AN INTRODUCTION TO LINE TRANSECT SAMPLING AND ITS APPLICATIONS

Bright Owusu
Department of Mathematical Sciences
Montana State University, Bozeman

April 16, 2019

APPROVAL

of a writing project submitted by

Bright Owusu

This writing project has been read by the writing project advisor and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

Date	John Borkowski Writing Project Advisor
Date	Mark C. Greenwood Writing Projects Coordinator

Contents

Abstract 2
1. Introduction 3
    1.1 Background 3
    1.2 Line Transect Sampling 3
    1.3 Selecting line Transect 4
    1.4 Detection 5
    1.5 Assumptions 5
2 Model Selection 7
    2.1 Models 7
    2.2 Model Selection 7
        2.21 Akaike Information Criterion(AIC) 7
        2.21 Likelihood ratio test 8
        2.23 Goodness of fit 8
    2.3 Abundance estimation 9
3 Simulation Study 10
    3.1 Data 10
    3.2 Research Questions 10
    3.3 Detection function 11
    3.4 Model fitting 11
    3.5 Abundance Estimation 14
        3.51 Using an exponential detection function 14
        3.52 Using a half-normal detection function 16
4. Results and Discussion 17
References 19

Abstract

Estimation of abundance and density of animals and plants population is essential for conservation and management in ecology. The study introduce the methods of line transect sampling, a type of distance sampling method used in ecology. Line transect sampling was first developed by R.T.Kings to estimate abundance of animals and plants in 1930s. Perpendicular distances of detected objects are used to estimate density and abundance of objects. The study describes randomly setting a single and multiple transect lines on the study region and introduce a detection function to determine the objects of interest that are detected based on the detection function. Model selection method using AIC and goodness of fit test is used to select the best models for the detection function. We describe how researchers can obtain estimates of abundance and density from the selected model using the "Distance" program in R. A simulation study of longleaf pine trees is used to demonstrate the method of line transect sampling. A transect line is first randomly placed at 120m on a 200m by 200m study region and three(3) distributions (Half-normal, Hazard-rate and Uniform with cosine adjustment) were considered to model the detection function. A 1000 simulated samples were used and estimates of abundance of longleaf pine trees were compared across the models. In addition, multiple transect lines are set on the study region and the simulated study is repeated.

1. Introduction

1.1 Background

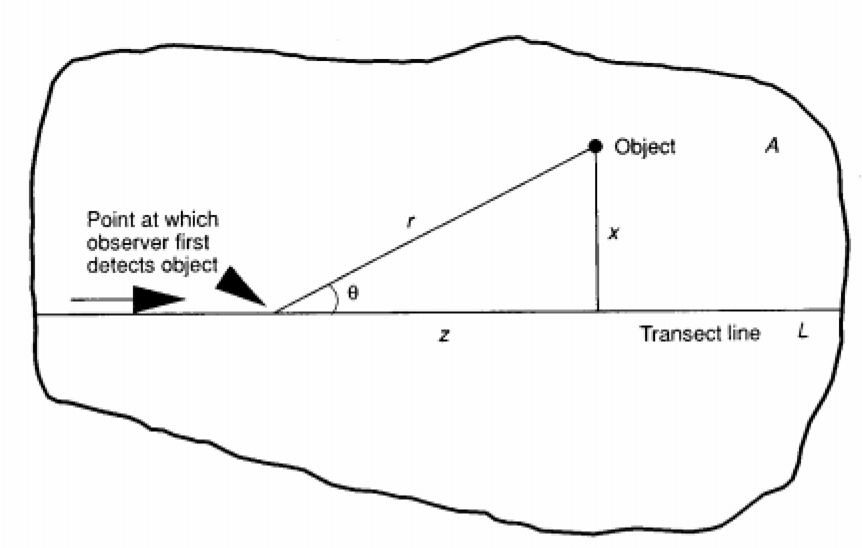
Several methods of estimating the density of plants and animals have been proposed. Although no method is bias free, the most accurate density estimates are obtained from complete counts [Davenport et al., 2007; McNeilage et al.,2001]. However, these methods require sampling effort that is often impractical, especially over large areas. The line transect sampling method is the most practical method [Plumptre,2000; Struhsaker,1997] in most cases. The line transect sampling provides a convenience method of estimating the number of objects in a study. The objects may be any species of animal or plant that is easily visible, at least at close range.

1.2 Line Transect Sampling

Distance sampling is a widely used group of closely related methods for estimating the density and/or abundance of biological populations. The main methods are line-transect sampling and point-transect sampling. In both cases, observer(s) perform a standardized survey along a series of randomly located lines or points, searching for objects of interest (usually animals or clusters of animals). For each object detected, they record the distance from the line or point to the object. Not all objects will be detected, but a fundamental assumption of the basic methods is that all objects that are actually on the line or point are detected. The key to distance sampling analyses is to use the observed distances to fit a detection function that describes how detectability decreases with increasing distance from the transect. The fitted function is used to estimate the average probability of detecting an object, this enables one to readily obtain point and interval estimates for the density and abundance of objects in the survey area.

There are two basic methods of distance sampling, line transect sampling and point transect sampling. In line transect sampling, a series of lines is distributed according to some design (usually a systematic grid of parallel lines and an observer travels along each line, searching for animals or animal clusters). For each animal or cluster detected, the observer measures or estimates the (perpendicular) distance x of the animal or cluster centre from the nearest part of the line. In many surveys, it is easier to measure or estimate the observer-to-animal (radial) distance r at the time the detection is made. If the sighting angle θ is also measured, then the perpendicular distance may be found by simple trigonometry, x = r $\times$ sinθ ($Fig.$1). The line transect sampling is associated with a detectability function g that indicates the detectability of an animal at a given location. The detectability function g is often a decreasing function of distance from the line transect and detectability on the line is assumed to be perfect. That is, if x is the perpendicular distance to the line, then $g(x)$ decreases as x increases and g(0) = 1. Under certain circumstances, animals may avoid the observer which can result in less than perfect detectability on the line (g(0) < 1) while the maximum detectability occurs at some distance x_{max} from the line. We use the recorded distance x of the detected animals from the line to model the *detection function*, $g(x)$, which is defined to be the probability of detecting an animal that is at distance x from the line.

Figure 1: A single transect line



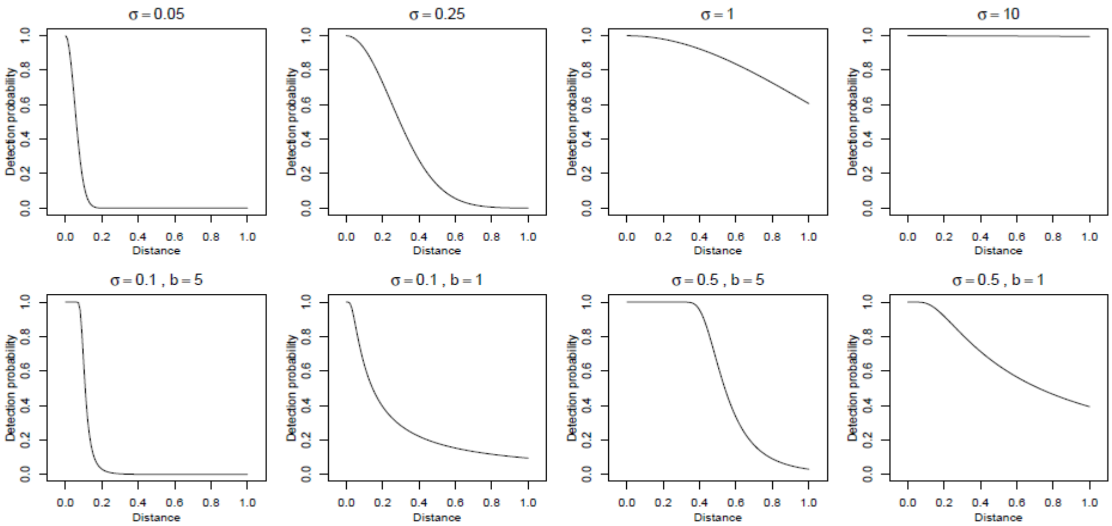
1.3 Selecting line Transect

The sampling design in a line transect study is the procedure by which the transect locations are selected. In line transects sampling, several sampling methods may be considered in selecting the line transects. Usually, single or multiple transect lines are randomly placed on the study region. Estimates of abundance from multiple transect lines are obtained by averaging the abundance from each transect line used.

1.4 Detection

In Line transect sampling, detection function is a key concept in estimating abundance of species. The detection function *g(x)* is the probability of detecting an object, given that it is at distance x from the transect line. The distance x refers to the perpendicular distance from the transect line. The detection function decreases with increasing distance but lies between 0 and 1. Thus, $0 \leq g(x) \leq 1$. When object is located on the line, then we assume perfect detection such that the detection function is $g(0) = 1$. This indicates that the object on the line is detected with certainty. *Figure 2* displays a graph of the detection function *g(x)* for a given distance x.

Figure 2: Graph of the detection function



The graph above represents half-normal (top row) and hazard-rate(bottom row) detection functions without adjustments, varying scale(σ) and (for hazard-rate) shape (b) parameters(values are given above the plots). On the top row from left to right, the study species becomes more detectable as the parameter, σ increases. The bottom row shows the hazard-rate model's with different parameters. In detection of objects, only small percentage of the object of interest are detected in the study region. Although this is the case, analysis of the associated distances allows reliable estimates of true density to be made.

1.5 Assumptions

In line transect sampling method, the design is associated with four main assumptions which includes;

- Animals are distributed independently of the lines.
- Objects on the line are detected with certainty.
- Distance measurements are exact.
- Objects are detected at their initial locations.

The assumption that animals are distributed independently of the line is characterized as one of the key design assumption. We assume that animals are distributed uniformly with respect to distance from the line. This assumption will hold based on a suitable randomized design (Strindberg et al. 2004). Thus, lines should be positioned according to a random design. However, we cannot ensure that assumptions related to the model holds by adopting a suitable design. Instead, we need to consider whether field methods can be adopted that will ensure low bias when they fail. For example, if animals show responsive movement, then field methods should if possible ensures that animals are detected and their locations recorded before they respond. This will ensure the assumption that objects are detected at their initial locations to be reasonable. Generally, movement independent of the observer causes no problems, unless the object is counted more than once on the same unit of transect line or if it is moving at at roughly half the speed of the observer or faster. Animal movement after detection is not a problem, as long as the original location can be established accurately and the appropriate distance measured. It is problematic when animals move to the vicinity of the next transect in response to disturbance by the observer. However, if movement is random, or at least not systematically in a single direction, then animals moving in one direction will tend to be compensated by animals moving in the other direction. It is also assumed that, objects on the line are detected with certainty. This assumption ensures that all objects at zero distance are detected such that $g(0) = 1$. Practically, detection on or near the line should be nearly certain. Distance measurements are assumed to be exact in line transect sampling. Thus, there is no measurement errors or recording errors of the distances and angles. Rounding errors in measuring angles near zero are problematic, most often in the analysis of ungrouped data. If errors in distance measurements are random and not too large, then reliable density measurements are still likely, given that the sample size is large. (Gates et al 1985).

2. Model Selection

2.1 Models

In the line transect sampling, several models may be considered for the detection function. However, the models for the detection function are expected to have the following propoerties (Buckland et al. 2015);

- *Shoulder*: We expect observers to be able to see objects near them, not just those directory infront of them. For this reason, we would expect the detection function to be flat near the line.
- *Non-increasing*: The non-decreasing property suggests that, observers are less likely to see distance objects than those nearer the transcript.
- *Model robust*: The model should be flexible enough to fit many different shapes.
- *Estimator Efficiency*: In modeling the detection function, we would like models to have low variance, given the above properties are satisfied.

In modeling the detection function, we would consider three(3) model types as key functions in this study. These models include, the Uniform, Half-normal and the Hazard-rate models. Adjustment terms would be included in the models to improve model fit. The adjustment terms considered for each model is based on the distribution of the distance data. If the survey results in good distance data, which exhibit a shoulder and with adequate sample size, the choice of model is unlikely to affect the abundance estimate much. However, if survey design results in poor distance data, models fitted to the data might yield different estimates. Generally, histogram of the distance data is plotted and visualize to identify any outliers, heaping, measurement errors, and movement prior to detection.

2.2 Model Selection

2.21 Akaike Information Criterion(AIC)

In the model selection approach, the study would considered the AIC method of model selection. The Akaike Information Criterion (AIC) provides quantitative method for model selection. The criterion is given by;

$$AIC \ -2log_e L \ 2q$$

where, L is the maximum likelihood and q is the number of estimated parameter. Once we have a set of plausible models, we can use the Akaike Information Criterion (AIC) to select between models. The model with the lowest AIC is selected for analysis and inference. Generally, if the difference between AICs is less than 2, we may investigate multiple "best" models, potentially resorting to the simplest of these models. This is because models with similar AICs would results in similar estimates probabilities of detection.The AICs between models would be compared for the same data with same truncation as AIC can only be used to compare models fitted to exactly the same data.

2.22 Likelihood ratio test

In testing between hierarchical models, the likelihood ratio test is used for choosing the number of adjustment terms to include in the model. Suppose that a fitted model(Model 1) has k_1 adjustment terms, the likelihood ratio test allows an assessment of whether the addition of k_2 terms improves the adequacy of a model. In hypothesis testing, the null states that Model 1 is the true model and the alternative states that Model 2 with all k_1 k_2 adjustment terms is the true model. The test statistics is given by;

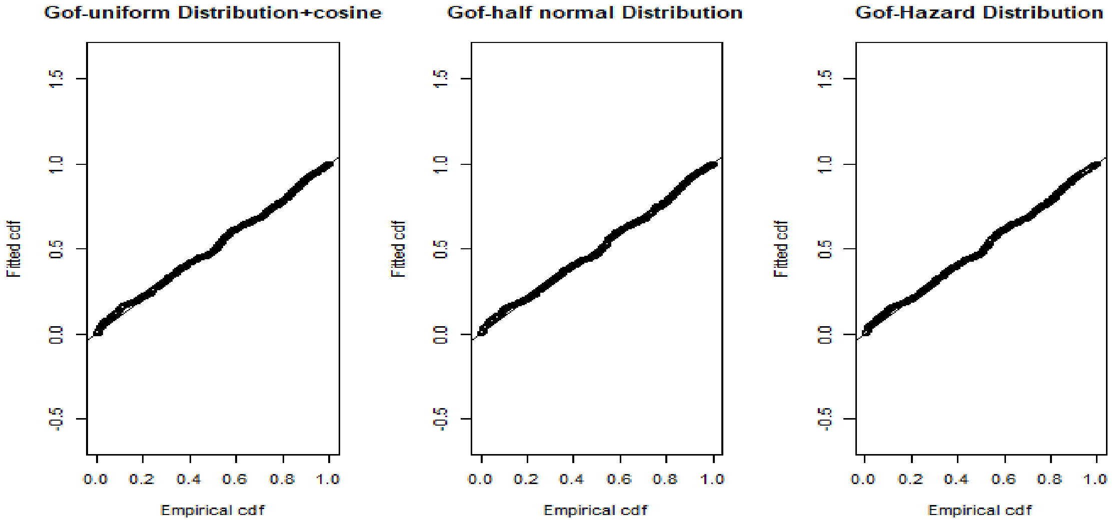
$$\chi^2 \ -2log_e(L_1/L_2) \ -2[log_e(L_1) - log_e(L_2)]$$

where L_1 and L_2 are the maximum value of the likelihood functions for Models 1 and 2 respectively. If model 1 is the true model, the test statistic follows a χ^2 distribution with k_2 degrees of freedom. In line transect sampling, models are fitted to the key function and lower order adjustment terms are fitted. If the adjustment term improve the fit, the next term is added and the likelihood ratio test is carried out. The process is repeated until the test is not significant.

2.23 Goodness of fit

In model selection for a line transect sampling, goodness of fit can be useful in assessing the quality of the distance data and understanding the general shape of the detection function. Models fitted to the distance data could have significantly poor fit. This need not be of great concern, as it provides a warning of a problem in the data or the selected detection model structure. This could be investigated by closer examination of the data or by exploring other models and fitting options. Often, a good model will give a significant goodness of fit statistics. Figure 3 below indicates example plots of goodness of fit to a distance data for Half-normal, hazard-rate and uniform with cosine adjustment models.

Figure 3: Graph of goodness of fit



2.3 Abundance estimation

The line transect sampling consider fitting the detection function as the primary modeling step followed by estimating density or abundance. The estimate of abundance is computed for the surveyed area based on the estimate of detection probability. Usually, researchers are interested in the number of objects estimated to be within the study region given a detection function. To begin, the density is estimated based on the number of objects detected(n), total area under the detection function (a) and the total length of the transect used(L). Mathematically, the density is given by;

$$D \quad \frac{n}{2La}$$

where $a \; \int_0^w g(x)dx$

The probability density function, f(x) of the detected distances is obtained as;

$$f(x) \quad \frac{g(x)}{a}$$

At zero distance of objects to the transect line, the detection function is g(0) = 1. Hence,

$$\Theta \quad \frac{1}{f(0)}$$

and density now becomes

$$D \quad \frac{nf(0)}{2L}$$

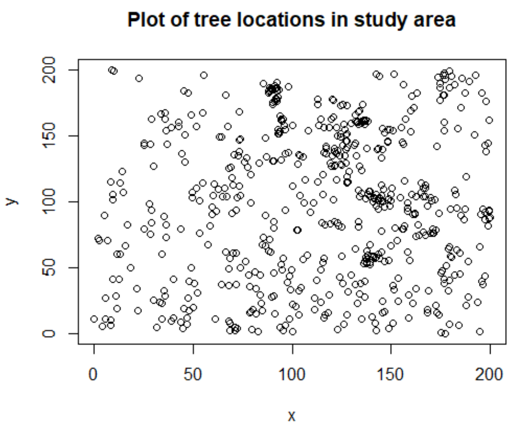
The estimate of abundance or population size is obtained as the product of the area under the study region(A) and the estimated density.

$$N \, A D \quad \frac{Anf(0)}{2L}$$

3. Simulation Study

3.1 Data

The data contains 584 longleaf pine tree locations (x,y) in a study region of 200m by 200m forest in Thomas County, Georgia. Perpendicular distances of trees from the transect line are measured as the variable of interest.



3.2 Research Questions

- When transect line is fixed at 120m, what is the estimated abundance of pine trees in the study region?
- When multiple transect lines are set at 60m,90m,120m and 150m, what is the average estimated abundance of pine trees in the study region?
- How does the estimated abundance of longleaf pine trees change for different detection functions?

3.3 Detection function

The detection of pine trees is assume to follow an exponential distribution function. The detected pine trees were obtained based on an exponential detection function with an estimated α parameter of 0.019. Mathematically, the detection function is given by;

$$g(x) \equiv \exp(-\alpha x)$$

where x is the distance of each pine tree from the transect line.

We applied this detection function to obtain detected pine trees when a single transect line and multiple transect lines were placed on the study region (200m by 200m forest in Thomas county, Georgia).

For further analysis, a half normal detection function with an estimated σ of 0.019 was also considered to obtain detected pine trees and distance of each detected pine tree to the transect line was recorded. Mathematically, the detection function is given by;

$$g(x) \equiv \exp(-\frac{x^2}{2\sigma^2})$$

where x is the distance of each pine tree from the transect line.

We applied this detection function to obtain detected pine trees when a single transect line and multiple transect lines were placed on the study region (200m by 200m forest in Thomas county, Georgia).

3.4 Model fitting

(A) Model fit using exponential detection function

Three(3) distributions (Half-normal, Hazard-rate, and Uniform with cosine adjustment) were considered to model the detection function based on the distances of detected pine trees. At a single transect line placed at 120m on the study region, the plots below indicates the fit of the three(3) distributions to the distances of the detected pine trees(based on an exponential detection function) for a given simulated case.

Figure 4: Model fit to detected pine trees for first simulated case

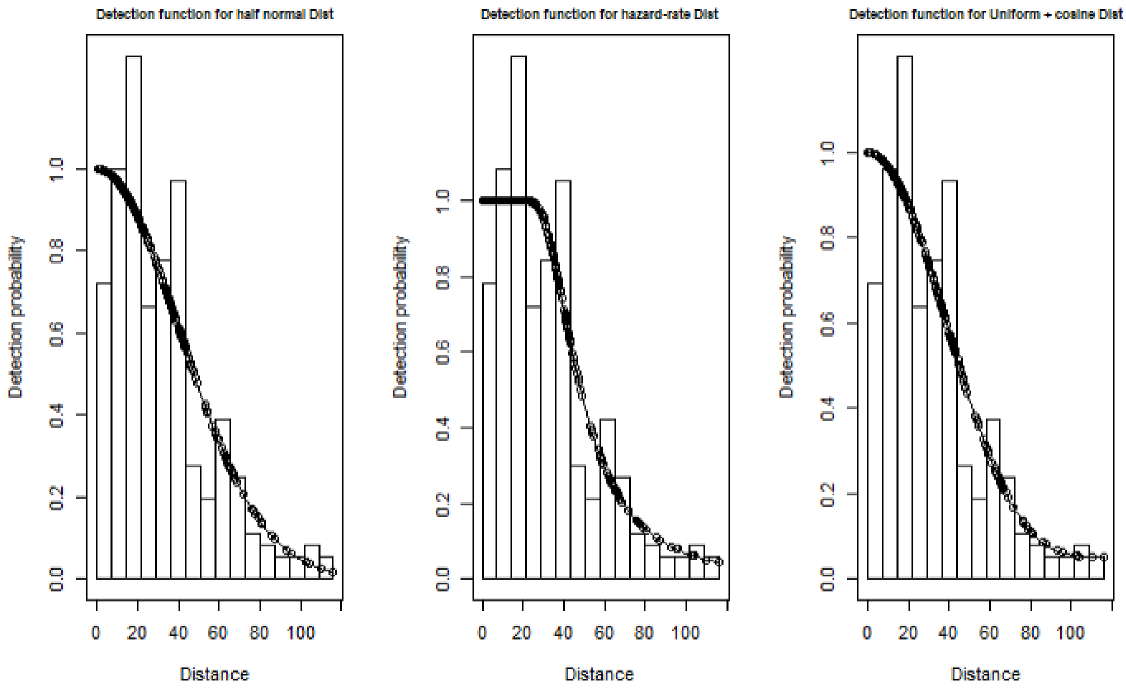
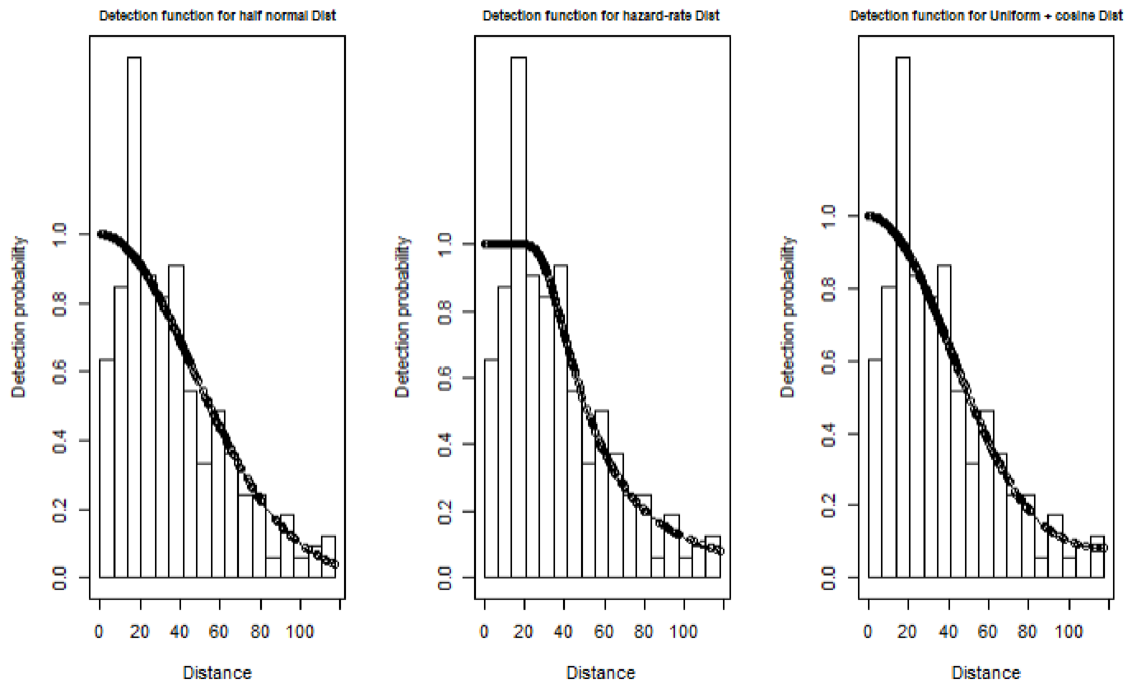


Figure 5: Model fit to detected pine trees for the tenth simulated case



(B) Model fit using half normal detection function

Three(3) distributions (Half-normal, Hazard-rate, and Uniform with cosine adjustment) were considered to model the detection function based on the distances of detected pine trees. At a single transect line placed at 120m on the study region, the plots below indicates the fit of the three(3) distributions to the distances of the detected pine trees(based on an exponential detection function) for a given simulated case.

Figure 6: Model fit to detected pine trees for first simulated case

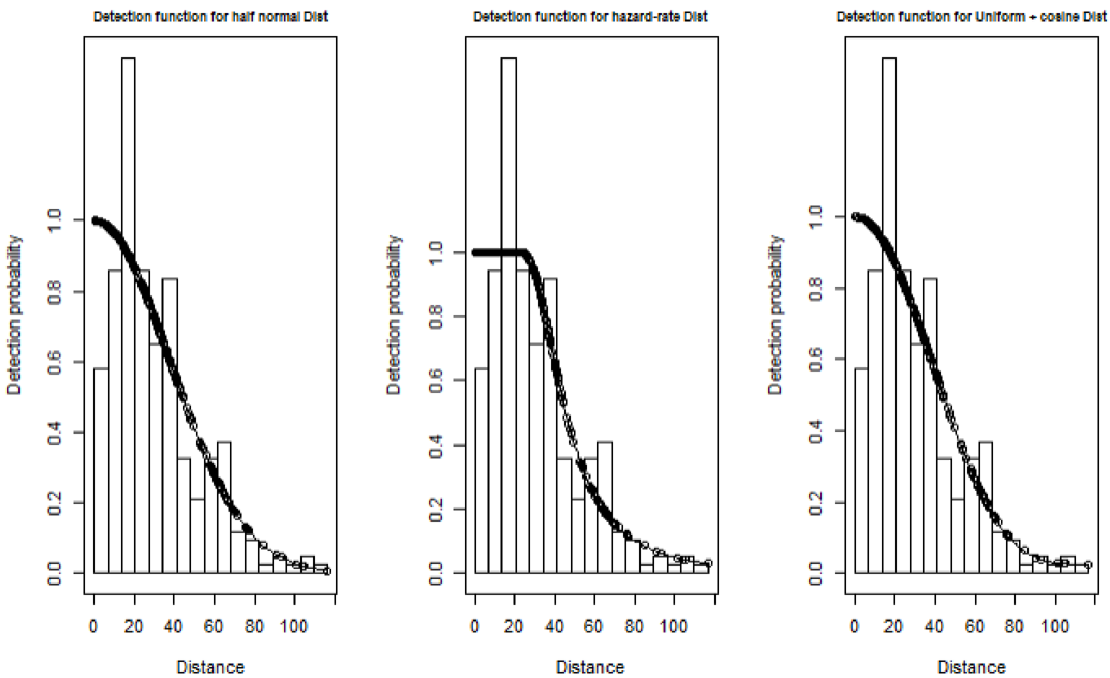
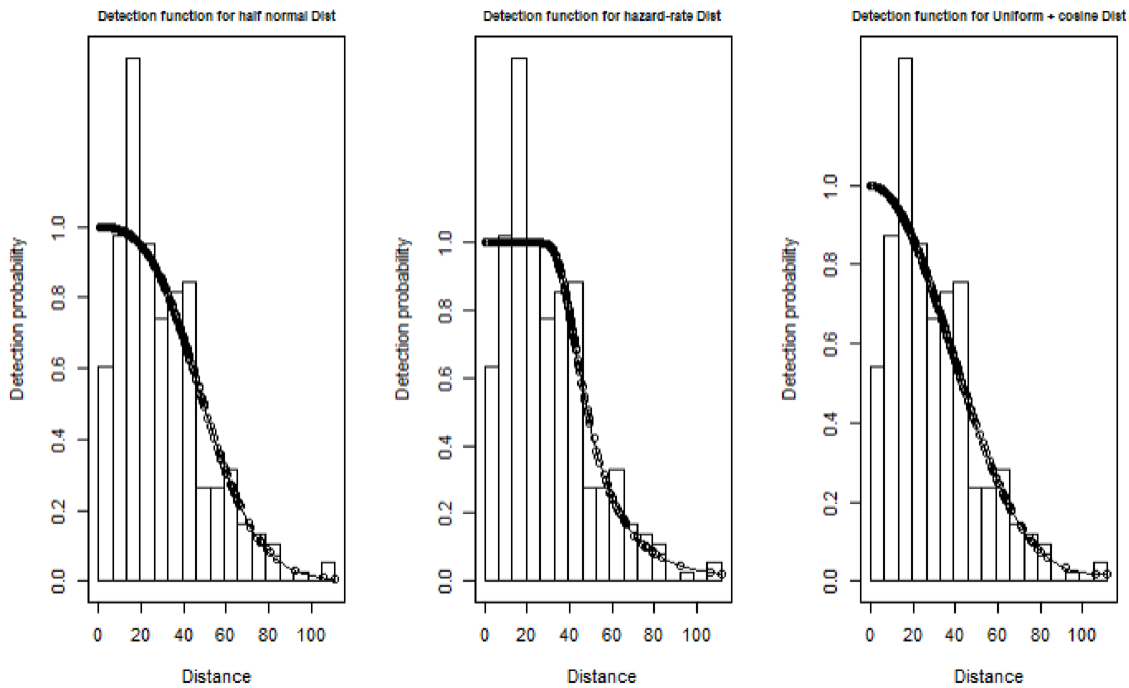


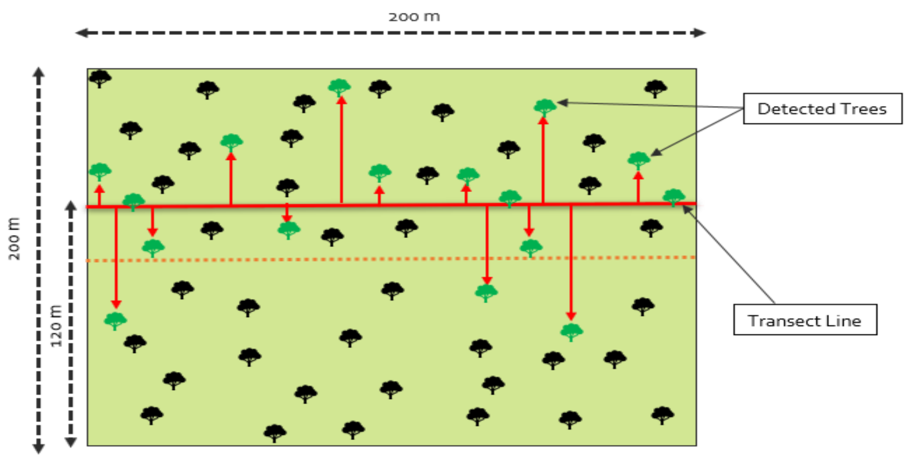
Figure 7: Model fit to detected pine trees for the tenth simulated case



3.5 Abundance Estimation

3.51 Using an exponential detection function

(A) Single transect line set at 120m.

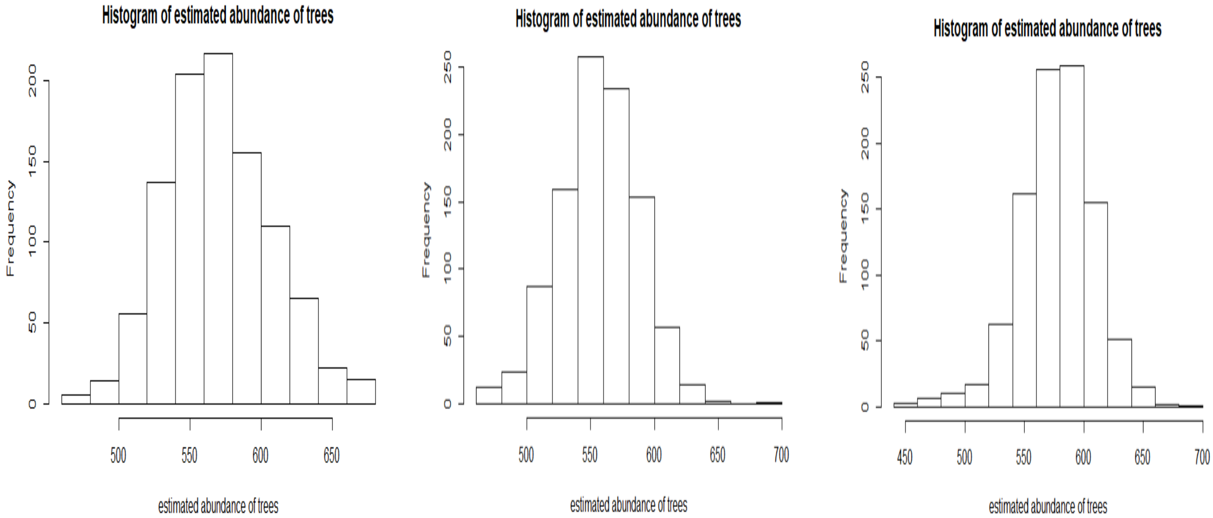


For 1000 simulated cases of re-sampling the same transect line set at 120m on the study region, the average abundance of the pine trees were estimated for the three(3) distributions. Table 1 below displays the average abundance, average standard deviation, average bias and a 95% confidence interval for each distribution.

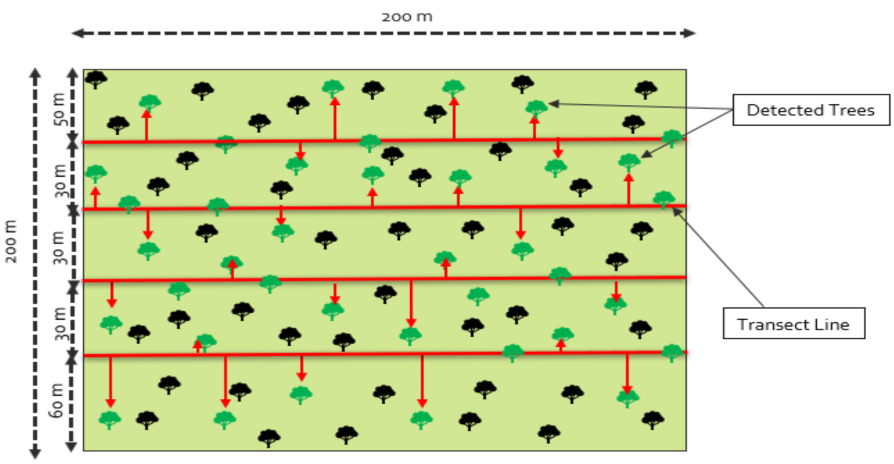
Table 1

Model	Avg.Ab	Avg.Bias	Avg.Sd	95% Conf.Int.
Half-normal	569.80	-14.20	45.76	(480.10, 659.49)
Hazard-rate	556.56	-27.44	47.14	(464.16, 648.97)
Uniform + Cos(1)	577.41	-06.59	40.74	(496.79, 658.04)

Figure 8 below displays the distribution of the average abundance from 1000 simulated cases for Half-normal, Hazard-rate and Uniform with cosine adjustment distributions.



(B) Multiple transect lines set at 60m,90m,120m and 150m.



For 1000 simulated cases of re-sampling multiple transect lines set at 60m,90m,120m and 150m on the study region, the average abundance of the pine trees were estimated for the three(3) distributions. Table 2 below displays the average abundance, average standard deviation, average bias and a 95% confidence interval for each distribution.

Table 2

Model	Avg.Ab	Avg.Bias	Avg.Sd	95% Conf.Int.
Half-normal	443.86	-140.14	40.48	(364.51, 523.21)
Hazard-rate	432.91	-151.09	44.80	(345.11, 520.71)
Uniform + Cos(1)	449.56	-134.44	34.93	(381.11, 518.02)

The average abundance of pine trees are underestimated for Half-normal, Hazard-rate and Uniform with cosine adjustment distributions.

3.52 Using a half-normal detection function

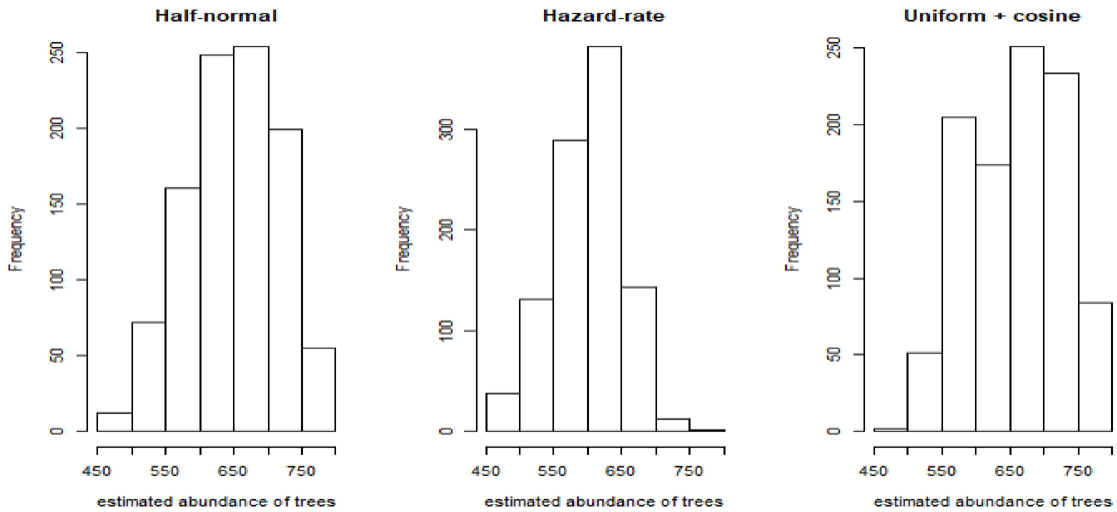
(A) Single transect line set at 120m.

For 1000 simulated cases of re-sampling the same transect line set at 120m on the study region, the average abundance of the pine trees were estimated for the three(3) distributions. Table 3 below displays the average abundance, average standard deviation, average bias and a 95% confidence interval for each distribution.

Table 3

Model	Avg.Ab	Avg.Bias	Avg.Sd	95% Conf.Int.
Half-normal	648.6	64.58	51.41	(547.8, 749.3)
Hazard-rate	600.7	16.68	38.98	(524.3, 677.1)
Uniform + Cos(1)	657.1	73.13	44.24	(570.4, 743.8)

Figure 9 below displays the distribution of the average abundance from 1000 simulated cases for Half-normal, Hazard-rate and Uniform with cosine adjustment distributions.



(B) Multiple transect lines set at 60m, 90m, 120m and 150m.

For 1000 simulated cases of re-sampling multiple transect lines set at 60m, 90m, 120m and 150m on the study region, the average abundance of the pine trees were estimated for the three(3) distributions. Table 4 below displays the average abundance, average standard deviation, average bias and a 95% confidence interval for each distribution.

Table 4

Model	Avg.Ab	Avg.Bias	Avg.Sd	95% Conf.Int.
Half-normal	648.43	64.41	51.19	(548.10, 748.75)
Hazard-rate	600.10	16.10	38.80	(524.05, 676.15)
Uniform + Cos(1)	657.23	73.23	43.86	(571.27, 743.18)

The average abundance of pine trees are overestimated for Half-normal, Hazard-rate and Uniform with cosine adjustment distributions.

4. Results and Discussion

Line transect sampling method was demonstrated from a simulated study of longleaf pine trees in a study region of 200m by 200m in Thomas County, Georgia. A single transect line set randomly at 120m with an exponential detection function was used to estimate the abundance of longleaf pine trees in the study region. Three models(Half-normal, Hazard-rate and Uniform with cosine(1) adjustment) were considered for the simulated study. We found that all models underestimated abundance of longleaf pine trees. However, on average the uniform with cosine(1) adjustment model had the largest estimate of abundance with the smallest bias. Again, when multiple transect lines were set at 60m, 90m, 120m and 150m, the uniform with cosine(1) adjustment model had the largest estimate of abundance with the smallest bias. Comparatively, we found that the average bias was higher for the multiple transect line than the single transect line. Generally, we cannot conclude this will always be the case as the difference could be attributed to lower estimates of abundance from transect lines closer to the end of the study region.

We also assessed abundance of logleaf pine trees using half-normal detection function. The simulated process was repeated with a single transect line set at 120m on the study region and the estimate of abundance using all models(Half-normal, Hazard-rate and Uniform with cosine(1) adjustment) were considered. The abundance of longlef pine trees were overestimated for all models. However, on average, the hazard-rate model provided the best estimate of abundance with the smallest bias. We also found that, the hazard-rate model provided the best estimate of abundance with the smallest bias when multiple transect lines were set at 60m, 90m, 120m and 150m. We noticed that, the uniform with cosine(1) model provided poor estimate of abundance with large bias among all the models when half-normal detection function is used.

From the simulated results, we observed that the model which provides the best estimate of abundance when single transect line is used also provides the best estimate of abundance when multiple transect lines are used on the study region. However, the model which provides the best estimate of abundance differ when different detection function is used. These results could change if several models were used in estimating the abundance of trees. It is therefore not certain to make a general conclusion on which model would be best for a given detection function when using a line transect sampling.

In summary, line transect sampling method appeared to provide a better estimate of abundance in a given study region. The estimate of abundance is dependent on which detection function is considered. The choice of the detection function and its parameter(s) may lead to underestimating or overestimating of abundance in a given study region. In addition, all assumptions are needed to be satisfied to ensure efficient use of informations to yield a better estimate of abundance. Again, several models are worth considering in modeling the detection function for a better estimate of abundance. Future studies may consider advance method of line transect sampling where truncation of the data are useful in estimating of abundance.

References

Anderson, D.R., Burnham, K.P. and Crain, B.R.(1985a). Estimating population size and density using line transect sampling. *Biomedical Journal*, 27, 723-31.

Anderson, D.R., Laake, J. L., Crain, B.R. and Burnham, K.P (1979). Guidlines for line transect sampling of biological population. *Journal of Wildlife Management*, 43, 70-8.

Brockelman, W. Y. (1980). The use of the line transect sampling method for forest primates, in *Tropical Ecology and Development* (ed. J.I.Furtado), The international Society of Tropical Ecology, Kuala Lumpur, Malaysia, pp. 367-71.

Buckland ST, Rexstad EA, Marques TA, Oedekoven CS(2015).*Distance Sampling: Methods and Applications.* Method in Statistical Ecology. Springer International Publishing.

DeVries, P.G. (1979a). Line transect sampling - statistical theory, applications, and suggestions for extended use in ecological inventory, in *Sampling Biological Populations* (eds R.M. Cormack, G.P. Patil and D.S. Robson), International Co-operative Publishing House, Fairland, MD, USA, pp. 1-70.

Eberhardt, L.L. (1978a). Transect method for popualtion studies. *Journal of Wildlife Management,* 42, 1-31.

Gates, C.E., Evans,W., Gober,D.R.,Guthery,F.S. and Grant, W.E. (1985). Line transect estimation of animal densities from large data sets.In *Game Harvest Management* (eds. S.L.Beasom and S.F Roberson), Caesar Kleberg Wildlife Research Institute, TX, pp. 37 - 50.

Rao, P.V., Portier, K.M. and Ondrasik, J.A. (1981). Density estimation using line transect sampling, in *Estimating Numbers of Terrestrial Birds.*. Studies in Avian Biology No. 6 (eds C.J. Ralph and J.M.Scott), Cooper Ornithological Society, pp. 441-4.

Strindberg, S., S. T. Buckland, and L. Thomas (2004).Design of distance sampling surveys and geographic information systems. pp. 190–228 in S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas (Eds.), *Advanced Distance Sampling* .