KFRI Research Report 142

EVALUATION OF METHODS FOR ESTIMATING THE ABUNDANCE OF HERBIVORES IN THE FORESTS OF KERALA

K. Jayaraman
P.S. Easa
E.A. Jayson

KERALA FOREST RESEARCH INSTITUTE
PEECHI, THRISSUR

March 1998 Pages: 47

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

The present study was undertaken with the objective of comparing the relative efficiency of some of the feasible methods of estimating the abundance of herbivores in the forests of Kerala and to suggest refinements in the existing methods wherever possible. Both empirical and theoretical investigations were undertaken in order to meet the above objective which lead to certain conclusions of practical value.

Data collected from eight sanctuaries during the wildlife census conducted in 1993 jointly by Kerala Forest Department and Kerala Forest Research Institute, were utilised to compare the relative efficiency of different detection function models for estimating the abundance of herbivores. The following species *viz.,* elephant, sambar, spotted deer, barking deer, wild boar and gaur were considered for the study. Univariate half normal distribution was found promising with respect to precision of the density estimates. The bivariate procedures were not effective as the size bias parameter was not significant for most of the species considered. The distribution of cluster size in the case of the six herbivore species considered for the study was found to be highly skewed. Arithmetic mean shall not be a good estimator of average cluster size in such cases. The use of median for average cluster size brought down the variance of the animal density estimates and also provided realistic values of the density since the median is unaffected by extreme values in the population.

An examination of the theory showed that for a given set of detections, overestimation of distances in the field would lead to underestimation of density in the case of line transect sampling and *vice versa* with half normal detection function. However, the coefficient of variation of the density estimate would remain unaffected by errors in distance measurement as the coefficient of variation is purely a function of sample size and not of distance values in the case of half normal model. With Fourier series model for detection function, the direction of effect on density estimate was found to be governed by the range and distribution of the distance measurements.

An *ex situ* trial was conducted to assess the agreement between actual distance and visual estimates made by the observers. Simple linear regression equation fitted through the origin showed that there was underestimation by 2 m for every 100 m of actual distance which is negligible. The mean bias in the visual estimation of actual distance was not significantly different from zero. However, the coefficient of variation of visual estimates of distance varied from 54 per cent in 0-20 m class to 34 per cent in 80-100 m class. Using the results of the field evaluation of bias in visual estimation at known distance values, the effect of increasing levels of random disturbance in distance measurements was investigated through simulation trials. For a given set of detections and transect length, increased disruption of distance values on an average was found to bring down the density estimates both in the case of half normal and Fourier series model.

The sampling intensity needed in line transect sampling to bring the coefficient of variation of density estimates to 20 per cent was estimated. The sampling intensity required was found different for the different species. On an average, one transect of 2 km was found necessary for every 5km^2 of the area sampled.

In line transect sampling, the form of the detection function is found to vary with the local conditions associated with the forest type, weather condition, observer's fatigue etc. The variation in detection function over a region can be considered as random and the detection function model can be brought under the framework of random parameter models. Hence a random parameter model was formulated taking the two parameter negative exponential model as detection function. The basic proposition was that apart from the estimation errors, the relation between perpendicular distance and cumulative density function of the number of sightings can have different parameters in different locations and these can be viewed as random deviations from population level parameters. The model was tested utilising data collected for the species sambar from 10 wildlife sanctuaries at different periods. The difference between the actual and .predicted cumulative density function values against each perpendicular distance was obtained. The mean and variance of the deviations revealed that bias is very negligible, variance decreased and R^2 (prediction) increased with increasing sample size as expected under the random parameter model. The method has the clear advantage of being able to develop density estimates based on very few observations from a location which would be impossible through traditional methods.

In the case of elephant and gaur, indirect evidence like dung density is a very strong indicator of the habitat use which is associated with animal density and therefore accurate estimation of dung density is important in the case of these species. An analysis of data on distance to dung piles, collected during the course of line transect sampling, indicated that Fourier series model is a good choice for detection function model in most of the vegetation types existing in the forests of Kerala. Other than being a flexible and robust nonparametric model, the use of the model resulted in the least coefficient of variation for dung density estimates.

The present study has shown that total count is inapplicable for estimation of animal abundance as it leads to heavy undercounting. Line transect sampling has a firm theoretical footing but suffers from low number of sightings arising from low density of animals or poor detection percentage. Calibration of detection functions using random parameter models shall go a long way in making localized prediction of animal density and hence future works should attempt to develop generalized prediction models based on random parameter models. The methods based on indirect evidences also hold promise for the future and works can be undertaken to convert indirect evidences to animal numbers. However, indices of abundance based on indirect evidences would serve most of the practical purposes in wildlife management.

1. INTRODUCTION

Estimation of animal abundance is of prime importance in wildlife management and in studies related to wildlife biology. Before the advent of many modern techniques the status of animals in an area used to be described in qualitative terms like "absent, rare, occasional, common or abundant" which is barely sufficient for present day studies. More accurate expressions of density were of need and a number of developments followed in this field both in the technology of detection of animals and in the theory of estimation. Modern methods used for estimation of animal abundance for open populations include aerial censusing, removal method, remote censusing and change in the ratio method (Seber,1992). For closed populations, capture-recapture method and home range mapping are also used. Most of the presently available methods for monitoring animals are those originally developed for the vast prairie lands of Africa. The undulating terrain and thick vegetation posed great difficulties while executing these methods in tropical forests. Several modifications and adaptations were required for their application in tropical forests the history of which is not well documented in the literature. A need for evaluating the efficiency of alternative methods possible in tropical forests for estimating animal abundance largely prompted this work.

The works under reference relate mostly to herbivores in the Kerala part of the Western Ghats in the Indian subcontinent. The common herbivores in the forests of Kerala are elephant *(Elephas maximus)*, gaur *(Bos gaurus)*, sambar *(Cervus unicolor)*, spotted deer (*Axis axis*), barking deer *(Muntiacus muntjak)* and wild boar (*Sus scrofa*). They are seen in most of the protected areas of the State. The present day thrusts of the management is centered on conserving as much of the animals possible. Periodical assessment of the density of animals is needed in order to monitor the status of the wildlife population in any specific area.

The most direct way to estimate the abundance of an animal population is to count all the individuals in an area of known boundaries. In such an approach, the size of the area sampled is known and therefore the population density can be assessed by dividing the number counted by the size of the area censused. These methods are referred as quadrat, plot or strip sampling methods. Establishing a plot and censusing all animals within it, will be time consuming. Such an approach is found to be impractical also in several instances. For example, the animals may be mobile or difficult to detect or scattered widely. For all these reasons, plot methods are considered generally unsuitable for estimating wildlife populations.

Line transect sampling came almost as a revolution in methods of estimating animal abundance. The basic advantage of the method was that not all the animals in the population need be sighted or counted. It is a direct sampling method which is cost effective and does not need killing or handling of animals. It can be applied easily for populations of animals which are difficult to count due to logistic problems. Although line transect sampling has a number of advantages it has some short comings too. Careful design of the study and rigorous statistical analysis of data will be required for applying line transect sampling method. Although complicated statistical methods have been developed, their application to estimating animal density in tropical forests is difficult mainly because of poor visibility and relatively low density of these populations resulting in inadequate sample sizes for statistically precise results. Random sampling cannot be carried out due to topographic features of habitat. Since the line transect estimator is based on the observed distances, size bias is one of its characteristics. Another disadvantage of this method is that the supporting statistical analyses account only for the influence of distance on detectability, where as many other factors also influence the chances that animals will be detected and counted. These factors are variability arising from different observers. differences in walking speed, heterogeneous habitat structure, slope of the area, size of the animal and size of the herd.

Yet another way of assessing the population of wild animals is based on indirect evidences. Pellet, dung, hoof marks, spoor, scat etc. are the indirect evidences left by wild animals when they visit an area. This method is successfully applied in India in the case of elephant and tiger. Methods based on indirect evidences have a number of advantages in the execution phase since no direct encounter with the animals is required. If perfected, the method holds considerable promise for future applications.

The present study specifically aimed to investigate the feasibility or otherwise of some of the standard methods used for abundance estimation in the case of herbivores and suggest suitable modifications. Comparison of some of the standard methods was effected and certain refinements possible in these methods were also investigated.

2. GENERAL EVALUATION OF FEASIBLE METHODS

Estimating animal abundance had attracted the attention of many workers in the past. Seber (1992) had given an extensive review of the statistical methodology involved in estimating animal abundance. He broadly grouped the methods as applicable to closed populations i.e., unchanging except for known removals during the period of investigation and those for open populations where migration, births and deaths occur. He had discussed the use of plots, strips, lines and points for estimating population density or for providing an index of density based on indirect signs. Seber (1992) reviewed adaptive sampling, use of negative binomial distribution, Taylor's power law and model based sampling in relation to quadrats and strip transects on the ground. The other methods discussed by Seber (1992) are those based on capture-recapture techniques and home range which most often require sophisticated equipments and are not readily applicable under Indian conditions. Methods based on aerial censusing are discussed but these are yet to prove their utility under conditions existing in tropical forests.

Sale and Berkmuller (1988) gave some general guidelines to be followed in wildlife surveys. Species occurring in relatively high density can be counted by direct sighting methods. He had provided a description of general methods, which can be applied for counting mammals in the forests. Species occurring in low density or which are difficult to be detected because of poor habitat visibility or cryptic behaviour, should be censused either by carefully planned intensive samples, or by indirect methods such as dung or pugmark counts. This is relevant to most of the carnivores and also to small or nocturnal mammals as well as to some large mammal populations in dense habitats. Most of the indirect methods are only suitable for obtaining relative indices of population size and only rarely yield a good estimate of actual population numbers.

Rodgers (1991) had described some procedures to be followed while counting animals especially under Indian conditions. Specifically, simple indices of population based on presence or absence recording, road side index, dung surveys and water-hole techniques were illustrated. Sample counts based on actual sightings such as block counts, vehicle based transects and 'foot counts' were also described. An overall assessment of the past works indicated that some of the feasible methods applicable to tropical forests are total count, line transect sampling and methods based on indirect evidences. Hence a detailed review of these methods is made here.

2.1. Total count

The most direct way to estimate the abundance of a biological population is to count all individuals in a known area. But methods based on that approach has severe limitations as animals are likely to be missed the counting and also an exhaustive count is impractical in large areas. Hence intensive counts in selected sample blocks or strips are resorted to in many cases. The population is usually stratified by habitat type. Blocks, strips or quadrats are taken as sampling units wherein exhaustive search is made for the animals. The sampling pattern to be used such as systematic or random. sampling intensity, method of estimation etc. are yet to be worked out in the case of many herbivore species, as applicable to the forests of Kerala. Most of the wildlife census carried out in Kerala in the past were total counts (Anonymous. 1993).

2.2. Line transect sampling

Line transect sampling has been a common method used for obtaining estimates of wildlife abundance since the early 1930's. This method was first developed for use in animals, especially upland game birds like grouse in North America (Leopold, 1933; Hayne, 1949).

Line transect method has the following general setting. Assume that one has an area of known boundaries and size A and the aim is to estimate the abundance of some biological population in the area. The use of line transect sampling requires that at least one line of travel be established in the area. The number of detected objects (s,) is noted along with the perpendicular distances (x,) from the line to the detected objects. Otherwise the sighting distance r, and sighting angle θ_i are recorded from which x, can be arrived at. Let n be the sample size. The corresponding sample of potential data is indexed by $(s_i, r_i\ \theta_i\ \ i = 1, ..., n)$. Four assumptions are critical to the achievement of reliable estimates of population abundance from line transect surveys viz., (i) points directly on the line will never be missed (ii) points are fixed at the initial sighting position and they do not move before being detected and none are counted twice (iii) distances and angles are measured exactly (iv) sightings are independent events. To infer about animal abundance from those data, one must have a conceptual model that relates the data to the abundance parameter to be estimated. The basic idea underlying such a general model is that the probability of detecting an animal decreases as its distance from the line increases. Mathematically the idea is represented by a function or curve g(x) called the detection function. The detection function g(x) gives the conditional probability of observing an object given that the object is at perpendicular distance x from the line. The marginal probability density function of any distance x is $f(x) = \dfrac{g(x)}{\mu}$ where $\mu = \int_0^w g(x)dx$ and w is the maximum perpendicular distance from the transect line. Much of the work reported on line transect sampling has dealt with defining the detection function more and more accurately. Some of the significant developments in this field are mentioned in the following.

Hayne (1949) provided the first estimator which was based only on sighting distances. Hayne's (1949) method was poor if sighting angle is not approximately 32.7'. After Hayne's paper, almost no significant theoretical advances appeared until 1968. Gates (1969) derived an estimator based on radial distances. Burnham and Anderson (1976) gave a general mathematical theory of line transects which supplied a frame work for either parametric or nonparametric density estimation based on either right angle or sighting distances. They developed the general result, D = nf(0)/2L, where D is the estimator of density ,n is the number of detections, L is the length of the transect and f(0) is the probability density function of the detection distance at zero distance from the line. Robust estimation from line transect data was proposed by Burnham *et al.* (1979). Specific parametric models such as the negative exponential or half normal were found not robust models.

Buckland (1982) discussed Fourier series model as a powerful procedure for analyzing line transect data. Three solutions for finding the confidence interval, one using Monte Carlo techniques, another making use of replicate lines and the third based on the Jackknife method were discussed and compared. Burnham and Anderson (1984) concluded that for reasons of efficiency and validity, transect count studies should record perpendicular distance data.

Zahul (1989) gave a model, for line transect sampling for the purpose of estimating animal population density which makes no assumption about the value of the detection probability along the transect or about the form of the detection probability function other than continuity. Asymptotic and small sample behavior was examined. The estimator was also applied to a field experiment. Fitting density functions with polynomials was tried by Buckland (1992). A key function was assumed as a first approximation to the density and the fit was improved by polynomial adjustments. A comparison with kernel estimator of the density revealed that polynomial and kernel fits are verysimilar.

Cook and Martin (1974), Quinn (1979), and Rao and Portier (1981) proposed estimation techniques to clustered populations. Drummer and McDonald (1987) introduced cluster size variable as a covariate in detection functions. A nonparametric approach using trignometric Fourier series estimator to size biased line transect sampling was given by Quang (1991). The standard errors of the estimators were computable either by calculating sample variances or by bootstrap resampling.

Line transect sampling method is practical, efficient and relatively inexpensive, but, while executing line transect method in the forests of Kerala, certain problems were encountered in the past (Varughese, (1992), Anonymous, (1993)). It was difficult to lay straight line transects in the dense tropical vegetation. Lack of visibility in the forests made the counting difficult. Unequal visibility on either sides of the observer's path was more serious in this respect. Lack of uniformity of distribution of different species of animal in the area could alter the precision of the estimates in unpredictable directions. The inability of the census parties to cover the whole area assigned to them was an additional problem. All these factors reduce the accuracy of the estimates. This calls for careful design of the census operations followed by rigorous statistical analysis.

2.3. Methods based on indirect evidences

Methods based on indirect evidences were in vogue, as early as 1940 (Bennet *et al.*, 1940). The method has the advantage that direct sighting of animals is not required. Recently many attempts were made to census Asian elephants using dung counts (Barnes and Jensen (1987), Dekker *et al.*, 1991). Similarly Ngampongsai (1981) had applied the method of pellet group counts for estimating sambar populations in Thailand. Spotted deer had been censused using the pellet group indices by Martin (1987).

Indirect methods usually provide only indices of animal abundance and not the actual size of the population. However, these indices can be used to estimate the actual numbers through the following methods. (i) direct conversion to a census method. (ii) calibration of population indices through ratio and regression methods following double sampling. (iii) making an improved index or a prediction equation by supplementing additional information to strengthen an index.

2.4. Comparison between methods

There were some attempts also to compare the relative efficiency of the above methods, in the past. Burnham *et al.* (1985) compared efficiency and bias in strip and line transect sampling. The results indicated a preference for the line transect method over strip transect on the basis of bias and efficiency. Comparison of transects and circular plots for estimating bobolink densities were made by Bollinger *et al.* (1988). Line transects consistently provided density estimates with smaller biases and higher correlations with true densities than did variable circular plots. Direct and indirect methods of counting elephants were done by Varman *et al.* (1995). They found that estimate of mean density from the direct count were higher (3.09 elephants/km^2) than that obtained by the indirect count (1.54 elephants/km^2) for average of seasonal densities.

In conclusion, there has been a general tendency to prefer line transect sampling over other methods. In the case of line transect sampling, developments occurred mainly in characterising the detection function through better models. Suitability of these detection function models in the case of herbivore species in tropical forests has not been extensively studied. The problem of variation in detection function parameters over sites has also not received much attention. Practical considerations like sampling intensity under specific situations in Kerala, also remains to be worked out. The following chapters address some of the above issues.

3. CHOICE OF DETECTION FUNCTION IN LINE TRANSECT SAMPLING

Modelling the detection function is an important task in estimating animal density through line transect sampling. A number of models originating from different contexts have been found proposed in the past for modelling the detection functions which involve both parametric and nonparametric approaches. Burnham *et al.* (1980) introduced several criteria to look for while choosing a model for detection function and promoted the idea of robust nonparametric estimation through Fourier series. Another major development in this area was the introduction of bivariate functions including cluster size as an additional variable to account for the size bias (Drummer (1991), Quang (1991), Laake *et al.* (1994)).Buckland (1992) compared the kernel estimator of Silverman (1982) with the hermite polynominal model having parametric key as normal, using the deer data from survey 11 of Robinette *et al.* (1974). An evaluation of different models used in line transect sampling in deciduous forests was reported by Varman and Sukumar (1995). A similar attempt is made here to identify a suitable model for detection function in the case of six herbivore species using line transect data from a typical tropical forest predominantly moist deciduous.

3.1. Materials and methods

Data collected during the wildlife census conducted in the State of Kerala, India, in 1993 were used for the present study. Data from eight wildlife sanctuaries which were predominantly of moist deciduous forests were used for this analysis. The sanctuaries were Wayanad, Aralam, Parambikulam, Peechi-Vazhani, Idukki, Peppara, Neyyar and Periyar Tiger Reserve (Figure 1). Due to the discontiguous nature, Wayanad Wildlife Sanctuary was considered as Tholpetty (northern part) and Wayanad (southern part). Their location and extent are given in Table 1. The total area of the sanctuaries was 1731.85 km^2 excluding the reservoirs. The species considered were elephant *(Elephas maximus)*, gaur *(Bos gaurus)*, sambar *(Cervus unicolor)*, spotted deer *(Axis* axis), barking deer *(Muntiacus muntjak)* and wild boar *(Sus scrofa)*. The census was carried out from 30th April to 3rd May in 1993. Line transect sampling was done on the first day followed by total count on the second day.

For the line transect sampling, randomly selected transects of about 2 km length were marked in the area map of each Forest Range with the help of forest officials. The number of transects in each sanctuary was proportional to the area of the sanctuary. The positions of the transects were identified in the field and laid by marking trees with paint. These transects were then covered on foot, recording the sighting distance (r) and the sighting angle (θ) to the geometric centre of the herds sighted between 6.00 hours to 10.00 hours. Ocular estimation of the sighting distance was made. The sighting angle (θ) was measured with a compass. The perpendicular distance (y) from the transect to the animal was then worked out using the formula $y = r \sin(\theta)$. The total length of the transects over the nine sanctuaries was 454.4 km. For the analysis,

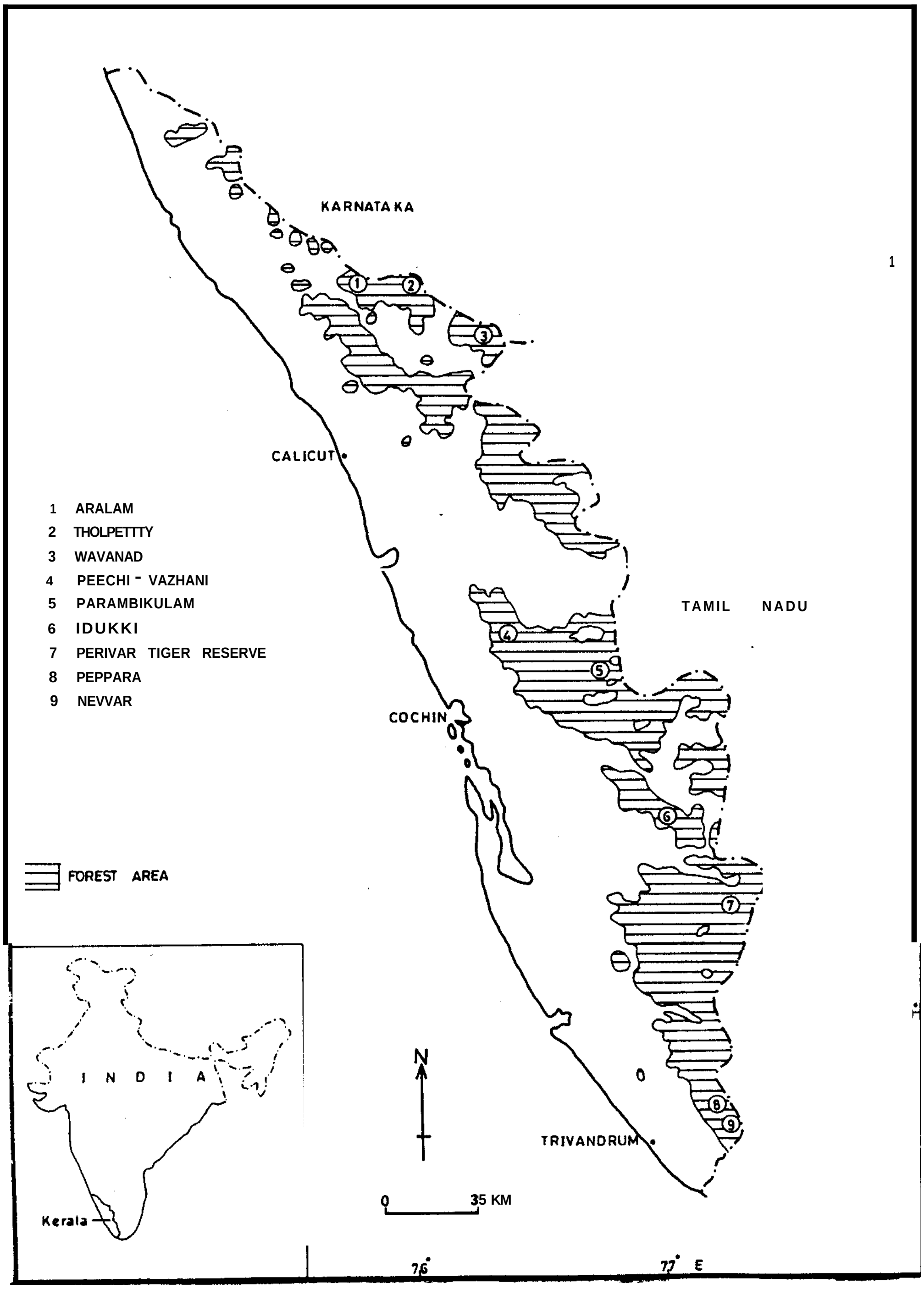

Figure 1. Map showing the location of study area

the data were truncated to a maximum perpendicular distance of 200 m. The data were analyzed using the programmes, SIZETRAN (Drummer, 1991), DISTANCE (Laake *et al.*, 1994) and NPARTRAN (Quang, 1991).

Table 1. The location and extent of the sanctuaries surveyed

Sanctuary	Latitude	Longitude	Altitude above msl (m)	Area excluding the reservoirs (km^2)
Wayanad	$11^{o}35'$to $11^{o}49'$N	$76^{o}13'$to $76^{o}27'$E	650 to 1150	266.77
Tholpetty	$11^{o}50'$to $11^{o}59'$N	$76^{o}02'$to $76^{o}07'$E	650 to 1150	77.67
Aralam	$11^{o}52'$to $11^{o}59'$N	$75^{o}47'$to $75^{o}56'$E	100 to 1598	55.00
Parambikulam	$10^{o}20'$to $10^{o}26'$N	$76^{o}35'$to $76^{o}50'$E	300 to 1224	258.00
Peechi-Vazhani	$10^{o}28'$to $10^{o}38'$N	$76^{o}18'$to $76^{o}28'$E	45 to 900	111.94
Idukki	$9^{o}46'$to $9^{o}53'$N	$76^{o}55'$to $77^{o}03'$E	500to 746	44.00
Peppara	$8^{o}34'$to $8^{o}42'$N	$77^{o}07'$to $77^{o}14'$E	100 to 1717	48.00
Neyyar	$8^{o}17'$to $8^{o}53'$N	$76^{o}40'$to $77^{o}17'$E	90 to 1868	119.00
Periyar Tiger Reserve	$9^{o}18'$to $9^{o}40'$N	$76^{o}55'$to $77^{o}25'$E	900 to 2019	751.54

In the programme SIZETRAN, both bivariate and univariate sighting models are employed for estimating probability density functions of perpendicular distances . The models used for the study were univariate negative exponential (UNE), univariate half normal (UHN), univariate Fourier series (UFS), bivariate negative exponential (BNE), and bivariate half normal (BHN). Modelling the bivariate detection functions was done by introducing the size covariate y into the univariate detection function via the ratio x/y^{α} The parameter α is referred as the size bias parameter. A value of 0 for α implies that size has no effect on the probability of detection. If the size bias has no effect on probability of detection, the mean cluster size will be used by the programme for the estimation of density of animals. A likelihood ratio test for the presence of size bias was performed. The test statistic has an asymptotic χ^2 distribution with one degree of freedom. A χ^2 goodness-of-fit test for the detection function was performed on the transformed data $z = x/y^{\alpha}$.

Semi parametric models are used in the programme DISTANCE. Specifically, the models are (Uniform + Cosine), (Uniform + Polynomial), (Half normal + Hermite) and (Hazard rate + Cosine). These are collectively referred to as Polynomial adjustment models (PAM) here. The AIC (Akaike's information criterion) was used for selecting between models. A regression equation was fitted between logarithm of cluster size and probability of detection of perpendicular distances, g(x). The estimate of cluster size was calculated at the point g(0) = 1.

In the programme NPARTRAN, bivariate detection function using Fourier series (BFS) is employed. This programme provides diagnostics for visibility bias and calculates bias-reduced estimates of both population density and group density.

Kernel estimator was introduced by Rosenblatt (1956) and Fryer (1977). If X_1, X_2, . . ., X_n are real observations from a probability density f, then the kernel estimate. f_n of f is defined by

$$f_n(x) = [(1/nh)] \sum_{j=1}^{n} K(x-X_j)/h \qquad (1)$$

where K is the kernel function and h is the smoothing parameter or window width. Both theory and practice (Epanechnikov, 1969) suggest that the choice of kernel is not crucial to the statistical performance of the method and therefore it is quite reasonable to choose a kernel for computational efficiency. The kernel used in the present case was the standard Gaussian density. The character of the estimate is mainly governed by the choice of window width, which determines how much the data are smoothed to obtain the estimate. The optimum window width depends on the unknown density being estimated, but it is worth noting that, if the data come from a normal distribution with standard deviation σ then the choice, $h = 1.06\sigma\, n^{-1/5}$, will, to a high degree of accuracy, minimise the integrated mean square error. Silverman's (1982) algorithm was applied to obtain a kernel estimator of the density. The side of the line that an animal is detected is not usually recorded in line transect surveys. To force the kernel method to fit a symmetric density each recorded perpendicular distance X was replaced by two values X and -X. In the case of kernel density estimation (KDE), the estimate of f(0) was found by running the program as given by Silverman (1982). Density estimate D was calculated by the formula , D = n f(0)/2L.

$V(\hat{f}(0))$ was found out from the following formula as given by Silvermart (1986).

$$VAR\left(\hat{f}(x)\right) \approx n^{-1}h^{-1}f(x) \int^{\infty} [K(t)]^2 dt \quad (2)$$

$$\text{Since } K(t) = \left[\frac{1}{\sqrt{2\pi}} e^{-t^2/2}\right]^{\infty} \quad (3)$$

$$VAR\left(\hat{f}(x)\right) \approx n^{-1}h^{-1}f(x) \int_{-\infty}^{\infty} (2\pi)^{-1} \left(e^{-t^2/2}\right)^2 dt \quad (4)$$

$$\approx n^{-1}h^{-1} \frac{\hat{f}(x)}{\pi} \int_0^{\infty} e^{-t^2} dt \quad (5)$$

$$\approx n^{-1}h^{-1} \frac{\hat{f}(x)}{2\sqrt{\pi}} \quad (6)$$

$$CV(\hat{D})^2 = var(n) / [E(n)]^2 + var(\hat{f}(0)) / (\hat{f}(0))^2 \quad (7)$$

In all the univariate procedures, the average cluster size was estimated by arithmetic mean with variance σ^2/n, where n is the number of clusters. Since the distribution of cluster size was found highly skewed, the average cluster density was estimated using median as well and the corresponding variance was computed by the formula reported by Kendall and Stuart (1977) which is Variance (median) = $1/4nf^2$ where n = number of clusters and f = the median ordinate. The animal density was then obtained by multiplying the cluster density with the average cluster size. The variance of the estimate of animal density was then arrived at using the formula given by Goodman (1960).

For the total count, the forest area under each sanctuary was divided into small blocks of convenient and manageable size. Each block was searched for the presence of large mammals on foot by a team consisting of a trained volunteer, a forest staff and a tribal tracker. Number of clusters and size of each cluster sighted were recorded during the survey. Density of each species was computed dividing the total number of animals sighted by the total area surveyed. The estimates obtained were used as a reference while comparing the different detection function models.

3.2. Results and discussion

3.2.1. Detection function models

Distribution of the perpendicular sighting distances for the six species is given in Figure 2. The patterns show no particula- signs of any stark irregularities like heaping or evasive movement. The density estimates and related statistics for the

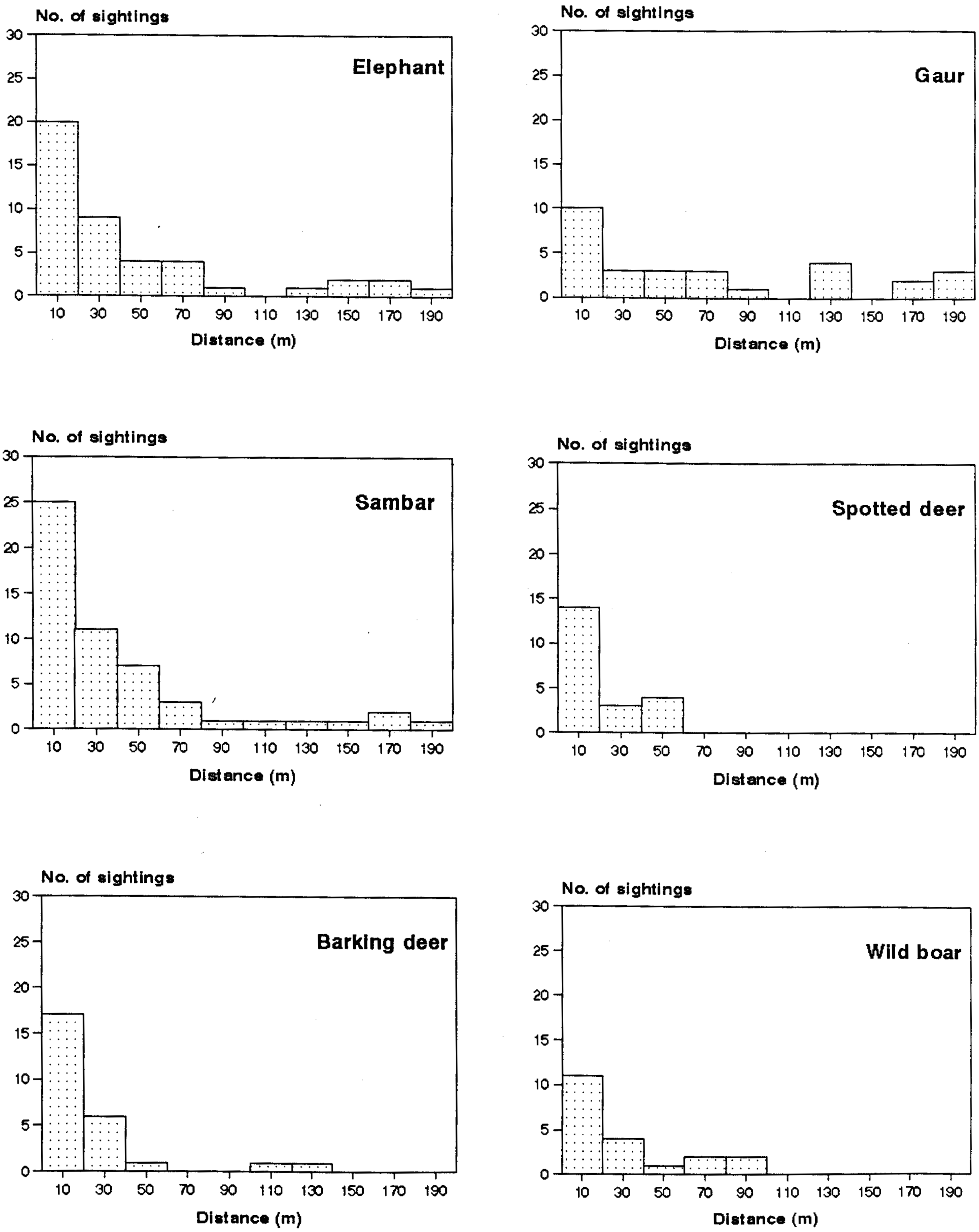

Figure 2. Distribution of perpendicular distances of sightings for the six species

different models fitted are shown in Table 2. The models having the least coefficient of variation with respect to the cluster density and also the density of individuals are shown in Table 3. Kernel density estimation was found to be good for estimating the cluster density whereas univariate half normal distribution with median as estimator for average cluster size was the best choice for estimating animal density for all species except spotted deer. The case of spotted deer being slightly different from the rest of the species could be attributed to the relatively low number of sightings for that species.

Although the UHN model was not a good fit in the case of elephant, gaur, sambar and barking deer, this model is to be preferred on account of the high precision of the estimates provided by the model. Moreover, the use of model fit as a criterion for choosing between models is deemphazised by Burnham *et al.* (1980). Varman and Sukumar (1995) found that cluster density estimates derived from Fourier series and half normal model had the lowest CV for the species considered by them.

The distribution of cluster size for'the six species is given in Figure 3. Since the distribution of cluster size was skewed for all the species, the arithmetic mean shall not be a good estimator of the average herd size. In such cases, the use of median shall be a better option as it is unaffected by extreme values of cluster size in the sample. In the present case, the estimators based on median were found to have better precision than those based on arithmetic mean.

3.2.2. Size bias pgrameter

The size bias parameter, the probability level and the coefficient of correlation between distance and group size are given in Table 4. The size bias parameter is significant in the case of elephant and sambar in certain models . Even though the size bias parameter was significant for certain species, there were no significant differences between densities obtained through the corresponding univariate and bivariate models ($P>0.01$). For gaur, wildboar, spotted deer and barking deer the size bias parameter was not significant. Varman and Sukumar (1995) found that there was no statistically significant relationship between detectability of a group and size of the group for any species.

3.2.3. Total count

The density estimates obtained through line transect sampling were in general higher when compared to that obtained through total count indicating the inefficiency of total count method in estimating animal numbers. Keeping the density estimate of animals based on median cluster size obtained through UHN model as a standard, the detection percentage in total count varied from just 5 per cent in the case of wild boar to 19 per cent in the case of spotted deer. These values are substantially low with

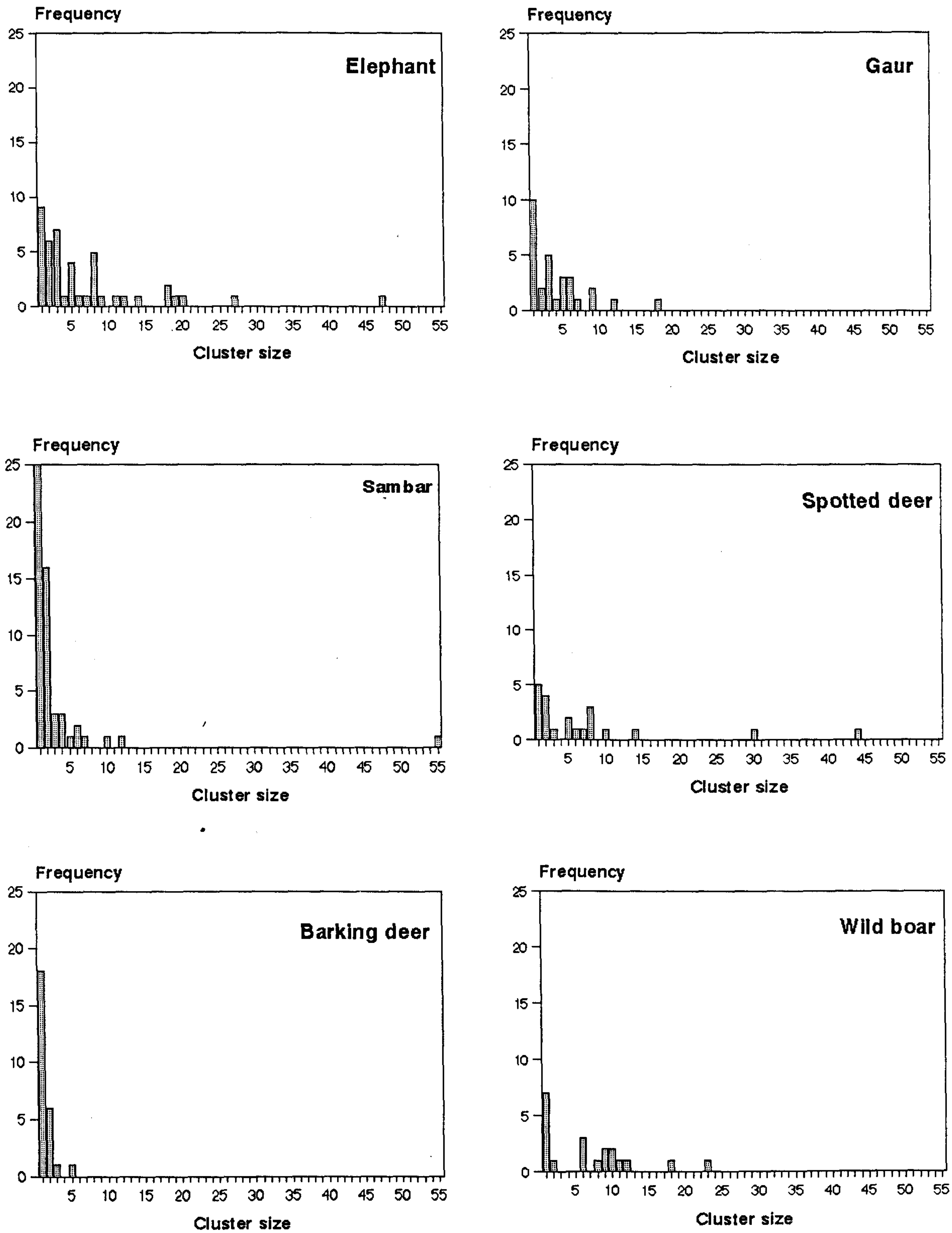

Figure 3. Distribution of cluster size for the six species

reference to the commonly held assumption that most of the animals are detected in total count.

The present study has however suffered from a limitation that due to the low number of sightings in individual sanctuaries, the data had to be pooled over different sanctuaries. These sanctuaries although dominated by moist deciduous forests carry mixed vegetation types in mosaic form in varying order giving rise to variations in local density of animal populations. Stratification of the population was not made because prior information on density of animals or any associated characteristics like vegetation type was not available. However, the results are of some validity and shall be useful in developing density estimates for populations comparable to the pooled structure of the sanctuaries. In a way, this study reconfirms some of the findings made by Varman and Sukumar (1995) in Karnataka but proposes an alternative estimator for average cluster size based on the precision of animal density estimates obtained for the six species considered in this work.

Table 2. Estimates of density obtained using different detection function models

Species	vlodel	Herd density (no.km-2)	Adjusted mean herd size	Animal density using mean (no.km-2)	Median herd size	Animal density using median (no.km-2)	$\chi 2$ Value
Elephant	UNE	1.03 (21.57)	7.14 (18.41)	7.36 (28.64)	3.50 (23.69)	3.61 (33.02)	12.94N
	JHN	0.55 (18.77)	7.14 (18.41)	3.94 (26.52)	3.50 (23.69)	1.93 (16.65)	24.99 *
	JFS	1.15 (22.94)	7.14 (18.41)	8.21 (29.72)	3.50 (23.69)	4.03 (37.94)	5.19 NS
	BNE	1.11 (28.77)	5.88 (17.07)	6.44 (32.16)	3.50 (23.69)	3.89 (41.06)	9.23 NS
	BHN	0.60 (22.80)	5.88 (17.07)	3.50 (26.95)	3.50 (23.69)	2.10 (19.66)	22.78 *
	BFS	1.15 (16.99)	7.26 (23.42)	8.35 (25.33)	3.50 (23.69)	4.03 (33.54)	----
	PAM	1.14 (32.15)	7.14 (18.41)	8.10 (37.04)	3.50 (23.69)	3.99 (45.41)	----
	KDE	0.74 (15.07)	7.14 (18.41)	5.26 (22.88)	3.50 (23.69)	2.58 27.34)	----
	TC	0.05	5.26	0.26	4.00	0.20	-----
Gaur	UNE	0.44 (26.74)	4.17 (17.56)	1.83 (32.34)	3.00 (18.00)	1.32 (14.16)	7.01 NS
	UHN	0.26 (23.29)	4.17 (17.56)	1.07 (29.46)	3.00 (18.00)	0.78 (7.57)	13.03 *
	UFS	0.51 (30.89)	4.17 (17.56)	2.15 (35.95)	3.00 (18.00)	1.53 (18.33)	18.06 *
	BNE	0.49 (35.68)	3.30 (16.70)	1.61 (37.30)	3.00 (18.00)	1.44 (19.48)	7.26NS
	BHN	0.27 (27.96)	3.63 (17.04)	0.99 (31.24)	3.00 (18.00)	0.81 (9.03)	14.51 *
	BFS	0.51 (23.86)	3.28 (26.22)	1.69 (25.74)	3.00 (18.00)	1.53 (15.31)	----
	PAM	0.82 (48.30)	2.42 (20.33)	1.98 (52.41)	3.00 (18.00)	2.46 (42.07)	----
	KDE	0.30 (18.57	4.17 (17.56)	1.25 (25.42)	3.00 (18.00)	0.90 (25.59)	----
	TC	0.03	4.66	0.14	3.00	0.09	

Note : NS-Not significant, *- Significant at p=0.05 level. The values in the brackets denote foefficient of variation of the estimates.

Table 2 Contd..,

Species	Model	Herd density (no.km^{-2})	Adjusted mean herd size	Animal density using mean (no.km^{-2})	Median herd size	Animal density using median (no.km^{-2})	χ^2 Value
Sambar	UNE	1.46 (19.62)	3.34 (31.16)	4.89 (37.32)	2.00 (11.38)	2.92 (33.03)	13.25 NS
	UHN	0.76 (17.05)	3.34 (31.16)	2.53 (35.91)	2.00 (11.38)	1.52 (15.48)	30.96 *
	UFS	1.62 (21.49)	3.34 (31.16)	5.39 (38.44)	2.00 (11.38)	3.24 (39.15)	3.89 NS
	BNE	1.58 (23.10)	2.30 (17.51)	3.63 (27.50)	2.00 (11.38)	3.16 (40.67)	11.46 NS
	BHN	0.78 (19.46)	2.74 (23.-82)	2.14 (29.78)	2.00 (11.38)	1.56 (17.55)	38.75*
	BFS	1.53 (17.40)	1.95 (30.33)	2.99 (22.58)	2.00 (11.38)	3.06 (31.69)	----
	PAM	1.64 (27.38)	1.74 (12.44)	2.86 (30.00)	2.00 (11.38)	3.28 (48.51)	----
	KDE	1.06 (13.74)	3.34 (31.16)	3.54 (33.86)	2.00 (11.38)	2.12 (17.41)	----
	TC	0.06	3.08	0.18	2.00	0.12	----
Spotted deer	UNE	1.05 (31.66)	7.67 (30.22)	8.03 (44.80)	2.00 (28.64)	2.10 (44.82)	8.38 NS
	UHN	0.67 (27.63)	7.67 (30.22)	5.17 (41.79)	2.00 (28.64)	1.34 (26.79)	8.18 NS
	UFS	0.67 (30.86)	7.67 (30.22)	5.14 (44.19)	2.00 (28.64)	1.34 (28.25)	10.77 NS
	BNE	1.11 (41.19)	8.51 (31.25)	9.44 (52.75)	2.00 (28.64)	2.22 (55.71)	7.70 NS
	BHN	0.71 (33.47)	8.58 (31.33)	6.09 (47.12)	2.00 (28.64)	1.42 (31.30)	8.36NS
	BFS	0.67 (19.60)	9.40 (24.68)	6.30 (23.87)	2.00 (28.64)	1.34 (23.30)	----
	PAM	0.46 (21.82)	7.67 (30.22)	3.54 (37.28)	2.00 (28.64)	0.92 (16.63)	----
	KDE	0.62 (21.82)	7.67 (30.22)	4.76 (37.82)	2.00 (28.64)	1.24 (23.39)	----
	TC	0.05	7.49	0.38	5.00	0.25	

Table 2 Contd...

species	Model	Herd density (no.km^{-2})	Adjusted mean herd size	Animal density using mean (no.km^{-2})	Median herd size	Animal density using median (no.km^{-2})	χ^2 Value
Barking deer	UNE	1.21 (28.31)	1.46 (12.14)	1.76 (30.99)	1.00 (14.00)	1.21 (38.20)	10.64 NS
	UHN	0.58 (24.64)	1.46 (12.14)	0.85 (27.63)	1.00 (14.00)	0.58 (16.46)	33.27 *
	UFS	1.07 (23.61)	1.46 (12.14)	1.57 (26.79)	1.00 (14.00)	1.07 (29.51)	10.32NS
	BNE	1.26 (31.15)	1.43 (11.37)	1.79 (32.88)	1.00 (14.00)	1.26 (43.01)	12.78"
	BHN	0.63 (26.63)	1.30 (8.62)	0.82 (26.79)	1.00 (14.00)	0.63 (19.00)	32.37*
	BFS	1.07 (14.07)	1.57 (14.18)	1.69 (21.18)	1.00 (14.00)	1.07 (21.38)	----
	PAM	1.12 (31.02)	1.46 (12.14)	1.63 (33.31)	1.00 (14.00)	1.12 (38.11)	----
	KDE	0.81 (19.61)	1.46 (12.14)	1.18 (24.47)	1.00 (14.00)	0.81 (25.66)	----
	TC	0.03	1.25	0.04	1.00	0.03	----
Wild boar	UNE	0.75 (32.49)	6.85 (20.16)	5.15 (38.79)	6.00 (12.42)	4.50 (26.16)	3.15 NS
	UHN	0.44 (28.34)	6.85 (20.16)	2.99 (35.25)	6.00 (12.42)	2.64 (13.56)	7.86 NS
	UFS	0.70 (27.61)	6.85 (20.16)	4.76. (34.64)	6.00 (12.42)	4.20 (21.09)	7.51 NS
	BNE	0.79 (41.82)	6.62 (20.48)	5.25 (46.16)	6.00 (12.42)	4.74 (34.59)	1.75 NS
	BHN	0.46 (33.34)	6.64 (20.45)	3.03 (38.67)	6.00 (12.42)	2.76 (16.24)	6.41 NS
	BFS	0.70 (15.14)	6.90 (22.03)	4.80 (25.84)	6.00 (12.42)	4.20 (13.68)	----
	PAM	0.72 (53.64)	6.85 (20.16)	4.90 (57.31)	6.00 (12.42)	4.32 (39.45)	----
	KDE	0.51 (22.36)	6.85 (20.16)	3.47 (28.24)	6.00 (12.42)	3.04 (23.34)	----
	TC	0.03	5.75	0.17	4.00	0.12	----

Table 3. Two best models obtained for the different species with respect to the precision of estimates

Species	CV of cluster density	CV of animal density
Elephant	KDE	UHN(median)
	BFS	BHN(median)
Gaur	KDE	UHN(median)
	UHN	BHN(median)
Sambar	KDE	UHN(median)
	UHN	KDE(median)
Spotted deer	BFS	PAM(median)
	KDE	BFS (median)
Barking deer	KDE	UHN(median)
	PAM	BHN(median)
Wild boar	. BFS	UHN(median)
	KDE	BFS (median)

Table 4. Influence of group size (size bias) in the detection of animal groups

Species	Model	Sample size	Size bias parameter	P value of size bias parameter	Correlation between perpendicular distance and cluster size
Elephant	BNE	44	0.18	0.09	0.15 NS
	BHN		0.18	0.03	
Gaur	BNE	29	0.31	0.09	0.12 NS
	BHN		0.18	0.12	
Sambar	BNE	53	0.32	0.05	0.05 NS
	BHN		0.15	0.16	
Spotted deer	BNE	21	-0.08	0.66	-0.12 NS
	BHN		-0.09	0.73	
Barking deer	BNE	26	0.10	0.43	-0.03 NS
	BHN		0.50	0.10	
Wild boar	BNE	20	0.04	0.42	0.05 NS
	BHN		0.03	0.40	

4. VISUAL ASSESSMENT OF DISTANCE AND SAMPLING INTENSITY IN LINE TRANSECT SAMPLING

In line transect sampling, measurement of distance to objects sighted is supposed to be made accurately using a tape stretched on the ground. Range finders are also in use for the purpose. In practice, these procedures are difficult and distances are assessed visually by the observers. Visual estimates are likely to be biased. Hence an examination was made as to the effect of inaccuracy in visual estimation of distance on density estimate of animals. The methods followed and the results obtained in this respect are explained in the following. The nature and magnitude of errors in visual estimation were first assessed by conducting a physical experiment and thereafter the effect of errors in distance measurement on density estimate was evaluated analytically and through simulation.

4.1. The extent of error in visual assessment of distance

An *ex situ* trial was conducted to assess the agreement between actual distance and visual estimate of distance made by the observers. Wooden poles were fixed at known distances ranging from 5 m to 100 m in random sequence on a flat ground and volunteers were asked to estimate the distance by visual observation. A total of 92 volunteers participated in the experiment. The results of field evaluation are given in Table 5. The mean bias in the visual estimation of actual distance was not significantly different from zero. However, the coefficient of variation of visual estimates of distance varied from 54 per cent in 0-20 m class to 34 per cent in 80-100 m class.

Table 5. The sampling variation in visual estimates of distance

Actual distance (m)	Mean bias of visual estimate of distance (m)	CV of visual estimate of distance (%)
0 - 20	1.55	54.36
20 - 40	2.29	40.07
40 - 60	-2.68	35.26
60 - 80	-2.17	35.66
80 -100	0.47	33.91
Mean	-0.54	39.85

Several regression models were fitted, taking visual estimate of distance as dependent variable and actual distance as the independent variable. In the absence of any bias, the ideal regression line will have zero intercept and unit slope. Regression equations fitted to the data obtained in this respect are given in Table 6. Simple linear regression function through the origin had the highest coefficient of determination of 0.88. The corresponding regression coefficient was 0.98 indicating an underestimation by 2 m for every 100 m of actual distance which is negligible compared to the error of estimation. The corresponding weighted linear regression also showed similar results with lesser bias. Addition of an intercept or quadratic term did not improve the fit. The logarithmic model also did not fare well.

Table 6. Regression statistics obtained for the relation between visual estimate of distance and actual distance

Regression equation	Remarks	Adj. R^2
$Y = 2.8600 + 0.9300X$ (0.9164) (0.0183)	Simple linear regression (SLR) model with intercept	0.61 (**)
$Y = 0.9800\ X$ (0.0088)	SLR model through origin	0.88 (**)
$Y = 0.9970\ X$ (0.0076)	Weighted linear regression model through origin	0.86 (**)
$Y = 1.9300 + 0.9800\ X + 0.0006\ X^2$ (1.6197) (0.0781) (0.0008)	Quadratic model	0.61(**)
$\ln Y = 0.0900 + 0.9600 \ln X$ (0.0556) (0.0152)	Double logarithmic model	0.71(**)

Note : Y = Visual estimate of distance, X = Actual distance measured by tape,

** - Denotes significance at P = 0.01 .The values in brackets denote standard errors of the estimates. In weighted linear regression, the weights were inversely proportional to the variance of errors at each X value.

The mode visual judgement of distance could be influenced by the surroundings and as such the same extent of error need not be expected under forest conditions. However, such effects have not been considered here since there is no effort made to develop any adjustment factor in density estimate for variation in distance assessment. Here the main concern has been to obtain an idea of the range and nature of errors and to see how the errors in visual judgement of distance would affect the density estimate.

4.2. Influence of errors in distance measurements on density estimate

The basic data for estimation of animal density through line transect sampling consist of distance measurements. The data on distance are utilized to estimate the parameters of the detection function and thereby the density. In the case of parametric models like half normal model, closed analytical form solutions are available for estimation of density. For nonparametric models like Fourier series, the equations change with the data depending on the number of terms in the series used for estimation of the detection function. The effects of errors in distance measurement on the density estimate and its standard error are evaluated in the following using both half normal and Fourier series models.

4.2.1. Half normal model

Let there be n sightings from a line transect sampling with a total transect length L. Let x_i, i = 1 ,..., n be the n perpendicular distance values recorded.
The half normal detection function is given by

$$g(x) = \exp\left(\frac{-x^2}{2\sigma^2}\right), \quad 0 \le x < \infty \tag{8}$$

With no truncation, the density function of detection distances is

$$f(x) = \frac{g(x)}{\mu} \text{ where} \tag{9}$$

$$\mu = \int_0^\infty g(x)dx = \int_0^\infty \exp\left(\frac{-x^2}{2\sigma^2}\right)dx = \sqrt{\frac{\pi\sigma^2}{2}} \tag{10}$$

Given n detections, the likelihood function is

$$L = \prod_{i=1}^{n} g(x_i)/\mu = \left\{\prod_{i=1}^{n} \exp\left(-x_i^2/2\sigma^2\right)\right\} \Big/ \mu^n \tag{11}$$

$$l = \log_e L = -\sum_{i=1}^{n}\left\{\frac{x_i^2}{2\sigma^2}\right\} - n\log_e\left\{\sqrt{\frac{\pi\sigma^2}{2}}\right\} \tag{12}$$

Differentiating l with respect to σ^2 and setting the result = 0 gives

$$\frac{dl}{d\sigma^2} = \sum_{i=1}^{n} \frac{x_i^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0 \tag{13}$$

so that

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{x_i^2}{n}. \text{ Then } \hat{f}(0) = \frac{1}{\hat{\mu}} = \sqrt{\frac{2}{\pi\sigma}} \tag{14}$$

By evaluating the Fisher information matrix, we get $\mathrm{VAR}\left(\hat{\sigma}^2\right) = \frac{2\sigma^4}{n}$ from which

$$\mathrm{VAR}\left\{\hat{f}(0)\right\} = \frac{1}{n\pi\sigma^2} = \frac{\{f(0)\}^2}{2n} \tag{15}$$

$$D = \frac{nf(0)}{2L} = \left[2\pi L^2 \sum_{i=1}^{n} \frac{x_i^2}{n^3}\right]^{-0.5} \tag{16}$$

$$VAR(\hat{D}) = \hat{D}^2\left\{(CV(n))^2 + (CV(\hat{f}(0))^2\right\} \tag{17}$$

$$[CV(\hat{f}(0))]^2 = \frac{VAR(\hat{f}(0))}{(\hat{f}(0))^2} = \frac{(f(0))^2}{2n(f(0))^2} = \frac{1}{2n} \tag{18}$$

$$[CV(n)]^2 = \frac{VAR(n)}{n^2} = \frac{1}{n} \tag{19}$$

$$VAR(\hat{D}) = \hat{D}^2\left(\frac{1}{n} + \frac{1}{2n}\right) \tag{20}$$

$$CV(\hat{D}) = \left(\frac{1}{n} + \frac{1}{2n}\right)^{-1/2} \tag{21}$$

which is independent of x, the distance values. The above derivation shows that for a given set of detections and transect length, the density estimate is inversely related to the distance values whereas the coefficient of variation of density estimate is unaffected by changes in distance values.

As a specific case of evaluation of the effect of over/under estimation of distance measures on density estimate, data on sightings of sambar collected during the Wildlife Census - 1993 were used. By increasing the distance from 1 m to 5 m and by decreasing the distance from 1 m to 5 m uniformly from the data points, the density estimate was evaluated using equation (16). The shape of the graph is shown in Figure 4. When distance increases from the reference set of values, density estimate decreases and when distance decreases, density estimate increases meaning that for a given set of detections, overestimation of distances would lead to underestimation of density and *vice versa.*

The above account clearly indicates the direction of error in the density estimate associated with a systematic error in distance measurement. In practice, the errors associated with visual estimate of distance are random. The effect of such errors can be assessed only through stochastic simulation. Again the data collected during the wildlife census conducted in 1993 in Kerala for the species sambar were used as the reference set. The data points of the reference set were subjected to increasing levels of random disturbance by adding normally distributed random variates generated from $N(0,\sigma)$ populations, σ varying from 1 m to 5 m. Appropriate change of origin was made to the actual distance to eliminate the negative values produced when simulations are performed. For each set of randomly distorted distance values, the corresponding density estimates were worked out using the programme SIZETRAN. Univariate half normal detection function was used to estimate the density. This exercise was repeated thirty times and the mean of the density estimates at each σ value was found out. The changes occurring in the density estimate due to the increasing level of disturbance on the distance measurements are depicted in Figure 5. With increasing level of disruption in distance measurements, there was a declining trend in the density values. This has occurred because the positive and negative deviations in distance will not have the same effect on squaring (see equation 16) although their effects may cancel out in a simple sum without squaring.

4.2.2. Fourier series model

The formulae leading to the density estimate and its CV with Fourier series model for detection function are the following.

$$\hat{D} = \frac{n\hat{f}(0)}{2L} \tag{22}$$

where $\hat{f}(0) = \frac{1}{w^*} + \sum_{k=1}^{m} \hat{a}_k$

$$\hat{a}_k = \frac{2}{nw^*}\left[\sum_{i=1}^{n} \cos\left(\frac{k\pi x_i}{w^*}\right)\right], \quad k = 1, 2, 3$$

w^* = maximum distance considered in line transect sampling

x_i = i th perpendicular distance

From the general theory of line transect sampling, we know that

$$\left[CV(\hat{D})\right]^2 = \left[(CV(n))^2 + \left(CV(\hat{f}(0))\right)^2\right] \tag{23}$$

Under the assumption of a random, (Poisson) distribution of objects,

$$(CV(n))^2 = \frac{1}{n} \tag{24}$$

$$\left(CV(\hat{f}(0))\right)^2 = \frac{VAR(\hat{f}(0))}{(\hat{f}(0))^2} \tag{25}$$

$$V\hat{A}R(\hat{f}(0)) = \sum_{j=1}^{m}\sum_{k=1}^{m} C\hat{O}V(\hat{a}_j, \hat{a}_k) \tag{26}$$

$$C\hat{O}V(\hat{a}_k, \hat{a}_j) = \frac{1}{n-1}\left[\frac{1}{w^*}\left(\hat{a}_{k+j} + \hat{a}_{k-j}\right) - \left(\hat{a}_k \hat{a}_j\right)\right], k > j > 1 \tag{27}$$

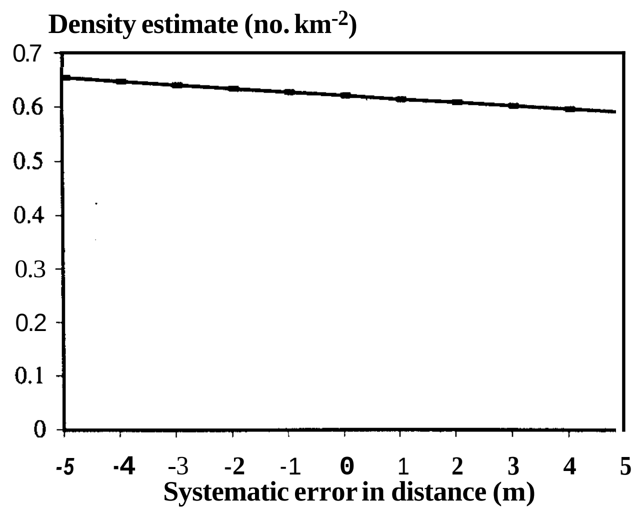

Figure 4. Effect of systematic error in distance measurement on density estimate using half normal detection function

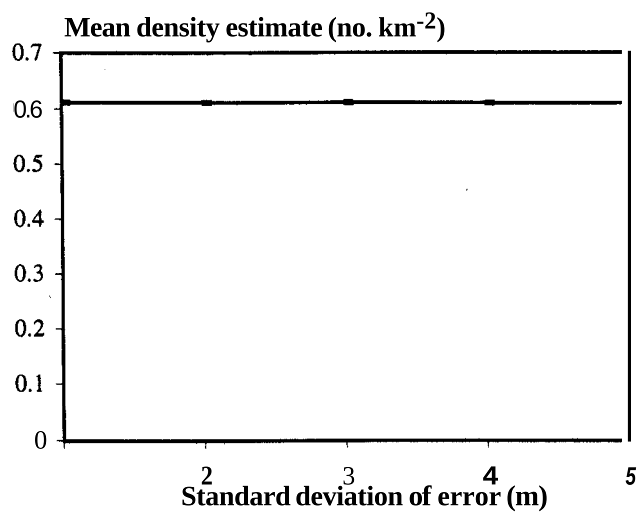

Figure 5. Effect of random error in distance measurement on density estimate using half normal detection function

Both the density estimate and its variance are complex functions of distance values and no specific direction can be identified because of the periodic nature of the cosine function which will increase or decrease depending on the specific range of x values present in the data. However, the effect of errors in distance measurement on the density estimate and its CV were investigated using a specific data set related to sambar, collected during the wildlife census conducted in 1993 the results of which are shown in Figures 6 and 7. The density estimate was found to decrease with increase in the distance values within the range considered both with systematic and random type of errors. The CV of the density estimate did not show any specific pattern although the general trend was on the decline.

4.3. Sampling intensity

The transect length needed and the number of sightings required to bring the coefficient of variation (CV) of the estimates to 20 per cent was estimated as described by Burnham *et al.* (1980). The coefficient of variation of D can be computed from equation (28)

$$[CV(D)]^2 = VAR(n)/(E(n))^2 + VAR(\hat{f}(0))/(\hat{f}(0))^2 \tag{28}$$

where VAR denotes sampling variance and E denotes expectation

To a first approximation VAR(n) = a1n and VAR ($\hat{f}(0)$) = $(f(0))^2$ a_2/n. The constants a, and a, are unknown parameters and in a given study they are independent (or almost so) of n, L and f(0). Thus replacing E(n) by just n

$$[CV(\hat{D})]^2 = (a_1 + a_2)/n = b/n \tag{29}$$

Where b is an unknown parameter. From the data of a previous survey, CV(D) and n can be obtained. Using these values, b can be calculated. Assume that the goal is to estimate D with a coefficient of variation of 20 per cent. The estimated sample size (n_1) can be obtained from the above equation and solving for line length by equating L/n = L_1/n_1, the transect length (L_1) needed to achieve 20 per cent coefficient of variation on D is worked out. L and n are the transect length used and the number of sightings obtained for each species in the pre-conducted survey data .

The number of sightings, transect length and sampling intensity needed to limit the CV to 20 per cent level for different herbivore species are given in Table 7 which were worked out utilizing the data of wildlife census conducted by the Forest Department during 1993. Sampling intensity for an area of 5 km^2 were worked out by dividing the transect length obtained for each species for limiting the CV to 20 per cent. by the area

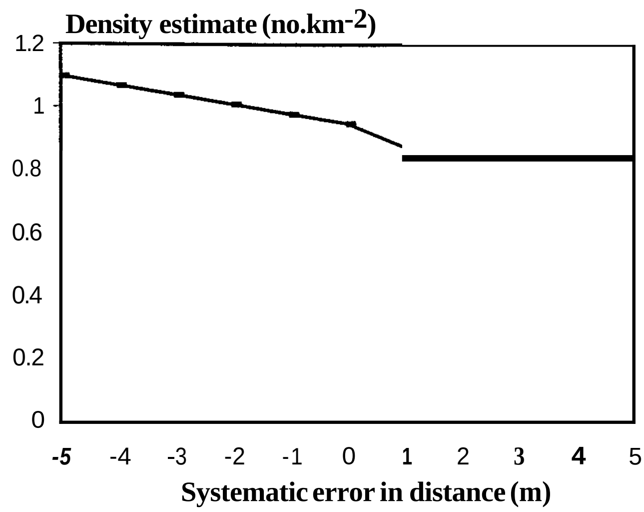

Figure 6a. Effect on density estimate

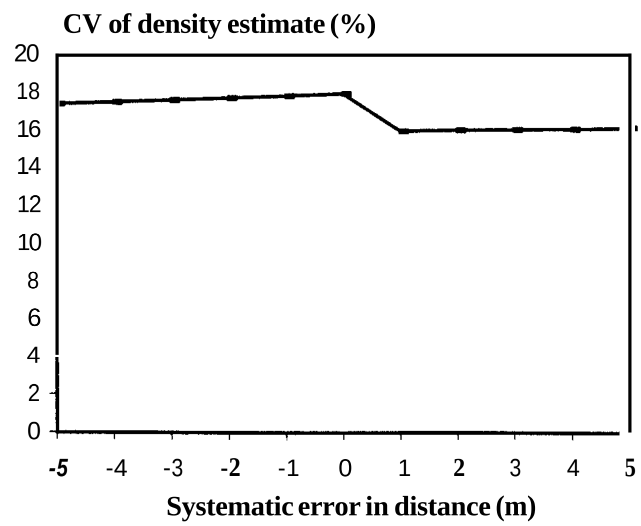

Figure 6b. Effect on CV of density estimate

Figure 6. Effect of systematic error in distance measurement on density estimate and its CV using Fourier series detection function

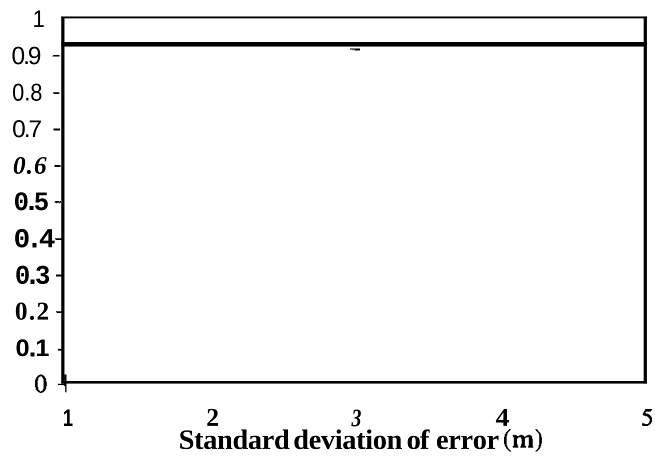

Figure 7a. Effect on density estimate

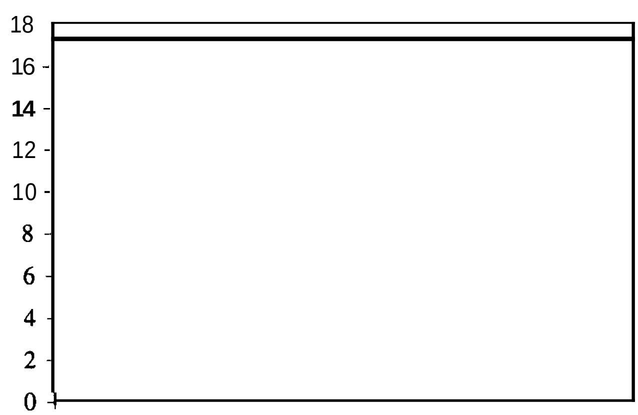

Figure 7b. Effect on CV of density estimate

Figure 7. Effect of random error in distance measurement on density estimate and its CV using Fourier series detection function

surveyed and multiplying by five. Sampling intensity varied for different species. On an average, one transect of 2km was necessary for every 5 km^2 of the area sampled.

Table 7. Number of sightings and transect length needed to reduce the CV to 20 per cent for each species

Species	Estimated sample size (sightings)	Estimated transect length (km)	Sampling intensity/ 5 km^2 (km)
Elephant	40	410.51	1.2
Gaur	38	599.34	1.7
Sambar	38	327.94	1.0
Spotted deer	41	892.57	2.6
Barking deer	41	712.18	2.1
Wild boar	39	891.76	2.6
Mean	--		1.9

5. PREDICTING THE VARIATION IN DETECTION FUNCTION IN LINE TRANSECT SAMPLING THROUGH RANDOM PARAMETER MODEL

Line transect sampling is one of the important methods for estimating animal abundance in tropical forests. It is a direct, cost effective method and involves only passive observations on the presence of animals. In the estimation process, the detection function plays an important role affecting the precision and accuracy of the density estimates. Major developments had taken place in this regard resulting in a number of models and techniques involving both parametric and nonparametric approaches (Burnham *et al.* (1980), Drummer and McDonald (1987), Quang (1991) Buckland *et al.* (1993)). However, certain limitations are inherent in these methods. For instance, a minimum number of 40 sightings are required for satisfactory estimation of the detection function in an area. The form of detection function is also found to vary with the local conditions associated with the forest type, weather conditions, observer's fatigue etc., the influence of which are difficult to be quantified. Hence there has to be a method by which we can calibrate (localize) the detection function for its variation from place to place, also allowing density estimation with a much lesser sample size. If the local conditions can be measured or at least can be categorized on a nominal scale, a generalised detection function can be formed including extraneous variables. The influence of such variables on the detection function parameters can then be assessed and localized predictions can be made based on measured local conditions. This may not be effective quite often and the alternative is to consider the variation in detection function over a region as random and bring the model under the framework of random parameter models. Results of efforts made with the latter approach are reported in the following.

5.1. Materials and methods

5.1.1. Model

General description of random parameter models can be found in Rao (1975), Graybill (1976) and also Vonesh and Carter (1987). Lappi and Bailey (1988) developed a height prediction model using random stand and tree parameters. Lappi (1991) used random parameter models to calibrate height-diameter and volume-diameter equations for trees of forest stands.

The model presented here is generally in line with the above referred works. Although several highly efficient functions have been proposed for modelling detection functions, only those which can be linearized can be used under the present approach. The two-parameter negative exponential model was chosen here as detection function model because of this requirement of linearity. The two-parameter negative exponential model can be described as

$$f(x; \mu, \sigma) = (1/\sigma) \exp[-(x-\mu)/\sigma], \quad 0 < x < \infty \tag{30}$$

where x = perpendicular distance
$f(x)$ = probability density function (pdf)
μ, σ are parameters

The cumulative density function (cdf), F(x) of the above function can be expressed as

$$F(x) = 1 - \exp[-(x-\mu)/\sigma] \tag{31}$$

and hence, $-\ln(1 - F(x)) = (1/\sigma)(x-\mu)$ (32)

Equation (32) can be written as

$y = A_0 + A_1 x$
where $A_0 = -\mu/\sigma$ and $A_1 = 1/\sigma$ (33)

Using the estimates of A_0 and A_1, μ and σ can be calculated and in turn f(0) which is

$$f(0) = (1/\sigma)\exp(\mu/\sigma) \tag{34}$$

An estimate of animal density (D) is obtained as

$$D = nf(0)/2L \tag{35}$$

where n = number of sightings
L = transect length

Under certain assumptions, variance of D is given by the following expression (Burnham *et al*(1980)).

$$VAR(\hat{D}) = (\hat{D})^2 \left\{ \left(\hat{CV}(n)\right)^2 + \left[\hat{CV}\left(\hat{f(0)}\right)\right]^2 \right\} \tag{36}$$

where $\hat{CV}(n)$ = Standard deviation of n /mean of n and
$\hat{CV}(\hat{f(0)})$ = Standard deviation of $\hat{f(0)}$ /Mean of $\hat{f(0)}$

The basic proposition here is that apart from the estimation errors, the between y and x in equation (33) can have different parameters in different locations and these can be viewed as random deviations from population level parameters. Hence the y value corresponding to the *i* th x value in location k can be described by the following model.

$$y_{ki} = A_0 + A_1 X_{ki} + a_{0k} + a_{1k} X_{ki} + e_{ki} \qquad (37)$$

where y_{ki} is the transformed cdf value for the *i* th x value in *k* th location.
x_{ki} is the *i* th perpendicular distance in the *k* th location.
A_0 and A_1 are fixed population parameters
a_{0k} and a_{lk} are random location parameters having zero expectations
e_{ki} is the random residual error attached to the observation y_{ki}

All random parameters of the same location are supposedly correlated with each other but not of different locations, and the residual errors are assumed to be independent with constant variance.

5.1.2. Parameter estimation

In the first phase of the analysis, the problem was to estimate the fixed parameters, the variances and covariance of random parameters, and the variance of residual errors. Later, a general form of prediction equation available in the theory of linear models was used to predict location level random parameters which effectively provided the location specific detection functions. The estimation of the model parameters proceeded as follows.

Step 1. The fixed parameters of model (37) were estimated first through Ordinary Least Squares (OLS) by combining y and x values from all locations.
Step 2. The fixed effects were eliminated from the model by computing y_{ki}- $\hat{A}_0$-- $\hat{A}_1 x_{ki}$ where $\hat{A}_0$ and $\hat{A}_1$ are OLS estimates of fixed parameters from step 1. The variances and covariance of the random parameters and the variance of residual errors were then estimated through the method proposed by Rao (1975). The details are provided in Appendix 1.
Step 3. The fixed parameters were then re-estimated through Generalised Least Squares (GLS) treating the random parameters as part of the error term and utilising their estimates from step 2. The details are given in Appendix 1.

After the estimation phase, the random parameters for any specific location could be predicted utilising a few y and x measurements from that location. The Best Linear Unbiased Predictor (BLUP) for the purpose is described below. Suppose that y_k , the set of transformed cdf values for the *k* th location is generated according to the random parameter model.

$$y_k = X_k a + Z_k b_k + e_k \qquad (38)$$

where $\mathbf{y}_k' = [\mathbf{y}_{k1}, \mathbf{y}_{k2}, ..., \mathbf{y}_{kn_k}]$

$$\mathbf{X}_k = \mathbf{Z}_k = \begin{bmatrix} 1 & x_{ki} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_{kn_k} \end{bmatrix}$$

$\mathbf{a}' = [A_0, A_1]$
$\mathbf{b}_k' = [a_{0k}, a_{1k}]$

b_k is a random parameter vector to be predicted, with $E(b_k) = 0$ and $var(b_k) = D_k$ and e_k is a Vector Of random errors With $E(e_k) = 0$ and $Var(e_k) = R_k$ and $COV(b_k, e_k') = 0$. E stands for expectation operator, VAR for variance and COV stands for covariance.

The $VAR(\mathbf{y}_k) = \mathbf{Z}_k\mathbf{D}_k\mathbf{Z}_k' + \mathbf{R}_k$ and $cov(\mathbf{b}_k, \mathbf{y}_k') = \mathbf{D}_k\mathbf{Z}_k'$

The BLUP of b_k is

$$\hat{\mathbf{b}}_k = \mathbf{D}_k\mathbf{Z}_k'(\mathbf{Z}_k\mathbf{D}_k\mathbf{Z}_k'+\mathbf{R}_k)^{-1}(\mathbf{y}_k-\mathbf{X}_k\hat{\mathbf{a}}) \tag{39}$$

This formula would require inversion of matrix with dimension equal to the number of observations. The predictor b_k and $var(\hat{\mathbf{b}}_k-\mathbf{b}_k)$ can equivalently be computed as

$$\hat{\mathbf{b}}_k = [\mathbf{Z}_k'\mathbf{R}_k^{-1}\mathbf{Z}_k + \mathbf{D}_k^{-1}]^{-1}\mathbf{Z}_k'\mathbf{R}_k^{-1}(\mathbf{y}_k-\mathbf{X}_k\hat{\mathbf{a}}) \tag{40}$$

$$VAR(\hat{\mathbf{b}}_k - \mathbf{b}_k) = [\mathbf{Z}_k'\mathbf{R}_k^{-1}\mathbf{Z}_k + \mathbf{D}_k^{-1}]^{-1} \tag{41}$$

These equations require inversion of matrix with dimension equal to the number of random parameters.

The calibrated detection function parameters for individual locations can then be obtained as

$$\hat{\mathbf{c}}_k = \hat{\mathbf{a}} + \hat{\mathbf{b}}_k = \begin{bmatrix} \hat{A}_0 + \hat{\mathbf{a}}_{0k} \\ \hat{A}_1 + \hat{\mathbf{a}}_{1k} \end{bmatrix} \tag{42}$$

$$\text{VAR}(\hat{\mathbf{c}}_k) = \text{VAR}(\hat{\mathbf{a}}) + \text{VAR}(\mathbf{b}_k) + 2\,\text{COV}(\hat{\mathbf{a}}, \text{b}_k) \tag{43}$$

where $\text{VAR}(\hat{\mathbf{a}}) = (\mathbf{X}_k'\mathbf{V}_k^{-1}\mathbf{X}_k)^{-1}$

$\text{VAR}(\hat{\mathbf{b}}_k) = \text{VAR}(\hat{\mathbf{b}}_k - \mathbf{b}_k)$

$\text{COV}(\hat{\mathbf{a}}, \hat{\mathbf{b}}_k) = -\text{VAR}(\hat{\mathbf{a}})\mathbf{X}_k'\mathbf{A}_k'$

$\mathbf{A}_k = (\mathbf{Z}_k'\mathbf{R}_k^{-1}\mathbf{Z}_k + \mathbf{D}_k)^{-1}\,\mathbf{Z}_k'\mathbf{R}_k^{-1}$

From the estimates of the transformed detection function for any sanctuary, the corresponding density estimates can be worked out using equations (33), (34) and (35). The methods described above are illustrated using data from a set of wildlife sanctuaries (locations) in Kerala. Widely separated temporal repetitions of a survey in the same location would also qualify for data from several locations.

5.1.3. Data

The data collected during the wildlife census conducted in the State of Kerala in 1993 through line transect sampling by direct sighting were used for the analysis. The species considered was sambar *(Cervus unicolor).* The line transect sampling was done on the 30th April, 1993. Data from eight Wildlife sanctuaries which were predominantly of tropical moist deciduous forests were taken for the analysis. The sanctuaries were Wayanad, Aralam, Parambikulam, Peechi-Vazhani, Idukki, Peppara, Neyyar and Periyar Tiger reserve. Their location and extent are described in Chapter 3 of this report. Additionally data from Chinnar Wildlife Sanctuary which has dry deciduous forests'were included in the data set. Chinnar is located between 10^{o} 05' and 10^{o} 22' W latitude and 77^{o} 05' and 77^{o} 17'E longitude. As the number of sightings was low in many sanctuaries, observations from Tholpetty, Aralam, Peppara, Peechi, Idukki, Neyyar and Parambikulam were pooled. This constituted three groups of sanctuaries *viz.*,Wayanad (Group 1), Thekkady (Group 2) and the rest (Group 3) for the purpose of estimation. Line transect data from the Chinnar Wildlife Sanctuary during the period from July 1993 to December 1994 were also taken. It was subdivided into four groups. Data for the period 1993 was taken as Group 4. Data obtained for the year 1994 was subdivided into three groups each group consisting of data of four months. They were Group 5, Group 6 and Group 7. Wildlife census data from the Chinnar Wildlife Sanctuary obtained in April 1993 through line transect sampling was taken as Group 8. An additional census was conducted in Parambikulam Wildlife Sanctuary during 1996. It was taken as Group 9. The total transect length from all the sanctuaries including multiple census operations were 798.60 km.

The procedure adopted for the laying of the line transects and measurement of parameters like sighting distance and sighting angle have been described in the materials and methods section of chapter 3. For the analysis. the data were truncated to a maximum perpendicular distance of 200 m. The values of x for the individual sanctuaries are given in Appendix 2. The transformed cdf values, $y = -\ln(1-F(x))$, were worked out for each x value and each group of sanctuaries. In total, there were 154 pairs of x and y values from all the 9 groups. For the purpose of validation of the model, 25 to 30 per cent of the data points were separated randomly from each group of sanctuaries. The validation data set consisted of 42 pairs of x and y values. The different parameters of the model (37) were estimated as described earlier, utilising the estimation data set.

5.1.4. Validation of the model

The random parameter model of equation (37) was tested by simulated calibration using the validation data set with a different number of sample points for each simulation. Sample points (pairs of perpendicular distance (x) and transformed cdf (y) values) were randomly selected from each group of sanctuaries and the parameters of the detection function were predicted. The calibrated detection function was applied to the remaining points in the same group of sanctuaries for which cdf measurements were available. The difference between the actual and the predicted cdf value against each perpendicular distance was obtained and the mean and variance of the deviations (residuals) were computed. The mean of the residuals would indicate any possible bias in the calibration and the variance of the residuals would measure the extent of deviation on repeated sampling. Additionally R^2 (prediction) was computed by dividing the sum of squares of the residuals with corrected sum of squares of the cdf values, and subtracting the result from 1. The R^2 (prediction) is a measure of the extent of agreement of the actual cdf with the predicted cdf on an average in relation to the variation in the actual cdf values. The computations were repeated 15 times using different seed values in the random number generator used to select sample point. The average values of the three statistics were computed. When no sample points were selected, cdf was predicted using the fixed population parameters over all the locations and the corresponding values of mean, variance and R^2 (prediction) were obtained.

5.2. Results and discussion

5.2.1. Parameter estimates

The estimates of the different parameters of the linear prediction equation of model (37) are given in Table 8. The figures in parenthesis are standard errors of the estimates. The estimate of $VAR(a_{0k})$ turned out to be negative. This is a common problem in variance components estimation. For the present analysis, the value of $VAR(a_{0k})$ was assumed to be a very small positive value in succeeding computations.

Table 8. Estimates of fixed parameters, variances and covariance of the random parameters and variance of residual error

Parameter	Estimate
Fixed parameters	
A_0	-0.03254
	(0.03083)
A_1	0.04021
	(0.00738)
Variance and covariance of random parameters	
VAR (a_{0k})	-0.00501
VAR (a_{1k})	0.00048
COV (a_{0k}, a_{1k})	-0.00121
Variance of residual error	0.04593

5.2.2. Validation of the model

The root mean squares of the residuals (sqrt(variance + observed bias^2)) and R^2 (prediction) obtained with different number of sample points are given in Table 9. The number of sample points had to be limited to 2 because there were at most three points for which x and F(x) values were available in the validation data set in the case of some sanctuaries. The bias seems to be negligible. The extent of bias is much lower when calibrated models are used to predict the detection function parameters. The root mean square of residuals decreased and (prediction) increased with increasing sample size as expected under the random parameter model. However, the advantage seems to have stopped with just one sample point used for calibration indicating that no more than one sighting is required to obtain a satisfactory prediction of the density in a region through random parameter model. This conclusion need not be generalised. With a different form of detection function and a different data set, the situation could be different. Further investigations in this line is warranted. But a conclusion which is evident is that an estimate of animal density in a region is impossible to obtain with just one sighting from that region and this has been made possible through the random parameter approach.

Table 9. Mean, Root mean square error(RMSE) and R^2(prediction) computed from the residuals for the validation data set

Number of sample points	Mean of residuals of cdf	RMSE of residuals of cdf	R^2(prediction)
0	-0.04312	0.01781	0.71734
1	0.00044	0.00800	0.88120
2	-0.01016	0.00746	0.88305

The detection function model used in the present case was not a good choice but was chosen only from the point of view of linearizable cdf. The negative estimate of variance of one of the random parameters was also partly due to the algebraic dependency of the μ and σ involved in the detection function. The study however has opened up an entirely new approach of dealing with variation in detection function over locations and brought out a method of developing animal density estimates with minimum possible observations in any particular location.

6. DETECTION FUNCTION MODELS FOR INDIRECT EVIDENCE

Evidences left by animals are a sure means to detect their presence. Sometimes such signs are successfully used for estimating the density of animal populations (Barnes and Jensen (1987), Varman *et al.* (1995)). In any case, accurate estimation of the density of indirect evidences will be the first step involved. Other than plot or strip sampling, quite often line transect methods are used for the purpose. As usual, the problem of identifying an appropriate detection function model arises in utilizing the data obtained through line transect sampling. An attempt is made here to identify a suitable model for detection function for data on dung piles obtained through line transect sampling in the case of elephant and gaur.

6.1. Materials and methods

Data collected during the course of the wildlife census conducted in the State of Kerala in April 1997 were used for the present study. The total forest area was divided into small blocks of about 6km^2 on an average, utilizing Survey of India maps. The total number of such blocks in the State was, 1506. The actual size of the blocks varied from 1 to 28 km2. These blocks formed the basic sampling units. About 36 per cent of the blocks in each Forest Division were selected for the survey through simple random sampling without replacement. In each selected block, a transect of 2 km was laid out and perpendicular distance to dung piles of elephant and gaur was measured using a tape. The dung piles were noted as belonging to three stages *viz.,* (i) fresh and moist, (ii) old and dry, and (iii) very old. The data pertaining to the first two stages were utilized for the present study.

Before proceeding with the estimation of dung density, the blocks were post-stratified as belonging to different vegetation types as follows. In each of the selected blocks visual estimation of percentages of area belonging to different forest types were made. Cluster analysis was carried out utilizing these data and the cluster membership of each block was found out. For the cluster analysis, the distance measure used was squared Euclidean distance and the clustering method used was Average linkage between groups (Norusis, 1988). The different clusters identified were predominantly (I) evergreen (II) moist deciduous (III) dry deciduous and scrub (IV) plantations (V) shola and grassland. The mean composition of the different clusters is indicated in Appendix 3. The block level data were pooled for each vegetation type and the combined density of dung belonging to fresh and old stages was estimated using programs SIZETRAN (Drummer, 1991) and DISTANCE (Laake *et al.*, 1994) for each vegetation type. The tansect length sampled for clusters I, II, III, IV and V were 371.30, 144.89, 184.99, 149.35 and 54.72 km respectively.

6.2. Results and discussion

The observed frequency distributions of distance values for the two species in different vegetation types are given in Figures 8 and 9. The dung density estimate and coefficient of variation obtained for elephant and gaur are given in Tables 10 and 11. The models

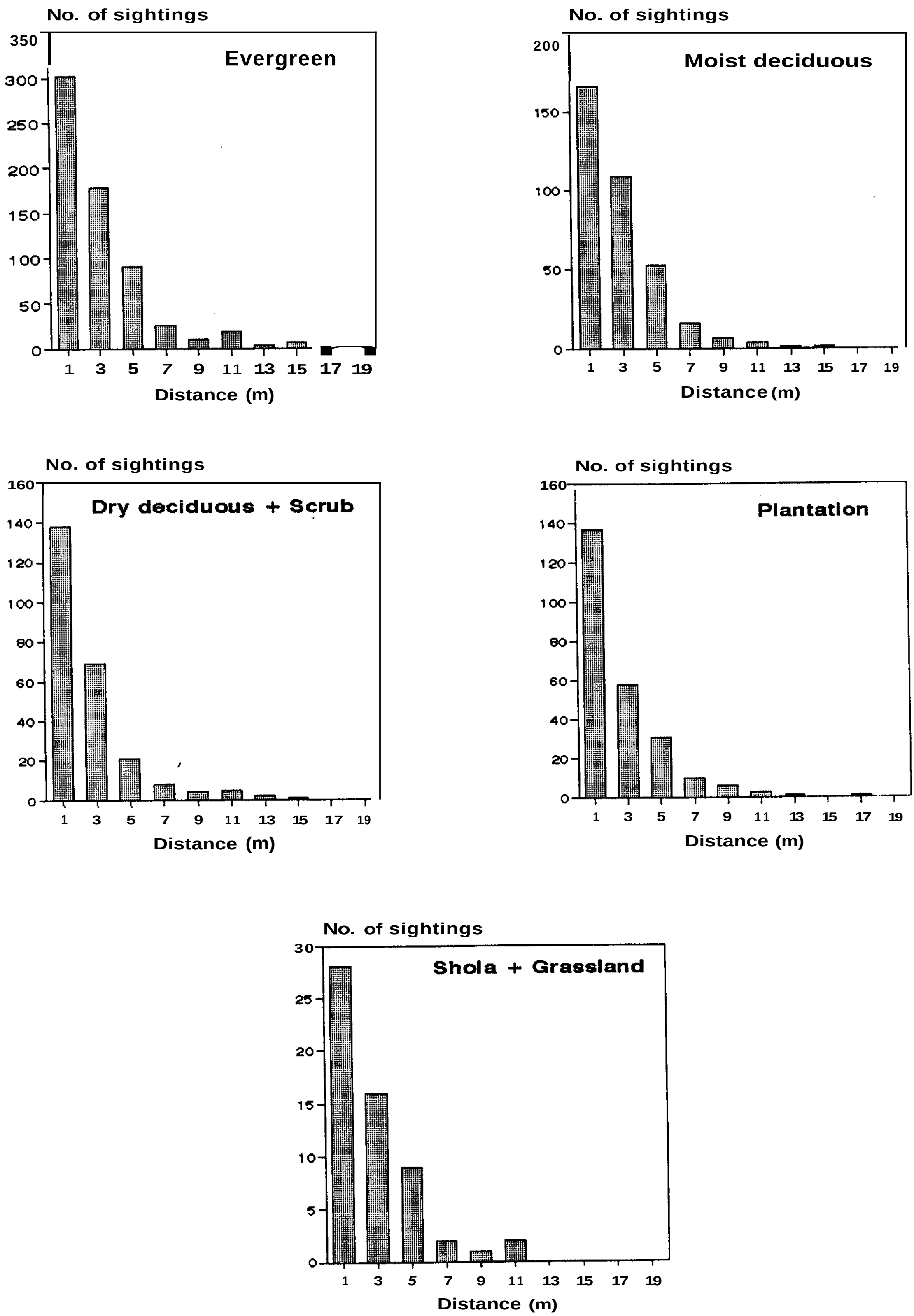

Figure 8. Distribution of perpendicular distance to dung of elephant in different vegetation types

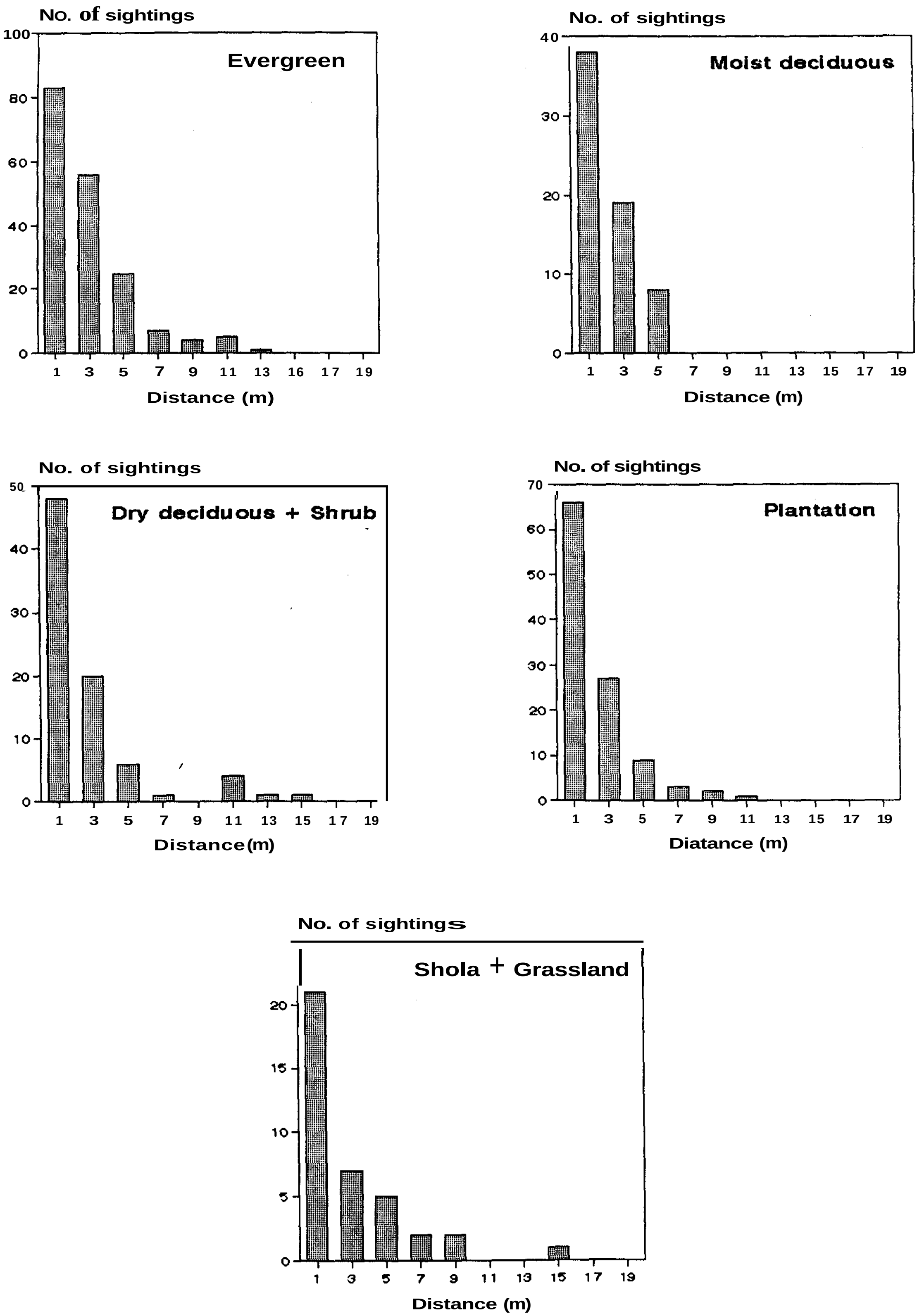

Figure 9. Distribution of perpendicular distance to dung of gaur in different vegetation types

used were Fourier Series, negative exponential, half normal and hazard cosine. In the case of elephant, least coefficient of variation (CV) was obtained with the Fourier series model in the first four clusters *viz.*, evergreen, moist deciduous, dry deciduous and scrub and plantations. In the fifth cluster (shola and grassland), half normal model had the least CV. In the case of gaur, the least CV was obtained for Fourier Series model in cluster I and cluster IV. In cluster II III and V least CV was obtained for half normal distribution.

Out of the four models tried, Fourier series is the most flexible model which can be made to fit a wide range of shapes for the detection function. It belongs to the class of nonparametric models and possesses several robustness properties (Burnham *et al.* . 1980). Since it has shown good results in the case of both elephant and gaur in most of the vegetation types considered here, it is recommendable as a detection function model to be used for estimating dung density in the case of these two species.

Table 10. Density estimate and coefficient of variation obtained for different detection function models in the case of elephant dung

Model	CL I	CL II	CL III	CL IV	CL V
Fourier Series	90.58 (4.01)	362.61 (6.90)	172.41 (6.94)	299.25 (9.17)	177.56 (19.02)
Negative exponential	218.51 (5.53)	486.77 (7.46)	267.46 (8.94)	346.35 (9.00)	215.85 (18.73)
Half normal	61.22 (6.78)	246.62 (9.15)	109.29 (10.96)	183.76 (11.03)	125.36 (16.26)
Hazard cosine	252.01 (5.93)	345.67 (7.76)	241.88 (9.54)	283.98 (10.00)	163.92 (21.67)

Table 11. Density estimate and coefficient of variation obtained for different detection function models in the case of gaur dung

Model	CL I	CL II	CL III	CL IV	CL V
Fourier Series	59.83 (8.58)	107.45 (23.92)	105.78 (16.49)	157.19 (12.89)	129.51 (22.21)
Negative exponential	87.97 (10.51)	130.86 (17.68)	100.34 (15.81)	195.02 (13.67)	126.16 (23.26)
Half normal	40.27 (12.89)	78.32 (15.40)	47.42 (13.71)	106.48 (16.77)	67.73 (20.21)
Hazard cosine	63.32 (10.87)	443.06 (33.78)	316.68 (25.43)	154.24 (15.77)	124.30 (31.00)

7. CONCLUSIONS

7.1. Choice of detection function

An important factor in line transect sampling is the estimation of detection function which describes the changes in the chance of detection of animals with increasing perpendicular distance from the transect. If the probability of detection varies with the herd size, size bias will also have to be taken into account while formulating the detection function. Data collected from eight sanctuaries during the Wildlife census conducted in 1993 jointly by Kerala Forest Department and Kerala Forest Research Institute, were utilised to compare the relative efficiency of different detection function models for estimating the abundance of herbivores. The following species *viz.,* elephant, sarnbar, spotted deer, barking deer, wild boar and gaur were considered for the study. Univariate half normal distribution was found promising with respect to precision of the density estimates. The bivariate procedures were not effective as the size bias parameter was not significant for most of the species considered.

7.2. Estimator for average cluster size

The distribution of cluster size in the case of the six herbivore species considered for the study was found to be highly skewed. Arithmetic mean shall not be a good estimator of average cluster size in such cases. The use of median for average cluster size brought down the variance of the animal density estimates and also provided realistic values of,the density since the median is unaffected by extreme values in the population.

7.3. Visual estimation of distance

An examination of the theory showed that for a given set of detections, overestimation of distances in the field would lead to underestimation of density in the case of line transect sampling and *vice versa.* An *ex situ* trial was conducted to assess the agreement between actual distance and visual estimates made by the observers. Wooden poles were fixed at known distances in random sequence on a flat ground and volunteers were asked to estimate the distance by visual observation. Simple linear regression equation fitted through the origin showed that there was underestimation by 2 m for every 100 m of actual distance which is negligible. The mean bias in the visual estimation of actual distance was not significantly different from zero. However, the coefficient of variation of visual estimates of distance varied from 54 per cent in 0-20 m class to 34 per cent in 80-100 m class. Increasing disruption of random nature in distance measurements was found to bring down the density estimate on an average, for a fixed set of detections and transect length.

7.4. Sampling intensity

The sampling intensity needed in line transect sampling to bring the coefficient of variation of density estimates to 20 per cent was estimated as described by Burnham *et al*(1980). The sampling intensity was different for different species. On an average, one transect of 2 km was found necessary for every 5 km^2 of the area sampled.

7.5. Calibration of detection function

In line transect sampling, the detection function plays an important role affecting the precision and accuracy of the density estimates. Major developments had taken place in this regard resulting in a number of models and techniques involving both parametric and nonparametric approaches. However certain limitations are inherent in these methods. For instance, a minimum number of 40 sightings are required for satisfactory estimation of the detection function in an area. The form of the detection function is found to vary with the local conditions associated with the forest type, weather condition, observers fatigue etc. Hence there has to be a method by which we can calibrate (localise) the detection function for its variation from place to place, also allowing density estimation with a much lesser sample size.

The variation in detection function over a region can be considered as random and the detection function model can be brought under the framework of random parameter models, Hence a random parameter model was formulated taking the two parameter negative exponential model as detection function. The cumulative density function of this distribution was linearisable. The basic proposition was that apart from the estimation errors, the relation between perpendicular distance and cumulative density function of the number of sightings can have different parameters in different locations and these can be viewed as random deviations from population level parameters. The model was tested utilising data collected for the species sambar from 9 Wildlife sanctuaries at different periods. The sanctuaries were Wayanad, Aralam, Parambikulam, Peechi-Vazhani, Idukki, Peppara, Neyyar, Periyar Tiger Reserve and Chinnar. Data obtained were separated into estimation data set and validation data set. The estimation data set was used to estimate the different parameters of the linear prediction equation. The random parameter model was tested by simulated calibration using the validation data set with different number of sample points for each simulation. The difference between the actual and predicted cumulative density function values against each perpendicular distance was obtained. The mean and variance of the deviations revealed that bias is very negligible, variance decreased and R^2 (prediction) increased with increasing sample size as expected under the random parameter model. The method has the clear advantage of being able to develop density estimates based on very few observations from a location which would be impossible through traditional methods.

7.6. Detection function for indirect evidence

In the case of elephant and gaur, indirect evidence like dung density is a very strong indicator of the habitat use which is associated with animal density and therefore accurate estimation of dung density is important in the case of these species. An analysis of data on distance to dung piles, collected during the course of line transect sampling, indicated that Fourier series model is a good choice for detection function model in most of vegetation types existing in the forests of Kerala. Other than being a flexible and robust nonparametric model, the use of the model resulted in the least coefficient of variation for dung density estimates.

7.7. Overview of methods

The present study has shown that total count is inapplicable for estimation of animal abundance as it leads to heavy undercounting. Line transect sampling has a firm theoretical footing but suffers from low number of sightings arising from low density of animals or poor detection percentage. Calibration of detection functions using random parameter models shall go a long way in making localized prediction of animal density and hence future works should attempt to develop generalized prediction models based on random parameter models. The methods based on indirect evidences also hold promise for the future and works can be undertaken to convert indirect evidences to animal numbers. However, indices of abundance based on indirect evidences would serve most of the practical purposes in wildlife management.

8. REFERENCES

Anonymous, 1993. *Wildlife Census-Kerala 1993, A Report.* Kerala Forest Research Institute and Kerala Forest Department Peechi, Kerala.

Barnes, R. F. W. and Jensen, K. L. 1987. How to count elephants in forests. *IUCN African Elephant and Rhino Specialist Group, Technical Bulletin,* 1, 1-6.

Bennet, L. J., English, P. F. and Mccain, R. 1940. A study of deer population by use of pellet group counts. *Journal of Wildlife Management,* 4,398-403.

Bollinger, E. K., Gavin, T. A. and McIntyre, D. C. 1988. Comparison of transects and circular plots for estimating bobolink densities. *Journal of Wildlife Management,* 52,777-786.

Buckland, S. T. 1982. A note on the Fourier series model for analysing line transect data. *Biometrics,* 38, 469-477.

Buckland, S. T. 1992. Fitting density functions with polynomials. *Applied Statistics,* 41,63-76.

Buckland, S. T., Anderson, D. R., Burnham, K. P. and Laake, J. L. 1993. *Distance Sampling: Estimating Abundance of Biological Populations.* Chapman and Hall, London.

Burnham, K. P. and Anderson, D. R. 1976. Mathematical models for nonparametric inferences from line transect data. *Biometrics,* 32, 325-336.

Burnham, K. P. and Anderson, D. R. 1984. The need for distance data in transect counts. *Journal of Wildlife Management,* 48, 1248-1254.

Burnham, K. P., Anderson, D. R. and Lakke, J. L. 1979. Robust estimation from line transect data. *Journal of Wildlife Management,* 43,992-996.

Burnham, K. P., Anderson, D. R. and Lakke, J. L. 1980. Estimation of density line transect sampling of biological populations. *Wildlife Monograph,* Number 72.

Burnham, K. P., Anderson, D. R. and Lakke, J. L. 1985. Efficiency and bias in strip and line transect sampling. *Journal of Wildlife Management,* 49, 1012-1018.

Cook, R. D. and Martin, F. B. 1974. A model for quadrat sampling with visibility bias. *Journal of American Statistical Association,* 69, 345-349.

Dekker, A. J. F. M., Dawson, S. and Desai, A. A. 1991. An indirect method for counting Asian elephants in forests. *In:* Ramakrishnan *et al.* (Eds.) *Censusing Elephants in Forests*. Proceedings of an International Workshop 1991. Asian Elephant Conservation Centre, Bangalore, 54-64.

Drummer, T. D. 1991. SIZETRAN: Analysis of size-biased line transect data. *Wildlife Society Bulletin*, 19, 117-118.

Drummer, T. D. and McDonald, L. L. 1987. Size bias in line transect sampling. *Biometrics*, 43, 13-21.

Epanechnikov, V. A. 1969. Nonparametric estimation of a multivariate probability density, *Theory of Probability and its Applications*, 14, 153 - 158.

Fryer, M. J. 1977. A review of some nonparametric methods of density estimation *Journal of the Institute of Mathematics and its Applications*, 20, 335 - 354.

Gates, C. E. 1969. Simulation study of estimators for the line transect sampling method. *Biometrics*, 25, 317-328.

Goodman, L. A. 1960. On the exact variance of products. *Journal of American Statistical Association*, 55, 708-713.

Graybill, F. A. 19,76. *Theory and Application of the Linear Model*. Duxbury Press, N. Scituate, MA.

Hayne, D. W. 1949. An examination of the strip census method for estimating animal populaticns. *Journal of Wildlife Management*, 13, 145-157.

Kendall, M. G. and Stuart, A. 1977. *The Advanced Theory of Statistics. Volume 1, Distribution Theory;* Fourth Edition, Charles Griffin and Co. Ltd.

Laake, J. L., Buckland, S. T., Anderson, D. R. and Burnham, K. P. 1994. *Distance User's Guide Version 2.1.* Colorodo Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins.

Lappi, J. 1991. Calibration of height and volume equations with random parameters. *Forest Science*, 37, 781-801.

Lappi, J. and Bailey, R. L. 1988. A height prediction model with random stand and tree parameters: An alternative to traditional site index methods. *Forest Science*, 34, 907-927.

Leopold, A. 1933. *Game Management*. Charles Scribner's Sons, New York.

Martin, C. 1987. Inter specific relationships between Barasingha and Axis deer in Kanha National Park, India and relevance to management. *In:* Wemmer, C. M. (Ed) *Biology and Management of the Cervidae,* Smithsonian Institution, 299-306.

Ngampongsai, C. 1981. The application of pellet group counts in estimating sambar population. *Proceedings The second seminar on Wildlife of Thailand,* Vol. 2, 316-320.

Norusis, M. J. 1988. SPSS/PC + Advanced Statistics Version 2.0. for the IBM PC/XT/AT and PS/2. SPSS Inc. Chicago.

Quang, P. X. 1991. A nonparametric approach to size biased line transect sampling. *Biometrics,* 47,269-279.

Quinn, T. J. II. 1979. The effect of school structure on line transect estimators of abundance. *In: Contemporary Quantitative Ecology and Related Ecometrics*. (Eds; Patil *G*. P. and Rozenweig, M. L.) International Co-operative Publishing House, Fairland. MD. USA, 473-491.

Rao, C. R. 1975. Simultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics,* 31, 545-554.

Rao, P. V. and K. M. Portier. 1981. Line transect sampling of clustered populations. *In: The 1981 proceedings of the American Statistical Association:* Statistical computing section.

Robinette, W. L., Loveless, C. M. and Jones, D. A. 1974. Field tests of strip census methods. *Journal of Wildlife Management,* 38, 81-96.

Rodgers, W. A. 1991. *Techniques for Wildlife Census in India. A Field Manual.* Wildlife Institute of India. Dehra Dun.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics,* 27, 832 - 837.

Sale, J. B. and Berkmuller, K. (Eds) 1988. *Manual of Wildlife Techniques for India.* Field document No. 11. Wildlife Institute of India, Dehra Dun.

Seber, G. A. F. 1992. A review of estimating animal abundance *International Statistical Review, 60,* 129-166.

Silverman, B. W. 1982. Kernel density estimation using the fast Fourier transform. Statistical Algorithm A S 176. *Applied Statistics,* 31, 93 - 97.

Silverman, B. W. 1986. *Density Estimation for statistics and Data Analysis* Chapman and Hall, London.

Varman, K. S. and Sukumar, R. 1995. The line transect method for estimating densities of large mammals in a tropical deciduous forest. An evaluation of models and field experiments. *Journal of Biosciences,* 20,273-287.

Varman, K. S., Uma Ramakrishnan and Sukumar, R. 1995. Direct and indirect methods of counting elephants. A comparison of results from Mudumali Sanctuary. *In: A Week with Elephants, International Seminar on Asian Elephants,* Bombay Natural History Society, 331-340.

Varughese, R. 1992. A review of quantitative studies of wildlife populations with reference to wildlife censuses of Silent Valley National Park, Kerala. Paper presented at the Indian Science Congress held at Cochin, Kerala.

Vonesh, E. F. and Carter, R. L. 1987. Efficient inference for random-coefficient growth curve models with unbiased data. *Biometrics,* 43,

Zahul, S. 1989. Line transect sampling with unknown probability of detection along the transect. *Biornetrics,* 45,453-470.

9. APPENDICES

Appendix 1. Estimation of covariance for random effects

Assume that we have m individuals, and for each individual i we have data which can be described by the linear model:

$$y_i = Xa_i + \varepsilon_i \qquad \text{(Al)}$$

Where a_i is a random parameter vector with mean a, and ε_i is the residual error which is independent of a_i and $COV(\varepsilon_i) = \sigma^2 I$ ($COV(\varepsilon_i)$ is the variance-covariance matrix of ε_i). The problem is to estimate $cov(a_i)$ and σ^2 For each individual i, a_i can be estimated unbiasedly by the OLS estimate $\mathbf{\hat{a}_i} = (X'X)^{-1}X'y_i$. The estimation error $\mathbf{\hat{a}_i - a_i}$ is independent of a, and $cov(\mathbf{\hat{a}_i - a_i}) = \sigma^2 (X'X)^{-1}$.

Thus $cov(\mathbf{\hat{a}_i}) = cov(a_i + \mathbf{\hat{a}_i} - a_i) = cov(a_i) + cov(\mathbf{\hat{a}_i} - a,) = cov(a_i) + \sigma^2 (X'X)^{-1}$, ie., the covariance matrix of parameter estimates is the covariance matrix of the parameters plus the covariance matrix of estimation errors.

$Cov(\mathbf{\hat{a}_i})$ can be estimated without bias by the usual sample covariance matrix of the $\mathbf{\hat{a}_i}$'s and σ^2 can be estimated by $\hat{\sigma}^2 = m^{-1}\Sigma\hat{\sigma}^2_i$, where $\hat{\sigma}^2_i$ is the usual estimate of the residual variance for individual i (sum of squared residuals divided by the degrees of freedom). Thus an unbiased estimate of the $cov(a_i)$ can be obtained by subtracting $\hat{\sigma}^2(X'X)^{-1}$ from the sample covariance matrix of $\mathbf{\hat{a}_i}$'s.

If the model matrix X is different for different individuals (X_i for individual i), then σ^2 can be estimated without bias by the weighted average of σ^2_i 's, the weights being equal to the degrees of freedom for each individual. Rao (1975) suggested further more that $cov(\mathbf{a}_i)$ is to be estimated by subtracting the average estimation error matrix $m^{-1}\Sigma\hat{\sigma}^2_i$ $(\mathbf{X_i'X_i})^{-1}$ from the sample covariance matrix of the $\mathbf{\hat{a}_i}$'s.

Generalised least squares

The model(37) can be written in the following form.

$$y = X a + Z b + e \qquad \text{(A2)}$$

where y represents the vector of dependent variables, a is the vector of fixed parameters, b is the vector of random parameters and X and Z are the corresponding incidence matrices. The vector e is the set of residual errors.

Assume that we have K classes and in each class, k, there are n_k observations. The model for class k is

$$y_k = X_k a + Z_k b_k + e_k$$

Next, let D = var(b_k), R = var(e_{ki})I, v_k = var(y_k) = $Z_k D Z_k' + R$. Treating the random parameters of the model as part of the error term, the generalised least squares estimate of a is

$$\hat{\mathbf{a}} = \left(\sum_{k=1}^{K} \mathbf{X}_k' \mathbf{V}_k^{-1} \mathbf{X}_k \right)^{-1} \left(\sum_{k=1}^{K} \mathbf{X}_k' \mathbf{V}_k^{-1} \mathbf{y}_k \right) \tag{A4}$$

$$\mathrm{VAR}(\hat{\mathbf{a}}) = \left(\sum_{k=1}^{K} \mathbf{X}_k' \mathbf{V}_k^{-1} \mathbf{X}_k \right)^{-1} \tag{A5}$$

V_k^{-1} can be easily computed in the following manner.

$$V_k^{-1} = R^{-1} - R^{-1} Z_k (Z_k' R^{-1} Z_k + D^{-1})^{-1} Z_k' R^{-1} \tag{A6}$$

Appendix 2. Perpendicular distance (x) and cumulative density function (F(x)) of the sightings obtained for each group of sanctuary described in the text

Group	x	F(x)	Group	x	F(x)
Group1	0.00	0.06	Group3	12.67	0.35
Group1	2.60	0.12	Group3	12.85	0.41
Group1	3.01	0.18	Group3	17.32	0.47
Group1	4.69	0.24	Group3	28.27	0.53
Group1	5.62	0.29	Group3	34.63	0.59
Group1	10.00	0.35	Group3	37.58	0.65
Group1	10.26	0.41	Group3	42.41	0.71
Group1	17.49	0.47	Group3	44.98	0.77
Group1	21.70	0.53	Group3	48.29	0.82
Group1	22.97	0.59	Group3	68.92	0.88
Group1	24.99	0.65	Group3	135.91	0.94
Group1	26.80	0.71	Group4	10.00	0.07
Group1	49.51	0.77	Group4	10.41	0.14
Group1	50.00	0.82	Group4	11.49	0.21
Group1	57.05	0.88	Group4	13.02	0.29
Group2	2.89	0.05	Group4	15.32	0.36
Group2	4.50	0.11	Group4	19.28	0.43
Group2	7.66	0.16	Group4	24.62	0.57
Group2	9.39	0.21	Group4	24.99	0.64
Group2	10.00	0.26	Group4	28.18	0.71
Group2	12.99	0.32	Group4	31.68	0.79
Group2	14.94	0.37	Group4	44.98	0.86
Group2	17.49	0.42	Group4	64.25	0.93
Group2	18.83	0.47	Group5	0.00	0.10
Group2	20.00	0.53	Group5	1.74	0.14
Group2	21.64	0.58	Group5	3.47	0.21
Group2	25.70	0.63	Group5	5.21	0.24
Group2	38.29	0.68	Group5	7.66	0.31
Group2	40.00	0.74	Group5	8.55	0.34
Group2	64.25	0.79	Group5	8.66	0.38
Group2	70.00	0.84	Group5	9.64	0.41
Group2	93.95	0.90	Group5	10.00	0.45
Group2	106.02	0.95	Group5	10.26	0.48
Group3	0.00	0.06	Group5	12.85	0.52
Group3	4.36	0.12	Group5	14.99	0.59
Group3	6.84	0.18	Group5	16.06	0.62
Group3	7.50	0.24	Group5	17.36	0.66
Group3	10.26	0.29	Group5	19.14	0.63

Contc ..

Appendix 2 Contd..

Group	x	F(x)	Group	x	F(x)
Group5	19.69	0.72	Group7	8.55	0.11
Group5	22.49	0.76	Group7	12.93	0.15
Group5	23.49	0.83	Group7	20.51	0.22
Group5	25.70	0.86	Group7	22.97	0.26
Group5	25.97	0.90	Group7	23.93	0.33
Group5	37.58	0.93	Group7	24.99	0.37
Group5	68.92	0.97	Group7	25.70	0.44
Group6	0.00	0.05	Group7	25.97	0.48
Group6	2.30	0.08	Group7	28.18	0.52
Group6	5.17	0.11	Group7	29.54	0.56
Group6	6.84	0.13	Group7	30.63	0.63
Group6	7.50	0.16	Group7	34.98	0.67
Group6	10.26	0.18	Group7	38.28	0.70
Group6	10.41	0.21	Group7	42.28	0.74
Group6	12.85	0.26	Group7	42.41	0.78
Group6	15.32	0.32	Group7	46.97	0.81
Group6	15.38	0.34	Group7	49.23	0.85
Group6	19.69	0.37	Group7	60.60	0.89
Group6	25.70	0.40	Group7	64.25	0.96
Group6	28.18	0.42	Group8	8.68	0.11
Group6	30.63	0.45	Group8	12.15	0.22
Group6	31.71	0.47	Group8	28.18	0.33
Group6	34.40	0.50	Group8	30.00	0.44
Group6	37.58	0.53	Group8	50.00	0.56
Group6	42.28	0.55	Group8	51.28	0.67
Group6	43.29	0.61	Group8	60.00	0.78
Group6	44.31	0.63	Group8	63.07	0.89
Group6	53.60	0.68	Group9	3.21	0.08
Group6	65.76	0.71	Group9	6.75	0.17
Group6	68.93	0.74	Group9	7.07	0.25
Group6	69.73	0.76	Group9	8.55	0.33
Group6	76.58	0.79	Group9	9.00	0.42
Group6	89.25	0.82	Group9	19.83	0.50
Group6	93.95	0.84	Group9	20.00	0.67
Group6	98.47	0.87	Group9	24.24	0.75
Group6	99.61	0.90	Group9	24.97	0.83
Group6	112.74	0.92	Group9	24.99	0.92
Group6	115.65	0.95			
Group6	120.00	0.97			
Group7	0.00	0.04			

Appendix 3. Mean composition of the different vegetation clusters identified

		Composition (Per cent area under different vegetation types)									
Cluster	Predominant vegetation	No. of Blocks	Ever-green	Shola	Moist Deci-duous	Dry Deci-duous	Grass land	Vay-als	Scrub	Plant-ations	Others
1	Evergreen	201	77.0	2.4	6.4	1.4	6.0	1.0	1.0	2.6	2.2
2	Moist Deciduous	82	6.8	1.2	76.5	3.3	5.1	1.5	1.7	3.7	0.2
3	Dry Deciduous	100	8.4	2.6	5.5	63.5	2.8	1.5	4.3	4.4	7.0
4	Plantations	79	1.4	0.8	7.9	7.7	2.2	0.8	0.8	76.7	1.7
5	Shola and grassland	27	12.2	37.8	1.5	1.5	40.0	2.2	0.4	2.4	2.0