Bret A. Collier

School of Renewable Natural Resources

Louisiana State University  
Baton Rouge, LA 70803

979/595/5076

[bret@lsu.edu](mailto:bret@lsu.edu)

RH: Heterogeneity and abundance estimation

**Understanding Heterogeneity in Wildlife Abundance Estimation**

Jeffrey L. Laake, Independent Researcher, Escondido, CA 92025, USA

Bret A. Collier, School of Renewable Natural Resources, Louisiana State University, Baton Rouge, LA 70803, USA

Correspondence: Bret A. Collier, School of Renewable Natural Resources, Louisiana State University, Baton Rouge, LA 70803. Email: bret@lsu.edu.

**Abstract**–Negative bias in mark-recapture abundance estimators due to heterogeneity in capture probability is a well-known problem. Various approaches for reducing said bias such as incorporating covariates have been developed and used extensively. However, unmodelled residual heterogeneity often remains because covariates causing heterogeneity cannot always be identified or measured properly. Heterogeneity is not as problematic for distance sampling and mark-resight methods because both techniques estimate capture probabilities with a known quantity. Distance sampling assumes detection probability at zero distance is perfect (and hence known) while mark-resight introduces a known number of marks into the population and through resighting the proportion of those detected is measured relative to the known number of marks. Herein, we describe how heterogeneity operates and leads to bias in mark-recapture abundance estimators. We show how introduction of a known number of planted and marked (e.g., with GPS or VHF tags) individuals can be used to remove residual heterogeneity. We provide a simulation example and a field based analysis of camera trapping data using a reintroduced population of wild turkeys of known size to show how negative bias can be addressed using planted known individuals.

**Keywords**: abundance, capture-recapture, distance sampling, heterogeneity, mark-recapture distance sampling, residual heterogeneity, planted

Understanding dynamics of wildlife populations often depends on the estimation of an unknown population size based on characteristics garnered from an observed sample. All abundance estimation methods rely on animal count statistics (*n*) which are used in conjunction with characteristics of the observation process to estimate abundance (*N*) by inferring what individuals were missed during the sampling (*N-n*). A suite of methodological approaches such as mark-recapture, distance sampling and mark-resight have been developed to provide information on the sample inclusion probability (e.g., detection, capture) used in estimating the proportion of animals that were not observed in the sample.

Going back to first principles of abundance estimation, consider the well-known Lincoln-Petersen estimator (Sebr 1982)for a two-sample mark-recapture survey. Animals captured or sighted in each sample are marked which can be detected or not detected in the subsequent sample. Let , be the counts in the first and second samples and let be the count of those detected in sample 2 that were also detected in sample 1. The detection probability estimator for each sample is and and the abundance estimate is which is the count divided by the sample inclusion probability which we will designate as detection probability hereafter.

The Lincoln-Petersen estimator assumes detection probability differs for each sample, but all animals have the same (homogeneous) detection probability within a sample. However, as homogeneity is rarely the case in field studies, it is well established that when detection probabilities are heterogeneous, abundance estimates will exhibit negative bias (Cormack 1966, Otis et al. 1978). Thus, an extensive suite of mark-recapture methods intended to provide ways to cope with heterogeneity have been developed, including techniques which incorporate observable covariates allowing heterogeneity in capture probabilities (Huggins 1989,1991), random effects following a specified distribution (Burnham and Overton 1978, Pledger and Efford 1998, and White and Cooch 2017), mixture distributions (Pledger 2000, Dorazio and Royle 2003), and a non-parametric jackknife estimator (Burnham and Overton 1979).

Distance sampling approaches, however, uses a different approach to estimating the sample inclusion probability and hence abundance. Under the assumption of random line placement relative to animal distribution, the expected number of observations at each perpendicular distance *x* from the line should follow a uniform distribution. All animals at 0 distance are assumed to be detected and conceptually the detection probability () at distance *x* is estimated by the decline in the number of detections relative to the number detected (at or near the line where *x=*0) where detection is assumed to be perfect () and thus the count is therefore known. Heterogeneity in detection probability is far less problematic for distance sampling based on the general property of pooling robustness (Buckland et al. 2001) which means estimators of overall abundance are robust to pooling over various conditions that affect detection probability.

However, under distance sampling, it is not always reasonable to assume that and composite mark-recapture and distance sampling (mrds) methods have been developed to estimate using double observers (platforms) (Butterworth et al. 1982, Butterworth and Borchers 1988, Hiby & Hammond 1989, Manly et al. 1996, Alpizar-Jara and Pollock 1996, Chen 2000). However, these early mark-recapture distance sampling based approaches assumed independence for all distances (full independence) when formulating the mark-recapture probability for the double observers and as such heterogeneity in detection probability introduced dependence and bias. Covariates can be included to reduce heterogeneity in detection but any remaining unmodelled (residual) heterogeneity would negatively bias the abundance estimator. Laake (1999) demonstrated that the independence assumption only needed to hold at *x*=0 (point independence) rather than for all *x* and in doing so, any residual heterogeneity at *x*>0 could be removed by using the decline in the number of detections at distance x, which is the fundamental basis of estimating detection probability in distance sampling. The early work by Laake (1999) was expanded for line transects (Laake and Borchers 2004, Borchers et al 2006, Laake et al 2008) and point transects (Laake et al. 2011) and in each case showed how residual heterogeneity in the observation process caused considerable underestimation of abundance.

As noted previously, mark-recapture has commonly relied on use of covariate or distributional assumptions to reduce the impacts of heterogeneity in the observation process. Contemporary work by Laake et al (2014) demonstrated how use of auxiliary marks could be used to accommodate dependence from residual heterogeneity in double tag loss when a subsample of permanently marked individuals are available. Further, Hennig et al. (2022) used the approach of Laake et al. (2014) to demonstrate how heterogeneity can be significantly reduced by including individuals with radio-collars who are located independently during double-observer aerial survey of wild burros. The approach by Hennig et al. (2022) accommodated heterogeneity via use of a Huggins model (Huggins 1989,1991) with a separate recapture probability which does not require a distributional assumption which may have problems with non-identifiability (Link 2003). As such, the approach outlined by Hennig et al. (2022) for accommodating heterogeneity was superior to the methods previously proposed by Griffin et al. (2013) which involved an additional covariate of marked status. The detection probability of marked groups is assumed to be lower than unmarked groups; thus, one can use the detection probability of the marked groups to estimate abundance of the unmarked groups to help account for residual heterogeneity.

The inclusion of known (planted) individuals into a population is not novel (Goudie 1995, Ashbridge and Goudie 2008) and has been used in human population sampling with the intent of using a single sample to adequately characterize population size (Laska et al. 1988). However, we have not seen any indication that those methods using known planted individuals have been used to accommodate heterogeneity. Mark-resight surveys (Bowden and Kufeld 1995, White 1996, McClintock et al. 2006) were created with the specific intent of estimating detection probability with the known marked sample during the resighting surveys and can accommodate residual heterogeneity from the known number of marked animals that were never detected. The approach used in Hennig et al. (2022) differs slightly from mark-resight surveys because unmarked animals in Hennig et al. (2022) have capture histories which are used in the detection probability estimation unlike mark-resight surveys that only use counts of unmarked animals without a capture history.

However, while the impact of heterogeneity is well-known, we do not believe that the literature has clearly detailed for non-statisticians the fundamental problem with heterogeneity and why residual heterogeneity induces dependence and bias for wildlife abundance estimation techniques using mark-recapture approaches. Thus, we will review and detail the fundamental issue of residual heterogeneity and outline the bias implications using a simple double-observer survey example. Then, we will demonstrate an approach using inclusion of a known component, wherein marked individuals are introduced into the population as in mark-resight surveys (McClintock et al. 2006) and show how this method can be used to remove some or all of the bias from heterogeneity irrespective of covariates. We also demonstrate the utility of using a known component to remove heterogeneity using data from a camera trap study with a known population size and a subset of telemetry marked Eastern wild turkeys (*Meleagris gallopavo silvestris).*

**Heterogeneity, Dependence and Bias**

A dictionary definition of heterogeneity is “the quality or state of consisting of dissimilar or diverse elements” and it is the antonym for homogeneity where everything is the same. In terms of wildlife abundance estimation, homogeneity would mean that all animals would have the same probability of being detected or captured and heterogeneity means that potentially you could have different probabilities ().

Independence is a simple concept that has important statistical implications with regard to heterogeneity. In the simplest terms, if two events are independent then the probability that both events occur is the product of the event probabilities. For example, with the two sample Lincoln-Petersen estimator the probability that an animal is detected in both samples is and the probability that animal is never detected is (1-)(1-). However, as detailed by Laake et al. (2011), two observers searching and detecting animals independently without cuing each other (e.g., Nichols et al. 2000) does not imply that those detections will be statistically independent. Any unmodelled (residual) heterogeneity in detection probability violates the assumption of statistical independence. In particular, the probability of observers either missing certain individuals or detecting certain individuals is much greater with residual heterogeneity.

Here we use a two-sample mark-recapture example which is relevant for mark-recapture (e.g., Lincoln-Petersen), double-observer sighting surveys and double tag loss approaches, to detail how heterogeneity induces dependence and the resultant impacts on abundance estimation. Following standard convention, individuals are given a 1 if they are detected and a 0 if they are not detected. There are 2 samples, so each detected individual can have an observation process represented by a pair of values, that is 10, 01, 11 if detected in first but not second, detected in second but not first and detected in both, respectively. This structure is typically defined as a capture or encounter history. We will denote to be the numbers observed in each capture history and denote as to be the number which are not observed and is what we want to estimate. In terms of the notation typically used for Lincoln-Petersen , , and .The sampling can be expressed as contingency table with the first sample (observer) as the rows and the second sample (observer) as the columns with values 0 (missed) or 1 (detected) for the entries (Table 1).

We simulated data under both homogeneity and heterogeneity scenarios to demonstrate the differences. For the homogeneity example, the probability for the first observer () is 0.4 and for the second observer the probability is 0.6 and we held each probability constant for the 1,000 simulated animals. For the heterogeneity example, we generated 1,000 normal random variables () with a mean of 0 and standard deviation of 2. The capture probability of the simulated animal for the first observer is } and for the second observer }. When (the mean value) the probabilities are equal to the homogeneity example. Most (95%) of the probabilities ranged from 0.029 to 0.987 for the simulated animals. Essentially, the homogeneity scenario assumes that the two observers each have their own inherent detection probabilities that remain constant regardless of the unique properties associated with each animal. The heterogeneity scenario assumes that the detection probabilities for each animal are the product of individual observer acuity and the properties affecting detection of each animal (i.e., group size, light level, etc.).

First, note the pattern of the observations (Table 2). As expected, there are more values than values because the detection probability is 50% greater for observer 2. But the most important comparison is between the homogeneity and heterogeneity values for the and values. Under homogeneity, only 47% are either missed by both or detected by both ( or ); whereas, under heterogeneity it is 69%. This result is typical when heterogeneity in detection exists because animals with low probabilities are typically missed by both observers () and animals with high probabilities are typically detected by both observers (. Hence, the detection probabilities for each observer are not independent – they are both influenced by the same factors. A chi-square test for independence between the outcomes of observer 1 and 2 is non-significant for the homogeneity example (=0, = 1) and significant for dependence with the heterogeneity example (=159.05, < 0.001).

We estimate by using the probability structure that assumes outcomes for the two observers are independent (Table 2). We estimate and ) as the proportion of animals detected by observer 1 of the total that observer 2 detected and vice versa. Herein lies the specific problem with heterogeneity because, by definition, the sample used to estimate the observation process will disproportionately include the most detectable animals. Thus, under the independence assumption, the probability that one or the other observer detected the animal is or and the estimate of abundance is where .

While conceptually simple, the implications for violations of the homogeneity assumption are tremendous. Consider the outcome from the homogeneous simulation wherein =752, =0.751 and an estimate of =1001.6 which is close to the true value of 1000. However, for the heterogeneous simulation, =657, =0.918 and the estimate of =715.5 which is biased low as is biased high from heterogeneity and the induced dependence it causes.

**Incorporating covariates to address heterogeneity**

If you can observe and measure the covariates that introduce heterogeneity into the observation process, then it is possible to reduce bias from heterogeneity by including the covariate(s) into the probability model. This approach underlies the work of Huggins et al. (1989,1991) for mark-recapture and for the sightability model approach to aerial surveys (Steinhorst and Samuel 1989). To demonstrate how covariates can be used to remove heterogeneity, we first fitted Huggins (1989, 1991) models implemented in the MARK software (White and Burnham 1999) with the RMark interface (Laake 2013) to MARK in the R software (R Core Team 2022). Using RMark/MARK (see online appendix) with separate observer probabilities for both the homogeneity and the heterogeneity simulated data detailed earlier, as expected the results matched those computed previously (Homogeneity: 1,001.6 [SE = 33.3] and Heterogeneity: 715.51 [SE = 11]). Next, we fitted a model with the observed covariate values that were used to generate heterogeneity in the simulated data. The resulting estimate (1,129.1, SE = 145) was closer to the true value and the 95% confidence interval (918.67, 1,508.8) included the true value of N = 1,000. However, the penalty for including the covariate was increased uncertainty (wider confidence intervals), which is an appropriate tradeoff as while the model without the covariate was more precise, it was also precisely wrong.

Even after incorporating covariates, residual (unmodelled) heterogeneity is likely to remain because we cannot identify nor appropriately model all conditions that create heterogeneity (Borchers et al 2006, Hennig et al 2022). Thus, it is important to consider ways to remove as much of the remaining residual heterogeneity as possible. For mark-recapture distance sampling that was accomplished by assuming point independence (and possibly limiting independence; Buckland et al. 2010 in some cases). However, for mark-recapture surveys focused on abundance estimation, other approaches for addressing residual heterogeneity are necessary. We suggest that the most obvious approach is to incorporate a known number of marked individuals in the population (McClintock et al. 2006, Hennig et al. 2022).

**Incorporating known individuals**

With a known planted sample of individuals with some type of location device (GPS, satellite) that can be located during a survey, researchers can assess the number of marked animals that were never captured (observed). In our double observer example, that would be equivalent to having information for a sample of animals which would then provide the opportunity to model dependence induced from heterogeneity.

With the from a sample of known animals in the population, then we can use the dependence probability structure (Table 3). From the complete table (including ), the probabilities can be computed as and where M is the number of known animals planted in the population. Under homogeneity the outcomes for the 2 observers are independent and and have the same expected value, but with heterogeneity, we expect and (Table 4).

MARK will not accept a capture history with all 0 values, so to incorporate the information for the sample of known animals we expanded the capture history with an added occasion at the end which has the value 1 for the sample of marked animals and the value 0 for any unmarked animals. Thus, for marked animals, the possible capture histories for a two sample survey are 001,011,101,111 and for unmarked animals are 010,100,110. The basic probability structure for a two-sample Huggins model with the expanded capture history using MARK notation is shown in Table 5. The capture probability for the third final occasion () is fixed to 1 for marked animals and 0 for unmarked animals, so it is never estimated and never enters the likelihood. The final occasion value simply allows the inclusion of a 00 capture history for marked animals who were never detected but were known to be present during the sampling process. Heterogeneity is incorporated by having a recapture probability () for the second observer that is different than the initial capture probability. By including a separate recapture probability (the definition for in the MARK notation becomes the conditional probability that an animal was detected on the second occasion given that it was missed on the first occasion (). It is fundamentally different than which is the unconditional probability of detection on the first occasion. Implementation of the Huggins model in MARK via RMark is described in the online electronic appendix. We note our notation is simplified from Hennig et al. (2022) because we have added the occasion at the end whereas Hennig et al (2022) added it to the beginning.

As an example, we randomly selected and planted (known) 50 animals of the 1000 in the heterogeneity simulation example above. We fitted the model using the structure in Table 5 and estimated =0.4564 (SE=0.046), =0.4113 (SE=0.07464), and =0.7868 (SE=0.0195) which are close to the values p1, p2|0, and c2 in Table 4. The estimate of abundance was = 966 (SE=91.3). The precision and ability to account for the heterogeneity will in part depend on the sample size of marked animals. To demonstrate, next we randomly selected and planted (known) 25 animals of the 1000 in the heterogeneity simulation example. We fitted the same model structure and got estimates of =0.4833 (SE=0.0617), =0.4581 (SE=0.1114), and =0.7868 (SE=0.0195). The estimate of abundance was = 912 (SE=112). As expected, the standard errors for the parameters increased.

**Turkey camera trap survey example**

Buckley et al. (in press) reported on a camera trapping experiment for 80 Eastern wild turkeys released on a wildlife management area in a region where no wild turkeys were present. All individuals in Buckley et al. (in press) were banded with an aluminum rivet bands on the left leg and up to 3 additional colored aluminum rivet bands in unique color combinations, totaling no more than 2 bands per leg. Additionally, Buckley et al. (in press) deployed 43 global positioning system (GPS/very high frequency (VHF); MiniTrack Backpack GPS units, Biotrack Ltd., Wareham, Dorset, UK or MicroGPS/VHF backpack, Sirtrack Wildlife Tracking Solutions, Havelock North, New Zealand) and 8 VHF (Advanced Telemetry Systems, Isanti, MN) backpack units on ~45% of released wild turkeys to monitor survival and movement ecology. At the time of camera survey data collection, 26 of the wild turkeys had a working GPS unit or VHF tag that enabled the wild turkeys to be located on the wildlife management area during the 30 day study capture period. The capture period was broken down into three 10 day periods, five 6 day periods and ten 3 day periods to construct capture histories for mark-recapture abundance estimates with various estimators including Huggins (1989,1991). Here we use the data from Buckley et al. (in press) to demonstrate that inclusion of the 26 known (GPS/VHF) wild turkeys into the estimation procedure can remove heterogeneity and provide an estimate that is closer to the true population size than other methods used.

Table 6 demonstrates the importance of the known individuals in reducing heterogeneity. Only 50% of the known animals were detected and this information is included in the estimation with the 000 observations. Likewise, 44.4% (24/54) of the unknown turkeys were never detected but we obviously can’t include those never detected (000). In the standard mark-recapture analysis in Buckley et al (in press), 50 of the 80 animals were never observed on the camera traps and the excess 000 observations due to heterogeneity could not be accounted correctly and the resultant abundance estimates were all much lower than the known population size of 80 wild turkeys.

In implementing our analysis, the capture histories were expanded to 4, 6, and 11 occasions with the final dummy occasion separating known (planted) wild turkeys and unknown wild turkeys as in the two-sample example above. The data and model fitting code are provided in the online appendix. We fitted the following 6 models with Huggins (1989,1991) model in MARK: 1) constant capture probability (~1), 2) time dependent capture probability (~time), 3) time dependence for initial capture probability and an additive recapture effect (~time+c), 4) separate time dependence in initial capture probability and recapture probability (~time+c:time) 5) time dependence plus an additive difference for known wild turkeys (type) as described in Griffin et al. (2013) (~time+type), 6) same as model 5 but with interaction of time and type (~time\*type). Inclusion of the recapture probability (c) accommodates the residual heterogeneity because the recapture probability () is greater than the initial capture probabilities ( and ). Models 3 and 4 are labelled as MR and model 5 as MH in Hennig et al (2022). Models 1 and 2 are similar but not the same as models fitted in Buckley et al. (in press) and the estimates are higher because the 26 known birds are included in the estimate without error. We expected models 3 and 4 to be the best models based on the experience of Hennig et al. (2022) and our understanding of how heterogeneity operates as explained previously.

For the 3-occasion capture history, the 95% confidence intervals only included the known value for models 3 and 4 (Figure 1). For each of the 3,5,10 capture occasion analyses (Tables 7-9) models 3 and 4 were clearly the best models with 95, 81, and 100 percent of the model weight respectively. We computed model averaged estimates of abundance (Table 10), the standard error and 95% log-normal confidence intervals for the 3, 5 and 10 occasion analysis. For 5 occasions the estimate is lower because of the weight assigned to models 5 and 6 which will tend to underestimate abundance.

DISCUSSION

Addressing heterogeneity in the detection process of wildlife abundance estimation has and remains an important for furthering our understanding of terrestrial and aquatic populations. As such, methodological advances in wildlife abundance estimation have focused on increasing prediction accuracy or reducing estimated biased relative to previous methods. However, underlying many advances in estimation methods are assumptions regarding independence, and how to address the lack of independence, within the detection process. As such, the impetus for our work herein was to further understanding of the impacts of residual heterogeneity on mark-recapture based abundance estimation methods. Mark-recapture analysis, widely applied in various forms, can be impacted by violations of statistical dependence (e.g., heterogeneity) in the detection process in situations where the sampling process is independent between observers or sampling periods (Laake et al. 2011). Consider, for example, double observer point count surveys as outlined by Nichols et al. (2000). Double observer (and similar) methods are ubiquitous in avian and terrestrial ecology, widely implemented, and are considered to yield precise estimates of abundance. However, as detailed herein, statistical dependence in the sampling process (e.g., residual heterogeneity) will induce positive covariance in the detection process and hence negatively bias avian point count survey data. However, methodological development of double observer methods (Nichols et al. 2000, Nichols et al. 2009) has been lax in addressing residual heterogeneity issues, even when the implications on abundance estimation are known (Laake et al. 2011).

As we have detailed, residual heterogeneity in mark-recapture methods will negatively bias the abundance estimator because the sample is composed only by detected individuals. This level of unresolved heterogeneity led Buckley et al. (in press) to conclude that a suite of mark-recapture approaches applied to camera trapping data produced biased estimates of abundance for a known population of wild turkeys. However, by using the wild turkeys with GPS and VHF transmitters as a known (planted) part of the study population, we demonstrated that abundance estimates could be produced that were closer to the known population size, as did Hennig et al. (2022) for wild burro surveys. Similar to our wild turkey abundance estimation example outlined above, Hennig et al. (2022) demonstrated a reduction in AIC by incorporating a different recapture probability in comparison to the mark-type model of Griffin et al. (2014). Griffin et al. (2014) proposed that the known planted animals with transmitters would have a lower probability of detection because they were a random sample of detected and undetected individuals or groups, whereas the unmarked animals were only included in the sample if they were seen by one or more observers (i.e., a biased sample). This mark‐type model may be warranted in populations where groups have zero chance of being detected by human observers, but it doesn’t address the fundamental impact of heterogeneity which is to increase the proportion of animals that are either not detected by either observer or detected by both observers. It is theoretically possible to develop a joint model which combines both approaches, but it is likely to have problems with parameter identifiability and would likely require an impractical number of known animals.

The concept of using planted animals we describe is very similar to mark-resight (Bowden and Kufeld 1995, White 1996, McClintock et al. 2006) except that in the applications we describe, a capture history for the unmarked (unknown) portion of the population is constructed and used in the estimation of detection probability. This was done in Hennig et al. (2022) because the capture history has 2 occasions (observers) and is recorded during the survey. The same approach could be used in a double-observer mark-recapture distance sampling survey (Laake et al. 2011) and was possible with the wild turkey example of Buckley et al (in press) because every wild turkey was carrying a unique visual tag for the camera trapping study. However, when applying our approach on a larger scale with unmarked animals that are not uniquely identifiable (e.g., natural marks), it will likely not be possible to construct a capture history for unknown animals. However, as a sample of animals could be tagged with less costly visual tags while others are fitted with transmitters, a future simulation study is needed to evaluate the information gained from incorporating the capture-history information from unknown animals versus the typical mark-resight survey. Such a study could also evaluate a broad range of population parameters and sample size of marked animals and the expected precision that could be achieved.

If covariates are going to be included in models for detection probability, it will be necessary to measure the covariates for all of the known planted animals which means those that are not detected have to be located to measure any relevant covariates not collected at capture. Covariates such as landscape-type metrics that can be remotely acquired from the animals known would be one example. However, for animals in groups as with the wild burro surveys in Hennig et al (2022), group size was an important covariate for visual detection and requires locating each group with a planted individual that was missed during the survey to assess the group size of the missed group.

Finally, we note that within wildlife and fisheries abundance estimation there are a variety of cases where use of known (planted) individuals could be used to address residual heterogeneity and reduce bias. , which may have unknown implications for trend detection and management thresholds for monitoring programs that attempt to address residual heterogeneity in the futureAdditionally, we acknowledge that the costs associated with the use of planted individuals includes not only increased uncertainty but also expenses associate with handling, tagging, and monitoring individuals. For our approach to be applicable, planted individuals need to be representative of the population as a whole which could be more difficult to accommodate in scaling up this approach to a regional or state-wide usage.

**Acknowledgements**

Funding and support for the wild turkey camera surveys were provided by the Texas Parks and Wildlife Department Upland Game Bird Stamp Fund, the U.S. Fish and Wildlife Service through the Federal Aid in Wildlife Restoration Act Grant W-1830-R-1, the Louisiana State University Agricultural Center, the School of Renewable Natural Resources at Louisiana State University. This manuscript is based on work supported by the National Institute of Food and Agriculture, United States Department of Agriculture, McIntire Stennis project number 7001494. We thank Jake Hennig for his review of a draft which led to some nice clarification to the manuscript.

Literature Cited

Alpizar-Jara, R. & K.H. Pollock 1996. A combination line transect and capture-recapture sampling model for multiple observers in aerial surveys. Env. Ecol. Stat. 3:311-327.

Ashbridge, J. and I. B. J.Goudie. 2008. Conditionally Unbiased Estimation of Population Size Under Plant-Capture, Communications in Statistics - Theory and Methods, 38:1, 1-12.

Borchers, D. L., J. L. Laake, C. Southwell, and C. G. M. Paxton. 2006. Accommodating unmodeled heterogeneity in double‐observer distance sampling surveys. Biometrics 62:372–378.

Bowden, D. C., and Kufeld, R. C. 1995. Generalized Mark-Sight Population Size Estimation Applied to Colorado Moose. Journal of Wildlife Management, 59: 840–851.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. 2001. Introduction to Distance Sampling: Estimating Abundance of Biological Populations, Oxford: Oxford University Press.

Buckland, S. T., J. L. Laake, and D. L. Borchers. 2010. Double‐observer line transect methods: levels of independence. Biometrics 66:169–177.

Burnham KP, and W.S. Overton. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika 65:625–633.

Burnham KP, and W.S. Overton. 1979. Robust estimation of population size when capture probabilities vary among animals. Biometrika 65:625–633.

Butterworth, D. S., and D.L. Borchers. 1988) Estimates of g(0) for minke schools from the results of the independent observer experiment on the 1985/86 and 1986/87 IWC/IDCR Antarctic assessment cruises. Report of the International Whaling Commission 38, 301–313.

Butterworth, D.S., P.B. Best and M. Basson 1982. Results of analysis of minke whale assessment cruise, 1980/1. Rep. Int. Whaling Comm. 32:819-834.

Chen, S.X. 2000. Animal abundance estimation in independent observer line transect surveys. Envir. Ecol. Stat. 7:285-299

Cormack, R.M. 1966. A Test for Equal Catchability. Biometrics 22(2):330-342.

Dorazio, R. M., and J. A. Royle. 2003. Mixture models for estimating the size of a closed population when capture rates vary among individuals. Biometrics 59: 351– 364.

Griffin, P. C., B. C. Lubow, K. J. Jenkins, D. J. Vales, B. J. Moeller, M. Reid, P. J. Happe, S. M. McCorquodale, M. J. Tirhi, J. P. Schaberl, and K. Beirne. 2013. A hybrid double‐observer sightability model for aerial surveys. Journal of Wildlife Management 77:1532–1544.

Goudie, I. 1995. A plant-capture approach for achieving complete coverage of a population. Communications in Statistics - Theory and Methods 24, 1293–1305.

Hennig, J,D., K.A. Schoenecker, J.W. Cain III, G.W. Roemer, and J.L. Laake. 2022. Accounting for residual heterogeneity in double‐observer sightability models decreases bias in burro abundance estimates. Journal of Wildlife Management 86:5 DOI: 10.1002/jwmg.22239.

Hiby, A.R. and P.S. Hammond 1989. Survey techniques for estimating abundance of cetaceans. Rep. Int. Whaling Comm. Special Issue 11:47-80.

Huggins, R. M. 1989. On the statistical analysis of capture‐recapture experiments. Biometrika 76:133–140.

Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. Biometrics 47: 725–732.

Laake, J. L. 1999. Distance sampling with independent observers: reducing bias from heterogeneity by weakening the conditional independence assumption. In ‘Marine Mammal Survey and Assessment Methods.’ (Eds G. Amstrup, S. Garner, J. Laake, B. Manly, L. McDonald and D. Robertson.) pp. 137–148. (Balkema: Rotterdam.)

Laake, J. L. 2013. RMark: an R interface for analysis of capture‐recapture data with MARK. AFSC Processed Rep 2013‐01.Alaska Fisheries Science Center, National Oceanic Atmospheric Administration, National Marine Fisheries Service, Seattle, Washington, USA.

Laake, J. L., M. J. Dawson, J. and Hone, J. 2008. Visibility bias: mark–recapture, line-transect, or both? Wildlife Research 35, 299–309.

Laake, J. L., and D. L. Borchers. 2004. Methods for incomplete detection at distance zero. Pages 108–189 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, United Kingdom.

Laake, J. L., D. S. Johnson, D. R. Diefenbach, and M. A. Ternent. 2014. Hidden Markov model for dependent mark loss and survival estimation. Journal of Agricultural, Biological, and Environmental Statistics 19:522–538.

Laake, J.L., B. A. Collier, M. L. Morrison & R. N. Wilkins, 2011. Point-Based Mark-Recapture Distance Sampling. Journal of Agricultural, Biological, and Environmental Statistics. 16(3), 389–408.

Laska, E. M., Meisner, M. & Siegel, C. (1988), ‘Estimating the size of a population from a single sample’, Biometrics 44, 461–472.

Link, W. A. 2003. Nonidentifiability of Population Size from Capture–Recapture Data with Heterogeneous Detection Probabilities. Biometrics 59: 1123– 30.

Manly, B.F.J., L.L. McDonald & G.W. Garner 1996. Maximum likelihood estimation for the double-count method with independent observers. J. Agr. Bio. Env. Stat. 1:170-189.

McClintock, B.T., G.C. White & K.P. Burnham. 2006. A robust design mark-resight abundance estimator allowing heterogeneity in resighting probabilities. JABES 11, 231. <https://doi.org/10.1198/108571106X129171>

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. The Auk 117: 393–408.

Nichols, J. D., L. Thomas and P. B. Conn. 2009. Inferences about landbird abundance from count data: recent advances and future directions. Pages 201-235 in E. G. Cooch et al., editors. Modeling Demographic Processes in Marked Populations, Springer.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62:3–135.

Pledger, S., and M. Efford. 1998. Correction of bias due to heterogeneous capture probability in capture-recapture studies of open populations. Biometrics, pp.888-898.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture-recapture models for mixtures. Biometrics 56: 434– 442.

R Core Team 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Seber, G. A. 1982. The estimation of animal abundance and related parameters. Griffin, London, UK.

Steinhorst, R. K., and M. D. Samuel. 1989. Sightability adjustment methods for aerial surveys of wildlife populations. Biometrics 45:415–425

White, G. C . 1996. NOREMARK: Population Estimation from Mark-Resighting Surveys. Wildlife Society Bulletin, 24: 50–52.

White GC, and E.G. Cooch. 2017. Population abundance estimation with heterogeneous encounter probabilities using numerical integration: individual heterogeneity and abundance estimation. J Wildl Manag 81:322–336

White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46:120–138.

| Table 1. Example contingency table showing the distribution of detected and missed observations and their associated cell count. | | |
| --- | --- | --- |
|  | Missed | Detected |
| **Missed** |  |  |
| **Detected** |  |  |

| Table 2. Example simulation realization of the homogeneity and heterogeneity conditions and the relative impact on detection frequency for a simple 2 sample mark-recapture survey. | | | | |
| --- | --- | --- | --- | --- |
|  | Homogeneity | | Heterogeneity | |
|  | Observer 2 | | Observer 2 | |
| Observer 1 | Missed | Detected | Missed | Detected |
| Missed | 248 | 376 | 343 | 216 |
| Detected | 150 | 226 | 94 | 347 |

| Table 3. Probability structure for the encounter histories of the complete table under independence and dependence due to heterogeneity. is the conditional probability that observer 2 detects an animal that observer 1 missed and is the conditional probability that observer 2 detects an animal that observer 1 also detected. | | | |
| --- | --- | --- | --- |
| Independence |  | Missed | Detected |
|  | **Missed** |  |  |
|  | **Detected** |  |  |
| **Dependence** |  |  |  |
|  | **Missed** |  |  |
|  | **Detected** |  |  |

| Table 4. Detection probabilities for homogeneity and heterogeneity simulation realizations in Table 2. | | |
| --- | --- | --- |
|  | Homogeneity | Heterogeneity |
|  | 0.38 | 0.44 |
|  | 0.60 | 0.56 |
|  | 0.60 | 0.39 |
|  | 0.60 | 0.79 |

|  |  |
| --- | --- |
| Table 5. Probabilities for each capture history of known (planted) and unknown animals with a two-sample (double observer) structure and an additional third dummy occasion to incorporate known marked animals. | |
| Marked Animals |  |
| 001 |  |
| 011 |  |
| 101 |  |
| 111 |  |
| Unmarked Animals |  |
| 010 |  |
| 100 |  |
| 110 |  |

| Table 6. Capture histories of known and unknown turkeys for three 10-day occasions from Buckley et al. (in press). Only known (planted) turkeys can have a “000” capture history. When they are analyzed the capture history for marked turkeys that were missed would be 0001 because 1 is appended at the end for all known turkeys. All unknown turkey capture histories would end with 0. | | |
| --- | --- | --- |
|  | known | unknown |
| **000** | 13 | 0 |
| **001** | 1 | 7 |
| **010** | 2 | 2 |
| **011** | 0 | 3 |
| **100** | 4 | 1 |
| **101** | 1 | 4 |
| **110** | 3 | 6 |
| **111** | 2 | 7 |

| Table 7. Model selection table from 3 occasions of 10 days each with wild turkey data. | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Model | # par | AICc | DeltaAICc | Weight | Deviance |
| **3** | p(~time + c)c() | 4 | 207.1 | 0 | 0.5403 | 283.6 |
| **4** | p(~time + c:time)c() | 5 | 207.7 | 0.5353 | 0.4134 | 282.1 |
| **5** | p(~time + type)c() | 4 | 213.3 | 6.178 | 0.0246 | 289.8 |
| **6** | p(~time \* type)c() | 6 | 213.6 | 6.461 | 0.02136 | 285.9 |
| **1** | p(~1)c() | 1 | 222.3 | 15.16 | 0.0002755 | 305 |
| **2** | p(~time)c() | 3 | 226 | 18.87 | 4.325e-05 | 304.6 |

| Table 8. Model selection table from 5 occasions of 6 days each with wild turkey data. | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Model | # par | AICc | DeltaAICc | Weight | Deviance |
| **4** | p(~time + c:time)c() | 9 | 333.8 | 0 | 0.5316 | 329.8 |
| **3** | p(~time + c)c() | 6 | 335 | 1.273 | 0.2813 | 337.4 |
| **5** | p(~time + type)c() | 6 | 337.1 | 3.344 | 0.09989 | 339.5 |
| **6** | p(~time \* type)c() | 10 | 337.4 | 3.625 | 0.08679 | 331.3 |
| **1** | p(~1)c() | 1 | 348.3 | 14.54 | 0.0003699 | 360.9 |
| **2** | p(~time)c() | 5 | 355.8 | 22 | 8.871e-06 | 360.2 |

| Table 9. Model selection table from 10 occasions of 3 days each with wild turkey data. | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Model | # par | AICc | DeltaAICc | Weight | Deviance |
| **3** | p(~time + c)c() | 11 | 540 | 0 | 0.7769 | 508 |
| **4** | p(~time + c:time)c() | 19 | 542.5 | 2.495 | 0.2231 | 493.6 |
| **5** | p(~time + type)c() | 11 | 561.7 | 21.66 | 1.538e-05 | 529.6 |
| **6** | p(~time \* type)c() | 20 | 567.8 | 27.72 | 7.415e-07 | 516.7 |
| **1** | p(~1)c() | 1 | 588.6 | 48.59 | 2.185e-11 | 577 |
| **2** | p(~time)c() | 10 | 589.9 | 49.89 | 1.14e-11 | 559.9 |

| Table 10. Model averaged abundance estimates for 3,5,10 occasions with wild turkey data. | | | | |
| --- | --- | --- | --- | --- |
| No. of occasions |  | Std. Error | 95% log-normal confidence interval | |
| 3 | 85.99 | 14.30 | 68.35 | 128.8 |
| 5 | 83.48 | 14.33 | 66.51 | 127.8 |
| 10 | 87.49 | 14.52 | 69.31 | 130.5 |

List of Figures:

Figure 1. Abundance estimates and 95% confidence intervals for models No. 1–6 with known animals included for 3 occasion capture history of turkeys from camera traps.

Chart, box and whisker chart

Description automatically generated