MEASUREMENT ERRORS
IN
HAWAIIAN FOREST BIRD SURVEYS
AND THEIR
EFFECT ON DENSITY ESTIMATION

Richard J. Camp

USGS Hawai'i Cooperative Studies Unit
Kīlauea Field Station, P.O. Box 44, Hawai'i National Park, Hawai'i

Technical Report HCSU-005

**MEASUREMENT ERRORS IN HAWAIIAN FOREST BIRD SURVEYS AND
THEIR EFFECT ON DENSITY ESTIMATION**

Richard J. Camp

USGS Hawai`i Cooperative Studies Unit
Kīlauea Field Station, P.O. Box 44, Hawai`i National Park, Hawai`i

Hawai`i Cooperative Studies Unit
University of Hawai`i at Hilo
Pacific Aquaculture and Coastal Resources Center (PACRC)
200 W. Kawili St.
Hilo, HI 96720
(808)933-0706

March 2007

Table of Contents

List of Figures

List of Tables

Abstract

Reliable count data are necessary for valid density estimation. Before each Hawaiian bird survey, observers go through a training and calibration exercise where they record measurements from a station center point to flagging placed about the station. The true distances are measured, and when an observer's measurements are within 10% of truth the observer is considered calibrated and ready for surveying. Observers tend to underestimate distances, especially for distant measures (e.g., true distance > 50 m). All proposed empirical distribution functions failed to adequately identify the function form of the calibration data. The effect of measurement errors were assessed with populations of known density in a simulation study. By simulations, using the true distances, the conventional estimator seems unbiased; however, in the presence of measurement errors the estimator is biased upward, resulting in overestimated population sizes. More emphasis should be made to minimize measurement errors. Observers' measurement errors should be small with deviances less than 10%, for example, and observers should recalibrate frequently during surveys. Truncation is not a surrogate for increased accuracy. When there are relatively large amounts of measurement error estimators to correct the errors should be developed and used. Measurement errors to birds heard but not seen needs to be calibrated, and adjustment parameters included in measurement error correction models.

Introduction

Point transect, or variable circular plot (VCP), sampling is one of the most widespread methodologies used for estimating densities of birds (Rosenstock et al. 2002, Thompson 2002), and this method provides the foundation for Hawaiian forest bird monitoring (Camp et al., In review). Counts of birds are adjusted for birds present but not detected. Unbiased abundance estimates, or density, may be calculated when factors affecting detection probabilities are accountable and key assumptions are supported. Parameters of a model assumed for the detection function are estimated from the data, leading to the following density estimator

$$\hat{D} = \frac{n\hat{h}(0)}{2\pi k}$$

where n is the sum of birds detected at k stations sampled, and $\hat{h}(0) = 2\pi/\hat{v}$ is the slope of the probability density function of detected distances ($\hat{v}$ is the effective area of detection). Buckland et al. (2001) describe the estimation procedures in detail.

Distance sampling relies on several design and model assumptions for reliable density estimation. Buckland et al. (2001:18) describe two design based assumptions: (1) birds are "spatially distributed in the area to be sampled according to some stochastic process;" and (2) "randomly placed points are surveyed and a sample of n objects is detected, measured and recorded." As long as the sampling points are randomly placed in the survey area the birds need not be randomly (i.e., Poisson) distributed. In addition, there are four model based assumptions (given in order from most to least critical): (1) species are correctly identified; (2) birds at the station center point are always detected with probability 1, or $g(0) \equiv 1$; (3) birds do not move in response to observers and are detected at their initial location; and (4) distances are measured accurately. Misidentification of species can be minimized through training and questionable detections should not be recorded. Alpizar-Jara and Pollock (1997) and Borchers et al. (1998) worked out methods that account for incomplete detections at $g(0)$, and Palka and Hammond (2001) accounted for responsive movements (see also Laake 1978). These methods were developed predominantly for marine and aerial surveys; nonetheless, these approaches

1

can be applied to forest bird sampling and analyses (Rosenstock et al. 2002, Buckland et al. 2004).

Error in distance measurements (i.e., imprecise object location) has been the subject of recent work in line transect sampling (e.g., Alpizar-Jara 1997, Hiby et al. 1989, Schweder 1996, 1997, Chen 1998, Chen and Cowling 2001, Marques 2004). It is common in bird studies to visually estimate distances (although this practice is not encouraged, instead distances should be measured; Buckland et al. 2001:265), and distances likely include stochastic measurement errors. Chen (1998) found that violations of this assumption result in underestimated densities (downward biased), and the downward bias occurs regardless of distance measurement bias and sample size (number of detections). Bias may differ in multiplicative measurement error models (Marques 2004, Borchers et al., In prep.).

The effect of measurement errors need to be either corrected or minimized in situations where discrepancies are large enough to induce bias. Borchers et al. (In prep.) developed an analytical approach that incorporates either a correction factor or maximum likelihood (ML) method to correct for the effect of measurement errors in line and point transect sampling. The correction factor method was easy to implement in program DISTANCE (a distance sampling software package; Thomas et al. 2006), as a multiplier in the multiple covariate distance sampling (MCDS) engine. The correction factor method did not perform as well as the ML method. The ML method produces approximately unbiased estimates (Borchers et al., In prep.), but it was more difficult to implement and has not been incorporated into DISTANCE. When measurement error was sufficiently small (CV < 10%) the abundance estimates were negligibly biased (Borchers et al., In prep.). Thus understanding the sources and minimizing measurement errors is the preferred approach to reduce bias (see Marques et al. 2006).

The objective for this study was to describe the effect of measurement errors in Hawaiian forest bird sampling. Simulations were used to illustrate bias in density estimation caused by stochastic measurement errors. Measurement errors can result when observers

round or heap measurements to convenient values (e.g., 5 and 10 m intervals). In this report I focused on measurement errors due to stochastic discrepancies, and did not directly address rounding and heaping effects. Estimator bias was discussed and recommendations were made to minimize measurement errors.

Calibration Trials

Observers participate in required training and calibration trials before conducting annual surveys for Hawaiian birds. This calibration data consists of eyeball estimated and true measurements between station center points and the horizontal distance to flagging. I used data from 5,284 survey calibration trials of paired measurements, estimated distance (y_{0k}) given the true distance (y_k), to directly specify the conditional distribution of the measurement errors. Assuming that distance estimation is unbiased, then $\pi(\epsilon|y_k)$ is based on the functional form for $\pi(y_{0k}|y_k)$ (Burnham et al. 2004:375). I make two additional assumptions: (1) calibration and survey data are comparable; and (2) heaping is an independent source of measurement error that has a negligible influence on the error distribution. (If heaping is present, Buckland et al. (2001:109) recommend analysis using grouped data.)

Discrepancies between actual and estimated measurements occurred at distances as close as 3 m from the station center point, and became more pronounced with increasing distance. The distribution of measurement errors was asymmetric and leptokurtic with mean -1.25 (5.850 SD; skewness = -0.581; kurtosis = 5.96; range -43, 39; Figure 1). No difference was observed between the slopes of the full calibration data set or truncation at 80 m, nor slopes between 50 or 30 m truncation ($P = 0.35$ and 0.88, respectively; Table 1). This implies that truncation will not eliminate the need to improve measurement accuracy, especially for distances within 30 to 50 m. This is the region in which most detections of Hawaiian birds occur (Camp et al., In prep.), and can strongly influence model selection and fit (Buckland et al. 2001).
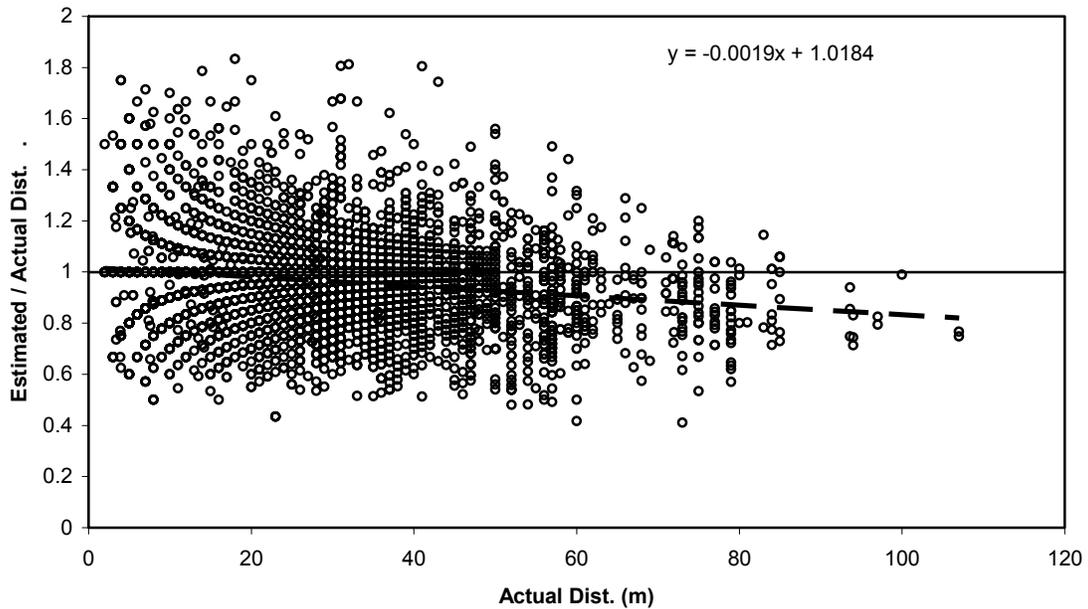
Figure 1. Relationship of estimated/actual distance measurements (open circles) plotted against actual distances. Solid line represents the 1:1 relationship (no measurement error). Dashed line represents the trend given by the regression model.

Table 1. Slope of actual versus estimated distances for various amounts of truncation. Slopes between no truncation and 80 m, and between 50 and 30 m were not different.

Truncation	Slope	SE	95% CI L	95% CI U
No truncation	0.889	0.0047	0.879	0.898
80 m	0.895	0.0049	0.885	0.904
50 m	0.937	0.0057	0.926	0.948
30 m	0.939	0.0074	0.924	0.953

The empirical distribution function filed to fit the cumulative distribution functions beta, gamma, Gaussian, lognormal and Weibull (goodness-of-fit tests; $P < 0.01$, respectively). Therefore, a Box-Cox transformation test was conducted to determine the power transformation value to correct for nonnormality. A power transformation value of 0.6 is sufficiently close to 0.5 to justify use of the square root transformation to normalize the data. Transforming the data, however, did not improve the fit of the proposed cumulative distribution functions (Table 2; Figure 2). Because the proposed distributional models did not support the empirical distribution function, measurement errors were resampled to contaminate simulated data sets.

Table 2. Fit of the empirical distribution (square root transformed estimated distances divided by actual distances) and proposed parametric cumulative functions. All proposed functions failed to adequately identify the distributional model.

| Test | Goodness-of-Fit Test for Distribution | | | | |
	Normal	Lognormal	Weibull	Gamma	Beta
Kolmogorov-Smirnov	0.098 <0.010	0.031 <0.010	—	0.048 <0.001	—
Cramer-von Mises	19.55 <0.005	1.06 <0.005	15.70 <0.010	4.50 <0.001	—
Anderson-Darling	121.5 <0.005	7.23 <0.005	107.8 <0.010	29.20 <0.001	—
Chi-square	105965 <0.001	278.5 <0.001	9179 <0.001	1267 <0.001	12043 <0.001

Curves: Normal(Mu=0.2289 Sigma=0.0923)   Lognormal(Theta=0 Shape=.37 Scale=-2)
Weibull(Theta=0 Shape=2.6 Scale=.26)   Gamma(Theta=0 Shape=7.28 Scale=0.03)
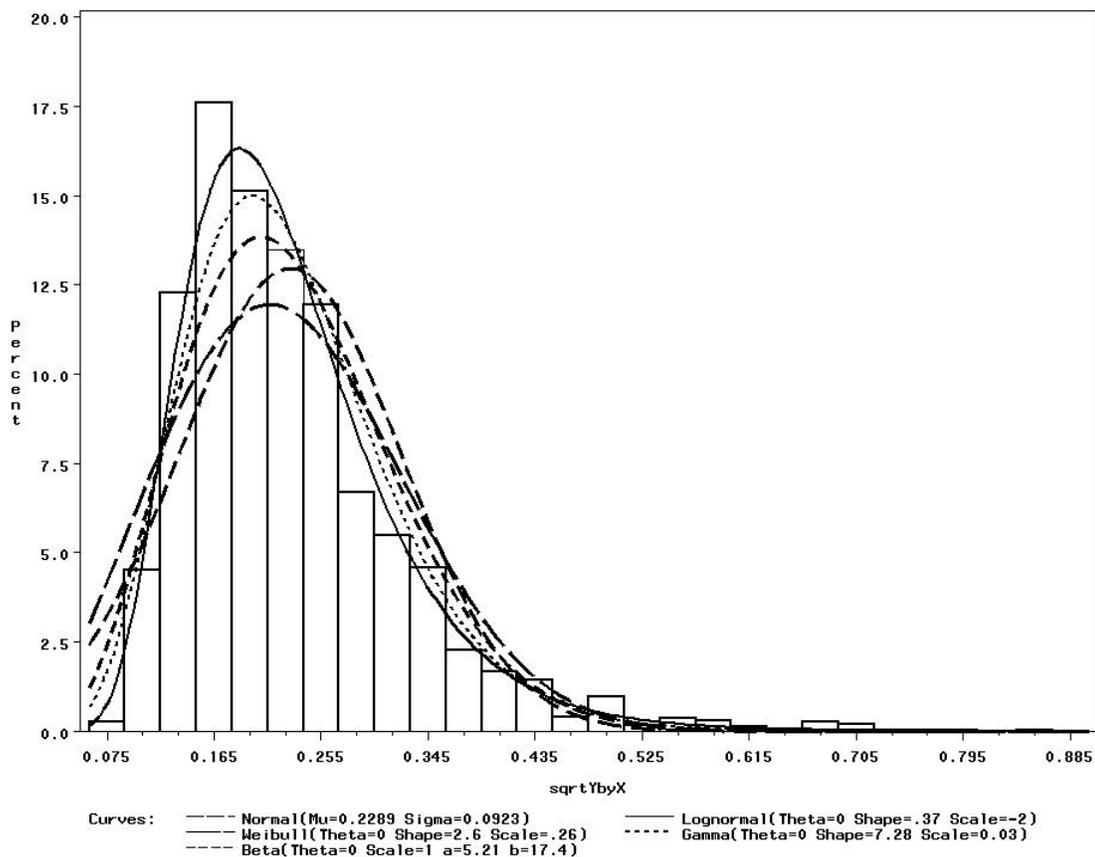Beta(Theta=0 Scale=1 a=5.21 b=17.4)

Figure 2. Histogram of the empirical distribution function (square root transformed estimated distances divided by actual distances) and the proposed parametric cumulative distribution functions. All proposed functions failed to adequately identify the distributional model (see Table 2).

Simulation Study

I evaluated the influence of measurement error on density estimation in point transect sampling with a simulation study. A square study space with sides of 200 m was delineated and populated with 1,000 "birds" following a Poisson random distribution (D = 250 birds/ha). Point transect sampling was conducted from a single station centered in the study space. The probability of detecting a bird was considered to be half-normal, with scale parameter 25. This scaling parameter value typically produced 80 to 120 detections, thus meeting the recommended sample size needed for reliable distance modeling (Buckland et al. 2001:241). Each bird was assigned a capture probability (p_i, based on the detection parameters) and a uniform (0,1) random number, u_i. A bird was considered detected when $p_i > u_i$. Radial distances were measured with an accuracy of 0.001, but were rounded to the nearest meter to reflect field sampling protocol. This procedure had nominal effects on density estimation. All design and model assumptions were realized.

Each simulated distance measurement y_i was contaminated with the additive model $Y_i' = y_i + d$. The y_i and associated calibration y_k were matched and a discrepancy value d was randomly selected from the observed calibration measurement errors associated with that y_i. Distances between 0 and 2 m were not contaminated. Distances between 2 and 50 m were contaminated with the associated calibration measurement errors. Measurement errors for distances > 51 m were pooled into 3 categories: x_i distances 51-60, 61-75, and > 75 m; and y_i were contaminated with the associated pooled calibration measurement errors. The advantages of this process were to contaminate Y_i with real measurement errors, and the pooling at distant measures avoided gaps in calibration measurements. In addition, the resulting measurements were dependent on the simulated distance and the functional distribution of the measurement errors, and Y' and d are independent.

Density was estimated with conventional distance sampling methods (Distance 5.0; Thomas et al. 2006), using (1) unaltered simulated distances, and (2) the distances contaminated with an error using the measurement error model (above). The detection function models half-normal and hazard-rate were used for density estimation, with cosine, Hermite polynomial and simple polynomial expansions series of order one, following Buckland et al. (2001:156) suggested model combinations. The detection function model that best-fit the data was chosen by the minimum Akaike's Information Criterion (AIC; Akaike 1973). Due to model convergence failures, 10% of the largest observed distance measurements were truncated (right-tailed truncation) following standard distance sampling methods (Buckland et al. 2001:151). The mean density and associated measures of precision were obtained from 1,000 replicates. Estimator reliability was assessed by the bias and coverage of the of density estimates compared to truth ($D = 250$). The average value of $\hat{D}$ was calculated over all replicates as

$Avg(\hat{D}) = \dfrac{1}{R}\displaystyle\sum_{k=1}^{R}\hat{D}_k$, where R is the number replicates. The sampling standard error of $\hat{D}$

was calculated over all replicates as $\sigma(\hat{D}) = \sqrt{\displaystyle\sum_{k=1}^{R}\left(\hat{D}_k - Avg(\hat{D})\right)^2 \Big/ (R-1)}$. The average

value of $\sqrt{\hat{\mathrm{var}}(\hat{D})}$ was calculated over all replicates as $Avg(\sqrt{\hat{\mathrm{var}}(\hat{D})}) = \dfrac{1}{R}\displaystyle\sum_{k=1}^{R}\sqrt{\hat{\mathrm{var}}(\hat{D})}$,

and the percent relative bias was calculated as $\%RB = 100[(Avg(\hat{D}) - D)/D]$. Coverage was the proportion of replicates in which the constructed confidence intervals covered the true value of D.

The simulated distance measures emulated the theoretical distribution expected by a half-normal detection function (panel a in Figure 3). Contaminating the distances with measurement errors resulted in a less smooth, or more "toothed" distribution, which was not as well fit by the true detection function (panel b in Figure 3). In addition, the maximum distance and the weight of the tail were reduced in the contaminated data set.

The average density estimator reliably estimated D in the absence of measurement errors (Table 3; see Buckland et al. 2001, 2004 for additional estimator performance reliability

and citations therein). For the 1,000 replicates, $\hat{D}$ was unbiased (Avg($\hat{D}$) – D = -0.49; %RB = -0.20). Confidence interval coverage was less than the constructed 95% CI level; only 88% of the estimator confidence intervals covered the true value of D.

In the presence of measurement errors the average density estimator was biased, overestimating the population size (%RB = 13.31; Table 3). The variability ($\sigma(\hat{D})$) in the density estimates were comparable between the uncontaminated and contaminated data sets ($F_{999,999}$ = 1.052, p = 0.210), and the coverage of the 95% confidence intervals were also comparable (88% coverage). Although the $Avg(\sqrt{Var(\hat{D})})$ was larger in the contaminated data set, the proportion relative to the $Avg(\hat{D})$ was comparable between the data sets (0.199 and 0.194, respectively).

Table 3. Summary of density estimates and estimator performance. See text for definitions and equations of column headings.

Data Set	D	$Avg(\hat{D})$	$\sigma(\hat{D})$	$Avg(\sqrt{Var(\hat{D})})$	Coverage	%RB
Uncontaminated	250	249.51	58.42	49.72	88.4	-0.20
Contaminated	250	283.27	56.98	54.89	88.8	13.31

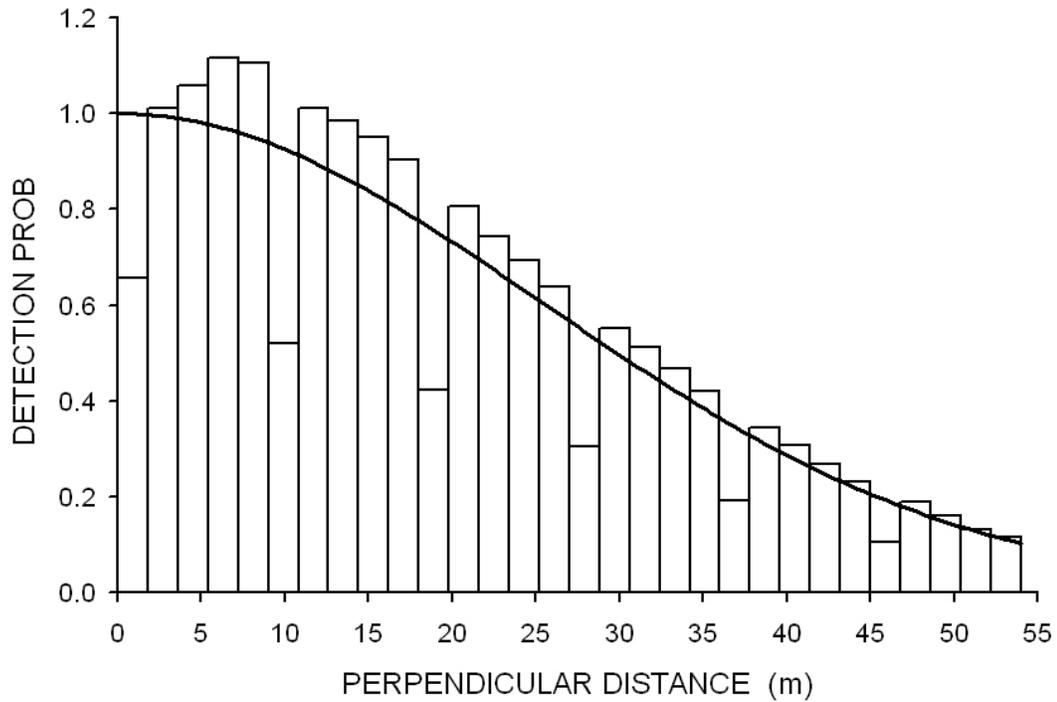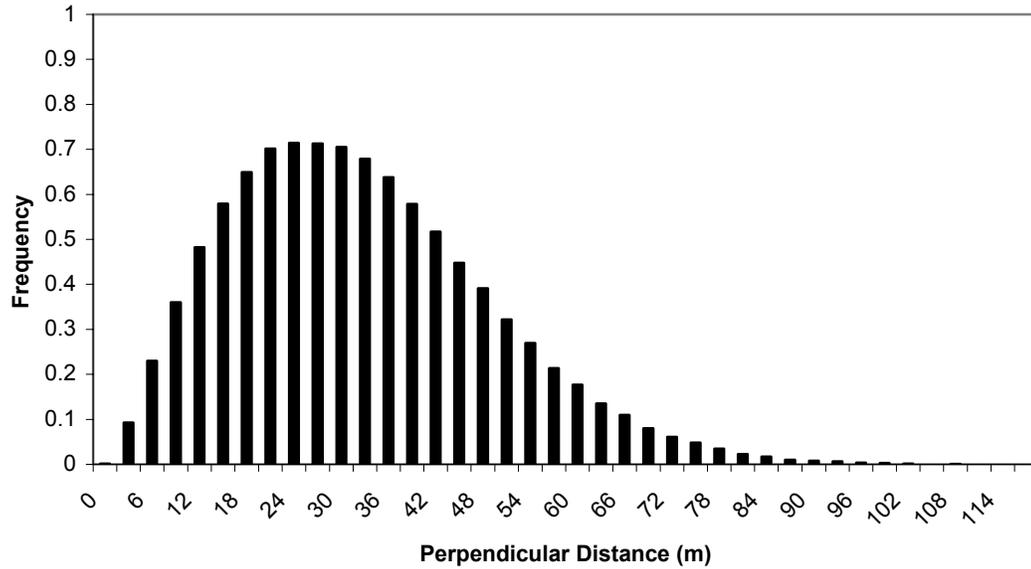**(a) Histogram of detections by distance for uncontaminated datasets**

Figure 3. Histogram of detections by distance for (a) uncontaminated and (b) contaminated data sets. A best fit half-normal detection function was modeled to each data set (solid line). Graphical representation of perpendicular distances resulted in the apparent small bars seen in the bottom panels.ls.

(b) Histogram of detections by distance for contaminated datasets
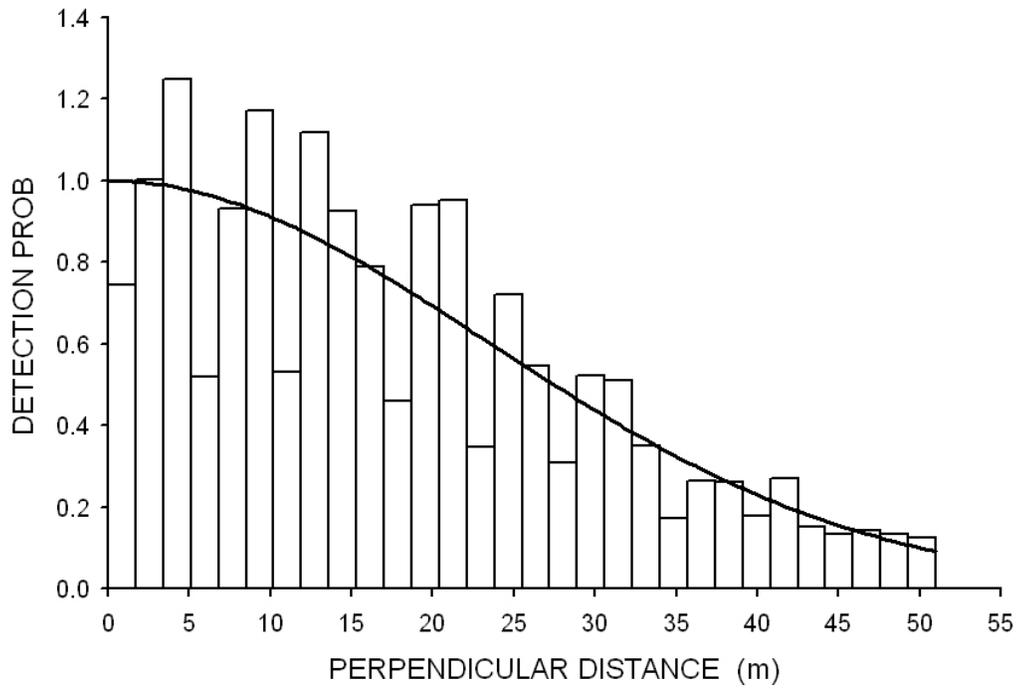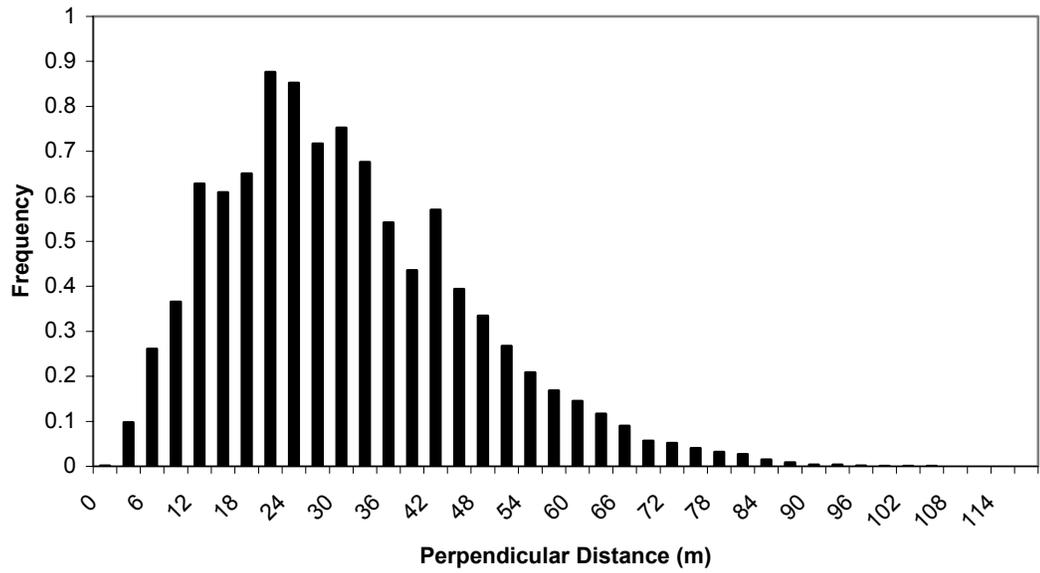
Discussion

Estimating distances to flagging is undoubtedly easier than distances to birds, which are often cryptic or detected aurally. For this reason it was inappropriate to develop a correction model based on calibration trial data. Furthermore, none of the proposed conditional functions was well adjusted to the functional distribution of the multiplicative model. Thus, a resampling procedure was used as the strength of the calibration trials lies in the ability to identify bias in distance estimates. Once identified, a correction model can be implemented to remove bias. Correction models require both parameter identification and model selection procedures. However, the poor fit of the proposed conditional functions hinders parameter identification.

If discrepancies in measurements can be reduced with calibration trials and experience, then density estimates will be more accurate and possibly more precise. Increased accuracy would result in more reliable (low levels of bias) abundance estimates. In addition, increased accuracy could eliminate the most errant measurement errors, which could result in increased estimator precision and increased power to detect population trends. Thus, discrepancies between actual and estimated measurements should be small (see Borchers et al., In prep.), and calibration trials should continue until distances are unbiased (<10%).

Auditory distance estimation is not currently practiced during the calibration trials, only visual measures are practiced. However, more than 80% of Hawaiian bird detections distance measures were recorded to birds that were heard but not seen during the counts (USGS – Pacific Island Ecosystems Research Center – Hawai`i Forest Bird Interagency Database Project, unpublished data). It is commonly assumed that measurements to visually located birds are more accurate than auditory based detections. Until auditory distance estimation is implemented in calibration studies, a necessary procedure to develop correction models, aurally detected birds should be placed in the environment and the horizontal distance estimated to the reference point recorded.

11

Further gains in accuracy and precision can be made by incorporating additional measurement aids (e.g., flagging and range finders) while sampling. Flagging is placed 10 m before and after each station along Hawaiian forest bird transects. In many situations this flagging is not visible, and does not help identify distances perpendicular to the transects. Additional flagging could be placed at the 5 m distance interval along the transects, and flagging at 5 and 10 m perpendicular to the transects. This procedure should help minimize errors. Range finders have been incorporated into many survey programs and are effective at reducing measurement errors. The use of range finders in dense vegetation is problematic; however, using them can help to ensure that model assumptions are met by measuring the horizontal distance to nearby trees and adjusting the location of detected birds appropriately.

Acknowledgements

References

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 267-281 *in* B.N. Petrov, and F. Csaki (editors). Second International Symposium on Information Theory. Akademiai Kiado, Budapest.

Alpizar-Jara, R. 1997. Assessing assumption violations in line transect sampling. Doctor of Philosophy Dissertation. North Carolina State University, Raleigh.

Alpizar-Jara, R., and K.H. Pollock. 1997. A combination line transect and capture-recapture sampling model for multiple observers in aerial surveys. Environmental and Ecological Statistics 3:311-327.

Borchers, D.L., T.A. Marques, and Th. Gunnlaugsson. In preparation. Distance sampling with measurement errors.

Borchers, D.L., W. Zucchini, and R.M. Fewster. 1998. Mark-recapture models for line-transect surveys. Biometrics 54:1207-1220.

Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas. 2001. Introduction to distance sampling: Estimating abundance of biological populations. Oxford University Press, Oxford, U.K.

Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas (editors). 2004. Advanced distance sampling. Oxford University Press, Oxford, U.K.

Burnham, K.P., S.T. Buckland, J.L. Laake, D.L. Borchers, T.A. Marques, J.R.B. Bishop, and L. Thomas. 2004. Further topics in distance sampling. Pages 307-392 *in* S.T. Buckland, D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas (editors). Advanced distance sampling. Oxford University Press, Oxford, U.K.

Camp, R.J., P.M. Gorresen, T.K. Pratt, and B.L. Woodworth. In preparation. Population trends of native Hawaiian forest birds: 1976-2005.

Camp, R.J., M.H. Reynolds, P.M. Gorresen, T.K. Pratt, and B.L. Woodworth. In review. Monitoring Hawaiian forest birds. Chapter 4 in T. K. Pratt, B. L. Woodworth, C. T. Atkinson, J. Jacobi, and P. Banko (editors). Conservation Biology of Hawaiian Forest Birds: Implications for insular avifauna.

Chen, S.X. 1998. Measurement error in line transect surveys. Biometrics 54:899-908.

Chen, S.X., and A. Cowling. 2001. Measurement errors in line transect surveys where detectability varies with distance and size. Biometrics 57:732-742.

Hiby, L., A. Ward, and P. Lovell. 1989. Analysis of the North Atlantic Sightings Survey 1987: aerial survey results. Report of the International Whaling Commission 39:447-455.

Laake, J.L. 1978. Line transect sampling estimators robust to animal movement. Master's Thesis, Utah State University, Logan.

Marques, T.A. 2004. Predicting and correcting bias caused by measurement error in line transect sampling using multiplicative error models. Biometrics 60:757-763.

Marques, T.A., M. Andersen, S. Christensen-Dalsgaard, S. Belikov, A. Boltunov, O. Wiig, S.T. Buckland, and J. Aars. 2006. The use of GPS to record distances in a helicopter line-transect survey. Wildlife Society Bulletin 34:759-763.

Palka, D.L., and P.S. Hammond. 2001. Accounting for responsive movement in line transect estimates of abundance. Canadian Journal of Fisheries and Aquatic Sciences 58:777-787.

Rosenstock, S.S., D.R. Anderson, K.M. Giesen, T. Leukering, and M.F. Carter. 2002. Landbird counting techniques: current practices and an alternative. Auk 119:46-53.

Schweder, T. 1996. A note on a buoy-sighting experiment in the North Sea in 1990. Report of the International Whaling Commission 46:383-385.

Schweder, T. 1997. Measurement error models for the Norwegian minke whale survey in 1995. Report of the International Whaling Commission 47:485-488.

Thomas, L., J.L. Laake, S. Strindberg, F.F.C. Marques, S.T. Buckland, D.L. Borchers, D.R. Anderson, K.P. Burnham, S.L. Hedley, J.H. Pollard, J.R.B. Bishop, and T.A. Marques. 2006. Distance 5.0. Release 5. Research Unit for Wildlife Population Assessment, University of St. Andrews, U.K. <http://www.ruwpa.st-and.ac.uk/distance/>.

Thompson,W.L. 2002. Towards reliable bird surveys: accounting for individuals present but not detected. Auk 119:18-25.