ORIGINAL ARTICLE

Application of a hybrid model to reduce bias and improve precision in population estimates for elk (*Cervus elaphus*) inhabiting a cold desert ecosystem

Kathryn A. Schoenecker [a],*, Bruce C. Lubow [b]

[a] U.S. Geological Survey, Fort Collins Science Center, and Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO 80523, USA
[b] Natural Resources Ecology Lab, Colorado State University, Fort Collins, CO 80523, USA

Abstract Accurately estimating the size of wildlife populations is critical to wildlife management and conservation of species. Raw counts or "minimum counts" are still used as a basis for wildlife management decisions. Uncorrected raw counts are not only negatively biased due to failure to account for undetected animals, but also provide no estimate of precision on which to judge the utility of counts. We applied a hybrid population estimation technique that combined sightability modeling, radio collar-based mark-resight, and simultaneous double count (double-observer) modeling to estimate the population size of elk in a high elevation desert ecosystem. Combining several models maximizes the strengths of each individual model while minimizing their singular weaknesses. We collected data with aerial helicopter surveys of the elk population in the San Luis Valley and adjacent mountains in Colorado State, USA in 2005 and 2007. We present estimates from 7 alternative analyses: 3 based on different methods for obtaining a raw count and 4 based on different statistical models to correct for sighting probability bias. The most reliable of these approaches is a hybrid double-observer sightability model (model M_H), which uses detection patterns of 2 independent observers in a helicopter plus telemetry-based detections of radio collared elk groups. Data were fit to customized mark-resight models with individual sighting covariates. Error estimates were obtained by a bootstrapping procedure. The hybrid method was an improvement over commonly used alternatives, with improved precision compared to sightability modeling and reduced bias compared to double-observer modeling. The resulting population estimate corrected for multiple

* Corresponding author at: USGS, Fort Collins Science Center, 2150 Centre Avenue, Building C, Fort Collins, CO 80526-8118, USA. Tel.: +1 970/226 9329; fax: +1 970/226 9230.
E-mail addresses: schoeneckerk@usgs.gov (K.A. Schoenecker), Bruce_L@Comcast.net (B.C. Lubow).

sources of undercount bias that, if left uncorrected, would have underestimated the true population size by as much as 22.9%. Our comparison of these alternative methods demonstrates how various components of our method contribute to improving the final estimate and demonstrates why each is necessary.

1. Introduction

Desert ecosystems cover 20–30% of the terrestrial surface of the earth (Hadley and Szarek, 1981) and another 25% is semi-desert (Pearson, 1965). These ecosystems support a myriad of wildlife and bird species, and are among the biomes with the highest number of threatened species globally (Baillie et al., 2004). Desert ecosystems contain high endemism and are at risk from the effects of global climate change (Baillie et al., 2004). Conservation of species in arid ecosystems is relevant to > 50% of the global terrestrial landscape, so developing better tools to support the conservation and management of desert species is essential.

One of the most elementary needs for resource managers and conservationists is obtaining accurate population estimates. Understanding population size, growth rates and fluctuations in natural populations is crucial for management, and provides the ability to make predictions and provide a solid foundation for conservation actions. Population estimates are vulnerable to numerous sources of bias, usually resulting in undercounting (Samuel and Pollock, 1981; McCorquodale, 2001).This is particularly applicable to species found in groups, such as ungulates, in which detection of groups during surveys can vary grossly based on the size of the group and other factors that influence the likelihood of their detection (sighting variables). Applying group-specific correction factors to observed animal groups can compensate for effects of detection bias when estimating abundance and composition. There are several methods for estimating group-specific detection probabilities in aerial surveys, including sightability models (Steinhorst and Samuel, 1989), distance models (Burnham and Anderson, 1984; Buckland et al., 2004), mark-resight models (White, 1996; Skalski et al., 2005), double-observer models (Graham and Bell, 1989), and methods that combine two or more such techniques often referred to as "hybrid" models (Quang and Becker, 1997; Buckland et al., 2010; Burt et al., 2014). These techniques all have individual strengths and weaknesses (Griffin et al., 2013). By combining several techniques to form a hybrid model, the biases of each individual technique can be minimized, while strengthening their collective precision (Griffin et al., 2013).

We conducted population estimation research on elk (Cervus elaphus) inhabiting a high elevation desert ecosystem in the southern Rocky Mountains, USA. Elk are native ungulates of North America and central Asia with a polygonous mating strategy, in which females remain in groups year round and are joined by males during the breeding season or "rut". Males form independent groups or remain solitary during the non-breeding season. The elk population we studied is free-roaming across a combination of state-, federal-, and privately-owned lands, each with very different management practices. To better manage the herd, an evaluation of the population on its entire range and across socio-political boundaries was needed to determine the whole-herd population size and support a strategy for cross-boundary management of elk.

We surveyed elk via helicopter and applied a hybrid population estimation model that combined 3 commonly used methods: (1) simultaneous double-observer (Cook and Jacobson, 1979; Seber, 1982; Pollock and Kendall, 1987; Bayliss and Yeomans, 1989; Graham and Bell, 1989; Rivest et al., 1995; Manley et al., 1996), (2) radiocollar mark-resight (Seber, 1982; Pollock et al., 1990; White and Burnham, 1999; McClintock et al., 2009), and (3) sightability correction models (Samuel et al., 1987; Steinhorst and Samuel, 1989). Estimating uncertainty for populations in which animals live in groups is problematic using available analytical formulae and software. Therefore we applied a numerical bootstrap algorithm following Wong (1996). The objective of our study was to estimate the elk population size and to analyze and present approaches for dealing with multiple sources of bias by evaluating the relative contribution of each bias correction to the final population estimate.

2. Study area

We conducted the present study in the San Luis Valley of Colorado, USA (37.7329°N, 105.5121W). The San Luis Valley is a high elevation arid ecosystem, often called a "cold desert." The valley encompasses Great Sand Dunes National Park and Preserve, containing the tallest active sand dunes in North America (Fig. 1). Precipitation averages 28 cm/year and falls mostly during monsoonal rains in July–September. Summers are warm and average daytime temperatures range from 26.5 °C to 29.5 °C. Winters are cold and dry with average valley daytime temperatures ranging from −9.5 °C to 1.5 °C. Elevations on the valley floor range between 2285 m and 2440 m, and up to 3570 m at higher elevations where elk over-summer. The elk population range is approximately 2500 km^2 on the east side of the San Luis Valley, and includes federally owned and protected lands within Great Sand Dunes National Park and Preserve, the Baca National Wildlife Refuge, and adjacent privately-owned lands. The study area extended from Great Sand Dunes west to Highway CO-17, south to Blanca Peak, and north to Poncha Pass including the western slopes of the Sangre de Cristo Mountains (Fig. 1).

Adjacent to the active dune field extend sandy plains of sand sheet and 'sabkha'or salt flats, which are stabilized by herbaceous vegetation and shrubs. Vegetation is dominated by greasewood (Sarcobatus vermiculatus)/rabbitbrush (Chrysothamnus nauseosus) shrub lands, with riparian corridors containing willow (Salix spp.) and narrow leaf cottonwood (Populus angustifolia). Dominant graminoids include arctic rush (Juncus balticus), needle and thread grass (Hesperostipa comata), Indian rice grass (Achnatherum
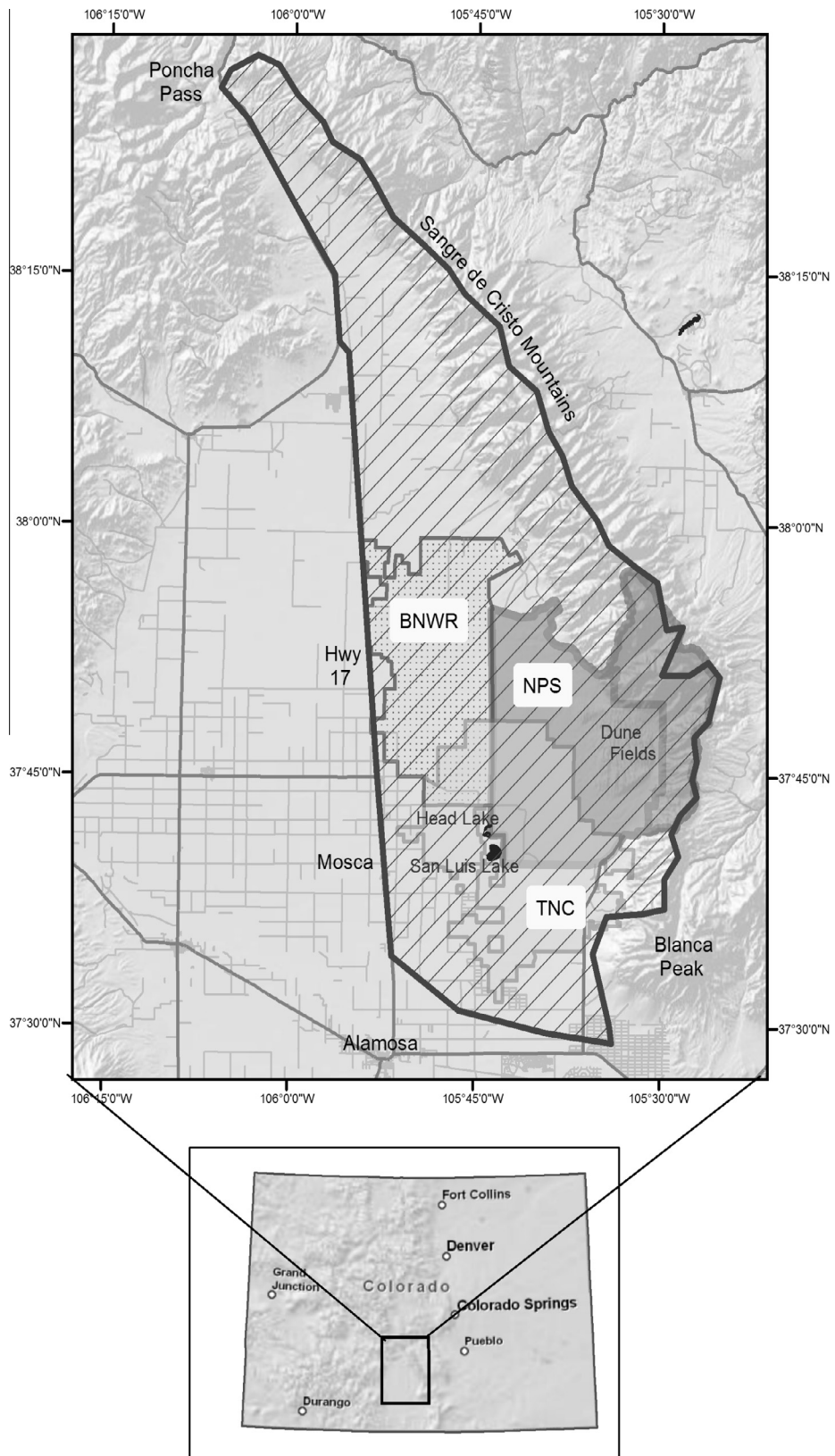
Figure 1 Elk population survey area (hatched) on the east side of the San Luis Valley, a high elevation desert in southern Colorado, USA. The insert box shows location of the study area within Colorado. Federally-protected lands include the Baca National Wildlife Refuge (BNWR – dotted), and National Park Service (NPS; Great Sand Dunes National Park – dark gray). The Nature Conservancy's (TNC) privately-owned lands are shown in light gray.

hymenoides), sand drop seed (*Sporobolus cryptandrus*), salt-grass (*Distichlis spicata*), slender wheatgrass (*Elymus trachy-caulus*), and beardless wild rye (*Leymus triticoides*).

The Sangre de Cristo Mountains are characterized by lodge pole pine (*Pinus contorta*) and quaking aspen (*Populus tremuloides*) on mid-elevation slopes, Engelmann spruce (*Picea engelmannii*) and subalpine fir (*Abies lasiocarpa*) dominating upper slopes, and bristlecone pine (*Pinus aristata*) are found at the highest elevations (Barbour and Billings, 2000). Areas above timberline contain alpine tundra grasslands and wetlands, interspersed with patches of spruce-fir krummholz and alpine willow (primarily short fruit willow [*Salix brachycarpa*] and diamond leaf willow [*S. planifolia*]).

The history of elk in the San Luis Valley parallels their history in other areas of the western USA. They were once numerous throughout the valley, but were displaced when the land was inhabited by European settlers in the 1800s and the population was mostly eradicated by overhunting (Swift, 1945). The elk population on the west side of the San Luis Valley recovered from an estimated 25 individuals in 1905 to about 1500 in 1940 (Swift, 1945). On the east side of the valley, elk moved into the area from neighboring game units where their numbers increased (B. Weinmeister, Colorado Parks and Wildlife [CPW], pers. commun.) and some mixing between the east and west populations likely occurred.

3. Methods

3.1. Field methods

In winter 2004–2005 we captured and radio-collared 65 cow elk with Colorado Parks and Wildlife (CPW) using a net-gun (Quicksilver Air, Colorado Springs, Colorado) and Clover and corral traps. All radio collars were made with very high frequency (VHF) radio beacons (Lotek Inc., Ontario, Canada and Advanced Telemetry Systems [ATS], Isanti, MN); some also included global positioning systems (GPS; Lotek Inc., Ontario, and ATS, Isanti, MN). To capture elk, we searched the study area from a helicopter, and placed collars on individual elk according to group size: 1 animal collared in group size of < 50, 2 animals each in groups of 50–99 animals, 3–4 in groups of 100 + . We attempted to saturate groups in the periphery of the population range to get an accurate representation of animals outside of the federally-protected core herd area that lies within the National Park and National Wildlife Refuge. When the capture crew could no longer locate new uncollared/unmarked groups in the periphery of the range, the remaining collars were placed west and just north of the Great Sand Dunes National Park dune field (Fig. 1).

We flew aerial population surveys on single days in March 2005 and March 2007 using 1 helicopter in 2005 (Bell 206B-III) and 2 helicopters (Bell 206B-III, AStar350B) in 2007. We simultaneously but independently collected data on elk radio collar locations via fixed-wing aircraft (Piper PA-18 Supercub) during helicopter surveys of 2005 and 2007. In 2007 we stratified the survey whereby one helicopter was flown on the valley floor while the other flew simultaneously in the mountains. We flew predefined transect lines, loaded ahead of time onto an onboard global positioning system (GPS). Transects were flown 1.5–2.5 km apart on the valley bottom, and 0.4–0.8 km apart along elevation contours in the mountains.

During surveys we recorded: (1) the size of groups (used digital photography to confirm the number of individuals, especially for large groups), (2) composition of groups in terms of animal age and sex, (3) the presence or absence of trees and shrubs, (4) vegetative canopy cover (%) following Unsworth et al. (1994), (5) snow cover (%), (6) time of day, (7) GPS location of each group, and (8) presence of one or more radio collars in the group. We also recorded whether each elk group was detected on the left, right, center, or both sides of the aircraft to evaluate which observers could detect the group. For example, if the elk group was to the right of the aircraft, it would not have been available to a rear observer on the left; if a group was directly under the aircraft, it would not have been available to a rear observer, unless it was large enough that it would be available to be seen on one or both sides of the flight path.

The simultaneous double observer method is a form of mark-resight survey in which the front and rear observers in the aircraft are treated as independent surveys where sighting of a group of animals by one observer constitutes the "mark" and sighting by the second observer is considered the "resight". To maintain independence, front and rear observers worked separately with no visual or verbal communication between them until each observer had adequate opportunity to detect the animal group independently. After the helicopter had passed a beam of the elk group, observers notified each other about their detection of the group, and asked the pilot to circle the group if they needed to obtain a better count and make observations of other covariates. After sharing information about independent detection of the elk group, all crew members participated in the count and took digital photographs of large groups. Helicopter surveys were conducted by 1 pilot and 2 observers; one front seat observer, and one back seat observer who switched sides in the aircraft to facilitate viewing/searching in the same compass direction of the flight path (e.g., always looking to the east). The pilot focused mostly on flying, but did search for elk groups as well. Therefore, any group seen by the pilot and/or front seat observer was part of the 'front' observations, and the single observer in the backseat comprised the 'rear' observations.

The fixed-wing aircraft, occupied only by the pilot, flew near but at a higher altitude than the helicopter to locate elk with radio collars. The pilot recorded the same covariates as the helicopter crews for all groups containing radio-collared elk. At the end of the survey we compared the group size, time of observation, presence of radio collars in a group, and Universal Transverse Mercator (UTM) locations to determine which groups with radio collars were seen or not seen by the observers in the helicopter.

Observers in the helicopter determined which elk groups had been independently detected by each observer. Along with records from the fixed-wing aircraft, this created a mark-resight dataset, in which groups could be sighted by 'front', 'rear' or 'fixed-wing' observers for radio collared groups; all elk groups with radio collars located in the study area during the survey were guaranteed to be sighted by the fixed-wing observer.

3.2. Analysis methods

To demonstrate the importance of each component of our method, we analyzed results for 7 alternative analysis models. The first 3 models use different methods for obtaining raw

counts without statistical adjustments for missed groups. The last 4 models apply various statistical sightability methods to correct raw counts for inherent biases. The first model is the most simplistic method, and each subsequent model adds an additional technique for improving the result. Comparisons among these models demonstrate the relative contributions, to the final population estimate, of the following elements: (1) adding additional observers, (2) using photographs to more accurately count large groups, (3) applying statistical models of sighting probability based on observed covariates, and (4) enhancing statistical models by adding information from radio collared elk to correct for residual sighting heterogeneity among groups that is not accounted for by the measured covariates. We used the Huggins (1989, 1991) closed capture estimator for mark-resight data with individual covariates in program MARK, as described by White and Burnham (1999) to fit models. The objects of interest in the mark-resight model were elk groups and not individual elk, because only groups were sighted independently. The model was structured with 3 capture occasions, corresponding to the three possible observers (front, rear, and fixed-wing pilot). The 7 analysis models are listed in order from the most negatively biased to the least biased (and most precise for those with error estimates).

1. M_F – *Raw count by front seat observers in the helicopter*. Population count obtained from a simple "census" using the pilot and a front seat observer simply reporting everything counted from the air.

2. M_C – *Raw count by all observers, front and back combined, in the helicopter*. Similar to M_F except that it included additional elk seen exclusively by the back seat observer.

3. M_P – *Corrected count based on photographs of large groups*. M_C but the number of elk in each group is based on a careful count of elk from photographs. During the survey we take multiple photos while circling the group because some individuals are hidden behind others. We evaluated all photos for the one with the most elk and used that as our final count for the group size.

4. M_D – *Double observer sightability model with individual group covariates*. Sighting probability models (see below) were fit to the data obtained from the observers in the helicopter, with reference to photographs to correct the counts of group size. This model uses all of the data that would be available for computing statistical corrections for sighting bias had radio collars not been present. Group size estimates are identical to those used in the M_P model.

5. M_D^+ – *Model M_D structure with all radio collar data included*. Identical to M_D except that it included radio collared groups not seen by helicopter observers but located by their radios. The additional data are expected to improve model fitting and resulting population estimates.

6. M_H – *Heterogeneity parameter added to M_D structure*. Similar to M_D^+ except that one additional model parameter is included to estimate the average difference in sighting probability between groups with and without radio collars. We presume that groups with radio collars are a random sample from the population during surveys (selected unconditionally), whereas groups seen by the helicopter observers are biased toward groups that are more easily seen or detected (i.e., their inclusion in the dataset is contingent on having been seen by helicopter observers). Without

accounting for this bias arising from contingently sighted groups or "residual heterogeneity", the overall sighting probability estimates will be too high, leading to statistical corrections and resulting estimates of the number of groups that are biased low. After fitting the M_H model to include radio collar data, the model is then applied only to groups actually seen by aerial observers (excluding those only known due to radios), thereby mimicking what would be done in the future when radio collars are unlikely to be present.

7. M_H^* – *Model identical to M_H but estimation for uncollared groups only*. M_H sighting models are used and fitted using the same data (i.e., the model parameters for M_H and M_H^* are identical), however, the sighting models are used to estimate only the portion of the population in groups with no radio collared animals. All radio collared groups, including those not detected initially by the helicopter observers, are included without variance. This estimation method allows for a potentially better estimate of the true population (because only a portion has to be estimated with statistical corrections and the rest is known without error). However, this method only applies to surveys in which radio collars are present. Results from this method provide the most accurate and precise estimates to which all others can be compared, but are not possible in future surveys if radio collars are unavailable (not deployed) in the population.

We did not choose 2 additional possible statistical models that could have been applied to these data. (1) A simple Lincoln–Peterson estimate (i.e., not including any covariates) based on the 2-occasion aerial observations (Graham and Bell, 1989) relies on all groups being equally likely to be seen, but the vast difference in elk group size along with other factors affecting sighting probability were expected to make this assumption implausible and resulting estimates unreliable. (2) We discuss the contrast between our methods with simple sightability models (Steinhorst and Samuel, 1989), but these offer no improvements over the other methods presented here, as has been previously argued and demonstrated empirically (see model M_S in Griffin et al., 2013), so we do not present estimates based on this method.

Our models M_D and M_H are structured similarly to models in Griffin et al. (2013), which should be consulted for additional details on model structure and fitting methods. We fit these same models to alternative subsets of the data, giving rise to M_D^+ and M_H^* estimates for comparison purposes.

We estimated sighting probabilities individually for each observer (front seat, back seat, and fixed wing). We fixed sighting probability at 0.0 for elk groups unavailable to a particular observer (i.e., groups without radio collars, for the fixed wing observer; groups on the opposite side of the aircraft, for the back seat observer) and fixed it to 1.0 for groups with radio collars observed by the fixed wing observer. Sighting probability for all other combinations of observer and elk groups were estimated from a logistic model following Griffin et al. (2013), but using a different list of covariates as predictors of sighting probability:

1. Incremental effect added for back seat observer.
2. Incremental effect added for 2007 valley survey.
3. Incremental effect added for 2007 mountain survey.

4. Elk group size effect.
5. Presence of vegetation (either shrubs or trees versus open).
6. Incremental effect added for groups with radios (residual heterogeneity effect, models M_H and M_H^* only).

All models also included an intercept. Alternative parameterizations of models M_D and M_D^+ were fit with all combinations of the first 5 covariates leading to $2^5 = 32$ alternative models. Model M_H was fit with all possible combinations of these 5 variables plus the residual heterogeneity covariate leading to $2^6 = 64$ models. We initially considered 2 additional covariates in our models: percent of the ground covered by vegetation and percent of the ground covered by snow. Both of these were found to have minimal predictive value or support in preliminary analyses and were dropped from further consideration before conducting the analyses presented here.

We computed separate elk abundance estimates for each of the 3 surveys conducted: 2005 combined valley and mountains, 2007 valley, and 2007 mountains. We further subdivided these abundance estimates into estimates for groups with and without radio collars. The subset of the population that had radio collars constitutes a subpopulation of known size. We segregated our estimates for each of the 7 models into estimates for the radio-marked and the unmarked subpopulations so that we could compare the estimates for the marked subset to a known correct answer to see how well each scenario performed. We also compared population estimates from the 7 models to explore the relative contribution of each component of the analysis.

In planning the survey, we attempted to define survey area boundaries that would encompass the entire elk population that uses the East San Luis Valley. However, several radio collars were not located in the defined study area covered by the aerial survey. We excluded these collared elk from the analysis reported here and made no additional adjustments to our estimates for these absent elk.

4. Results

For the 2005 survey, 27 groups with radio collars were present in the study area, 44 elk groups (with or without radio collars) were detected by the helicopter crew, and an additional 4 elk groups with radio collars that were present in the survey area were not detected by the helicopter crew. Group size ranged from 1 to 667 elk. In 2007, there were 12 groups with radio collars in the valley portion of the survey area and 5 in the mountain portion. The helicopter crew detected 24 elk groups in the valley and 22 in the mountains and missed 3 radio collared groups in each of these sections of the survey area. Group size ranged from 1 to 865 elk in the valley and 1 to 525 elk in the mountains. In both years, there were a total of 44 radio collared groups available for detection by the helicopter crew, which represents the sample size available for estimating the residual heterogeneity parameter. The sample size for models M_D^+ and M_H includes all 100 groups known to be present during any of the 3 surveys, including both those seen by the helicopter crew and those only known due to radio collars.

Over the course of the 3 surveys, back seat observers saw 3961 elk, whereas front seat observers (model M_F) saw 6052 (52.8% more), reflecting the fact that there was only a single back seat observer surveying on one side of the aircraft. The back seat observers saw 88 elk that the front seat observers did not, increasing the total seen by the helicopter crew (model M_C) to 6,140 elk, an increase of 1.5% over the front seat observer's count. When photographs of the larger groups were scrutinized following the flights (model M_P), an additional 361 (5.9%) elk were detected that had not been counted by helicopter observers from the air.

Model weights for the 32 parameter variants of model M_D showed very strong support for a positive effect of group size and a negative effect of vegetation (Table 1). Support was moderate among these models for differences in sighting probability for the back seat observer and for differences among the 3 survey occasions. When the additional groups missed by the helicopter and known only due to radio collars were added and used to fit these same models (model M_D^+), the strength of support for the 2007 valley and 2007 mountain effects increased substantially, whereas support for a back seat observer effect decreased, and support for group size and vegetation were reduced slightly (Table 1).

Model weights for the 64 parameter variants of model M_H were similar to those for model M_D^+, but also revealed moderate support for a residual heterogeneity effect. After controlling for other effects (survey occasion, group size, and vegetation), groups with radio collars (i.e., those selected randomly, or unconditionally) were more difficult to detect on average than those without radio collars (i.e., those detected by helicopter observers). This difference indicates that elk groups in the visual observers data set were not representative of the total population of elk groups available to have been detected, but were biased toward more easily detected groups, even after accounting for the effects of the covariates we recorded.

In addition to the strength of support for each effect, the relative magnitude and direction of each is important. For a baseline example, consider the 2005 survey with a group size of 12 elk (the median value) and no vegetation present. In this instance the front observers had an estimated sighting probability of 85.5% for groups without radio collars and 68.9% for groups with radio collars (first row of Table 2). The back seat observer had higher sighting probabilities (Table 2). Bear in mind that sighting probability for groups on the opposite side of the aircraft from the back seat observer were fixed at zero, so estimated sighting probabilities for the back are conditional on the group being on the same side. Average sighting probabilities were lower for the 2007 valley survey and even lower for the 2007 mountain survey, relative to the baseline example for 2005. Estimated sighting probability rises gradually with group size and drops substantially when vegetation is present (Table 2). Of the groups detected, the lowest sighting probability estimated by model M_H was 11.7% for a group of 2 elk in the 2007 mountain survey that was not on the same side as the back seat observer and was found in a vegetated area. Across the 3 surveys, 21% of groups had sighting probability <30% and 35% had sighting probability $\geqslant$ 80%.

In order to look closely at the relationship between the differing models and examine the contribution of each component of information, we pooled data from all 3 surveys as if they were one large survey. The sum of the population estimates across the 3 surveys based on the model-weighted average across the 32 fitted M_D models (Fig. 2) was 7076 (SE = 1196; CV = 16.9%). This is 575 (8.8%) more elk than the raw count obtained from the helicopter crew after using

Table 1 Model-weighted parameter estimates and total model weight supporting each parameter for the 3 alternative models of elk group sighting probability for elk in the San Luis Valley, Colorado, USA. Values for model M_H^* are identical; however, M_H^* is applied only to groups without radio collars.

Model parameter	Parameter estimate (SE)			Model Weight		
	MD	MD+	MH	MD (%)	MD+ (%)	MH (%)
Intercept	0.805 (0.754)	1.699 (0.455)	1.665 (0.535)	100.0	100.0	100.0
Back seat observer[1]	0.341 (0.211)	−0.206 (0.251)	0.114 (0.243)	53.6	28.3	58.2
2007 Valley Survey	−0.204 (0.591)	−1.657 (0.531)	−0.804 (0.574)	30.0	65.1	69.9
Residual heterogeneity	n/a	n/a	−1.075 (0.271)	0.0	0.0	52.7
2007 Mountain Survey	−0.473 (0.862)	−2.124 (0.642)	−2.125 (0.551)	41.1	91.9	98.4
Group size	0.051 (0.016)	0.002 (0.002)	0.009 (0.002)	97.4	92.9	100.0
Vegetation (Shrub or tree)	−2.085 (1.007)	−1.430 (0.500)	−0.735 (0.565)	92.3	89.6	56.6

[1] Effect of back seat observer on the sighting probability for elk groups on the same side of the aircraft as the observer; sighting probability for the back seat observer with elk groups on the opposite side is fixed at 0.0.

Table 2 Estimated sighting probability for front- and back-seat observers viewing groups with and without radio-collars, based on model weighted average values of parameters from 64 fitted variants of model M_H. Estimates are shown for illustration purposes of 6 alternative sets of conditions, as indicated in the first 3 columns.

Covariate Values			Groups With Radio Collars			
Survey	Group Size	Vegetation	No		Yes	
			Front (%)	Back[1] (%)	Front (%)	Back[1] (%)
2005, All	12	No	85.5	86.9	68.9	69.4
2007, Valley	12	No	72.6	74.8	49.8	50.3
2007, Mtn.	12	No	41.4	44.2	20.9	21.3
2005, All	1	No	84.2	85.7	66.7	67.1
2005, All	100	No	93.2	93.8	83.6	83.9
2005, All	12	Yes	73.9	76.1	51.5	52.0

[1] Back seat observer sighting probabilities for elk groups on the same side of the aircraft as the observer; sighting probability for the back seat observer with elk groups on the opposite side is fixed at 0.0.

the photographs of large groups to account for undercount of these groups in the air (model M_P). Model M_D^+, fitted to the dataset that included groups with radio collars missed by the helicopter crew, produced an estimate of 7365, which is 290 (4.1%) more elk than model M_D, and with better precision (SE = 929; CV = 12.6%). Model M_H, which accounted for the difference in sighting probability of conditionally detected groups (without radio collars) and unconditionally detected groups (with radio collars), increased the estimate by another 372 elk (5.1%) to a total of 7737, with better precision than model M_D^+ (SE = 845; CV = 10.9%). When the same model was applied to estimate only groups without radio collars while those with radio collars were included as a known quantity (model M_H^*), the final estimate was another 116 elk (1.5%) higher and more precise: the final estimate for the sum across surveys for model M_H^* was 7,853 elk (SE = 697; CV = 8.9%). Using this best model, the elk population in the San Luis Valley was 3534 elk (SE = 71.5; CV = 2.0%) in 2005 and 4319 elk (SE = 698; CV = 16.2%) in 2007 (Fig. 2).

During surveys, the back seat observer saw only 50.4% of the estimated total number of elk, and the front seat observers alone saw only 77.1% of the final estimate. The combined counts by both front and back seat observers, supplemented by more accurate counts from photographs raised the raw count to 82.8% of the final value. This adjusted raw count still missed 1352 elk that our best statistical estimates were able to reveal. Results for only those groups with radio collars, for which the true population was known, are similar: 91.6% of all elk in these groups were detected by the combined aerial observers after supplementing aerial observations with photographs (Fig. 2).

The best statistical model that could be estimated without radio collars (M_D) resulted in a population estimate that was 90.1% of the final value (M_H^*) and 95.4% if considering only the radio collared portion of the population. When radio collar data were incorporated to increase the size of the data set and to estimate the residual heterogeneity parameter (difference between sightability of randomly selected and conditionally selected elk groups; M_H), the best estimate was 98.5% of the final value (M_H^*) and 98.1% for the radio collared portion of the population.

5. Discussion

Each element of our method provided incremental gains in bias correction for population estimates. Adding a second, independent observer in the back seat to the aerial survey enabled us to reduce detection bias in 2 ways. First, the second observer saw elk groups not seen by the first. More importantly, using statistical models with data from multiple observers (model M_D) provided estimates of the different sighting probabilities for each unique observer specific to these surveys, which enabled us to estimate both the number of elk not observed by either observer and the precision of those esti-
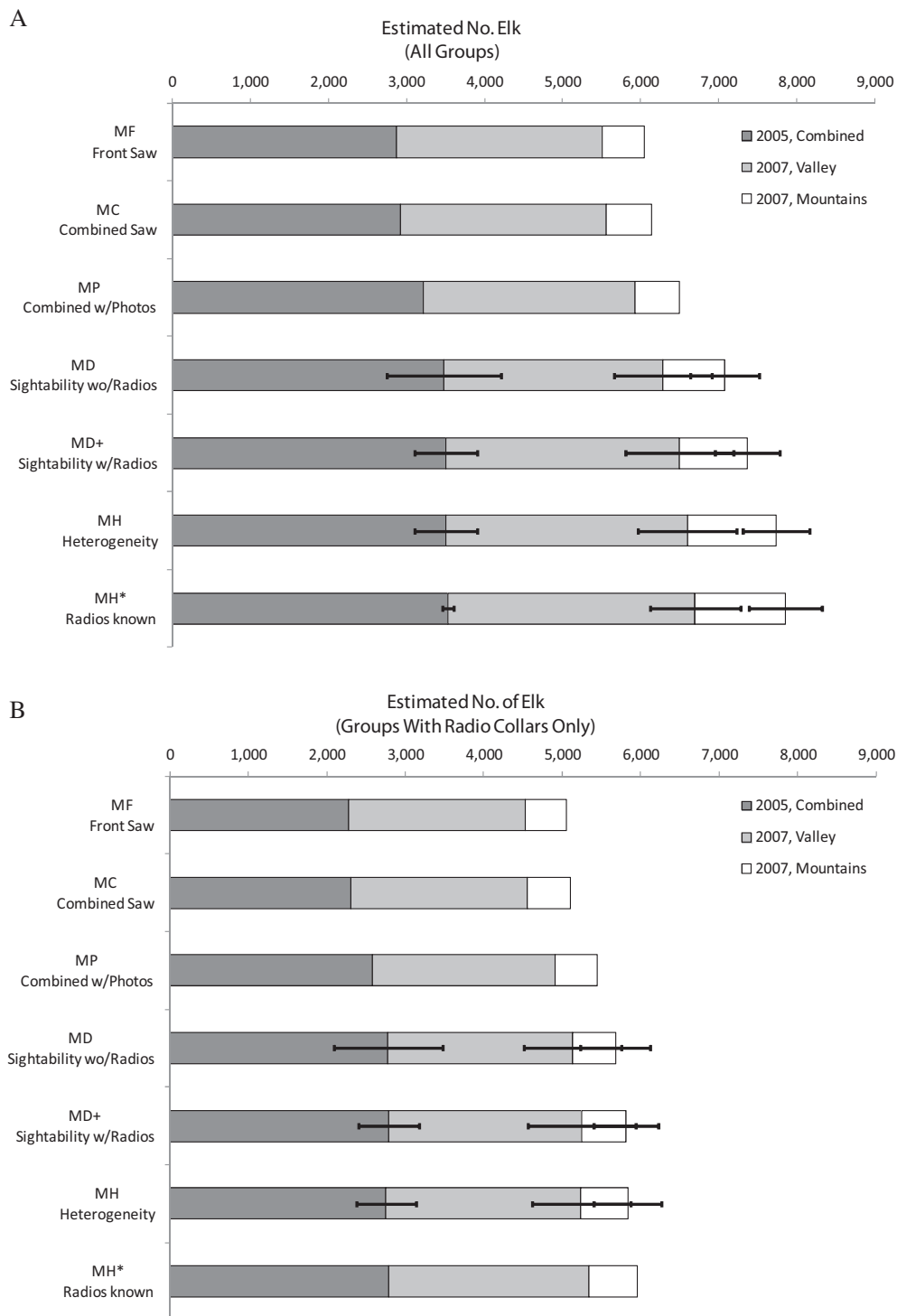
Figure 2 Elk population estimates for 3 surveys in the east side of the San Luis Valley, based on 3 methods of obtaining a raw count uncorrected by statistical estimation of sighting probability, and estimates based on 4 statistical sighting probability modeling methods. See text for explanation of the 7 methods. Panel A shows estimates for all elk groups; panel B shows estimates only for radio collared elk groups. The final bar (Model M_H^*) in panel B is the known, true population of radio collared elk. The true population in panel B can be compared to the other model estimates to observe the bias of each. The 2005 estimate covers both valley and mountain portions of the survey area, whereas the 2007 estimates for these portions of the study area were separated. The estimates from the 2 surveys were pool for this comparison, which focuses on the relative bias of each method. The error bars on the 4 statistical estimates are 1 standard error (SE) and apply to each of the 3 individual surveys, not to the cumulative total.

mates. However, sighting corrections based on just 2 observations are only unbiased if all groups have equal sighting probability or if the sighting covariates explain all of the variation in sighting probability among groups (Burt et al., 2014). This is only a plausible assumption where sighting probabilities are very high or sighting conditions are nearly uniform. However, a bias is likely when heterogeneity in sighting probability among groups is present and not adequately modeled. Therefore, we attempted to use covariates that we hypothesized would explain some of this variability in sighting probability among groups, to correct for several sources of bias that would have resulted in undercounting the true population.

Elk groups in this population during winter are often very large. Our largest group contained 865 elk. The combined motion of the elk and the helicopter makes accurate counting of groups this large unlikely without the use of high-resolution photographs. Consequently, using photography for all large groups improves accuracy. However, we acknowledge that even with the photographs, some negative bias (undercount) could still remain, but the magnitude of that bias is reduced substantially by using high-quality photography.

The sighting probability estimates calculated from double-observer data (models M_D and M_D^+) enabled us to correct for elk not seen by either aerial observer. This bias can be corrected through the regular use of 2 (or, preferably 3 if both seats in the back are occupied to provide coverage on both sides of the aircraft) aerial observers so that sightability correction models can be calculated based on the observers and field conditions specific to each survey. Data from subsequent surveys conducted under similar conditions can usually be pooled to obtain a larger sample size of observations, enabling better estimates of sighting probability and the effects of individual covariates.

Unless all of the difference among groups is modeled with the selected covariates, however, a residual bias will persist. With the radio collars, we were able to estimate a residual heterogeneity parameter to account for the otherwise unmodeled (unaccounted for by our selected covariates) differences between randomly (unconditionally) selected groups (those with radio collars) and groups conditioned on being seen by the helicopter crew (those without radio collars). We concluded by comparing model M_D^+ with M_H^* that a significant negative bias averaging 6.2% remained that was only correctable through this method, unless a better set of predictive covariates could be found and tested. This finding is consistent with similar results reported in Griffin et al. (2013).

Traditional sightability correction models attempt to correct for missed groups using individual covariates, however this correction is based on initial calibration of the sightability model when radio collars are present but is not updated or supplemented by accumulating data from subsequent surveys when radio collars are no longer available in the population. Thus, with traditional sightability models, future population estimates depend entirely on the assumption that the sightability corrections estimated from initial flights are universally applicable on later flights that have different observers and different environmental conditions. Our surveys used a small number of observers, all with extensive aerial survey experience, and the same aircraft. Despite this, we found strong evidence for different sighting probabilities for each of the 3 surveys. For this reason, we caution against applying the population estimation model fitted in this study to future surveys,

even in the same location and with the same observers. Data from future surveys at this site could be pooled with surveys from 2005 and 2007 to obtain a larger sample size and greater precision, if statistical evidence supports such pooling. Sighting bias estimation as a function of observer and environmental covariates should always be included in future surveys, at least to confirm that the current corrections still work, and at best to continue to improve them over time as more elk group observation data accumulates.

Data collected on future elk surveys of this study area, when radio collars are no longer present, will add to the cumulative dataset available for fitting sightability models of the model M_H type. However, without radio collars, no additional information will be gathered about the residual heterogeneity parameter, only for the other parameters in the model. This means that, like a sightability model, it must be assumed that this residual heterogeneity effect remains constant in the future. However, this assumption is limited only to this single parameter, which explains only a small fraction of the bias – most of the bias is modeled by the parameters that can be updated after future surveys are conducted without radio collars. This leaves much less room for error due to unmeasured changes in this single parameter. To the extent that other parameters can be found to minimize the residual heterogeneity effect, errors caused by future changes in this residual heterogeneity parameter would be correspondingly reduced. There is some loss of accuracy in any survey without radio collars because model M_H^* cannot be applied, but the correction due to this additional information was small in our study (1.5%), although it also improved precision from CV = 10.9% to 8.9%. Repeating the radio collar study periodically to test for changes in and to update the estimate of the heterogeneity parameter would be desirable. This is especially true in this case, given that we were only able to apply a sample size of 44 groups to use in fitting the residual heterogeneity parameter.

As in most modeling, there are some uncertainties even in what we find to be our best modeling approach, and our estimates may rely on some extrapolation beyond observed conditions. We only placed collars on elk found in the survey area, which may not accurately represent all elk that occasionally use the area. Our estimates only apply to elk present in the survey area during the survey and do not account for other elk that may use the survey area intermittently. A far more extensive radio-collaring effort would be required to estimate the portion of the potentially larger population that occasionally uses the survey area that were actually present during a given survey.

A second limitation of our bias corrections is that all collars were placed on adult female elk. We cannot determine from our data (or any other available) whether these corrections apply equally to other age and sex classes. Calves are found in groups with females, so it is less likely that a different bias applies to this segment. Another work has documented a negative sighting bias for male elk in aerial surveys (McCorquodale, 2001), leading us to believe that our population estimate from model M_H^* may be low.

We acknowledge that several improvements could be made in our methods for future surveys. It is far better to have 2 observers in the back seat so that all elk groups are subject to sighting and resighting, not only those on one side of the aircraft with the single back seat observer. This would not only

increase the raw count reducing the magnitude of the statistical correction, but also would provide a larger sample size for estimating sighting probabilities, which can only be done when ≥ 2 observers are available to observe a particular group. We suspect that some of the residual heterogeneity that could only be estimated through the use of radio collars is due to variation in the distances from observers to elk groups. Collecting at least crude distance estimates could improve the ability to estimate sighting probability without depending as much on the residual heterogeneity parameter. Additional covariates that might also help explain visibility differences among groups could be considered, such as indicators for lighting condition (e.g., full sun, patchy sun, and overcast sky) and activity of elk (Griffin et al., 2013; Samuel et al., 1987; Anderson and Lindzey, 1996; Gilbert and Moeller, 2008; McIntosh et al., 2009). If different observers are used in the future, covariates to account for individual observer differences could improve precision and reduce bias. Separating the valley and mountain portions of the survey area (stratifying) in 2007 improved survey results, given the significant differences in sighting probabilities estimated for these distinctly different habitat types. To apply this method to other locations and species, experienced biologists and aerial observers who know what factors affect their ability to detect the target species in that area must carefully consider and select the list of covariates to be collected and tested. Statistical analysis requires that values for all of those candidate covariates must be recorded for each observed animal group.

We recognize the possibility that a larger dataset would likely support additional parameters to explain more of the differences in sighting probability among elk groups. In particular, interactions among main effects variables in our models are a possibility. For example, front and back seat observers may not respond identically to group size, vegetation, survey occasion, or the unknown heterogeneity effects captured by our heterogeneity parameter. The unknown heterogeneity might also not be constant, but might interact with group size, vegetation, or survey occasion. We were not able to model these potential interactions in our study without a larger sample size.

Surveys should be designed to hold as much constant as possible and repeated using consistent methods, pilots, aircraft, and observers over multiple years. By minimizing the variation that needs to be explained and increasing the sample size to facilitate estimation of more components of the remaining variation, even better estimates are possible. Nonetheless, using every available tool to correct for biases, as we did here, is an improvement over methods that rely on less information. No population estimation method will ever be perfect, but we've demonstrated that at least the following components of the survey design are important: (1) multiple observers, (2) counts for large groups based photographs, (3) statistical correction models incorporating measured covariates that influence sighting probability, and, when budgets permit, (4) additional statistical correction for unmodeled heterogeneity based on radio collar data.

Simple raw counts, often incorrectly referred to as a "census," can produce severely biased estimates (undercounts) of ungulate populations, and provide no measure of precision with which to evaluate the usefulness of the estimate. In this study, 22.9% of the estimated population would have been missed by a simple raw count using a single aerial observer

and pilot. The validity of a traditional sightability model, calibrated for different observers, locations, conditions, and populations is questionable. Simple double-observer methods (i.e., a Lincoln–Peterson estimator, as in Graham and Bell, 1989) without sighting covariates is likely to be substantially biased by heterogeneity (differences in sightability) among elk groups (Burt et al., 2014).

Using hybrid methods that combine double observer data with sighting covariates greatly improves the reliability of population estimates. Much accuracy and precision can be gained using the double-observer method (model M_D) alone, but the use of radio collars during an initial study period enables the use of model M_H, which can further improve results and increase confidence in their reliability.

The hybrid method we applied (model M_H or M_H^* when collars are present) is suitable for most aerial surveys of large terrestrial wildlife species. It may be particularly applicable in arid environments, where there tends to be relatively little vegetation that obstructs aerial observers' view of wildlife. The field methods are only slightly more complex and generally no more costly than for other aerial survey methods. We do recognize, however, that the analytical techniques are complex, particularly the bootstrap modeling required to estimate confidence intervals (Griffin et al., 2013). To date, data analysis has been developed specifically for each study and requires an experienced statistical analyst. Development of reusable software to facilitate analyses of this type is an important area for future work.

Acknowledgements

References

Anderson, C.R., Lindzey, F.G., 1996. Moose sightability model developed from helicopter surveys. Wildl. Soc. Bull. 24, 247–259.

Baillie J.E.M., Hilton-Taylor, C., Stuart, S., 2004. 2004. IUCN Red List of Threatened Species: A global species assessment. In: IUCN, Gland, Switzerland and Cambridge, UK.

Barbour, M.G., Billings, W.D., 2000. North American Terrestrial Vegetation. Cambridge University Press, New York, New York, pp. 434.

Bayliss, P., Yeomans, K.M., 1989. Correcting bias in aerial survey population estimates of feral livestock in northern Australia using the double-count technique. J. Appl. Ecol. 26, 925–933.

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

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L., 2004. Advanced Distance Sampling. Oxford University Press, Oxford, U.K..

Burnham, K.P., Anderson, D.R., 1984. The need for distance data in transect counts. J. Wildl. Manage. 48, 1248–1254.

Burt, M.L., Borchers, D.L., Jenkins, K.J., Marques, T.A., 2014. Using mark-recapture distance sampling methods on line transect surveys. Methods Ecol. Evol. 5, 1180–1191.

Cook, R.D., Jacobson, J.O., 1979. A design for estimating visibility bias in aerial surveys. Biometrics 35, 735–742.

Gilbert, B.A., Moeller, B.J., 2008. Modeling elk sightability bias of aerial surveys during winter in the central Cascades. Northwest Sci. 82, 222–228.

Graham, A., Bell, R., 1989. Investigating observer bias in aerial surveys by simultaneous double-counts. J. Wildl. Manage. 53, 1009–1016.

Griffin, P.C., Lubow, B., Jenkins, K.J., Vales, D., Moeller, B.J., Reid, M., Happe, P., McCorquodale, S., Tirhi, M., Schaberl, J., Beirne, K., 2013. A hybrid double-observer sightability model for aerial surveys. J. Wildl. Manage. 77, 1532–1544.

Hadley, N.F., Szarek, S.R., 1981. Productivity of desert ecosystems. Bioscience 31, 747–753.

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.

Manley, B.F.J., McDonald, L.L., Garner, G.W., 1996. Maximum likelihood estimation for the double-count method with independent observers. J. Agric. Biol. Environ. Stat. 1, 170–189.

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

McCorquodale, S.M., 2001. Sex-specific bias in helicopter surveys of elk: sightability and dispersion effects. J. Wildl. Manage. 65, 216–225.

McIntosh, T.E., Rosatte, R.C., Hamr, J., Murray, D.L., 2009. Development of a sightability model for low-density elk populations in Ontario, Canada. J. Wildl. Manage. 73, 580–585.

Pearson, L.C., 1965. Primary productivity in grazed and ungrazed desert communities of eastern Idaho. Ecology 46, 278–285.

Pollock, K.H., Kendall, W.L., 1987. Visibility bias in aerial surveys: a review of estimation procedures. J. Wildl. Manage. 51, 502–510.

Pollock, K.H., Nichols, J.D., Brownie, C., Hines, J.E., 1990. Statistical inference for capture-recapture experiments. Wildl. Monogr. 107.

Quang, P.X., Becker, E.F., 1997. Combining line transect and double count sampling techniques for aerial surveys. J. Agric. Biol. Environ. Stat. 2, 230–242.

Rivest, L.-P., Potvin, F., CrÅpeau, H., Daigle, G., 1995. Statistical methods for aerial surveys using the double-count technique to correct visibility bias. Biometrics 51, 461–470.

Samuel, M.D., Pollock, K.H., 1981. Correction of visibility bias in aerial surveys where animals occur in groups. J. Wildl. Manage. 45, 993–997.

Samuel, M.D., Garton, E.O., Schlegel, M.W., Carson, R.G., 1987. Visibility bias during aerial surveys of elk in north-central Idaho. J. Wildl. Manage. 51, 622–630.

Seber, G.A.F., 1982. The Estimation of Animal Abundance and Related Parameters. Griffin, London, United Kingdom.

Skalski, J.R., Millspaugh, J.J., Spencer, R.D., 2005. Population estimation and biases in paintball, mark-resight surveys of elk. J. Wildl. Manage. 69, 1043–1052.

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

Swift, L.W., 1945. A partial history of the elk herds of Colorado. J. Mammal. 26, 114–119.

Unsworth, J.W., Leban, F.A., Leptick, D.J., Garton, E.O., Zager, P., 1994. Aerial survey: user's manual: Boise, Idaho Fish and Game Department.

White, G.C., 1996. NOREMARK: population estimation from mark-resighting surveys. Wildl. Soc. Bull. 24, 50–52.

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

Wong, C., 1996. Population size estimation using the modified Horvitz-Thompson estimator with estimated sighting probabilities. Dissertation, Colo. State Univ., Fort Collins, USA.