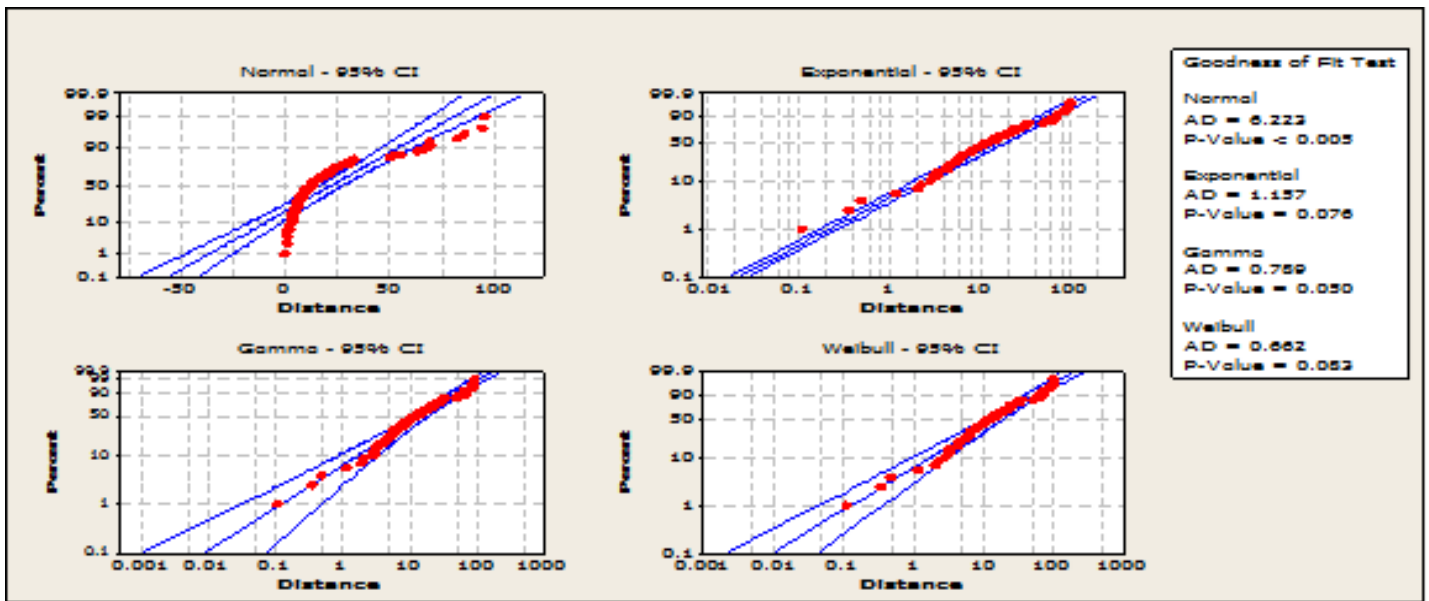


Figure 5. Plots of identified probability distributions of distances of detected elephants.

an associated probability greater than the selected significant level as observed in Figure 4.

In Figure 5, it can be observed that the Exponential, the Gamma and Weibull appear somewhat good to explaining the distribution of the elephant observations in the Mole National Park. The pdf plots of these functions are

represented in Figure 5.

Test of goodness-of-fit of the identified distributions

In Figure 6, the observed distances of the detected

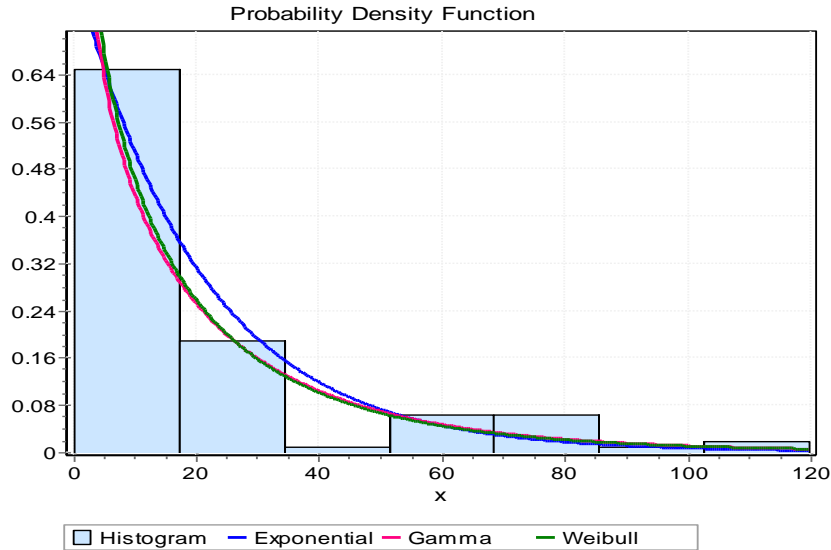


Figure 6. pdf plot of the most three important distributions of the elephant data.

Table 1. The Weibull distribution.

Kolmogorov-Smirnov					
Sample Size	111				
Statistic	0.07446				
P-Value	0.54461				
Rank	11				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.10184	0.11608	0.1289	0.14408	0.15462
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size	111				
Statistic	0.85124				
Rank	10				
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Degree of Freedom	6				
Statistic	4.8843				
P-Value	0.55874				
Rank	11				
Critical Value	8.5581	10.645	12.592	15.033	16.812
Reject?	No	No	No	No	No

elephants has clearly shown in the plots not to follow the normal distribution. It appears to somehow fit the Gamma and Weibull distributions. Even though the plotted points of the Exponential distribution appears to form a straight line, some points fell outside the confidence level. The smaller Anderson-Darling statistic of the Weibull with its greater probability appears more favorable to the

Gamma. Figure 5 reveals that the three identified distributions that best explain the distributions of elephants in the MNP all indicate a decreasing probability of detection of units as distance increases and satisfies basic assumptions of distance sampling.

From the goodness-of-test results as revealed in Tables 1, 2 and 3, it is evident that the Weibull

Table 2. The Exponential distribution.

Kolmogorov-Smirnov					
Sample Size	111				
Statistic	0.12223				
P-Value	0.06656				
Rank	26				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.10184	0.11608	0.1289	0.14408	0.15462
Reject?	Yes	Yes	No	No	No
Anderson-Darling					
Sample Size	111				
Statistic	2.6398				
Rank	17				
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	Yes	No	No
Chi-Squared					
Degree of Freedom	6				
Statistic	9.2302				
P-Value	0.16104				
Rank	22				
Critical Value	8.5581	10.645	12.592	15.033	16.812
Reject?	Yes	No	No	No	No

Table 3. The Gamma distribution.

Kolmogorov-Smirnov					
Sample Size	111				
Statistic	0.09289				
P-Value	0.27612				
Rank	19				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.10184	0.11608	0.1289	0.14408	0.15462
Reject?	No	No	No	No	No
Anderson-Darling					
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Degree of Freedom	6				
Statistic	14.511				
P-Value	0.02442				
Rank	27				
Critical Value	8.5581	10.645	12.592	15.033	16.812
Reject?	Yes	Yes	Yes	No	No

distribution actually fits the elephant data well as hypothesis confirmed that there exists no evidence to reject the statement that the Weibull distribution fits the

distance data better at all levels of significance.

Figure 7 represent a simulated observation which appears almost in the same form in shape as that

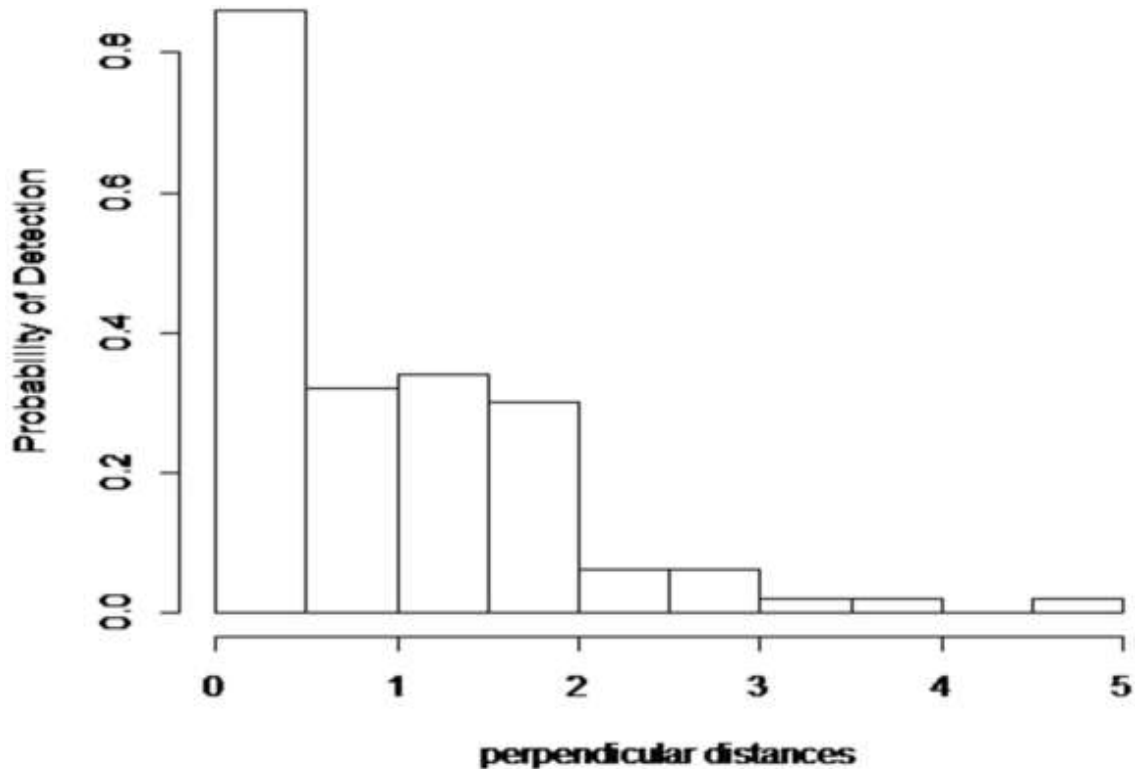


Figure 7. Simulated observation with the Weibull distribution.

of the identified Weibull of the observed elephants in the MNP. This confirms that any random observation over space and time within any region of study of the elephant species can be explained by the Weibull function.

Explored probability distributions of other observed mammals in the MNP

Figure 8 demonstrates combined plots of all possible functions capable of explaining the distributions of the Hartebeests, Waterbucks and Warthogs, respectively in the Mole National Park of Ghana. Testing for the best appropriate distribution using the Anderson-Darling, the Chi-square and the Kolmogorov statistics as previously observed, it was found that different functions appear to distinctively describe individual species. The Hartebeests and Waterbuck detections however appear to be described by the 3p-Weibull distribution whereas that of the warthogs fits well in the Johnson SB distribution which is assumed to follow the normal distribution after some transformation process.

Regression analysis

This statistical approach is adopted to determine or describe the relationship between a response variable

and one or more explanatory variables defined as covariates in the context of this study. With an assumption of linearity and dichotomous nature of this variable of interest, the regression took the approach of the general regression form to determine how detection of objects can be influenced by distances at various factor level combinations of the covariates.

Analysis of un-augmented observation of elephant data

The regression equation as earlier indicated exhibits the relationship between cluster size (number of detections) conditional on distance only. As observed in the equation, as distances increases, detection decreases and for every unit increase in distance, ability to detect is decreased by about 0.5%. In this model, only about 3% of the variability is explained by the presence of the predictor. The constant value in the regression only exhibits statistical significance at 95% level of confidence as Table 4 reveals. The analysis of variance predicts that there is no statistical evidence to suggest that the detection of elephants within the survey area is actually related to the distances at which detection can occur.

With inclusion of other categorical predictors, ability to detect is assumed to be conditional on distance, and the other categorical variables provided separate

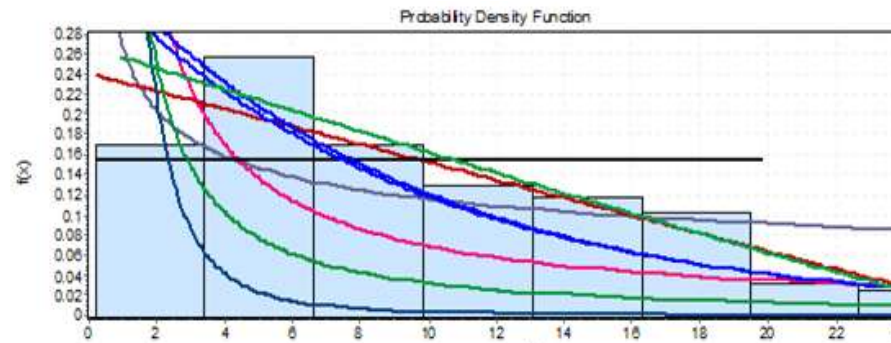


Figure 8a: Plots of observed Warthogs.

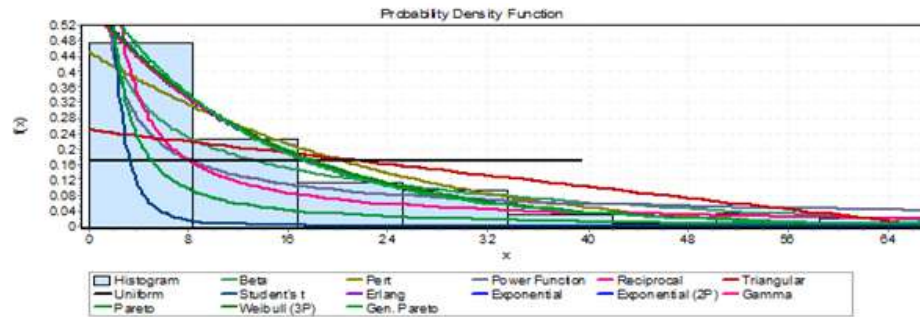


Figure 8b: Plots of observed Hartebeests

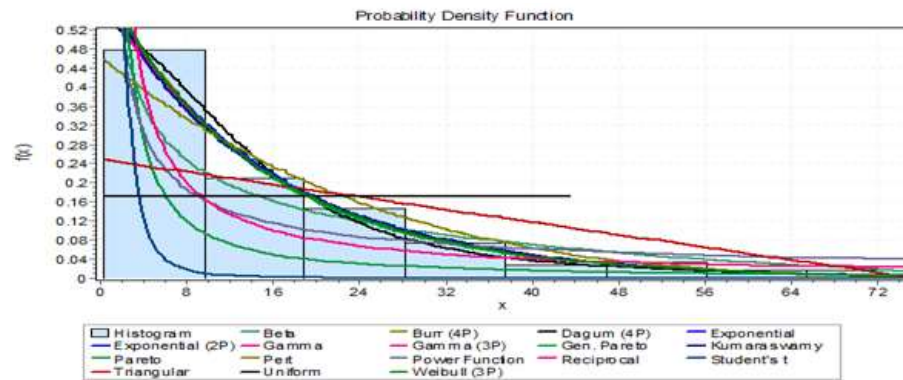


Figure 8c: Plots of Observed Waterbucks

Figure 8. pdf plots of possible functions of the observed Hartebeests, Waterbucks and Warthogs.

Table 4. Coefficients of un-augmented analysis.

Term	Coef	SE Coef	T	P	95% CI	VIF
Constant	1.71684	0.124493	13.7906	0.000	(1.46835, 1.96533)	
Distance	-0.00479	0.003590	-1.3354	0.186	(-0.01196, 0.00237)	1

Regression equation: $y = 1.71684 - 0.0047948 * \text{distance}$

Summary of model: S = 0.785451, R-Sq = 2.59%, R-Sq(adj) = 1.14%, PRESS = 43.8207; R-Sq(pred) = -3.27%.

Table 5. Coefficients of continuous and categorical predictors combined.

Term	Coef	SE Coef	T	P	95% CI	VIF
Constant	1.45859	0.232167	6.28248	0.000	(0.99350, 1.92367)	
Distance	-0.00572	0.003774	-1.51459	0.135	(-0.01328, 0.00184)	1.11963
Transect						
1	-0.05202	0.235775	-0.22062	0.826	(-0.52433, 0.42030)	2.37410
2	-0.12521	0.297361	-0.42108	0.675	(-0.72090, 0.47047)	2.82910
3	-0.10434	0.319820	-0.32624	0.745	(-0.74501, 0.53634)	3.07293
4	0.56585	0.320043	1.76805	0.082	(-0.07527, 1.20698)	3.07721
5	-0.73617	0.712195	-1.03367	0.306	(-2.16287, 0.69052)	9.92545
6	0.11741	0.260315	0.45104	0.654	(-0.40406, 0.63889)	2.54558
7	0.37943	0.376646	1.00740	0.318	(-0.37508, 1.13395)	3.68783
8	0.07337	0.374763	0.19577	0.845	(-0.67737, 0.82411)	3.65104
9	-0.38696	0.378675	-1.02187	0.311	(-1.14553, 0.37162)	4.02123
Session 1	0.22625	0.135681	1.66752	0.101	(-0.04555, 0.49805)	1.48653
Season 1	0.11488	0.173759	0.66113	0.511	(-0.23320, 0.46296)	1.55249

	Transect	Session	Season	
	1	1	1	Cluster size = 1.7477 - 0.00571612 distance
	3	2	2	Cluster size = 1.01312 - 0.00571612 distance
	4	1	1	Cluster size = 2.36557 - 0.00571612 distance
Regression equations:	10	2	2	Cluster size = 1.38609 - 0.00571612 distance

Summary of model: S = 0.780252; R-Sq = 19.66%; R-Sq(adj) = 2.44%.

regression equations for each level of each categorical variable in the model as indicated Table 5 wherein more units are likely to be detected at factor combination 4, 1, 1 for transect, session and season, respectively which exhibits statistically significance at not less than 10% level of significance as seen in Table 5. With this adjustment based on other categorical predictors, more of the variations appear to be explained by the model than their nonexistence as seen in the values of the R-Sq with detections made at transect line 4 indicating some statistical significance at the 1% significance level. There is also a change in the value of the S statistic from S = 0.785451 to S = 0.780252, indicating an improvement in the model with the inclusion of the categorical predictors and a greater reduction in ability to detect per every unit increase in distance (from 0.0047948 to about 0.00571612). The inclusion of the categorical predictors has also resulted in an increase of the Durbin-Watson Statistic from 2.12984 to 2.25769. The S statistic

represents the standard distance that data values fall from the regression line and for a given equation, the lower the S, the better the performance of the model in predictions.

Regression analysis of augmented observation

Augmentation is a process of data recording involving more than one species of observed units whilst adjusting for unobserved units of interest over space and time. In augmentation, it is assumed that if a particular species is observed at a particular distance in time T, then there exist same number of other species that appear undetected at a different location beyond the point where detections of one particular species is recorded. In distance sampling, a unit may appear detected or not with unobserved units not signaling total absence at that particular time. Augmentation thus allows all unobserved

Table 6. Coefficients of un-augmented analysis.

Term	Coef	SE Coef	T	P	VIF
Constant	1.01019	0.0002805	3601.36	0.000	
x	-0.00050	0.0000001	-3499.43	0.000	1

Regression equations: $y = 1.01019 - 0.000504838 * \text{distance}$

Summary of model: S = 0.00292523; R-Sq = 99.98%; R-Sq(adj) = 99.98%; PRESS = 0.0172989; R-Sq(pred) = 99.98%.

units to be catered for in data recording process.

Table 6 confirms that both constant and coefficient appear significant with the augmentation process, with a percentage of about 99.98 of variation being explained by the inclusion of the predictor variable as against 2.59% with the un-augmented case. This indicates a significant support of the augmented than the un-augmented form.

CONCLUSION AND RECOMMENDATIONS

The study has thus revealed that the four specified functions in the application software appear insufficient in modeling all animals and plants of different species since differences in biological characteristics of different species can be manifested in their detections and detectability processes.

In addition, data recordings and presentation in this type of study requires more generalization in a form described in this study as “augmentation” process. Detectability of the elephant species is observed to be explained mathematically as $g(x) = 1 - e^{[-(x/6.4)^{-1.5}]}$ in the classical perspective with the DISTANCE application software. With covariates inclusion, the detectability appears as $g(x) = 1 - \exp(-(x/16.33957)^{-2.507})$ wherein a change can be observed in both the scale and shape parameters as with the case of Marques et al. (2007) who indicated that performing a multiple covariate analysis affects only the scale parameter. With further exploration with possible functions that can perform an equal role or better based on probability of detection produced, the Weibull distribution appear to better explain the distribution of the elephant species than the hazard rate functions.

Finally, observation of units of interest conditional on distances has been modeled in a more general perspective with additional categorical predictors wherein all regression equations are provided at various factor level combinations with about 99.98% of explained variation in the model by the data augmentation process and less than 20% of explained variability for un-augmented case. It is therefore recommended that:

1) Further exploration should be performed prior to analysis and estimation to determine appropriate distribution that seems to appropriately describe specific units of interest under investigation.

2) Expansion of the functional space in the application DISTANCE software.

3) Application of data augmentation process in data representation with LTM involving observation with more than one species at a time.

4) Data augmentation is a necessary tool for data organization especially when a complete coverage of the entire study region is or may appear impossible or difficult.

5) Regression in a generalized form appears more appropriate in modeling with both continuous and categorical variables combined. It allows for separate equations to be provided at various factor level combinations of the categorical variables.

CONFLICT OF INTERESTS

The author has not declared any conflict of interests.

REFERENCES

- Anderson DR, KP Burnham (1976). Population ecology of the mallard: VI. The effect of exploitation on survival. U.S. Fish and Wildlife Service.
- Anderson DR, Pospahala RS (1970). Correction of bias in belt transects of immotile objects. J. Wildl. Manage.
- Buckland ST, Rexstad EA, Marques TA, Oedekoven CS (2015). Distance Sampling: Methods and Applications. Springer Cham Heidelberg New York Dordrecht London.
- Buckland ST, Anderson DR, Burnham KP, Laake JL (1993). Distance sampling Estimating abundance of biological populations. Chapman & Hall, London.
- Burnham KP, Anderson DR, Laake JL (1980). Estimation of density from line transects sampling of biological populations. Wildlife Monographs 72:1-202.
- Borchers DL, Stevenson BC, Kidney D, Thomas L, Marques TA (2015). A Unifying Model for Capture–Recapture and Distance Sampling Surveys of Wildlife Populations, J. Am. Stat. Assoc. 110(509):195-204.
- David L, Borchers, Samantha Strindberg (2002). Distance Sampling, John Wiley and Sons, Ltd, Chichester, in Encyclopedia of Environmetrics (ISBN 0471 899976). 1:544-552
- Eberhardt (1968). A preliminary appraisal of line transects. J. Wildl. Manage. 32(1):82-88
- Gates CE (1968). Simulation study of estimators for the line transect sampling method. Biometrics 25(2):317-328.
- Emlen JT (1971). Population densities of birds derived from transect counts. Auk 88:323-42.
- Hayne DW (1949). An examination of the strip census method for estimating animal populations, J. Wildl. Manage. 32(2):145-157.
- Hayward MW, de Tores PJ, Dillon MJ, Fox BJ, Banks PB (2005). Using faecal pellet counts along transects to estimate quokka (*Setonix brachyurus*) population density. Wildl. Res. 32:503-7.

- Johnson DS, Laake JL, Ver HJM (2010). A model-based approach for making ecological inference from distance sampling data. *Biometrics*, 66:310-318.
- Laake JL, Buckland ST, Anderson DR, Burnham KP (1993). *Distance Users Guide*. Version 2.0. Colorado Cooperative Fisheries and Wildlife Research Unit, Colorado State University, Colorado.
- Marques TA, Thomas L, Fancy S, Buckland ST (2007). Improving estimates of bird density using multiple-covariate distance sampling. *The Auk*, 124:1229-1243.
- Robinette WL, Jones DA, Gashwiler JS, Aldous CM (1974). Field tests of strip census methods. *J. Wildl. Manage.* 38(1):81-96.
- Sadler LMJ, Webbon CC, Baker PJ, Harris S (2004). Methods of monitoring red foxes *Vulpes vulpes* and badgers *Meles meles*: are field signs the answer? *Mammal Rev.* 34:75-98.
- Sen AR, Tourigny J, Smith GEJ (1974). On the line transect sampling method. *Biometrics* 30(3):329-340.
- Seber GAF (1986). A review of estimating animal abundance. *Biometrics* 42:267-92.
- Sullivan BJ, Baxter GS, Lisle AT (2002). Low-density koala (*Phascolarctos cinereus*) populations in the mulgalands of south-west Queensland. I. Faecal pellet sampling protocol *Wildl. Res.* 29:455-62.
- Buckland ST, Oedekoven CS, Borchers DL (2016). Model-Based Distance Sampling. *J. Agri. Biol. Environ. Stat.* 21:1-58
- Thomas L, Buckland ST, Rexstad EA, Laake JL, Strindberg S, Hedley SL, Bishop JRB, Marques TA, Burnham KP (2010). *Distance software: design and analysis of distance sampling surveys for estimating population size*. *J. Appl. Ecol.* 47:5-14. DOI: 10.1111/j.1365-2664.2009.01737.x
- Thompson SK (2002). *Sampling*, 2nd edn, Wiley, New York.

Related Journals:

