Technical Report HCSU-049

MONITORING HAWAIIAN WATERBIRDS: EVALUATION OF SAMPLING METHODS TO PRODUCE RELIABLE ESTIMATES

Richard J. Camp[1], Kevin W. Brinck[1], Eben H. Paxton[2], and Christina Leopold[1]

[1]Hawai'i Cooperative Studies Unit, University of Hawai'i at Hilo,
P.O. Box 44, Hawai'i National Park, HI 96718
[2]U.S. Geological Survey, Pacific Island Ecosystems Research Center, Kīlauea Field Station,
P.O. Box 44, Hawai'i National Park, HI 96718 USA

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

We conducted field trials to assess several different methods of estimating the abundance of four endangered Hawaiian waterbirds: the Hawaiian duck (*Anas wyvilliana*), Hawaiian coot (*Fulica alai*), Hawaiian common moorhen (*Gallinula chloropus sandvicensis*) and Hawaiian stilt (*Himantopus mexicanus knudseni*). At two sites on Oʻahu, James Campbell National Wildlife Refuge and Hamakua Marsh, we conducted field trials where both solitary and paired observers counted birds and recorded the distance to observed birds. We then compared the results of estimates using the existing simple count, distance estimates from both point- and line-transect surveys, paired observer count estimates, bounded count, and Overton estimators. Comparing covariate recorded values among simultaneous observations revealed inconsistency between observers. We showed that the variation among simple counts means the current direct count survey, even if interpreted as a proportional index of abundance, incorporates many sources of uncertainty that are not taken into account. Analysis revealed violation of model assumptions that allowed us to discount distance-based estimates as a viable estimation technique. Among the remaining methods, point counts by paired observers produced the most precise estimates while meeting model assumptions. We present an example sampling protocol using paired observer counts. Finally, we suggest further research that will improve abundance estimates of Hawaiian waterbirds.

INTRODUCTION

Current procedures to monitor endangered Hawaiian waterbirds—Hawaiian duck (*Anas wyvilliana*), Hawaiian coot (*Fulica alai*), Hawaiian common moorhen (*Gallinula chloropus sandvicensis*), and Hawaiian stilt (*Himantopus mexicanus knudseni*)—use area search and count techniques (herein we use the Recovery Plan bird names; U.S. Fish and Wildlife Service [USFWS] 2011). This sampling approach provides only indices of species' presence and relative abundance and is not robust compared to other census techniques widely used. Specifically, population counts from area searches and direct counts do not account for imperfect detection of presence, with simple counts widely known to underestimate relative abundances by some unknown amount (Anderson 2001, USFWS 2011). Importantly, while changes in the occurrence and numbers of birds counted across years are assumed to be a close surrogate for true abundance, detection probabilities can change between years and sites making such trend analysis questionable (Anderson 2001; see also DesRochers *et al.* 2008). Thus, the census techniques currently being employed for Hawaiian waterbirds are not the best approaches for managers who need to consider minimum viable population size, likelihood of extinction, and trends, all of which are necessary for downlisting or delisting and recovery actions (USFWS 2011). Methods do exist to calculate detection probabilities and convert indices to direct population measurements of occupancy and abundance (Thompson *et al.* 1998, Anderson 2001), but the information needed to do so currently is not collected during waterbird surveys in Hawaiʻi.

Population monitoring for Hawaiian waterbirds is conducted biannually by the State of Hawaiʻi Division of Forestry and Wildlife (DOFAW) and consists of visits to wetlands on a single day each winter and summer—the third Wednesday of January and August, respectively. The number of detected individuals of all waterbird species at each wetland is recorded. In addition, counters collect ancillary information on water level, vegetation cover, weather conditions, and disturbances. Again, while it is believed that these counts are "probably fairly accurate

1

population estimates for Hawaiian coot and Hawaiian stilt in most years" (USFWS 2011:107), it is known to be negatively biased by some unknown amount because some individuals will not be detected, and detection probability will vary by species and location. This bias is likely to be greatest for the common moorhen, as they are secretive and use densely vegetated areas, and for the Hawaiian duck because they use montane stream habitats that are not surveyed (Robinson *et al.* 1999, Bannor and Kiviat 2002, Engilis *et al.* 2002, Pratt and Brisbin 2002, USFWS 2011).

Estimating the numbers of individuals—abundance—is based upon detection probabilities that involve "individual"-level detection or encounter histories (individuals may be any distinctly recognizable unit, such as specific animals, breeding pairs, or territories). Quantitative sampling techniques to estimate abundance include double sampling (Cochran 1977, Bart and Earnst 2002), double-observer sampling (Nichols *et al.* 2000), mark-recapture/resight methods (Borchers *et al.* 2002, McClintock *et al.* 2009), distance sampling (Buckland *et al.* 2001), and partitioned point-count sampling (Farnsworth *et al.* 2002).

In this report we review and evaluate the existing waterbird survey methods and describe potential sampling protocols to provide a statistically robust estimate of the true number of birds. During a pilot study in March 2013, we field-tested a variety of sampling methods to produce measures of waterbird species' detectability and density. These sampling methods involved collecting auxiliary information needed to estimate detection probabilities and the effective area surveyed, which were used to calculate densities and abundances at our survey stations. We were unable to observe any pure Hawaiian ducks, and encountered too few ducks to analyze properly, but we did analyze surveys for the non-endangered, native, black-crowned night-heron (*Nycticorax nycticorax hoactli*). Testing for constant detection probability across years requires more than one year of study, which was beyond the scope of this study; however, in this report we tested that assumption for other factors such as observer and time of day which was a first step in evaluating the importance of considering detection probability.

Bird Species
Historically Hawai'i supported a large and diverse waterbird fauna that has been reduced to only six endemic species, all of which are listed as endangered—Hawaiian duck, Laysan duck (*Anas laysanensis*), Hawaiian coot, Hawaiian common moorhen, Hawaiian stilt, and Hawaiian goose (*Branta sandvicensis*). In this report, we address sampling only for the Hawaiian coot, Hawaiian common moorhen, Hawaiian stilt, and Hawaiian duck. Historically, all four species were found on all of the main Hawaiian Islands except Kaho'olawe and Lana'i.

The koloa maoli (*Anas wyvilliana*), or just koloa, Hawaiian duck, is closely related to the mallard (*A. platyrhynchos*) and possesses many similar characteristics (Engilis *et al.* 2002). In the course of our trial surveys on O'ahu we did not detect any pure Hawaiian duck. The Hawaiian duck is secretive, spends part of its time in areas not surveyed (mountain streams), and we were not able to adequately assess a best method for census; therefore, we do not consider the duck in our analysis.

The 'alae ke'oke'o (*Fulica alai*), or Hawaiian coot, is closely related to the American coot (*F. americana*). The Hawaiian coot is a large, fully aquatic rail with dark slate-gray plumage and a white bill and bulbous frontal shield (Pratt and Brisbin 2002). The following description is a summary of information in the Hawaiian coot Birds of North America account (Pratt and Brisbin 2002). Coot use a variety of wetland habitats, including natural, modified, and man-made marshes, ponds, estuaries, reservoirs, agricultural fields, irrigation ditches, and sewage

treatment ponds. Coot forage mainly in shallow water; in deeper water they forage along emergent vegetation. They rest in open water and along emergent vegetation and on banks, dikes, berms, and levees. Where persecuted or threatened they are skittish and move to cover quickly when disturbed. They can occur in pairs or singly and form large flocks of many tens of individuals. Coot are usually silent, but may utter loud *keck-keck* and *keek* calls when alarmed.

The 'alae 'ula (*Gallinula chloropus sandvicensis*), or Hawaiian common moorhen, is an endemic subspecies of the common gallinule or common moorhen (*G. galeata*) of North America (Bannor and Kiviat 2002). The following description is a summary of information in the Hawaiian common moorhen Birds of North America account (Bannor and Kiviat 2002). The moorhen is an aquatic rail the size of a smallish duck with blackish, dark slate-gray to olive-brown plumage, posterior underparts, and tail-covert feathers tipped with white, and a bright-red-and-yellow bill. Moorhen use a variety of wetland habitats, including natural, modified, and man-made marshes, ponds, estuaries, reservoirs, agricultural fields, irrigation ditches, and sewage treatment ponds. Moorhen forage in deep to shallow water, along emergent vegetation, and on land adjoining water. They rest in open water, along emergent vegetation, and on banks, dikes, berms, and levees. Individuals in human-dominated areas are habituated and are often seen begging for handouts. When startled they take cover in emergent vegetation and cover along water banks. They usually occur in pairs or singly and are often secretive. Moorhen make a wide variety of vocal sounds with individuals frequently responding to one another. Their calls are loud, harsh, and easily heard over vehicle traffic.

The ae'o (*Himantopus mexicanus knudseni*), or Hawaiian stilt, is a Hawaiian endemic subspecies of the black-necked stilt (*H. mexicanus*). The stilt is a large and conspicuous waterbird with a long black bill, marked black above and white below the body, and stand on very long reddish legs (Robinson *et al.* 1999). The following description is a summary of information in the Hawaiian stilt Birds of North America account (Robinson *et al.* 1999). Stilt use a variety of shallow wetland habitats with emergent vegetation including natural, modified, and man-made marshes, ponds, estuaries, reservoirs, agricultural fields, irrigation ditches, and sewage treatment ponds. They wade in shallow water while foraging. They rest in shallow open water, along emergent vegetation, and on land adjacent to water. Stilt are conspicuous and become aggressive and vocal when approached by other stilts or predators or when disturbed. They may be semicolonial, occur in pairs or singly, and form small groups when foraging and resting. Stilt make a variety of calls, but their alarm calls are loud and piercing and easily heard over surf and vehicle traffic. Individuals make frequent, quieter calls to one another. Stilt are unmistakable with other birds in Hawai'i.

As we were unable to observe the Hawaiian duck, we instead examined the 'auku'u (*Nycticorax nycticorax*), or black-crowned night-heron. Night-heron are found world-wide, with a resident, native population in Hawai'i. They are conspicuous and relatively large compared to other waterbirds in Hawai'i, with black on the top of their head with a gray back and wings and are white to pale gray underneath. They have a stocky build with a long, black bill, and relatively short yellow-green legs. According to their account in the Birds of North America (Hothem *et al.* 2010), they are opportunistic foragers; eating fish, amphibians, large invertebrates, or small mammals. Night-heron prefer shallow, weedy pond margins, creeks, and marshes. They capture prey by grasping with their bills from a standing position, but are also known to dive and swim. Their common call is a guttural squawk, and sometimes they rattle their bills as part of social interactions.

Survey Locations

Waterbird surveys have been conducted biannually on the island of Oʻahu since the mid-1950s. A review of the existing waterbird data and consultation with USFWS indicated that James Campbell National Wildlife Refuge (NWR; managed by USFWS; Universal Transverse Mercator zone 4 Q 608200 E 2398500 N) and Hamakua Marsh (managed by DOFAW; 4 Q 630400 E 2365900 N) were appropriate study sites to compare sampling methods. Both sites are easily accessed and actively managed. Water levels are controlled at James Campbell NWR, while water levels at Hamakua Marsh fluctuate according to local rainfall with occasional managed tidal inundations. These two sites provide a range of geophysical, hydrological, and management attributes that are generally representative of wetlands on Oʻahu and in Hawaiʻi. Furthermore, both sites have large populations of coot, moorhen, and stilt. The duck, or hybrid form, occurred only at James Campbell NWR. In our surveys we only had sufficient numbers of three endangered species (coot, moorhen, and stilt) to allow comparison among sampling methods.

Within James Campbell NWR, counts were conducted in the Kiʻi Unit, a wetland of ponds of varying intensities of water level management (Figure 1). Portions of the wetlands in the Kiʻi Unit are blocked from sight by emergent vegetation; however, most of the wetlands are visible from different vantage points. Stilt, coot, and moorhen have become habituated to humans at Hamakua Marsh (Figure 2). Most of Hamakua Marsh is visible from all vantage points, although duck, coot, and moorhen hide in the pickleweed (*Batis maritime*) mats, and trees and shrubs along the eastside can obscure some portions of the wetland.

Figure 1. Waterbird survey stations at Kiʻi Unit, James Campbell National Wildlife Refuge, island of Oʻahu, Hawaiʻi.

Figure 2. Waterbird survey stations at Hamakua Marsh, island of Oʻahu, Hawaiʻi.

SAMPLING METHODS

Between 24 and 29 March, 2013, we conducted field-tests to compare a variety of sampling methods to produce measures of waterbird species′ density. We sampled waterbirds using the standard DOFAW count method, multiple-observer and repeat visit methods, point- and line-transect distance methods using both single and double observers, and mapping the locations of detected birds. Surveys were conducted three times a day: early morning (08:00–10:00), late morning (10:00–12:00), and early afternoon (12:00–14:00). Each team of two counters had 20 minutes to conduct trial surveys at each station, 30 minutes for line-transect surveys and 30 minutes of transit and set-up time between stations; thus allowing each team to survey three stations and the line-transect in each time block. Survey teams rotated through each block (Table 1).

We conducted the standard DOFAW waterbird count where each counter scanned the entire area visible from the survey station for six minutes and tallied the numbers of individuals of each species observed. Following the direct count, counters created a reference map showing the survey site, identifying prominent features, areas of open water, emergent vegetation, and areas of the wetland that were blocked by vegetation. Paired counters compared their maps to ensure consistency. During a six-minute point-transect count, each counter measured the distance to birds detected using a laser rangefinder, recorded the species, number of individuals, distance and detection type (1 – heard first; 2 – seen first; or 4 – heard first and subsequently seen), and plotted the location on their own map. After completing the DOFAW

5

Table 1. Daily schedule of surveys conducted. Each pair of observers was assigned to three (consecutive) survey stations for each day. On the final survey day counts were halted at 12:00.

Time period	Station	Surveys conducted
Early morning (08:00–10:00)	First station	Direct count (record covariates)
		Station map
		Double-observer distance count
	Second station	Direct count (record covariates)
		Station map
		Double-observer distance count
	Third station	Direct count (record covariates)
		Station map
		Double-observer distance count
	Line-transect distance survey of wetland	
Late morning (10:00–12:00)	First station	Direct count (record covariates)
		Station map
		Double-observer distance count
	Second station	Direct count (record covariates)
		Station map
		Double-observer distance count
	Third station	Direct count (record covariates)
		Station map
		Double-observer distance count
	Line-transect distance survey of wetland	
Afternoon (12:00–14:00)	First station	Direct count (record covariates)
		Station map
		Double-observer distance count
	Second station	Direct count (record covariates)
		Station map
		Double-observer distance count
	Third station	Direct count (record covariates)
		Station map
		Double-observer distance count
	Line-transect distance survey of wetland	

and point-transect counts at three stations, counters traversed one edge of the wetland perimeter recording data for a one-sided line-transect survey. The data collected were the same as that recorded for the point-transect counts, including mapping the location of detected birds.

Observers also recorded environmental covariates that might affect bird behavior: cloud cover, vegetation cover, rainfall, wind and gust speed, water level, and the degree of human influence. Cloud cover was estimated as a continuous percentage between 0 and 100 by tens. Vegetation cover was ranked in discrete categories from 0 to 3: 0 = open water, 1 = 26–50% cover, 2 =

51–75% cover, and 3 = ≥75% cover. Rainfall was recorded in discrete categories of 0 = no rain, 1 = mist or fog, 2 = drizzle, and 3 = light rain. Wind and gust speed were recorded as Beaufort categories: 0 = no wind, 1 = smoke drifts (4–7 mph), 2 = wind felt on face, and 3 = leaves, small twigs in constant motion (8–12 mph). Water level was recorded as a discrete category ranging from 0 to 3, where 0 = dry, 1 = lower than normal, 2 = normal, and 3 = higher than normal. Human impact was recorded as ranging from 0 to 2: 0 = indirect, 1 = moderate, and 2 = heavy.

ANALYSIS

We calculated six different estimates of abundance from the four different data collection methods detailed in Table 1, as well as evaluating the inter-observer consistency of covariates collected during the direct count phase. Direct counts of bird abundance were compared, and variation was calculated between pairs of simultaneous counters and among all counts at a station during a two-hour survey block. Double-observer sampling estimates were produced from the station maps and the double-observer distance counts. Repeated survey sampling estimates using bounded count and Overton estimators were calculated from the direct counts. Finally two distance-based estimates were calculated: point-transect estimates from the double-observer distance count and line-transect estimates from the line-transect survey of the wetland.

Covariate Consistency
Many of the estimators we examined used covariates (e.g., vegetation cover, observer) to adjust detection probability as a function of environmental conditions, bird behavior, or observer differences. Doing so can improve the estimation of true abundance by taking into account factors that increase or decrease detection probability. In the models below, we consider the strength of multiple analysis methods incorporating covariates to demonstrate the strength and weakness of different approaches. However, it is important that the covariates are recorded in the same way by all observers, and we compared the covariates recorded by pairs of observers as a measure of observer consistency in recording these covariates.

The effect of behavioral or survey condition changes throughout the day were assessed between survey teams. Time-of-day effects on surveys were also assessed by comparing among teams that re-surveyed the same set of stations in each time block (early morning, late morning, and early afternoon).

Direct Counts
Based on the simple area count comparable to the current survey methodology, this index was calculated following standard DOFAW procedures by tallying the numbers of detected individuals of all waterbird species during each count at each sampling site. As all of the counts were conducted over a short period, we did not adjust the relative abundances based on survey or wetland conditions (i.e., time of day, water level, etc.). Counts were averaged to produce a mean relative abundance per period, and variance in counts was calculated for each period among counters. The assumptions of the standard counts are:
1. Probability of detection is 1; i.e., all individuals that are present are counted;
2. Individuals are not counted more than once during the count;
3. The sampling scheme provides complete coverage of the survey region;

4. The population is closed during the course of the count; i.e., individuals are not flying in and out of the area during the count.

Because we included replicate counts, we made two additional assumptions:
5. Replicate counts are independent;
6. The detection probability is constant across all replicate counts.

Double-observer Sampling

Double-observer sampling requires that two or more observers map the locations of detected birds at each sampling site. For our trial surveys we increased sample size by treating each survey station as a separate count; the number of birds available represents the average number of individuals at a survey station for that wetland and day. Using the maps recorded during the point- and line-transect counts, we were able to construct independent paired observations with individuals detected by the first observer, the second observer, or both. The sampling method is logistically more demanding to collect and process but produces more information. The data were analyzed in a maximum likelihood framework to estimate detection probabilities for each observer as well as the base abundance of birds present in the surveyed area using methods described by Royle *et al.* (2005) and implemented in the "unmarked" library (Fiske and Chandler 2011) running in the R statistical language version 3.0.1 (R Core Team 2013). The maximum likelihood model we used relies upon independent pairs of observations, meaning that the counters do not cue off of or otherwise influence each other, and that the ability of each observer to see a bird is the same. The model also assumes the population is closed over the survey area and period of inference.

The double-observer models were used to identify the ability of observers to detect the birds that are there and assess which of the measured covariates had a significant effect upon the number of birds present and detected in the survey area. We used AIC to select the best model from a candidate pool of models including all our measured covariates plus the wetland and survey day (i.e., first or second visit to the wetland). Once significant main effects were identified, they were then tested with remaining (non-selected) covariates to look for interaction effects. In cases where the AIC of the two best models differed by less than two units the simpler model was chosen following the suggestion of Arnold (2010).

The simplest double-observer estimator described by Nichols *et al.* (2000) uses "dependent" sampling, where one observer relays detections to a second observer, who records those detections as well as his/her own detections of individuals that were missed by the first observer. This method requires communication between the observers; if there are a large number of individuals to be counted, or the individuals are grouped into clusters, the second observer may have difficulty determining whether or not the first observer missed individuals.

By collecting independent observations from two observers more information is available to the analysis. Rather than using data from only the first observer as in dependent sampling, the independent double-observer model uses data from both observers and allows for the estimation of detection probability for both observers. This method assumes that individuals detected by both observers are correctly matched on the two maps. Surveys should be of an agreed upon area that is small enough so that counts can be done quickly to minimize the effects of animal movement.

Repeated Survey Sampling

From the repeat counts we can estimate abundances with bounded count and Overton

estimators. Regier and Robson (1967; see Skalski *et al.* 2005) proposed the bounded-count method for estimating abundance when it is conceivable that all individuals could be enumerated during a single canvass and repeat counts were conducted. The abundance is calculated using the formula:

$$\widehat{N} = 2x_m - x_{m-1} \qquad \text{(Equation 1)}$$

Where x_m is the largest of the m repeat counts and x_{m-1} is the second largest count obtained. The approximate upper bound of a 95% confidence interval is:

$$20 \cdot (x_m - 0.95 \cdot x_{m-1}) \qquad \text{(Equation 2)}$$

assuming a normal distribution. The lower bound is x_m, the maximum number observed. The assumptions of the bounded-count method are:
1. The detection probability is 1 or nearly 1, such that it is possible to enumerate all individuals;
2. The m repeat counts are independent;
3. The detection probability is constant across all replicate counts;
4. Individuals are not counted more than once during the count;
5. The population is closed during the course of the m replicate counts.

Repeat count data were also analyzed using a binomial method-of-moments or Overton estimator (Overton 1969; see Skalski *et al.* 2005) with the same independence assumptions as the bounded count estimator. This estimator performs well only if there are a large number of replicate counts (i.e., $m > 20$) with high detection probabilities (i.e., $p > 0.60$). This amount of effort makes it impractical for large-scale waterbird surveys, but we included it in our analysis as a point of comparison.

Point-transect and One-sided Line-transect Distance Sampling
Distance data were analyzed to model bird detectability as a function of increasing distance (Buckland *et al.* 2001). After counts were adjusted for detection probability they were used to estimate mean bird density in the surveyed area or along the surveyed transect line. Species-specific density estimates (birds ha^{-1}) were calculated using program DISTANCE (version 6.0, release 2; Thomas *et al.* 2010), following methods described in Buckland *et al.* (2001), Camp *et al.* (2009), and Thomas *et al.* (2010). Survey effort was adjusted by the number of counts, and line-transect effort was halved to account for the one-sided count. Candidate detection function models were limited to half normal and hazard-rate detection functions with expansion series of order two (Buckland *et al.* 2001:361, 365). Species, habitat characteristics, water level, time of day, weather, and individual observer were all potential covariates that might affect detection probability as a function of distance. Data were truncated at a distance where detection probability was <10%. The model with the lowest second-order Akaike's Information Criterion corrected for small sample sizes (AIC$_c$) was selected as the best approximating model (Buckland *et al.* 2001, Burnham and Anderson 2002). The assumptions of distance sampling are:
1. All birds are detected with certainty at the station's center point or along the center line;
2. Individuals are detected prior to any responsive movement;
3. Distances are measured without error.

Comparison of Methods

The coefficient of variation (CV) is a useful metric for assessing the amount of uncertainty in estimates across different sampling methods. CV is the ratio of the standard error to the mean, and ranges from 0 to infinity, where a standard error equal to the mean yields a CV of 1. Sampling methods that produce small CVs are more precise than methods with large CVs. For our assessment, we assume that—provided the necessary assumptions are met—sampling methods that produce more precise estimates are superior to less precise methods. We considered sampling methods reliable when they produced estimates with an acceptable degree of precision (CV less than about 20%). This would imply that the 95% confidence interval (CI) equals the mean plus and minus up to 40% of the mean. Using the CV also circumvents the necessity of producing a standard abundance metric (i.e., birds per unit area, density, versus total numbers of birds in an *a priori* defined area, abundance) for each method. This allows us to compare among methods that sampled different portions of each wetland.

RESULTS

Our intention with these test surveys was to evaluate different sampling methods rather than produce estimates of abundance at a particular wetland. Accordingly we treated each sampling station within a wetland as an independent survey of that station only, rather than combining stations to produce a wetland-wide count.

Covariate Consistency

Recording environmental covariates allows for modeling varying detection probabilities from site-to-site and day-to-day using DISTANCE and double-observer methods. On both days that we surveyed the Hamakua Marsh, water levels were much higher than is typical (James Cogswell, personal communication). At the Kiʻi Unit of James Campbell NWR, management activity moved water from our focal area to a neighboring wetland between our two survey visits, noticeably affecting the number of birds present.

A prerequisite of using covariates is that their values must be recorded consistently by all observers at all sites and counts. Figure 3 shows the percentage of the sky covered by cloud as recorded by pairs of observers at the same place and at the same time. Ideally there would be 100% correlation between the pairs of values, but the figure demonstrates that is not the case. Some small differences are inevitable and expected, but there are multiple instances where the two observers disagreed by more than 50%.

A smaller number of coarser categories makes it easier for observers to be consistent in their classifications. Table 2 shows the distribution of pairs of observations for vegetation thickness, human influence, and rainfall. It is worth noting that, although the data sheet listed human influence as ranging from 0 to 2, there were eight survey stations where the observer recorded human influence as "3." Wind speed was often not recorded and showed very little variation, so is not included in this analysis. Vegetation classification showed 100% correlation; all observers agreed in all cases. Human influence showed close agreement with no pairs of observers more than one category apart. Observers were free (and encouraged) to communicate and agree upon categories, so the consistency of these results may be an overestimate of how well independent observers would agree.

Rainfall was less informative because the majority of observation periods had no rain, which was easy to classify. The sometimes severe (category 0 vs. category 3 in one case)
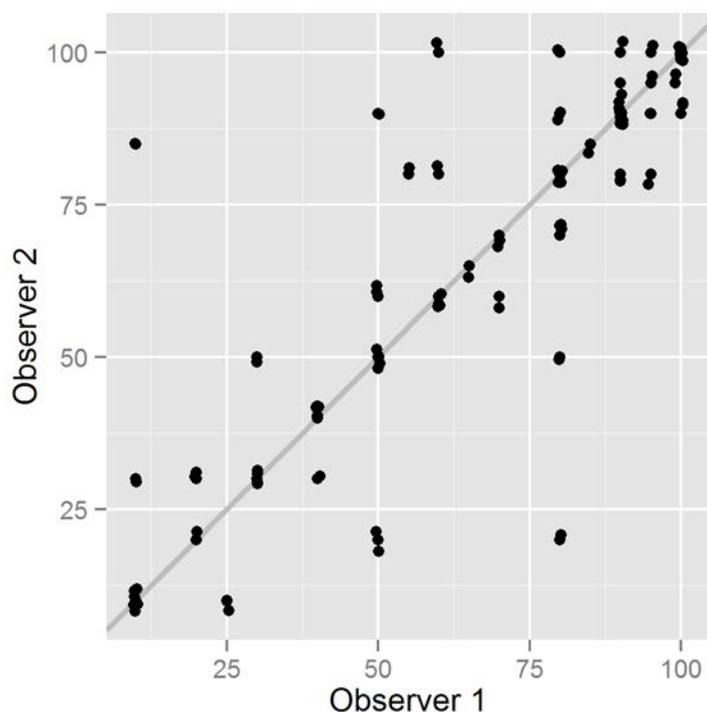
Figure 3. Percent cloud cover recorded by pairs of observers surveying at the same time and place. Overlapping points were offset slightly to make the distribution more evident. The grey line marks one-to-one agreement between the observers.

discrepancies may have been due to ambiguity as to how to record rainfall when a sudden downpour swept over the sampling area.

Direct Counts
By replicating the direct counting approach of current surveys with observers counting simultaneously and making repeat visits, we were able to estimate the accuracy of simple counts. As expected, observers, even two people counting waterbirds from the same place at the same time, varied in the number of birds detected (%CV ranged between 34 and 90). Table 3 details the mean variation both between paired observers and across a two-hour survey period for night-heron, coot, moorhen, and stilt, using all observations where at least one of the observers detected at least one bird.

We also examined the variation in the number of birds counted from the same station throughout a two-hour survey period (i.e., as many as six observers [three pairs of two] at each station in a period). The variances were larger than the within-pair variation because they encompass both observer variation and variability due to movement of waterbirds over the course of the survey period. This variability is more representative of the expected variability of a single count of an area.

Double-observer Sampling
We used the double-observer models to identify the ability of observers to detect the birds that

Table 2. Consistency between observers recording categorical covariates. Observers classified vegetation, human influence, and rainfall intensity on a scale of 0 to 3. Table cells record the number of times the first (horizontal) and second (vertical) observers recorded the category at the same time and place. If there were perfect agreement between the observers all combinations would be on the diagonal, as is the case for vegetation.

Vegetation category

	0	1	2	3
0	7	0	0	0
1	0	18	0	0
2	0	0	51	0
3	0	0	0	17

Human influence category

	0	1	2	3
0	36	9	0	0
1	0	10	3	0
2	0	1	11	8
3	0	0	0	0

Rainfall category

	0	1	2	3
0	78	0	1	1
1	2	0	0	0
2	0	0	2	0
3	0	1	0	0

are there and assess which of the measured covariates had a significant effect upon the number of birds present and detected in the survey area. In all but two cases the model with the next lowest AIC_c was separated by more than four units. In two cases where the AIC_c of the two best models differed by less than two units and the simpler model was chosen. In general the covariates wetland and survey day, and their interaction, were the most significant predictors of abundance, accounting for the differences between Ki'i and Hamakua and the difference in raw abundance between the first and second survey days at Ki'i (likely driven by a managed reduction in water level in our focus wetland). Time of day was also a significant factor in abundance in most models, with bird abundance declining as the day progressed. For the Hawaiian stilt, cloud cover had a significant effect on abundance in the point surveys, with greater numbers detected as cloud cover increased, and human influence had an effect in the line surveys, with greater number of birds detected at higher levels of human influence. This last effect may have been driven by the survey logistics at the Hamakua wetland where one line transect had a high human influence score and was elevated 3 m above the water level, possibly making more stilts available to the observers from an improved vantage. Table 4 lists the best models for each species using the line and point double-observer data.

Table 3. Variation in counts by different observers at the same sites. Within-pair variation is between two observers at the same survey station counting for six minutes; paired counts is a tally of detected individuals by observer pairs for each species. Within site per period variation is among all observers who counted at a station during the two-hour survey period. Range is the difference between the minimum and maximum number of birds counted between paired observations or among all counts at a site during a period. As is common in wildlife abundance measures the variance increased along with the average number of individuals detected.

Species	Paired counts	Between paired observer variation				
		Mean count	Mean variance	Mean SD	%CV	Range
Night-heron	92	3.0	1.05	1.02	34	3
Coot	91	5.6	25.19	5.02	90	25
Moorhen	85	2.9	1.24	1.11	39	5
Stilt	102	3.4	2.19	1.48	43	8
		Variation within 2 hr sampling period				
		Mean count	Mean variance	Mean SD	%CV	Range
Night-heron		3.5	2.03	1.42	70	8
Coot		10.0	62.22	7.89	13	77
Moorhen		2.7	2.71	1.65	61	9
Stilt		5.4	4.83	2.20	46	11

In all models observer was a significant factor in detection probability, indicating that individual heterogeneity among the individual surveyors affected their counts. Time of day was also a factor in many models, indicating that detection probability decreased as the day progressed. This decline might be due to bird behavior, observer fatigue, or perhaps the changing sunlight. Vegetation category also affected the ability to detect birds, with increasingly thick vegetation decreasing the probability of detecting birds. Table 4 presents the mean detection probability, averaging over all observers and environmental conditions. Line-transect models had higher (~+15%) detection probabilities than point-transect models, but it is possible this difference is due to the greater ease of matching individual birds on the maps produced by paired observers for the one-sided line-transect surveys where both observers were looking in the same direction at approximately the same time than for the point-transect survey. In point-transect models, stilt were more likely to be detected than the other species.

Table 5 presents the model predictions for each species using line and point double-observer methods. All models except for stilt had wetland and survey day as significant predictors of abundance. Survey day was not significant for stilt, which is why the results are identical for both days. Stilt also had human influence (line-transect models) and cloud cover (point-transect models) as abundance covariates, which were set to zero to produce the estimates in Table 5.

Repeated Survey Sampling
We used the counts recorded during the line transects by each observer during each sampling period as independent observations of a closed population on the wetland for a particular day.

Table 4. Factors affecting detection and abundance in double-observer surveys. Observer and some environmental covariates had effects on detection probability, but the overall means are presented here.

Model type	Species	Total detections	Simultaneous detections	Factors affecting detection	Factors affecting abundance	Detection probability	SE	% CV
Line transect	Night-heron	220	86	Observer, vegetation, time of day	Wetland, survey day, time of day	0.56	0.034	6.1
	Coot	432	209	Observer, vegetation, time of day	Wetland, survey day, start time	0.65	0.022	3.3
	Moorhen	129	65	Observer, vegetation	Wetland, survey day	0.67	0.039	5.8
	Stilt	243	125	Observer	Wetland, human influence	0.68	0.028	4.1
Point transect	Night-heron	301	118	Observer, vegetation, time of day	Wetland, survey day, time of day	0.56	0.029	5.2
	Coot	979	326	Observer, vegetation, time of day	Wetland, survey day, time of day	0.50	0.017	3.4
	Moorhen	225	76	Observer, vegetation, time of day	Wetland, survey day, time of day	0.50	0.036	7.1
	Stilt	554	266	Observer	Wetland, cloud cover	0.66	0.019	3.0

Table 5. Predicted number of birds available to be counted by wetland and survey day with 95% confidence intervals by line- and point-transect methods. Additional covariates (cloud cover and human influence) were set to zero.

Model type	Species	Wetland	Survey day	Abundance	SE	% CV
Line transect	Night-heron	Ki'i	First	13.5 (11.0–16.6)	1.41	10.4
			Second	9.0 (6.9–11.9)	1.27	14.0
		Hamakua	First	4.6 (3.3–6.5)	0.79	17.1
			Second	1.6 (0.7–3.5)	0.63	40.9
	Coot	Ki'i	First	3.7 (2.4–5.5)	0.75	20.6
			Second	8.0 (5.8–11.1)	1.32	16.5
		Hamakua	First	1.2 (0.8–1.8)	0.26	21.2
			Second	0.5 (0.2–1.1)	0.20	42.2
	Moorhen	Ki'i	First	2.3 (1.3–4.3)	0.74	31.6
			Second	4.8 (2.8–8.3)	1.33	27.6
		Hamakua	First	8.2 (6.5–10.4)	1.00	12.2
			Second	4.6 (3.0–7.0)	0.99	21.4
	Stilt	Ki'i	First	1.7 (0.9–2.5)	0.42	24.4
			Second	1.7 (0.9–2.5)	0.42	24.4
		Hamakua	First	1.9 (1.6–2.1)	0.11	6.1
			Second	1.9 (1.6–2.1)	0.11	6.1
Point transect	Night-heron	Ki'i	First	4.2 (3.0–5.7)	0.67	16.1
			Second	2.6 (1.8–3.8)	0.51	19.3
		Hamakua	First	2.5 (1.8–3.5)	0.43	17.0
			Second	3.3 (2.1–5.1)	0.73	22.2
	Coot	Ki'i	First	17.7 (14.8–21.3)	1.65	9.3
			Second	15.3 (12.8–18.4)	1.43	9.3
		Hamakua	First	2.0 (1.4–2.8)	0.34	16.8
			Second	1.7 (1.1–2.8)	0.41	23.6
	Moorhen	Ki'i	First	2.4 (1.5–3.7)	0.55	23.0
			Second	2.0 (1.2–3.3)	0.53	26.9
		Hamakua	First	2.9 (2.0–4.2)	0.55	19.1
			Second	3.1 (2.1–4.6)	0.61	19.6
	Stilt	Ki'i	First	2.8 (2.2–3.6)	0.34	12.1
			Second	2.8 (2.2–3.6)	0.34	12.1
		Hamakua	First	5.4 (4.2–7.0)	0.69	12.6
			Second	5.4 (4.2–7.0)	0.69	12.6

Assuming a closed population over the six hours of surveys is questionable, but making such an assumption was needed to get a sufficient number of samples to apply the repeat count estimator.

Table 6 details the bounded count estimators of abundance, an approximate 95% confidence interval (the confidence interval has a lower bound equal to the maximum number of birds detected), along with the number of surveys used to calculate the estimate and interval. The binomial method-of-moments Overton estimator is known to be poorly behaved (Skalski *et al.* 2005) and at least 30 independent samples are recommended. When applied to the repeat counts (as described above) for the purposes of species abundance estimation, all but one of the samples had a count variance larger than the mean count, which results in a (nonsensical) negative estimate. The Overton estimator is not appropriate if the variation observed in this study is characteristic of the sample variation to be expected in regular surveys.

Table 6. Bounded count estimators with associated 95% confidence intervals, calculated with the assumptions that each line transect surveyed by each observer was an independent count of a closed population. The coefficients of variation are calculated assuming a normal distribution of the uncertainty.

		Night-heron				Coot		
Kiʻi	n	Abundance		% CV	n	Abundance		% CV
03/26/13	23	39	(27–255)	149	25	132	(86–960)	169
03/29/13	10	12	(11–30)	40	12	41	(35–149)	71
Hamakua								
03/27/13	20	7	(6–25)	69	24	7	(6–25)	69
03/28/13	8	11	(8–65)	132	11	7	(6–25)	69

		Moorhen				Stilt		
Kiʻi	n	Abundance		% CV	n	Abundance		% CV
03/26/13	17	7	(5–43)	138	16	22	(16–130)	132
03/29/13	7	5	(4–23)	97	9	12	(9–66)	121
Hamakua								
03/27/13	23	11	(10–29)	44	18	19	(18–37)	26
03/28/13	12	10	(8–46)	97	11	12	(11–30)	40

Point-transect and One-sided Line-transect Distance Sampling
Contrary to the typical pattern in forest bird surveys, detection type (visual vs. auditory) was not a significant factor in any of the waterbird models. This may be because the vast majority of detections were visual (type 2) and made up 96% of 3,121 detections on the line transects and 97% of 4,083 detections on point transects.

In general the best-fitting detection functions were still poor fits to the observed distance data (Table 7). With the exception of the point-transect model for moorhen, none of the models passes a Kolmogorov-Smirnov goodness-of-fit test. This failure is not a fatal flaw and indicates that there are violations of the modeling or sampling assumptions, or perhaps that there are significant covariates affecting detection that were not recorded. Time of day influenced detectability of all four species for point-transect models and stilt using line-transect models. Vegetation was a significant covariate of night-heron and moorhen density for line-transect

Table 7. Factors affecting detectability in point- and line-transect models, along with estimated densities and associated 95% confidence intervals. Densities are birds per kilometer of transect for a one-sided line transect, and birds per hectare for point-transect models. Dn is the Kolmogorov-Smirnov test statistic with associated p-value of the null-hypothesis that the fitted detection function explains the observed distribution of distances.

Model type	Species	Truncation distance	Covariates	Dn	p	Density	% CV
Line transect	Night-heron	107	Vegetation	0.135	0.0000	8.8 (7.0 - 11.1)	11.6
	Coot	120	Survey date	0.342	0.0000	19.3 (15.7 - 23.8)	10.5
	Moorhen	88	Vegetation	0.139	0.0006	7.5 (5.8 - 9.6)	12.6
	Stilt	95	Time of day	0.084	0.0005	11.1 (9.6 - 12.8)	7.1
Point transect	Night-heron	138	Time of day	0.128	0.0000	0.57 (0.53 - 0.61)	3.8
	Coot	195	Time of day, Observer	0.129	0.0000	0.97 (0.89 - 1.06)	4.4
	Moorhen	127	Time of day	0.048	0.5134	0.73 (0.67 - 0.81)	4.9
	Stilt	121	Time of day	0.059	0.0224	1.00 (0.96 - 1.06)	2.5

models. Interestingly the covariate wetland did not influence the detection models, but this could be a limitation of our sampling only high-density wetlands.

In Figure 4 two sample detections are shown against the background of distances used to fit the model. In the line transect for coot we see a paucity of detections at shorter distances, perhaps reflecting an uneven distribution of birds in the environment due to limitations on where the observers can sample or movement away from the observers. The point-transect detection function for moorhen shows an anomaly at very close distances, perhaps due to movement towards the observers. Because point transects have relatively much less area to survey at short distances, the method is less sensitive to lack of fit closer to the observer than for the line-transect method. Regardless, the general lack of fit of both point- and line-transect models results in larger CVs than would be expected from data that better fit the models.

DISCUSSION

We sampled waterbirds using the existing index-count survey method and compared it to a variety of sampling methods that allow for estimating detectability and absolute abundance. Using index counts to estimate abundance requires a measure of the detection probability to account for birds that are not detected (Anderson 2001, Thompson 2002). Robust measures of the detection
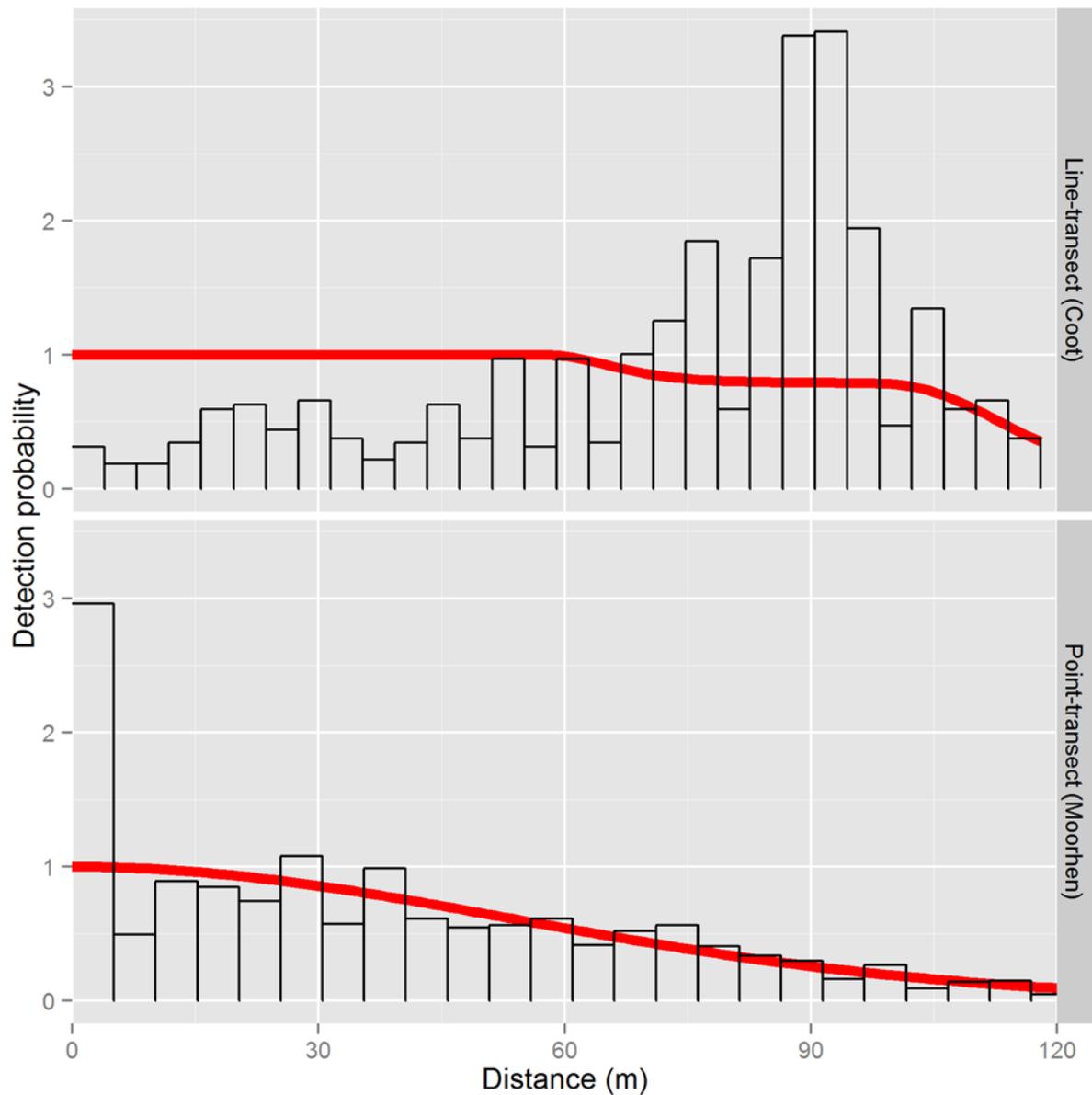
Figure 4. Fit detection functions for line- and point-transect methods for two species with similar truncation distances. Note that for the line-transect method there is a paucity of observations at shorter distances relative to the fit model.

probability can be used to calibrate indices into abundance estimates, but require some form of replicated sampling either by repeat counts or repeated measures of distances to detected birds—either as part of a continuing survey methodology or applied later as a correction factor that was derived from an intensive repeat survey.

The variability among single visit simple counts can be used to show that the counts are not exact measurements (Table 3). We show that experienced observers will count different

numbers of birds throughout a two-hour time period, or even from the same place and time. This variance in detection probability is not taken into account with the current survey protocol, yet has large implications for estimates of population size. For example, the coefficient of variation for the number of stilt counted was 46%, meaning that—all things being equal—a total count of 1,000 stilts should be evaluated as having some uncertainty, with a 95% confidence interval of the count $\pm$ (1.96 $\cdot$ 0.46 $\cdot$ 1,000) or between 100 and 1,900 individuals. Different species of birds had different detectability. From Table 4 the mean detectability for coot was 0.50, while that of moorhen was 0.66, meaning moorhen were easier to detect. All other things being equal, this means that a count of 15 coots would translate to an abundance estimate of 30 (15 / 0.50), while counting 15 moorhens would lead to an abundance estimate of 22.7 (conservatively rounded down to 22) birds.

Additionally, our analyses identified several factors (e.g., time of day, vegetation, and water level) that affect the ability of observers to detect waterbirds, as well as the simple realization that some species are easier to detect at some wetlands than at others. As an example, all other factors held equal, observers RJC and KWB were modeled as having different detection probabilities for moorhen. In circumstances where observer RJC had a 60% chance of detecting a bird, observer KWB would have a 75% chance. This implies that if both observers counted 10 moorhen, RJC's estimate would imply there were 16.6 birds (10 / 0.60) while KWB's count would translate to 13.3 (10 / 0.75) birds. Therefore, the intrinsic variability due to observer idiosyncrasies, survey conditions, and uncontrollable factors that change over relatively short time scales mean that simple counts are not reliable and the index may not be proportional to true abundance. The use of sampling techniques that can account for differences in detectability of each species and covariates will greatly improve population estimates.

Both DISTANCE and double-observer methods can use habitat and sampling covariates to improve estimates of detection probability under varying conditions (Marques and Buckland 2004, Royle *et al.* 2005), but they assume that covariates are measured consistently by all observers at different times and places. Our data show how, even between two people working together during the same count and site, covariates are not always measured in the same way, although standardization of covariate recording can improve estimator precision (Thompson 2002).

Evaluation of Alternative Survey Methods
In addition to assessing the direct count method currently used to monitor Hawaiian waterbirds, we evaluated the performance of several alternative surveying methods. We tested abundance estimation methods that could be performed through simple observation of birds from a distance without identification of individuals. There are other techniques which can provide more accurate estimates as well as information about population demographics and animal movement, but they require that a significant proportion of the population be marked in an individually identifiable fashion. However, a program to band large numbers of waterbirds would be costly, time consuming, and does not currently exist. Therefore, we limited our evaluation to methods using unmarked individuals.

With standardized training of waterbird observers, direct-count based indices could be improved, but would still fall short of being measures of abundance. A double-observer protocol such as the one described below could be used to model the effects of covariates and imperfect detectability to estimate the number of individuals of endangered Hawaiian waterbirds. Here we provide an assessment of the current survey method, address assumption violations of the

survey methods used in the pilot study, propose a standard monitoring protocol using the double-observer method to estimate bird abundances, and recommend future research that will improve the protocol.

Assessment of Method Assumptions

All counting protocols and abundance estimators are based on underlying model assumptions. While some minor violation of model assumptions is inevitable, and certain models are more robust to violations of assumptions than others, we identified some problems that would affect the precision and accuracy of the abundance estimates.

The detection probabilities for even the most conspicuous waterbirds are less than 100%. Relying on direct count methods to estimate abundance is therefore not feasible because they assume 100% detectability of all birds present (White and Bennetts 1996, Anderson 2001). Furthermore, detection probabilities differed among site and sampling covariates (including habitat, observer, time of day, and wetland), making the proportionality assumption questionable for both the repeated and canvass survey methods (Burnham 1981, Anderson 2001). In addition, temporal differences in detectability indicated that the detection probability varies over even short periods, thus violating another assumption of the repeated survey method. The repeated survey method also assumes that the counts are independent (i.e., that the observers do not inform or cue each other). From the mapping exercise, our results indicated that assumption was likely to be met. There were situations, however, where observers would point out birds or cue off of each other. This lack of independence is likely to be more problematic at wetlands with lower bird abundances. Thus, it is unlikely that direct count and repeated survey method assumptions can be met for Hawaiian waterbird surveys. The range of the confidence interval should inform expectations of estimates from comparable survey efforts, although the assumption of a closed population is specious and renders estimates from the repeated survey method inaccurate.

Both distance sampling and double-observer methods rely on differing critical assumptions to produce reliable estimates. Distance methods assume that birds are, on average, evenly distributed in the environment (Buckland *et al.* 2001). In practice, waterbirds are generally seen on or in the water, and humans must observe from nearby dry land, meaning there is usually some unoccupied area between the observer and the populated habitat the birds occupy. This is less of a problem with point transects because the radial area close to the observer is small relative to the total area surveyed, but for line transects this phenomenon can reduce the precision of the detection function and create a positive bias in the estimate by including unoccupied habitat in the area of inference. Distance sampling methods also assume that all birds are detected with certainty at the station's center point or along the center line. The placement of both the stations and lines on the boundary of the wetlands instead of within the wetland precludes birds from occurring at the station or line center. Placement of the sampling units within the wetlands would be very disruptive to the birds and logistically not feasible. Distance sampling methods assume that birds are detected prior to any responsive movement. Normally observers are trained to move carefully and quietly to avoid scaring birds away from the survey point or line. We noticed that birds were more sensitive to humans at the Kiʻi Unit than at Hamakua Marsh, which abuts a commercial center with constant traffic. At the Kiʻi Unit, Hawaiian waterbirds were observed moving, and some birds were sneaking away as observers approached stations and walked along transects. In contrast, at Hamakua Marsh, some birds—particularly moorhen—approached observers, begging for food. It was not possible to identify the original location of individuals prior to their movement, nor was it possible to note with

certainty which birds had or had not moved. The use of laser rangefinders ensured that distances were measured without error, thus meeting the final distance sampling assumption. If observers were forced to estimate distances without rangefinders, the performance of the method would likely degrade. Thus, the assumption of uniform distribution and two of the three critical distance sampling assumptions will likely not be met for Hawaiian waterbird surveys due to logistical constraints and bird behavior.

Double-observer methods rely on the three assumptions that detections are made independently by observers, that both observers are physically able to see the bird, and that the population is closed during the course of the survey. Although there was initial communication between observers to standardize maps, to the extent possible, observers detected and recorded birds during each count independently. Birds flying overhead and moving into or out of the sampling sites were recorded but not used in analyses. This procedure helped meet the closed population assumption. The closed population assumption can also be met by keeping the length of the island-wide survey short and by sampling all wetlands within a relatively short period. Knowing the amount of movement among wetlands will be useful in evaluating this assumption. Our double-observer methods depend upon coordinating the maps drawn by each observer to identify which birds were seen by both of them, and which by just one observer. Hand-drawn maps with different landmarks made this step sometimes difficult, and matching the detections was an additional step in processing the data. When there was ambiguity in determining if individual birds were the same, we erred on the side of not identifying them as being seen by both observers. This bias has the effect of under-estimating the detection probabilities and inflating the estimate of underlying abundance; however, doing so did not violate the method assumptions. Therefore, it is likely that the double-observer method assumptions can be met for Hawaiian waterbird surveys, especially if the recommended changes are made to coordinating observer maps (see below).

Comparison of Precision among Survey Methods
We showed that there is a substantial amount of variability (CVs were greater than 20%) both among observers and over a two-hour interval. This indicates that in order to estimate true abundance a method needs to be able to account for covariates including observer, time of day, and site and sampling conditions. This requirement precludes the use of bounded count and binomial method-of-moments Overton estimators. Both of these methods also produced CVs that were generally much larger than other methods. The mean CV for the bounded count estimators was 91% and ranged from 26% to 169%. Figure 5 shows the relation between survey method precision and bias, where bias is a qualitative measure based on the degree of assumption violation.

CVs for both distance sampling and double-observer methods were typically within the desired level of certainty. Distance sampling methods yielded more precise estimates than double-observer methods (mean CVs across species and models = 7.2% and 19.0%, respectively). For both methods point models were more precise than line models (see Tables 4 and 5).

Our surveys were conducted in two wetlands with relatively high densities of waterbirds. It is likely that CVs will be much larger for wetlands with low waterbird densities, making any sampling method less reliable. Further, distance sampling methods are known to break down when insufficient numbers of birds are detected to fit a detection function (Buckland *et al.* 2001). Reliably modeling detection functions requires >75 detections, and this may be difficult to obtain for wetlands with low waterbird densities. Pooling detections across wetlands and
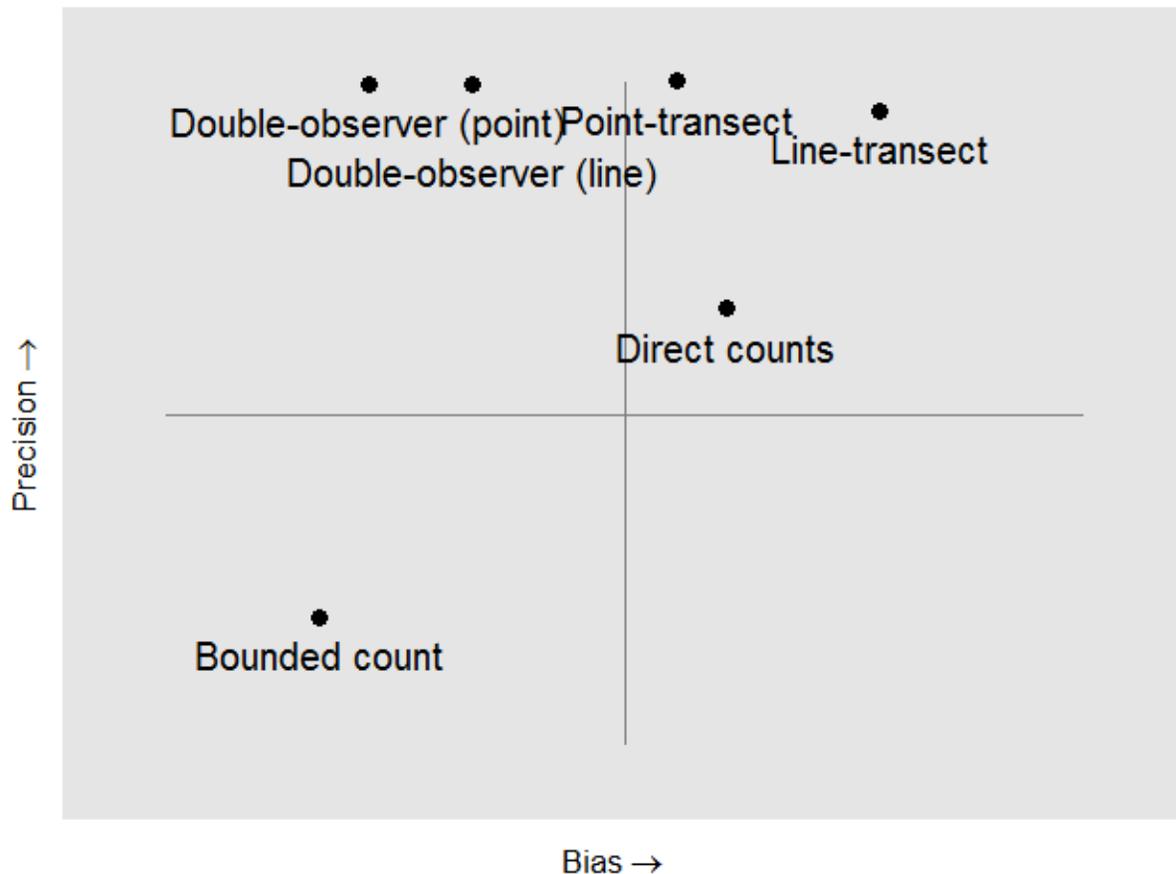
Figure 5. Relationship between survey method precision and bias. Desirable methods have high precision and low bias and are situated in the upper left quadrant. Bias was based on qualitative measure of the degree of method assumption violations.

counts may be used to meet the required numbers of detections; however, population estimates will not be reliable for wetlands with only a few detections regardless of pooling. The double-observer method also behaves poorly for species that occur in low numbers, although we believe the double-observer method will perform better than the distance sampling method in this situation.

Proposed Approach for Future Surveys
In order to estimate abundance, rather than an index of abundance, a more sophisticated estimator that accounts for imperfect detection under a range of covariate values is required. Among the methods we have examined in this document only double-observer sampling meets the requirements of the underlying assumptions of the method without requiring an excessive amount of sampling effort. In the process of conducting this pilot study we identified three concerns that influence the reliability of all counting methods. The following actions would improve survey reliability: 1) formalize a study design, 2) standardize training of observers, and 3) standardize definitions of covariates. In addition, knowing the degree of between-wetland bird movement will influence the duration of the survey period.

The spatial extent of the surveys needs to be carefully considered. The current monitoring includes counting birds at all of the important wetlands identified in the Hawaiian waterbirds Recovery Plan (USFWS 2011), and this practice should be continued. In the future, accessibility to all of the wetlands may not be feasible for a variety of reasons. If that is the situation, statistical inference cannot be extended to the portions of wetlands that were not sampled. Randomly selecting among the wetlands to be surveyed will be necessary if logistics further limit the number of wetlands that can be sampled. Before this happens, the degree to which birds move among wetlands should be determined as subsampling wetlands may result in estimator bias.

Additionally, the exact area within each wetland should be carefully demarcated for consistency across years. For waterbirds it seems intuitive that the sampling area to estimate waterbird abundance should match and coincide with wetland boundaries. However, a loosely defined sampling area that does not explicitly delineate the area sampled can result in confusion among observers and unknowingly change over time. The boundary of each wetland and the area sampled can be demarcated using GIS to delineate the inference area. Before sampling commences, a reconnaissance should be conducted to ground truth that all portions within the inference area can be accessed and sampled. Any discrepancies should be noted and the area adjusted accordingly. During subsequent sampling, any changes to the sampling frame (e.g., management or landscaping creating new habitat) and inference area should be noted, the protocol modified, and future abundance estimates adjusted accordingly.

Covariates are only useful if they accurately measure the same values across observers, wetlands, and years. Covariates can be used to adjust the detection probability given different conditions. This process will improve estimator reliability by reducing variability in abundances. Missing values add statistical noise making covariates less informative; therefore, observers should record all covariates for each count and record them consistently and accurately. We showed that observers can have widely divergent estimates of covariate values (such as cloud cover; Figure 3), which points to the need to carefully train observers in the data they collect. Discussions with the observers revealed they were using different definitions of cloud cover. The exact definition is less important, so long as it is consistent, but we suggest that cloud cover pertain only to the sky immediately overhead—everything more than 45° above the horizon, excluding any landscape or buildings that obscures that part of the sky. In that portion of the sky, estimate how much is obscured by clouds. If the clouds are partially transparent, estimate the amount of light that is being blocked.

Wind (and gust) strength and rainfall intensity must have similar standards. Beaufort category descriptors are already distributed with datasheets, so each observer has a reference. Rainfall category is more difficult to demonstrate, but a consistent description (e.g., none, fog, mist, droplets, steady rain) is already being used. Similarly the covariates for human influence and vegetation need to have common standards; if observers cannot be calibrated at a common location, photographic examples might be distributed instead.

Water level is particularly difficult, as it is a relative measure at each wetland. Recording whether the water level is below, at, or above normal requires some experience with the wetland history. A set of descriptions or submerged landmarks at each wetland could be assembled and distributed with the data recording sheets for each wetland. Water level is particularly important for survey areas with a strong tidal influence, where the level is expected to vary throughout the day.

Individual observers must be distinctly identified so their particular detection abilities are accounted for across wetlands and years. Ideally there should be a list of names and contact information with a unique, easy to record, difficult to confuse code that can be used across years. Three-letter initials are often used, with rare letters such as "X" or "Q" to fill in missing names or avoid duplicates. With care taken to note name changes, a consistent code can be used to account for an observer's particular effect across decades of data collection.

The degree to which birds move among wetlands is unknown but current studies of marked birds are underway. From a sampling perspective this is important information as it will determine the sampling window, the number of days over which counts can be conducted, and if a subset of wetlands can be sampled instead of counting all wetlands. This is because all of the survey methods either implicitly or explicitly assume that the population is closed over the survey area and period. If birds move freely among wetlands then counts should be conducted over as short a period as possible, ideally within a single day. If birds are resident at wetlands for relatively long periods, at least several days, then the sampling window can be lengthened, and counts can occur over a few days. Sampling over a few days will allow the counts to occur when the birds are most detectable and to halt sampling before detectability declines in late morning and afternoon. If birds move freely among wetlands then there is a risk that subsampling the wetlands will induce an estimator bias. This can arise when abundances at sampled wetlands are not representative of the entire population. Until the degree to which birds move among wetlands has been determined, all of the important wetlands designated in the USFWS (2011) should be sampled.

There are several logistic issues involved with double-observer sampling that should be addressed before the survey occurs. Each wetland to be surveyed needs to be divided into one or more segments that are small enough for the observers to count and map all the birds they detect in a short amount of time. The segments must cover the entire wetland area of inference and should have access points where observers can see the entire segment—keeping in mind that shorter observers might have obstructions to the line of sight that taller observers do not. If possible, there should be natural barriers or landmarks that allow the observers to count and map only birds within the segment (to avoid double-counting in adjacent segments).

Post-processing the observers' maps to identify which detections were in common with each other and which were unique is sometimes difficult when each observer draws a map. Aerial photographs might be applicable in some cases and not in others for the same reason the maps needed to survey a particular wetland may vary from survey to survey. Water levels may change, vegetation may grow up or die back altering visibility, or access paths may change. Ideally each wetland to be surveyed should be visited a short time before the actual survey. During this pre-visit, the wetland can be partitioned into suitable segments, survey points identified, and maps drawn and labeled, photocopied, and distributed to the observers.

During the survey, observers should attempt to map the birds in a segment as quickly as possible before moving on to the next segment, until the entire wetland is surveyed. The exact time spent at each segment should be standardized, despite the fact that the size of the survey area and density of birds is variable. A minimum of six minutes per station is used in forest bird surveys. Until the sampling duration can be formally tested, counts should last no longer than six minutes. Keeping the count period short will help meet the requirement that observers have the same opportunity to see each bird. Birds that are seen moving from segment to segment should be noted and recorded or mapped in their source segment to avoid double-counting.

Fly-ins from outside the wetland should be recorded separately. In areas with extreme bird densities (e.g., the Kailua-Kona water treatment plant) it may be necessary to survey each endangered species separately. Time of day had a strong effect on both the ability of observers to detect birds and the availability of birds to be counted. Some of this variability could be avoided by restricting survey hours to morning conditions, such as halting counts at 11:00 am.

While mapping the birds it is important that the observers do their best not to cue off of each other nor communicate in any other way. This is very difficult, but the method assumes that the detection probability of each does not depend upon whether or not the other observer detected a particular bird. Careful positioning of observers so that they stay out of each other's field of vision will help to ensure that observers do not cue off of each other. To this end it is sometimes helpful if observers pretend to write down observations—especially if counts are slow—to help each other practice ignoring the other observer.

After the survey is complete, the maps are compared and individual birds that were detected by both observers identified. For each wetland the total number of birds (summed across all segments of the wetland) seen by only the first observer, only the second observer, and by both observers (along with covariates) constitute the data for that species at that wetland. With an entire island (or state) worth of data, the underlying abundance of birds can be estimated using a hierarchal model such as that described by Royle *et al.* (2005) and implemented by Fiske and Chandler (2011).

Double-observer methods can easily be expanded to incorporate more observers so long as it is logistically feasible for all the observers to count in the same time and place, and so long as they do not cue from each other to maintain the assumption of independent detections. In this current trial of methods we increased sample size to fit the double-observer models by treating each survey station as a separate count, therefore the estimates here represent the average number of individuals observable for that wetland and day. In a practical survey design where each count represents the survey area of interest, double-observer models can be used to estimate the abundance of birds available to be counted. Provided that day-to-day migration between wetlands is negligible, repeat surveys of a wetland done close in time may help to quantify variability due to bird availability.

Further Research

Occupancy surveys
In occupancy surveys, repeat visits allow for estimating the probability of detecting a species given that it is actually present (MacKenzie *et al.* 2006). Occupancy surveys are usually less expensive and simpler than collecting the data needed for abundance estimation, and can be particularly useful for rare or elusive species (Thompson 2004). Occupancy data are commonly used to describe population distribution and geographic range. In many cases, occupancy and abundance metrics are closely related. In our test surveys, occupancy analysis (estimating the probability that a site is occupied, and the chance of detecting birds given it is occupied) was not appropriate because all four species were detected at all but one observation site. The method does not reliably estimate occupancy when all sites are occupied (MacKenzie *et al.* 2006). In the context of a larger-scale, island-wide survey, occupancy of surveyed wetlands might be useful for determining the distribution and habitat utilization of endangered waterbirds. Occupancy surveys do not provide estimates of abundance, which is the current management objective of monitoring waterbirds, but can be valuable to assess where waterbirds occur and link occurrence with important factors such as type of wetland, level of

management, and so on. Surveys over a larger range of potential locations will be needed to assess the value of this additional method. The estimated detection probabilities in the double-observer models are similar in interpretation to those obtained in occupancy analyses and can be illustrative of what might be expected for occupancy analysis where there are lower bird abundances.

Hawaiian ducks

We detected no pure Hawaiian ducks during our trial surveys, and only a small number of hybrids of indeterminate purity. Because Hawaiian ducks are rare, and it is difficult to differentiate among Hawaiian ducks, mallards, and hybrids, we did not include them in our analyses. However, from our pilot survey we believe simple counts in wetlands are not an appropriate method to measure duck abundance. More widespread occupancy surveys (as described above) of their entire habitat, including mountain streams, might suffice to quantify their distribution. Alternately, individually marking a significant proportion of the population might be a prerequisite for true estimates of abundance via mark-recapture techniques. Further research is needed to design an effective survey protocol for Hawaiian ducks.

Availability

It is possible for a bird to be present at a wetland and yet not be detectable. A coot that is hiding in emergent vegetation during a survey will not be counted by any method because it is unavailable to observers. If the average rate of movement in and out of the vegetation—transitioning between available and unavailable states—is shorter than the survey time, availability will not have an effect on abundance estimates. If it is longer than the survey time, some birds will be hidden throughout the survey making them impossible to detect, and abundance will be underestimated because it does not account for the unavailable fraction of the population. During the nesting season, birds are more likely to be brooding upon concealed nests. Conducting surveys outside of the nesting season will reduce the effects of unavailability due to brooding behavior. Radio-telemetry studies would allow one to quantify the fraction of unavailable birds. If the radio signal of a tagged bird can be detected, while the bird itself cannot be seen or heard, it is known to be present at the wetland yet unavailable to the survey.

Movement between wetlands

It is unknown how frequently waterbirds move between distinct wetlands, both at a local scale and across large distances within and between islands, and whether movement is continuous throughout the year or seasonal. Currently waterbird surveys are conducted in a single day to avoid the possibility of double-counting a bird if it moves between wetlands. If it could be shown that inter-wetland movement is infrequent during a survey, then sampling could take place over multiple days.

Quantifying such movement requires individually marking some waterbirds and tracking them as they move about the island. Marking individuals with leg or neck bands could allow observers to report the locations of birds through systematic resight surveys and incidental sightings. Radio transmitters or GPS recording tags could be used to track individuals in real time to determine how far and how often the birds move from wetland to wetland, which would not only help improve survey design but also provide important information on how wetlands are connected by the birds.

LITERATURE CITED

Anderson, D. R. 2001. The need to get the basics right in wildlife field studies. Wildlife Society Bulletin 29:1294–1297.

Arnold, T. W. 2010. Uninformative parameters and model selection using Akaike's information criterion. Journal of Wildlife Management 74:1175–1178.

Bannor, B. K., and E. Kiviat. 2002. Common Gallinule (*Gallinula galeata*). *In* A. Poole (editor). The Birds of North America Online. Ithaca: Cornell Lab of Ornithology. Available at: http://bna.birds.cornell.edu/bna/species/685. doi:10.2173/bna.685

Bart, J., and S. Earnst. 2002. Double sampling to estimate density and populations trends in birds. Auk 119:36–45.

Borchers, D. L., S. T. Buckland, and W. Zucchini. 2002. Estimating animal abundance: closed populations. Springer-Verlag, London, UK.

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, UK.

Burnham, K. P. 1981. Summarizing remarks: environmental influences. Studies in Avian Biology 6:324–325.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach, second edition. Springer-Verlag, New York, NY.

Camp, R. J., P. M. Gorresen, T. K. Pratt, and B. L. Woodworth. 2009. Population trends of native Hawaiian forest birds, 1976–2008: the data and statistical analyses. Hawaiʻi Cooperative Studies Unit Technical Report HCSU-012. University of Hawaiʻi at Hilo. Available at: http://www.uhh.hawaii.edu/hcsu/publications.php.

Cochran, W. G. 1977. Sampling techniques, 3rd edition. John Wiley and Sons, New York, NY.

DesRochers, D. W., H. K. W. Gee, and J. M. Reed. 2008. Response of Hawaiian Moorhens to broadcast of conspecific calls and a comparison with other survey methods. Journal of Field Ornithology 79:448–457.

Engilis, Jr., A., K. J. Uyehara, and J. G. Giffin. 2002. Hawaiian Duck (*Anas wyvilliana*). *In* A. Poole (editor). The Birds of North America Online. Ithaca: Cornell Lab of Ornithology. Available at: http://bna.birds.cornell.edu/bna/species/694. doi:10.2173/bna.694

Farnsworth, G. L., K. H. Pollock, J. D. Nichols, T. R. Simons, J. E. Hines, and J. R. Sauer. 2002. A removal model for estimating detection probabilities from point-count surveys. Auk 119:414–425.

Fiske, I., and R. Chandler. 2011. Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. Journal of Statistical Software 43:1–23. http://www.jstatsoft.org/v43/i10/.

Hothem, R. L., B. E. Brussee, and W. E. Davis. 2010. Black-crowned Night-Heron (*Nycticorax nycticorax*). *In* A. Poole (editor). The Birds of North America Online. Ithaca: Cornell Lab of Ornithology. Available at: http://bna.birds.cornell.edu/bna/074. doi: 10.20173/bna.74

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Academic Press, Burlington, MA.

Marques, F. F. C., and S. T. Buckland. 2004. Covariate models for the detection function. Pp. 31–47 *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, UK.

McClintock, B. T., G. C. White, M. F. Antolin, and D. W. Tripp. 2009. Estimating abundance using mark–resight when sampling is with replacement or the number of marked individuals is unknown. Biometrics 65:237–246.

Nichols, J. D., J. E. Hines, J. R. Sauer, F. W. Fallon, J. E. Fallon, and P. J. Heglund. 2000. A double-observer approach for estimating detection probability and abundance from point counts. Auk 117:393–408.

Overton, W. S. 1969. Estimating the numbers of animals in wildlife populations. Pp. 403–455 *in* R. H. Giles (editor). Wildlife Management Techniques, 3rd edition. The Wildlife Society, Washington, DC.

Pratt, H. D., and I. L. Brisbin, Jr. 2002. Hawaiian Coot (*Fulica alai*). *In* A. Poole (editor). The Birds of North America Online. Ithaca: Cornell Lab of Ornithology. Available at: http://bna.birds.cornell.edu/bna/species/697b. doi:10.2173/bna.697b

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/.

Regier, H. A., and D. S. Robson. 1967. Estimating population number and mortality rates. Pp. 31–66 *in* S. D. Gerking (editor). The biological basis of freshwater fish production. Blackwell Scientific Publications, Oxford, UK.

Robinson, J. A., J. M. Reed, J. P. Skorupa, and L. W. Oring. 1999. Black-necked Stilt (*Himantopus mexicanus*). *In* A. Poole (editor). The Birds of North America Online. Ithaca: Cornell Lab of Ornithology. Available at: http://bna.birds.cornell.edu/bna/species/449. doi:10.2173/bna.449

Royle, J. A., J. D. Nichols, and M. Kéry. 2005. Modelling occurrence and abundance of species when detection is imperfect. Oikos 110:353–359.

Skalski, J. R., K. E. Ryding, and J. J. Millspaugh. 2005. Wildlife demography: analysis of sex, age, and count data. Elsevier Academic Press, Boston, MA.

Thomas, L., S. T. Buckland, E. A. Rexstad, J. L. Laake, S. Strindberg, S. L. Hadley, J. R. B. Bishop, T. A. Marques, and K. P. Burnham. 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. Journal of Applied Ecology 47:5–14.

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

Thompson, W. L. 2004. Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters. Island Press, Washington, DC.

Thompson, W. L., G. C. White, and C. Gowan. 1998. Monitoring vertebrate populations. Academic Press, San Diego, CA.

USFWS [U.S. Fish and Wildlife Service]. 2011. Recovery plan for Hawaiian waterbirds, second revision. U.S. Fish and Wildlife Service, Portland, OR.

White, G. C., and R. E. Bennetts. 1996. Analysis of frequency count data using the negative binomial distribution. Ecology 77:2549–2557.