RT&A, No 4(80)
Volume 19, December, 2024

A NEW COMPACT DETECTION MODEL FOR LINE TRANSECT DATA SAMPLING

Ishfaq S. Ahmad[1,] , Rameesa Jan[2]*,

•

[1] Department of Mathematical Sciences,
Islamic University of Science and Technology, Awantipora, J&K, India;
[2] Department of Statistics,
Government Degree College Sopore, Baramulla, J&K, India;
[1] peerishfaq007@gmail.com [2] bhatrumaisa2@gmail.com

Abstract

A new parametric model is proposed in line transect sampling for perpendicular distances density functions. It is simple, compact and monotonic non increasing with distance from transect line and also satisfies the shoulder condition at the origin. Numerous interesting statistical properties like shape of the probability density function, moments, and other related measures are discussed. Method of Moments and Maximum Likelihood Estimation is carried out .Applicability of the model is demonstrated using a practical data set of perpendicular distances and compared with other models using some goodness of fit tests.

Keywords: Line transect; shoulder conditions; detection function; maximum likelihood estimation; perpendicular distance.

1. Introduction

Line transect approach is a key technique for determining the population abundance(D) or density of objects in a study region(A). These objects may be species of animals, birds or plants that are easily visible at close range (Buckland et al. [**?**], Buckland et al. [**?**] and Barabesi [**?**]). It is the easiest, most useful and inexpensive of all the population abundance estimation. In a typical application of Line transect method, an observer walks a straight path of length L, noting all the animals seen (n) and their right-angle distances from the transect line (x).

In order to determine D from this data, a model is required which can be mathematically represented by conditional function $h(x)$ knows as detection function which is defined as :

$$h(x) = \Pr(\text{an object is detected given its perpendicular distances x from line})$$

where $0 < x < \zeta$, and ζ is the limiting value of perpendicular distance at which the observations are made. To demonstrate the fruitfulness of a detection model, numerous assumptions are to be made [Buckland et al. [**?**] and Miller and Thomas [**?**]]. It is logical to say that objects which are far away from transect line have least chances of detection and therefore in mathematical terms we can say that h(x) is assumed to be monotonically non increasing with respect to x. Furthermore, $h(0) = 1$, implies "with probability 1 objects on the path will be spotted" and the detection probability should approach to 1 at a distance approaching to 0. Additionally, tangent slope is 0 at $x = 0$ (i.e.,$h\prime(0) = 0$, indicating flat at zero distance) depicting horizontal tangent thereby h(x) satisfying the shape rule. These are the shoulder conditions which must exist in any detection model. Buckland [?] and Buckland et al. [?] mentioned some other prominent features of line transect sampling :

- N entities are randomly distributed over A with $D = N/A$.
- $h(0) = 1$.
- Entities are found at the initial observing place.
- No entity is counted included twice.
- Perpendicular distances are noted without errors.
- Entities are distributed independently of the line.
- Some, perhaps many, entities will be missed.

The elementary relation for evaluating the density of entities in a particular area [Burnham and Anderson [?] and Seber [?]] can be stated as

$$D = \frac{E(n)j(0)}{2M}, \tag{1}$$

where $E(n)$ is the expected value of the number of spotted entities. Burnham and Anderson [?] showed the general estimate for D as:

$$\hat{D} = \frac{n\hat{j}(0)}{2M}, \tag{2}$$

$\hat{J}(0)$ is a suitable sample estimator of $j(0)$ based on 'n' examined distances $x_1, x_2, \ldots, x_n$. When objects are observed from a line transect with a detection function $h(x)$, the distance X to the observed object from a randomly placed transect will tend to have a pdf j(x) of the same shape as h(x) , but scaled so as the area under j(x)=1 i.e,

$$j(x) = \frac{h(x)}{\kappa}, \tag{3}$$

where $\kappa = \int_0^{\zeta} h(x)dx$ is the normalizing constant and ζ is taken to as ∞. $j(x)$ satisfies the shoulder conditions iff $f^{'}(0) = 0$ and and $f(x)$ is monotonically non increasing[Eberhardt [?]]. This condition is one of the most important criteria for a robust estimation of $j(0)$ which is related to the properties of the proposed model for $j(x)$ [Crain et al. [?]]. Numerous parametric and non parametric methodologies have been proposed to estimate j(0). This article focuses on the parametric method to estimate the parameters using MLE. Hence, an estimator of j(0) and D is obtained.

The layout of the article is outlined as: In Section **??**, a new single-parameter detection model (SPDM) satisfying the shoulder conditions has been introduced. Some intriguing properties have been discussed in Section **??**. All the related expressions of this model have closed forms and hence easy to work out. Section **??** deals with estimation of the parameters and the practical application of the model is being described in Section **??** . Lastly, the article is completed with some remarks in Section **??** .

2. The Proposed Model

Suppose the detection function of SPDM with parameter β ($\beta > 0$) is given by

$$h(x;\beta) = \left(3 - 2e^{-\frac{x}{\beta}}\right) e^{-\frac{2x}{\beta}}, \qquad 0 \leq x < \infty, \quad \beta > 0. \tag{4}$$

The detection function (**??**) satisfies all the shoulder conditions; $h(0) = 1$ making it impeccable for detection on the transect line path. The first derivatives of (**??**) w.r.t x are, respectively, given by

$$h\prime(x) = -\frac{6}{\beta}\left(e^{\frac{x}{\beta}} - 1\right) e^{-\frac{3x}{\beta}},$$

$\implies h\prime(0) = 0 \; \forall \beta$. Since $e^{-\frac{3x}{\beta}} > 0 \quad \forall \quad x \quad \in (0, \infty)$, implies that $h\prime(x) = -\frac{6}{\beta} < 0 \quad \forall \quad \beta > 0$ which means that (**??**) is monotonically decreasing $\forall x \in (0, \infty)$. Figure **??** confirms all the shoulder conditions of the detection function.

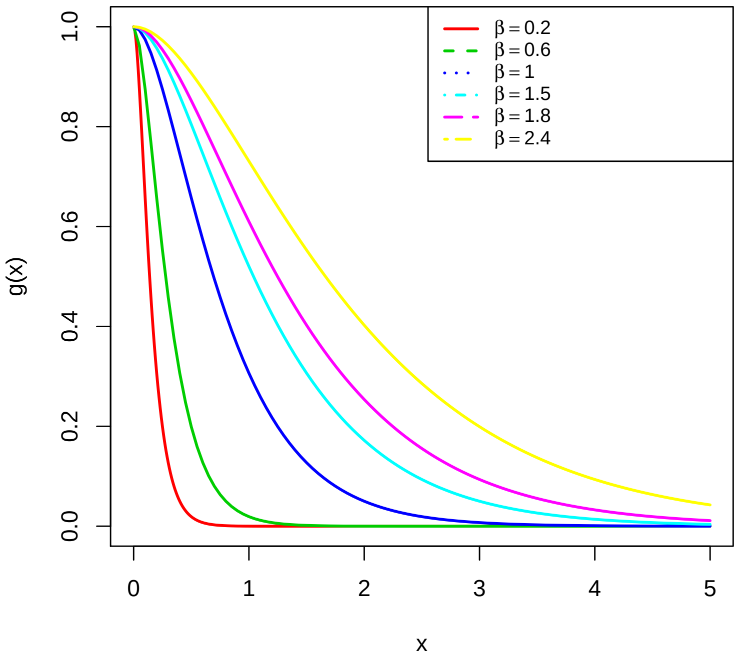

Figure 1: *Plot of detection function for different choices of parameter.*

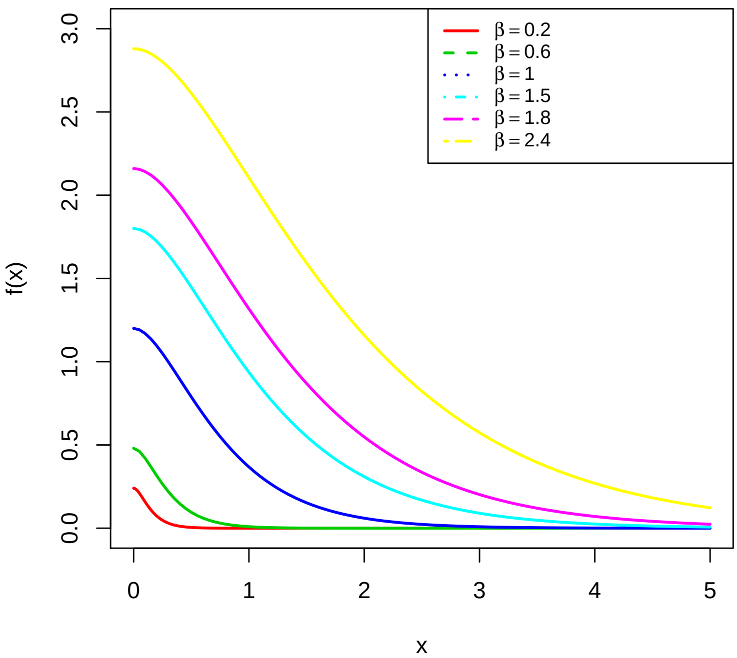

Figure 2: *pdf plot for several choices of parameter.*

Now the corresponding pdf is obtained by substituting (??) in (??) as:

$$j(x;\beta) = \frac{6}{5\beta}\left(3 - 2e^{-\frac{x}{\beta}}\right)e^{-\frac{2x}{\beta}}, \quad x > 0, \quad \beta > 0. \tag{5}$$

The pdf plot of SPDM model for different choices of parameter β is exhibited in Figure **??**. The cummulative distribution function (cdf) corresponding to pdf (**??**) is :

$$J(x;\beta) = 1 - \frac{1}{5}\left[9 - 4e^{-\frac{x}{\beta}}\right]e^{-2\frac{x}{\beta}}, \quad x > 0, \quad \beta > 0. \tag{6}$$

Since $h(0) = 1$ and if we substitute $x = 0$ in (**??**), we will get

$$j(0) = \frac{6}{5\beta}, \quad \beta > 0. \tag{7}$$

In the light of above expression, the pdf of SPDM can be phrased as

$$f(x;\beta) = f(0)\left(3 - 2e^{-\frac{x}{\beta}}\right)e^{-\frac{2x}{\beta}}, \quad x > 0, \quad \beta > 0, \tag{8}$$

which is a function of β and $j(0)$ and will serve as the base of the MLE maximum of β and $j(0)$.
The first derivative of pdf (**??**) w.r.t x is

$$\frac{\partial j(x;\beta)}{\partial x} = -\frac{36e^{-\frac{3x}{\beta}}\left(e^{x/\beta} - 1\right)}{5\beta^2}, \quad x \geq 0, \quad \beta > 0. \tag{9}$$

Since from expression (**??**), it is clear that h(x) $\propto$ j(x). Therefore, j(x) possess some attributes similar to h(x), such as $j\prime(0) = 0 \quad \forall \quad \beta > 0$ and the property of being monotonically decreasing. These characteristics are displayed in Figures **??** and **??**, and in turn the proposed model introduces a robust estimator for $f(0)$, named as "Shape Criterion" []Burnham et al. [**?**]]. It is also evident from the plots of pdf and detection function that as we move away from transect line (i.e., $x \to \infty$) , the probability of observing an object diminishes (i.e., all plots decays slowly to 0), that is one of the preferred character of a detection model.
Besides

$$D = \frac{E(n)j(0)}{2M},$$

substituting (**??**) in above expression, we obtain

$$\hat{D} = \frac{3\hat{n}}{5M\hat{\beta}}. \tag{10}$$

For estimating β, we will use MLE technique. Thereupon, we estimate j(0) and D which will be addressed at length in the subsequent section.

3. Statistical Properties

For SPDM model it is easy to prove the following properties:

1. The moment generating function (mgf): $M_X(t) = \frac{6(5-\beta t)}{5(\beta^2 t^2 - 5\beta t + 6)}$.
2. The rth moments: $E(X^r) = \frac{\Gamma(r+1)\left(3^{r+2} - 2^{r+2}\right)}{5(6\beta)^r}$.
3. $E(X) = \frac{19\beta}{30}$, $Var(X) = \frac{289\beta^2}{900}$ and coefficient of variation (C.V)=0.89 and skewness=1.69. The mean and variance for different choice of parameter β are exhibited in Table **??**. However, the C.V, skewness and kurtosis are independent of parameter β.

RT&A, No 4(80)
Volume 19, December, 2024

Table 1: *Mean and variance for different choices of parameter of SPDM Model*

Parameter $\beta \downarrow$	Mean $\downarrow$	Variance $\downarrow$
0.2	0.1267	0.0128
0.6	0.3800	0.1156
1.2	0.7600	0.4624
2	1.2667	1.2844
2.6	1.6467	2.1707
3.5	2.2167	3.93362

4. Assume a random sample $X_1, X_2, \ldots, X_n$ of size n drawn from SPDM pdf (**??**), then the Fisher information measure about the parameter β is given by

$$\mathbf{I}(\beta, n) = -nE\left[\frac{\partial^2 \log j}{\partial \beta^2}\right] = \frac{1.7957n}{\beta^2}.$$

- If $\hat{\beta}_{MVUE}$ is the MVUE for the parameter β, then

$$Var(\hat{\beta}_{MVUE}) = \frac{\beta^2}{1.7957n}. \tag{11}$$

Note that, this is the lower limit of Cramér?Rao inequality related to $\mathbb{NDM}(\beta)$.

4. Estimation

Here we will consider two methods of estimation: MOM and MLE for estimating the parameters of SPDM Model which are being discussed one by one in the following subsections.

4.1. Method of Moments

Suppose $x_1, x_2, \ldots, x_n$ be the observed values of a random sample (r.s) taken from model (**??**). Moment estimators consists of equating first m sample moments with corresponding m population moments, and solving the resulting system of simultaneous equations. Thus

$$m_1 = \frac{19}{25j(0)},$$

and

$$m_2 = \frac{13\beta}{15j(0)},$$

where m_1 and m_2 are first and second sample moments. Solving for $f(0)$ and β, we get

$$\hat{j}(0) = \frac{19}{25m_1}, \tag{12}$$

and

$$\hat{\beta} = \frac{57m_2}{45m_1}. \tag{13}$$

By substituting values of m_1 and m_2 from the sample, we can calculate the parameter estimates of $f(0)$ and β directly without involving any non-linear approximation. Both the estimates derived here can be taken as initial guesses for parameters to be estimated via MLE method.

4.2. Maximum Likelihood Estimation

Assume $\underline{\mathbf{x}} = \{x_1, x_2, \cdots, x_n\}$ be a r.s of size n from (**??**). The likelihood function is obtained as

$$L(j(0), \beta|\underline{\mathbf{x}}) = \prod_{i=1}^{n} j(x_i) = \prod_{i=1}^{n} j(0) \left(3 - 2e^{-\frac{x_i}{\beta}}\right) e^{-\frac{2x_i}{\beta}}. \tag{14}$$

The log-likelihood function analogous to (**??**) is obtained as

$$\log L(j(0), \beta|\underline{\mathbf{x}}) = n \log[j(0)] + \sum_{i=1}^{n} \log\left[3 - 2e^{-\frac{x_i}{\beta}}\right] - 2\sum_{i=1}^{n} \frac{x_i}{\beta}. \tag{15}$$

The ML Estimates $\hat{j}(0)$ of $j(0)$ and $\hat{\beta}$ of β, can be derived as:

$$\frac{\partial \log L}{\partial j(0)} = 0, \quad \text{and} \quad \frac{\partial \log L}{\partial \beta} = 0.$$

where

$$\frac{\partial \log L}{\partial j(0)} = \frac{n}{j(0)},$$

and

$$\frac{\partial \log L}{\partial \beta} = 2\sum_{i=1}^{n} \frac{x_i}{\beta^2} - \sum_{i=1}^{n} \frac{3x_i e^{x_i/\beta}}{\beta^2 \left(3e^{x_i/\beta} - 2\right)}.$$

As the above equation is not in closed form, hence cannot be solved explicitly. Using an iterative procedure to find the estimates of β through `maxLik()` function in `R` would do the job.

The Fisher information matrix is given as

$$\mathbf{V_x} = \begin{bmatrix} -\mathbb{E}\left(\frac{\partial^2 \log L}{\partial j(0)^2}\right) & -\mathbb{E}\left(\frac{\partial^2 \log L}{\partial j(0)\partial \beta}\right) \\ -\mathbb{E}\left(\frac{\partial^2 \log L}{\partial \beta \partial j(0)}\right) & -\mathbb{E}\left(\frac{\partial^2 \log L}{\partial \beta^2}\right) \end{bmatrix}$$

which can be approximated and written as

$$\mathbf{V_x} \approx \begin{bmatrix} V_{j(0)j(0)} & V_{j(0)\beta} \\ V_{\beta j(0)} & V_{\beta\beta} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 \log L}{\partial j(0)^2}\Big|_{\hat{j}(0),\hat{\beta}} & \frac{\partial^2 \log L}{\partial j(0)\partial \beta}\Big|_{\hat{j}(0),\hat{\beta}} \\ \\ \frac{\partial^2 \log L}{\partial \beta \partial j(0)}\Big|_{\hat{j}(0),\hat{\beta}} & \frac{\partial^2 \log L}{\partial \beta^2}\Big|_{\hat{j}(0),\hat{\beta}} \end{bmatrix}$$

where $\hat{j}(0)$ and $\hat{\beta}$ are the ML estimators of $j(0)$ and β respectively. Hence, when n is large and under some mild regularity conditions, we have

$$\sqrt{n} \begin{pmatrix} j(0) - \hat{j}(0) \\ \beta - \hat{\beta} \end{pmatrix} \overset{a}{\sim} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{V_x}^{-1}\right),$$

$\mathbf{V_x}^{-1}$ is the inverse of $\mathbf{V_x}$. The approximate confidence intervals for the parameters are; $\hat{j}(0) \pm z_{1-\alpha/2} se(\hat{j}(0))$ and $\hat{\beta} \pm z_{1-\alpha/2} se(\hat{\beta})$ for $j(0)$ and β. Here, se is the asymptotic standard error of the parameters that can be derived as a square root of the diagonal element of $\mathbf{V_x}^{-1}$, $z_{(1-\alpha/2)}$ indicate the $(1-\alpha/2)$ quantile of standard normal distribution.

5. Numerical Illustration

To check the practical potentiality of the suggested model, it has been analyzed with already existing models using some goodness of fit tests. The existing models with their detection functions h(x) and pdfs j(x) over the support $0 \leq x \leq \infty$, are given as under :

1. Two parameter model (NDM) (Bakouch et al [?]):

$$h(x) = (1+\lambda x^{\beta})e^{-\lambda x^{\beta}}, \quad j(x) = \frac{\beta^2 \lambda^{1/\beta}}{(\beta+1)\Gamma(1/\beta)}(1+\lambda x^{\beta})e^{-\lambda x^{\beta}}; \quad \lambda, \beta \geq 0.$$

2. Negative exponential model (NEM) (Gates et al. [?]):

$$h(x) = e^{-\lambda x}, \quad j(x) = \lambda e^{-\lambda x}; \quad \lambda \geq 0.$$

3. Exponential power series model (EPSM) (Pollock [?]):

$$h(x) = e^{-(x/\lambda)^{\beta}}, \quad j(x) = \frac{e^{-(x/\lambda)^{\beta}}}{\lambda\Gamma(1+1/\beta)}; \quad \lambda, \beta \geq 0.$$

4. Reverse logistic model (RLM) (Eberhardt [?]):

$$h(x) = \frac{(1+\gamma)e^{-\alpha x}}{1+\gamma e^{-\alpha x}}, \quad j(x) = \frac{\alpha\gamma(1+\gamma)e^{-\alpha x}}{(1+\gamma)\log(1+\gamma)(1+\gamma e^{-\alpha x})}; \quad \alpha, \gamma \geq 0.$$

5. Weighted exponential model (WEM) [Ababneh and Eidous [?]]:

$$h(x) = \left(2 - e^{-\theta x}\right)e^{-\theta x}, \quad j(x) = \frac{2\theta}{3}\left(2 - e^{-\theta x}\right)e^{-\theta x}; \quad \theta \geq 0.$$

The data set here has been reported by Burnham et al. [?], Barabesi [?], Bakouch et al. [?] and corresponds to a number of perpendicular distances, assumed to be in meter(mtr), of wooden stakes in a sagebrush meadow east of Logan with $D = 0.00375$ stake/mtr. Walking a single path of length L=1000 meters, out of population size N=150 stakes, a number (sample) of objects n=68 stakes are detected and their corresponding perpendicular distances are recorded, constituting the data $x_1, x_2, \ldots, x_n$. The data are: 2.02, 2.90, 11.82, 4.85, 3.17, 15.24, 1.27, 9.10, 1.23, 4.97, 0.45, 8.16, 14.23, 1.47, 7.10, 3.47, 13.72, 3.25, 1.67, 3.17, 10.40, 6.47, 2.44, 18.60, 10.71, 3.05, 6.25, 8.49, 4.53, 7.67, 3.61, 5.66, 1.61, 0.41, 3.86, 7.93, 3.59, 6.08, 3.12, 18.16, 0.92, 2.95, 31.31, 0.40, 6.05, 18.15, 9.04, 0.40, 3.05, 4.08, 1.00, 3.96, 6.50, 0.20, 6.42, 10.05, 7.68, 9.33, 6.06, 3.40, 0.09, 8.27, 11.59, 3.79, 4.41,4.89, 0.53, 4.40.

The ML estimates of the data for all the given detection models have been obtained and presented in Table **??**. As displayed previously, the functioning of the detection model is directly proportional with its pdf. For checking the performance of the given models, different tests such as Akaike's Information Criterion (AIC) [?], Bayesian information criterion (BIC) [?], Kolmogorov-Smirnov statistics (K-S) and associated p-value (p-value) have been carried out and results have been shown in Table **??**.

Table 2: *ML Estimates and LL values*

Model	ML Estimates	LL
$SPDM(\beta)$	$\hat{\beta} = 9.594$	-190.009
$NDM(\beta, \lambda)$	$\hat{\beta} = 1.00941$, $\hat{\lambda} = 0.239$	-190.021
$NEM(\lambda)$	$\hat{\lambda} = 0.164$	-190.967
$EPSM(\beta, \lambda)$	$\hat{\beta} = 1.313$, $\hat{\lambda} = 8.306$	-190.22
$RLM(\alpha, \gamma)$	$\hat{\alpha} = 0.221$,$\hat{\gamma} = 2.292$	-190.048
$WEM(\theta)$	$\hat{\theta} = 0.192$	-190.044

From these Tables , it has been found that the proposed model outbeats the models in comparison in terms of Log-Likelihood (LL) values, AIC, BIC, K-S and p-values. Thus, the proposed model can be considered as a powerful competitor among other detection models.

Table 3: *Goodness of fit values*

Model	AIC	BIC	K-S Value	p-value
$SPDM(\beta)$	382.018	381.851	0.10437	0.4493
$NDM(\beta,\lambda)$	384.042	383.707	0.1115	0.4137
$NEM(\lambda)$	383.934	383.767	0.14306	0.1236
$EPSM(\beta,\lambda)$	384.44	384.105	0.1530	0.035
$RLM(\alpha,\gamma)$	384.096	383.761	0.1502	0.3703
$WEM(\theta)$	382.900	381.921	0.1917	0.0135

6. Conclusion

This manuscript focuses on the introduction of new one-parameter detection model which satisfies the shoulder conditions of the detection model and has more flexible shapes of detection model. Methods like MOM and MLE are used to estimate the parameters of model. Applicability of this model has been tested using perpendicular distances data set, therefore can be expected to appeal wide range of real life situations.

References

[1] Ababneh, F.; Eidous, O. M. (2012). A weighted exponential detection function model for line transect data. *Journal of Modern Applied Statistical Methods*, 11(1) : 475–478.

[2] Akaike, H.(1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control,* 19(6), 716–723.

[3] Buckland, S. T.; Anderson, D. R.; Burnham, K. P.; Laake, J. L.; Borchers, D. L.; Thomas, L.(2001).*Introduction to distance sampling: estimating abundance of biological populations* Oxford university press

[4] Buckland, S. T.; Anderson, D. R.; Burnham, K. P.; Laake, J. L.; Borchers, D. L.; Thomas, L. 2004. . *Advanced distance sampling,* Oxford University Press,

[5] Barabesi, L. Environmetrics. 2000. The official journal of the International Environmetrics Society. *Off. Jour. of Int. Evs. 11*, 413–422.

[6] Miller, D. L.; Thomas, L. 2015. Mixture models for distance sampling detection functions. *Plos One*. *10*, 413–422.

[7] Burnham, K. P.; Anderson, D. R. 1976. Mathematical models for nonparametric inferences from line transect data. *Biometrics*, 325–336.

[8] Seber, G. A. F. 1982. *Estimation of animal abundance and related parameters,* 2nd ed.; Publisher: London: Griffin.

[9] Bakouch, H. S.; Chesneau, C.; Abdullah, R. I.2022. A pliant parametric detection model for line transect data sampling. *Communications in StatisticsTheory and Methods*, 51 (21), 7340–7353.

[10] Buckland, S. T.; Rexstad, E. A,; Marques, T. A.; Oedekoven, C. S. 2015. *Distance sampling: methods and applications*, **431**, Publisher: Springer.

[11] Buckland, S. 1985. Perpendicular distance models for line transect sampling. *Biometrics*, 41 (1), 177–195.

[12] Burnham, K. P.; Anderson, D. R.; Laake, J. L.. (1980). Estimation of density from line transect sampling of biological populations. *Wildlife Monographs,* (72):3–202.

[13] Eberhardt, L. Transect methods for population studies. 1978. *The Journal of Wildlife Management*, 42(1): 1–31.

[14] Crain, B.; Burnham, K.; Anderson, D.; Lake, J. (1979). Nonparametric estimation of population density for line transect sampling using fourier series. *Biometrical Journal* 21 (8): 731–48.

[15] Gates, C. E.; Marshall, W. H.; Olson, D. P.1968. Line transect method of estimating grouse population densities. *Biometrics*, 135–145.

[16] Pollock, K. H. 1978. A family of density estimators for line-transect sampling. *Biometrics*, 475–478.

[17] Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics*, 6(2): 461–464.