EVALUATING POPULATION ESTIMATES

OF MOUNTAIN GOATS BASED ON CITIZEN SCIENCE

By

JAMI JEAN BELT

Bachelor of Philosophy, Miami University, Oxford, Ohio, 1997

Thesis

presented in partial fulfillment of the requirements
for the degree of

Master of Science
in Wildlife Biology

The University of Montana
Missoula,

May 2010

Approved by:

Perry Brown
Graduate School

Paul. R. Krausman
Wildlife Biology

Joel Berger
Wildlife Biology

Michael S. Mitchell
Montana Cooperative Wildlife Research Unit

Belt, Jami, M.S., May 2010 Wildlife Biology

Evaluating population estimates of mountain goats based on citizen science

Graduate advisor: Paul R. Krausman

Citizen science programs that use trained volunteers may be a cost-effective method for monitoring wildlife at large spatial and temporal scales. However, few studies have compared inferences made from data collected by volunteers to professionally collected data. In Glacier National Park (GNP), Montana, I assessed whether citizen science is a useful method to monitor mountain goat (*Oreamnos americanus)* populations. I compared estimates of mountain goat abundance by volunteers at 32 sites throughout GNP to estimates by biologists and raw counts from aerial surveys at a subset of 25 and 11 sites, respectively. I used multiple observer surveys to calibrate the indices of abundance for the effect of observer variation between volunteers and biologists. I used N-mixture models, which calculated detection probability through patterns of detection and non-detection to obtain estimates of abundance. Population estimates made by citizen science overlapped estimates by biologists and estimates from previous research. Density estimates from aerial surveys were lower, possibly due to imperfect detection during aerial surveys or due to violation of the assumption of population closure. Mean detection probability from multiple observer surveys for biologists was significantly higher and less variable than that of volunteers, but was not a suitable correction factor, because it was not consistent across all densities of mountain goats. Volunteer experience did not significantly influence detection probability or abundance estimates. Abundance estimates by volunteers were influenced by number of site visits. More frequent site visits balanced out lower detection probability by volunteers and resulted in abundance estimates that were less variable than those of biologists. When large spatial and temporal coverage can be achieved, citizen science can provide mountain goat population estimates that are statistically similar to those of biologists. However, neither estimates by volunteers or biologists had sufficient statistical power to detect a 30% decline in mountain goat population size over 10 years. Power by volunteers could be increased by reducing the number of sites and increasing surveys/site or by continuing monitoring over a longer time frame (i.e., 30 years). Citizen science programs can contribute to long term monitoring when properly designed.

ACKNOWLEDGEMENTS

The encouragement and support of S. J. Hejl, who created Glacier National Park's citizen science program, was invaluable. Special thanks to the citizen science volunteers who dedicated many hours and enthusiastically braved the fickle weather of Glacier National Park's high country. T. Carolin, M. Sladek, S. Sindt, B. Thomas, and P.Ollig at the Crown of the Continent Research Learning Center provided logistical and moral support. S. Gniadek, J. Waller, D. Chadwick. and Montana Fish, Wildlife & Parks biologists provided guidance on mountain goat research and aerial surveys. I appreciate the assistance of J. Potter and personnel of the Science and Resources Division of GNP. T. Shepherd, D. Kotter and R. Menicke assisted with the database and GIS. I thank T. Gibbins for assistance with site mapping and J. Soye and S. Kintner for data entry. L.Bate, J. Davis, D. Elwood, S. Ollig, H. Oswald, and C. Drake assisted with data collection.

I thank my major advisor, .P. R. Krausman, for his support and guidance in planning field work, and for reviewing this thesis. My graduate committee, J. Berger and M. Mitchell provided ideas and reviews. The Boone & Crockett Fellows also contributed helpful suggestions I am grateful to my parents for their support. I thank T.Rosenkoetter for his support, humor, and field assistance. B. Douglas, G. Delp, and the community of Missoula provided me a home away from home. The Glacier National Park Fund, the Crown of the Continent Research Learning Center at GNP, and the Boone & Crockett Program in Wildlife Conservation at the University of Montana funded this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDIX

INTRODUCTION

Research and monitoring that relies on volunteers to collect data without direct supervision is called "citizen science" (Trumbull et al. 2000). The use of citizen science for long-term ecological data collection is increasing (Newman et al. 2003, Danielsen et al. 2005, Greenwood 2007, Cohn 2008) as ecosystem-level disturbances (e.g., global climate change; Morisette et al. 2009),and public participation in resource management increases (Yung 2007), and funding for ecological monitoring declines (Pilz et al. 2005). In addition, many granting organizations (e.g., National Science Foundation) often require grantholders to incorporate public participation in research and monitoring (Silvertown 2009).

Careful training and sampling design may allow citizen science programs to yield results that are as reliable as those from professionals monitoring programs (Hochachka et al. 2000, Yoccoz et al. 2003, Gouevia et al. 2004). Citizen science data are often collected on spatial scales beyond the reach of most research budgets (Cooper et al. 2007, Greenwood 2007, Cohn 2008). Because funding required for citizen science programs is lower, they may also be conducted over a longer term than professional monitoring programs with larger funding needs (Danielsen et al. 2005). For example, the National Audubon Society's Christmas Bird Count, a citizen science effort, which began in 1900, has yielded the longest unbroken record of bird diversity and distribution (Root and Alpert 1994) with the broadest temporal and geographic coverage of North America's avian fauna (Dunn et al. 2005).

However, the question of whether citizen science is a scientifically robust approach remains unanswered. Citizen science sampling often represents a compromise between ideal design and design that maximizes participation (Greenwood 2007). Many of the assumptions underpinning statistically rigorous ecological surveys (e.g., skill levels of observers, consistent survey effort, homogeneity of temporal variables) are violated with citizen science (Danielsen et al. 2005). The impact on the validity of the inferences that can be drawn from the data caused by these departures from traditional survey design needs to be assessed. A major obstacle to understanding the effectiveness of citizen science is the paucity of studies that have compared data and results from volunteers to professionally collected data (Fitzpatrick et al. 2009).

In 2005, managers at Glacier National Park (GNP) established a citizen science program to conduct needed baseline monitoring for common loons (*Gavia immer).* Volunteers ($\bar{x}$ / year =117 ± 21 [SD]) participated each year from 2005 through 2009, gathered observational data (effort > 8,000 hours), and contributed information on nest location and estimates of chick hatch dates. Volunteers detected 86% of loon chicks known by biologists to be present in GNP (Jami Belt, Glacier National Park, unpublished data). Due to these successes, GNP managers decided to expand the citizen science program to monitor mountain goats (*Oreamnos americanus*).

Declines in goat numbers at a prominent mineral lick in GNP (Steve Gniadek, Glacier National Park, unpublished data), and throughout the neighboring Bob Marshall Wilderness (Koeth 2008), and uncertainty about mountain goat response to climate change (Pettorelli et al. 2007) generated concern about the stability of mountain goat populations in GNP. Earlier studies yielded density estimates of 1.16 mountain goats/ km^2 in a 310 km^2 central portion of GNP (Chadwick 1977) and 2.9 mountain goats/ km^2 in a 32.7 km^2 area (Singer and Doherty 1985) of the park's 4,081 km^2 area. More recent and broad-scale abundance estimates of mountain goats in GNP are needed for future trend monitoring. Given National Park Service interest in non-invasive monitoring methods due the sensitivity of mountain goats to trapping (Côté et al. 1998) and the costs, safety concerns, and potential impacts to visitor experience associated with traditional mark- recapture methods, citizen science was suggested as a monitoring approach.

Aerial survey by helicopter are the primary non-invasive technique used to census mountain goats (Shackleton 1997) but the high cost of this method means that coverage or replication is often sacrificed. A combination of aerial surveys and observational ground counts may improve precision of population estimates (Festa-Bianchet and Côté 2008). Mountain goats are an ideal candidate species for ground counts due to the high visibility of their exposed habitats (Veitch et al. 2002). Ground counts are rarely conducted however, as they are not typically cost-efficient due to the rugged and remote places inhabited by mountain goats (Festa-Bianchet and Côté 2008). Studies using the number of ungulates seen/day by hunters and outfitters as an estimate of ungulate density (Pettorelli et al. 2007, Veitch et al. 2002, Ericsson and Wallin 1999) have reported promise as viable long-term monitoring techniques.

The large number of visitors to GNP (approx. 2 million/ year) similarly provides a potentially useful resource for monitoring mountain goat populations. In spring 2008 GNP created the High Country Citizen Science (HCCS) program to train volunteers to conduct observational surveys of mountain goats. These surveys take place on a park-wide scale and focus on data collection at backcountry locations (i.e., data that have been logistically difficult and costly for biologists to collect). The goals of the program are to estimate the distribution and abundance of mountain goats in GNP and to establish protocols for long-term trend monitoring.

My objective was to determine whether citizen science is a viable method for long-term population monitoring of mountain goats. As a model for comparing inferences from data derived by volunteers to those from biologists, I compared citizen science estimates to estimates from data collected by biologists following the same survey protocols over a smaller area of GNP. To determine whether these mountain goat population estimates were similar to estimates derived from other methods, I compared volunteer and biologist estimates to aerial surveys, and to the earlier density estimates of Chadwick (1977) and Singer and Doherty (1985).

Given a limited budget, a tradeoff is necessary in professional monitoring between spatial coverage and temporal coverage. A team of volunteers can achieve large spatial and temporal coverage, but the reliability of inferences that can be drawn from the data is unknown. Do the benefits of large spatial and temporal sample sizes attainable using citizen science balance the limitations that result from varied skill levels and heterogeneous survey effort? Volunteers with varied skill level often underestimate abundance (Newman et al. 2003, Delaney et al. 2008) but estimates are consistent and have a linear relationship to density estimates obtained from other sources (Kindberg et al. 2009). Therefore, I predicted volunteers would detect less mountain goats than biologists and due to differences in skill level among volunteers, would have more variable detection rates. Abundance estimates of mountain goats by volunteers at each site would be more variable and biased low compared to counts by biologists. Because volunteers would be likely to survey sites more often, I predicted that they would be more likely than biologists to capture a higher minimum count of mountain goats as the number of site visits increased. I, therefore, expected that estimates of mountain goats at specific sites from citizen science volunteers would have a larger negative degree of bias and higher variation than estimates from biologists, but that abundance estimates across all survey sites would be similar to estimates from biologists due to larger spatial and temporal coverage.

STUDY AREA

Glacier National Park contains 4,081 km² of federally protected land and is situated in the northern Rocky Mountains, Montana, USA. The park is divided lengthwise by the Continental Divide. Elevations range from 945 m to 3,200 m. Timberline east of the Continental Divide is 244 m lower than on the western side. Forest canopy cover is dominated by Engelmann spruce (*Picea engelmannii*), Douglas fir (*Pseudotsuga menziesii*) and subalpine fir (*Abies labioscarpa*). Over one-third of park is within the alpine zone, which is dominated by sparse rock outcroppings (Chadwick 1977). Alpine vegetation includes ledge, talus, meadow and krummholz communities that grade into subalpine coniferous forests at lower elevations (Hop et al. 2007).

Glacier National Park provides habitat to an unhunted native population of mountain goats of undetermined size. Density estimates in limited sections of the park have ranged from 1.16 mountain goats/ km² (Chadwick 1977) to 2.9 mountain goats/km² (Singer and Doherty 1985). Glacier National Park maintains an extensive network of >1,127 km of hiking trails. My study area included all portions of the Livingston and Lewis mountain ranges that had slope angles ≥ 25° and were within 3.2 km of a trail (Fig. 1). Surveys were conducted at 32 sites with an area of 1,311km² (32.1% of GNP).

METHODS

Proximity to escape terrain is a strong determinant of habitat use by mountain goats (Brandborg 1955, McFetridge 1977, Haynes 1992, Hamel and Côté 2007). Escape terrain (i.e., precipitous terrain with cliffs and rocky ledges and slopes angles ≥25 º [Chadwick 1976, Varley 1994, Gross et al 2002]) is used to evade predators. A model based solely on distance to escape terrain with slopes angles ≥33º correctly classified 87% of mountain goat observations from a study in Colorado (Gross et al. 2002). Therefore, I modeled escape terrain from a 30m digital elevation model using slope angle classifications of 25 to 32, 33 to 39, and 40 to 90º (Gross et al. 2002). The area of GNP with slope angles ≥25º was 1,653 km^2 (40.5% of GNP). I then created 8 km by 11 km grid cells over the escape terrain to systematically locate observation sites. The grid cell size was large enough to encompass the maximum home range size of a mountain goat (Rideout 1977), to minimize the likelihood that mountain goats would move from one grid cell to another.

Most (76.6%) of the escape terrain used by mountain goats in GNP exceeded 40º (Chadwick 1977). Therefore, I identified all grid cells with ≥40º escape terrain ≤3.2 km from a hiking trail for observation sites. The distance of 3.2 km is the maximum line of sight distance for reliable detection and identification of mountain goats (S. Gniadek, Glacier National Park, personal communication; D. Chadwick, National Geographic, personal communication) which I verified through field tests. Other studies have used distances of up to 10 km to observe ungulates (Krausman et al. 2004) that, like mountain goats , are highly visible in exposed habitats . I then divided each trail into 2 km segments, randomly selected 1 segment/cell, hiked each segment and recorded Universal Transverse Mercator (UTM) coordinates of all points from which slope features were visible. I finally randomly selected from all points a single observation point (i.e., site) in each grid cell from which the largest area of escape terrain ≥25 º was visible.

I selected 32 sites but topographic features blocked portions of each site. I calculated the portion of each site that was visible from each observation point using the Spatial Analyst Viewshed function in ArcGIS (Environmental Systems Research Institute, Redlands, California). We used the Spatial Analyst Extract by Mask function in ArcGIS to calculate the area of escape terrain in the viewshed of each site ($\bar{x}$ =4.7 km² ± 2.3 [SD]. The area of escape terrain within viewsheds was 149 km² (9.0% of escape terrain in GNP) and the area of escape terrain at sites was 727 km² (43.9% of escape terrain in GNP). Most escape terrain was above treeline.

I recruited volunteers for the HCCS program using press releases, newspaper articles, public presentations, and flyers. In 2008 and 2009 selected volunteers attended a standardized 6 hour training session where they learned ecology of mountain goats, field identification relative to co-occuring ungulates (e.g., bighorn sheep [*Ovis canadensis*], mule deer [*Odocoileus hemionus*], white-tailed deer [*O. virginianus*], elk [*Cervus canadensis*], and moose [*Alces alces*]), and classification of age and sex following criteria developed by Smith (1988). Volunteers then worked in the field with HCCS staff to learn survey protocol, data form completion, and the use of survey equipment (e.g., binoculars, spotting scopes, Global Positioning System (GPS), and compasses). Volunteers were asked to record the power and field of view of optical equipment on their data sheet.

All volunteers in 2009 completed a participant information sheet (Appendix 1) detailing their experience with spotting scopes and viewing wildlife. I used these answers as a self-evaluation measure to assess experience level (Martin 1997, Scott et al. 2005). I also assessed the experience level of biologists who conducted mountain goat surveys in GNP using the same questionnaire (scores could range from 9- 42). I pooled separate scores for volunteers and biologists, calculated the quartiles for the pooled data, and ranked each participant as: novice (minimum to first quartile, 9-20.75 points); some relevant experience (first quartile to mean, 20.76-26 points); moderate (mean to third quartile: 26.01- 31 points); and skilled (third quartile and above ,31.01- 42 points).

Once trained, volunteers conducted surveys at selected sites based on their schedule, hiking ability, and preference. Volunteers navigated to each site using a GPS unit and site map. Photos of the observation point and the views due north and due south were provided to ensure that volunteers could locate the correct site despite GPS error (+/- 10m). Volunteers conducted a 1 hour survey, recorded the number, age and sex of mountain goats detected, time of initial detection, and group size. Volunteers were also asked to take photos of each group of mountain goats with a digital camera through a spotting scope, and to submit photos for verification purposes.

Volunteers also recorded temperature, cloud cover, weather, time of day, and behavior that may affect detection probability Volunteers recorded behavior of individual mountain goats upon detection (e.g., bedded, standing, foraging, walking), which I converted into the percentage of mountain goats that were moving upon detection. Beginning in 2009 volunteers also documented visibility (as a proxy for distance estimation) and habitat use. I used a 2 step process to estimate visibility.

Volunteers first recorded how they detected each group of mountain goats (visible with the naked eye, visible with binoculars or visible only with spotting scope). I then scaled this visibility information into a single value for each survey by weighting the percentage of mountain goats seen in each category (naked eye by 1, binoculars by 2, spotting scope by 3), then summed the total. Habitat use was estimated by recording landscape features where mountain goats were detected. Landscape features recorded were those that may influence the distribution of mountain goats in GNP (Chadwick 1976) and included permanent snow or icefields, ledges, talus-scree-moraine, meadows, shrubs-krummholz, forests, roads, and trails. I identified the dominant landscape feature in which the majority of mountain goats were seen from these data. I chose these covariates and the covariates listed in the previous paragraph because they were the factors most likely to influence detection probability that could also easily be recorded by volunteers.

Surveys were conducted between the second week of June and the last week of October, 2008 and 2009, after parturition and before the rut. During this time mountain goats are more likely to remain within their home ranges (Festa-Bianchet and Côté 2008) and I assumed that the population was closed to changes in occupancy (MacKenzie et al. 2003). My use of sites that were larger than maximum home range size estimates for mountain goats also made the assumption of population closure during sampling periods viable. The goal was for volunteers to conduct ≥ 3 surveys at each site. I sent periodic emails to volunteers to inform them of sites that had been surveyed most recently and sites that needed to be surveyed. Due to the voluntary nature of the program, however, I could not assign survey locations. Because volunteers chose their own survey locations and schedule, and individual volunteers rarely surveyed the same site ≥ 1X, potential sources of heterogeneity from observer and time of day effects were minimized (MacKenzie and Royle 2005).

Biologists in GNP who have >1 year experience monitoring mountain goats conducted observational surveys at a subset of sites to compare to data from volunteers. Biologists conducted ≥ 3 surveys following the same protocols as volunteers at each of 14 sites that were randomly selected from all sites accessible within 1 day of travel (one-way distance). The order of site visits was rotated to avoid the introduction of systematic variation (MacKenzie and Royle 2005). Although mountain goats are most active in the morning (0700-1000) and late afternoon (1500-2000) (Rideout 1977, Singer and Doherty 1985), biologists conducted surveys during times of day that volunteers most commonly conducted surveys. I also asked biologists to conduct surveys at additional sites whenever possible.

Detection probability is rarely constant at all sites and times and not all covariates can be measured, so direct estimation of detection probability is an important part of monitoring (Alldredge et al. 2006). Experience level differences among observers can also influence detection probability ($\hat{P}_i$), biasing abundance estimates (Nichols et al 2000, Genet and Sargent 2003). In this case $\hat{P}_i$ refers to the probability that mountain goats were detected at a site given they were present, rather than the probability that a mountain goat was present at the site. I conducted independent multiple observer surveys (Nichols et al. 2000) . to directly estimate differences in $\hat{P}_i$ between biologists ($\hat{P}_{biol}$) and volunteers ($\hat{P}_{vol}$). Multiple observer approaches enable use of mark-recapture methodology to move point counts from indices to estimates of abundance (Nichols et al 2000, Johnson 2008). Constant use of spotting scopes and binoculars during the 1 hour survey inhibited the ability for observers to cue off detections of others and ensured that observers maintained independence during multiple observer surveys (Nichols et al 2000).

One biologist conducted 76 multiple observer surveys simultaneously with volunteers. I selected a number of volunteers from each experience rank proportionally to the number of volunteers in that experience rank to ensure that $\hat{P}_i$ was measured for volunteers with all levels of experience,. Each biologist also conducted ≥2 multiple observer surveys on 2 separate occasions with each other to measure differences in $\hat{P}_i$ among biologists. I used the Lincoln-Petersen estimator to obtain an estimate of abundance for each survey, then divided the observer's count by that estimate to determine $\hat{P}_i$ for each observer (Nichols et al. 2000).

Results from multiple observations where no mountain goats were seen were omitted from analysis. I used Welch's t-approximation (Welch 1947) to test for differences between volunteers and biologists, and between biologists. I divided high counts at each site from volunteers and from biologists by mean $\bar{P}_i$ for each group ($\bar{P}_{biol}$ for biologists and $\hat{P}_{vol}$ for volunteers) to get corrected counts.

I developed sets of logistic regression models, using data from volunteers and data from biologists to test the importance of covariates with the potential to affect $\hat{P}_{biol}$ and $\hat{P}_{vol}$. Covariates tested in the models were observer experience, size of largest group of mountain goats detected, total number of mountain goats, temperature, binocular power, binocular field of view, scope power, scope field of view, start time of survey, wind speed, and weather. I used Akaike's Information Criteria for small sample sizes

(AIC_c; Burnham and Anderson 2002]) to evaluate models and select the top model. I used R statistical software (http://www.r-project.org/) for all data analysis.

I used the highest count obtained at each site as the count statistic for that site. Density for each site was estimated by dividing the high count of mountain goats by the area of escape terrain at each site. To derive more robust estimates of density at each site from these index counts I divided density estimates by $\hat{P}_i$ for each observer class (i.e. volunteer or biologist). I then divided density estimates by the average density to obtain relative density estimates for each observer class (i.e., volunteer, biologist, and aerial observers). I determined quartiles of relative density estimates for each observer class and assigned each site a density rank of no mountain goats, low (first quartile), moderate (second quartile), high (third quartile), and very high (fourth quartile). I considered relative density estimates between observer classes to be in agreement if they were within 1 quartile of one another.

Because all or nearly all members of a mountain goat group occasionally travel together (Chadwick 1977, Singer and Doherty 1985), these estimates represent the highest density of mountain goats within the viewshed of each site. It is unlikely, however, that all mountain goats occupying a site will be available for detection in the viewshed simultaneously. Therefore, density estimates based on high counts were likely biased low and do not accurately reflect the number of mountain goats inhabiting the surrounding survey area that were not detected.

I also needed estimates that incorporated the probability that a mountain goat was present at the site to enable estimation of abundance beyond the viewshed at each site,. Patterns of detection and non-detection during spatially replicated counts can be used to adjust for biases in counts that are caused by false absences (MacKenzie et al. 2002, MacKenzie and Kendall 2002). I used N-mixture models to derive an estimate of average abundance (λ) and occupancy (ψ) across all survey areas (Royle 2004). I analyzed volunteer data and biologist data separately to estimate λ and ψ for each group. N-mixture models assume that site-specific abundance influences detection (or non-detection) of animals at a site (p), that distribution of animals across survey sites is random, and can be described by a Poisson distribution (Royle and Nichols 2003). An estimate of λ is derived by integrating the binomial probabilities of detecting a certain count of animals at each site over the possible values of abundance for that site (Royle 2004).

I developed a series of models of covariates with the potential to influence λ and p but used only 2009 data; the number of visits to each site in 2008 by biologists was too low (≤ 2) to obtain adequate estimates of p. Covariates tested in relation to p included all of the variables included in models for $\hat{P}_i$ from multiple observer surveys except estimated abundance of mountain goats. Additional covariates tested for p included the percentage of mountain goats that were moving when detected, visibility of mountain goats, and the dominant landscape feature. Covariates with potential to influence λ included viewshed area (km^2), area of escape terrain within viewshed (km^2), area of escape terrain at site (km^2), and number of site visits. I then used AIC to select the top model and evaluated the goodness-of-fit of our fitted model using parametric bootstrapping. I estimated average density of mountain goats in GNP using the following equation:

$$\overline{density}_{GNP} = \frac{\lambda * \ number\ of\ sites\ surveyed}{area\ of\ escape\ terrain\ in\ all\ sites} \qquad (1)$$

To extrapolate abundance to all areas with escape terrain ≥25°, I used the following equation:

$$\hat{N}_{GNP} = \frac{\lambda * number\ of\ sites\ surveyed * area\ of\ escape\ terrain\ in\ GNP}{area\ of\ escape\ terrain\ in\ all\ sites} \quad (2)$$

I conducted N-mixture analysis, model development and model selection using the "Unmarked" package (http://r-forge.r-project.org/projects/unmarked) in R.

We conducted aerial surveys of the entire site (including area outside of the viewshed) at 11 sites (total 450.56 km²; 11% of GNP) to obtain estimates of the number of mountain goats at each site against which to compare our estimates. Montana Fish, Wildlife & Parks personnel, who had extensive experience with aerial mountain goat surveys, conducted aerial surveys by helicopter during 2 days in August 2009 at minimum above ground elevations of 150m. Locations were recorded using GPS for all mountain goats observed. I overlaid all goat locations onto our site viewshed maps to determine the count of mountain goats at each site that were within viewsheds. I developed a density estimate for viewsheds by dividing this count by the summed area of escape terrain in all viewshed surveyed during aerial surveys. I used regression analysis to compare raw counts within site viewsheds from aerial surveys to raw and corrected high counts from volunteers and biologists. I estimated density from aerial surveys for the entire survey area by dividing the sum of counts at each site by the area of escape terrain at all sites surveyed during aerial surveys. I compared this density estimate to N-mixture model density estimates from volunteers and biologists to determine whether the aerial survey estimate fell within the confidence intervals of either estimate. Other studies have reported detection probabilities of mountain goats during aerial surveys from 0.55 to 0.84 (Gonzalez-Voyer et al. 2001) and 0.75 to 0.91 (Rice et al. 2009).

The long-term goal of the HCCS program is to detect trends in mountain goat populations. I conducted a power analysis to determine whether the levels of precision in λ and standard error of λ from estimates by volunteers and biologists were adequate to detect a population decline. The International Union for Conservation of Nature criteria for upgrading mountain goats from their current status as least concern to vulnerable include a 30% level of reduction in population size over 3 generations, or 10 years (Shackleton 1997). I used methods proposed by Field et al. (2005) to assess the power of surveys by volunteers and biologists for identifying a population decline of 30% over 10 years.

RESULTS

During 2008 and 2009, 140 volunteers were trained and 104 volunteers conducted ≥1 survey. Volunteers and biologists spent 4,401.3 and 1,219.1 hr., respectively, conducting mountain goat surveys (Table 1). Experience ranks of biologists ($\bar{x}$ = 34.9 ± 1.64 [SD], range = 33- 37) were higher and less varied than those of volunteers ($\bar{x}$ = 24.5 ± 7.09, range = 10- 39). All biologists were in the skilled experience rank. The proportion of volunteers in each experience rank (including volunteers who conducted ≥ 1 survey) varied: novice = 27%, some relevant experience = 31%, moderately skilled = 27%, skilled = 15%. The proportion of volunteers in each experience rank that accepted our invitation to conduct multiple observer surveys also varied: novice = 20%, some relevant experience = 38%, moderately skilled = 35%, skilled = 7%.

Mean $\hat{P}_i$ for GNP biologists and the GNP biologist who conducted multiple observer surveys with volunteers was not significantly different (t = 1.5344, df = 35.2, *P* = 0.93). Therefore I combined $\hat{P}_i$ from all multiple observer surveys by biologists into mean $\hat{P}_{biol}$ for all biologists. Mean $\hat{P}_{biol}$ (0.809 ± 0.249) was significantly higher (t = 3.1609, df = 81.5, P = 0.001) than mean $\hat{P}_{vol}$ (0.647 ± 0.317). No misidentifications of other species as mountain goats or other false positives were reported during multiple observer surveys. Examination of verification photos made during single observer surveys by volunteers and biologists showed no evidence of misidentifications or false positives. Verification photos were submitted for 15% of groups of mountain goats detected by volunteers and for 74% of groups of mountain goats detected by biologists.

Group size of the largest group of mountain goats (GroupSize), and total number of mountain goats (TotalGoats) were the most influential covariates for predicting detection probability by volunteers and biologists (Table 2). A model for $\hat{P}_{biol}$ adding observer experience (ExpRank) to the model had the lowest AIC score, but was competing with the 2 parameter model. I considered the contribution of observer experience to be negligible, however, because it was not a significant predictor (P = 0.237), and inclusion in models resulted in marginal reductions of residual deviance (< 2) compared to competing models. In the best supported models, $\hat{P}_{vol}$ increased by 0.16 with an increase in GroupSize and decreased by 0.09 with an increase in TotalGoats (Table 4), and $\hat{P}_{biol}$ increased by 0.24 with an increase in GroupSize and decreased by 0.13 with an increase in TotalGoats (Table 5). Inclusion of GroupSize and TotalGoats as predictors of $\hat{P}_{vol}$ resulted in a 41% reduction in residual deviance and for $\hat{P}_{biol}$ a 42% reduction in residual deviance. Both models underestimated $\hat{P}_i$ at high values of the combined predictor variables.

Raw high counts (uncorrected for $\hat{P}_i$) from volunteers had a strong statistical relationship in 2008 with counts from biologists in 2008 (R^2_{adj} = 0.66, df = 19, P < 0.001) but not in 2009 (R^2_{adj} = 0.38, df = 23, P < 0.001) (Fig. 2). Density estimates followed the same pattern (Fig. 3). In 2008, 64% of the variation in density estimates by volunteers was explained by density estimates by biologists (R^2_{adj} = 0.64, df = 19, P < 0.001) while in 2009 the correlation was 49% (R^2_{adj} = 0.49, df = 23, P < 0.001). However, when 2009 density estimates for sites surveyed < 3 times by volunteers and biologists were excluded, and 1 strong leverage point (Cook's distance > 1.0) removed, regression results had high explanatory power (R^2_{adj} = 0.84, df = 11, P < 0.001). Raw counts from aerial surveys were not correlated with estimates from 2009 corrected high counts by volunteers (R^2_{adj} = 0.20, df = 8, P < 0.06) nor with estimates by biologists (R^2_{adj} = 0.07, df = 6, P = 0.26) (Fig. 4). Density estimates for aerial surveys were also poorly correlated with 2009 density estimates from volunteers (R^2_{adj} = 0.47, df = 8, P < 0.001) and biologists (R^2_{adj} = 0.45, df = 6, P < 0.04).

Aerial counts in survey viewsheds and 2009 density estimates from uncorrected high counts by volunteers and biologists in survey viewsheds were similar (1.99 mountain goats/ km², 1.91 mountain goats/ km² and 1.87 mountain goats/ km² respectively). Mean density estimates for all escape terrain at sites, based on corrected high counts by volunteers (0.54 to 0.72 mountain goats/ km²) and biologists (0.48 to 0.55 mountain goats/ km²) were lower than density estimates from the aerial survey counts (0.95 mountain goats/ km²). Density estimates by volunteers were higher and more variable than density estimates by biologists, but confidence intervals overlapped. When calculated only for sites visited ≥ 3 times in 2009, however, density estimates by volunteers remained nearly the same (0.56 to 0.74 mountain goats/km²), but density estimates by biologists were higher (0.71 to 0.8 mountain goats/km²). Relative density estimates (Table 5 and Fig. 5) were in agreement between volunteers and biologists at 18 of 25 sites, between volunteers and aerial surveys at 10 of 11 sites, between biologists and aerial surveys at 6 of 9 sites, and between all 3 (volunteers, biologists and aerial) at 5 of 9 sites.

The best supported N-mixture model for 2009 volunteer data included number of site visits (SiteVisits) as a predictor of mean abundance at sites (λ), and group size of largest group of mountain goats (GroupSize), and landscape feature in which the majority of mountain goats were seen (DomFeat) as predictors of detection of mountain goats at sites (p). In the top model for 2009 data from biologists, area of escape terrain within viewshed (ViewshedEscape) was the most influential predictors of λ. GroupSize and DomFeat were also influential predictors, but the addition of visibility of mountain goats (Visibility) as an additional predictor of p improved the model performance. Goodness-of-fit tests for the selected models for volunteer data ($P = 0.96$) and biologist data ($P = 0.455$) yielded small differences between observed residual deviance and expected residual deviance indicating that the N-mixture models fit the data. N-mixture models were not developed for 2008 data due to the low number of site visits by biologists.

The best N-mixture model for 2009 volunteer data (Table 6) estimated abundance at 23.44 to 32.3 mountain goats/ site ($\lambda = 27.52 \pm 2.25$ [SE], $p = 0.06 \pm 0.41$, $\psi = 0.96$). Multiplying λ by the number of sites surveyed ($N = 32$) and dividing by escape terrain at sites (km^2) yielded a density estimate of 1.23 ($\pm$ 0.195) mountain goats/km^2. The best N-mixture model for 2009 data from biologists estimated abundance at 26.13 to 45.51 mountain goats/site ($\lambda = 32.95 \pm 3.89$, $p = 0.094 \pm 0.17$, $\psi = 0.97$), yielding a density estimate of 1.56 ($\pm$ 0.42) mountain goats/km^2 ($N = 25$). Extrapolating these density estimates from volunteer and biologists models to all escape terrain in GNP yielded an estimate of 1,705 to 2,349 mountain goats by volunteers and or 1,885 to 3,269 mountain goats by biologists.

The N-mixture density estimate by biologists and volunteers were higher than the aerial survey estimate (0.95 mountain goats/ km^2). Estimates by volunteers and biologists overlapped the estimate of 1.16 mountain goats/ km^2 by Chadwick (1977). Volunteer and biologists estimates were lower than the estimate of 2.9 mountain goats/ km^2 by Singer and Doherty (1985).

I conducted a power analysis using N-mixture model estimates of λ and SE, and assuming no change in SE in future years of monitoring to evaluate the probability of detecting a population change of 30% over 10 years. When significance level (α) was 0.05, the power to detect a 30% decline over 10 years by volunteers was 0.28 and for biologists was 0.12. Increasing α increased the power for volunteers and biologists [α = 0.1, power (volunteers) = 0.40, power (biologists) = 0.20; α = 0.2, power (volunteers) = 0.56, power (biologists) = 0.33], but did not near the goal to detect a decline with high probability (e.g., power > 0.8).

DISCUSSION

Developing baseline estimates of species density or abundance over entire management areas, (e.g., National Parks) requires large sample sizes, often precluding the use of traditional, but more costly, monitoring strategies (e.g., monitoring by biologists, aerial surveys, mark-recapture). The use of volunteers in monitoring programs offers 1 solution, but few programs are able to assess the relative quality data of data collected by biologists and volunteers (Fitzpatrick et al. 2009, Kindberg et al. 2009). I compared population estimates of mountain goats from data collected by citizen science volunteers at sites throughout GNP to population estimates from data collected by biologists at a subset of sites. Our conclusions are broadly applicable to citizen monitoring programs, and have particular relevance to programs monitoring common and highly visible species.

Population estimates from citizen science data were similar to those from data collected by biologists. Uncorrected counts in survey viewsheds by volunteers and biologists were similar and were close to counts from aerial surveys. Confidence intervals of our density estimates from high counts by volunteers and biologists overlapped, despite a lower mean detection probability by volunteers. Density estimates by volunteers and biologists from N-mixture models also overlapped each other and those from earlier research (Chadwick 1977). Volunteer estimates provide similar baseline information compared to biologists for planning future monitoring and research.

A few discrepancies in the estimates raise important considerations. The average density and relative abundance estimates were considerably lower than estimates by Singer and Doherty (1985)(i.e., 1.03- 1.42 mountain goats/ km^2 versus 2.9 mountain goats/ km^2). This may be partially explained by the small area encompassed by their study area and its proximity to a heavily-used mineral lick. Conversely, our N-mixture model estimates were higher than aerial survey counts. I found no evidence of misidentification of mountain goats to suggest that our estimates were falsely inflated. A simple explanation may be that detection probability during aerial surveys was less than perfect. I did not measure detection probability during aerial surveys but other studies have reported detection probabilities from 0.55 to 0.84 (Gonzalez-Voyer et al. 2001) and

0.75 to 0.91 (Rice et al. 2009). Using this range (i.e., 0.55- 0.91) as a hypothetical correction factor for our aerial survey count yields an estimate of 1.03 to 1.38 mountain goats/ km^2 which overlaps our N-mixture density estimates by volunteers and biologists (Fig. 6). An alternative explanation for the discrepancy between our estimates and aerial counts is that that our assumption of closure may have been violated. Mountain goat home range sizes (e.g., from 6.3 km^2 [Singer and Doherty 1985] to 24 km^2 [Rideout 1977]) vary widely throughout their range. Due to a lack of specific information about mountain goat home range sizes and locations in GNP, I assumed a rectangular home range. Home ranges may, in fact, be more linear, reflecting fidelity to escape terrain (Brandborg1955, Hamel and Côté 2007). If home ranges are in fact linear, mountain goats may have moved from one survey site to another, potentially inflating population estimates from ground counts. Ensuring closure using linear home ranges would require a more terrain-specific approach with specific knowledge about movement patterns.

Observer experience for volunteers was lower and more varied than for biologists, as I expected, but did not correlate with detection probability or directly influence estimates. The proportion of volunteers in the novice and skilled experience ranks who accepted our invitations to conduct multiple observer surveys was lower than the proportion of overall volunteers in these experience ranks. This potential underrepresentation of volunteers at the lower and upper ends of observer experience in our multiple observer sample may have negatively affected the correlation between experience and detection probability.

Other citizen science programs use scores from observer experience to weight data from volunteers (Silvertown 2009). This approach may be effective if skill level is correlated with experience, but my results suggest that weighting data in this manner may not be valid and I caution against this practice. The adequacy of the chosen tool for measuring observer experience is also an important consideration for other programs. In our study, low correlation between experience and estimates of detection probability and abundance may have resulted from failure of our participant survey to accurately measure experience level. A few relevant metrics of experience were not included because they are difficult to quantify (e.g., the degree to which volunteers have a search image, the amount of investment volunteers have in surveying). Factoring in the number of surveys conducted by each observer may also improve the measurement of experience rank. Observer bias, such as lower detection probability, generally decreases as observers become more experienced (Delaney 2008) and number of surveys may provide a promising avenue for exploring the relationship between experience and observer bias.

I used multiple observer surveys to correct for observer variation between volunteers and biologists and to calibrate our indices of abundance. Incorporating detection probability measured by multiple observer approaches has been proposed as a correction factor for data with high inter-observer variation (Nichols et al. 2000, Alldredge et al. 2006). Such a correction factor could be useful in calibrating data from citizen science programs to make it comparable to data from biologists. Correcting data according to detection probability differences is only valid, however, if it is consistent across survey conditions, or if the variation can be adequately modeled.

My results suggest that detection probabilities from multiple observer surveys are not sufficient to correct counts because they are not consistent across all levels of mountain goat abundance. If mean detection probability is used as a correction factor for volunteer data in future monitoring, abundance of mountain goats would be overestimated when group size was large and underestimated when the total number of mountain goats was high. The models estimated to explain the variation provided some insight into the factors that influenced detection probability, but failed to explain most of the variation, reducing residual deviance by <50%. This suggests that unmeasured variables (e.g., volunteer effort) may have contributed to differences in detection probability. Detection probabilities that were inconsistent across mountain goat densities may explain the low correlation between density estimates by volunteers and biologists, but this fails to explain the low correlation between high counts by volunteers and biologists that were not corrected by detection probability.

To obtain estimates of true abundance I used N-mixture models, which incorporated the probability that a mountain goat was present at the site, but was not detected because it was either outside of the viewshed during the survey period or missed due to observer error. Similar to multiple observer surveys, N-mixture models resulted in lower detection probabilities and more variation among volunteers than among biologists. Detection probability in N-mixture models for volunteer and biologist data was again influenced by mountain goat group size. Habitat use also influenced detection probability for volunteers and biologists while visibility was influential only for biologist data. Habitat use and visibility may have similarly influenced multiple observer models but these parameters were not tested because data was not available for 2008.

N-mixture model estimates were higher than estimates from high counts in viewsheds that were corrected for detection probability, most likely because not all mountain goats were available for detection in viewsheds during survey periods. The high probability of occurrence ($\geq$0.96) and low detection probability ($\leq$0.094) estimated by N-mixture models confirm this explanation and suggest that mountain goats inhabiting sites were frequently absent from survey viewsheds. Abundance estimates by biologists were influenced by area of escape terrain in viewsheds, a result supported by previous research that identified escape terrain as the best predictor of mountain goat occurrence (Gross et al. 2002, Hamel and Côté 2007). While the area of escape terrain in viewsheds also influenced abundance estimates by volunteers in several of the highest ranking models, the number of site visits alone best explained the variation and had the largest influence on abundance estimates by volunteers. Other studies have reported that volunteers underestimate abundance due to lower detection probability (Newman et al. 2003, Delaney et al. 2008). In our study, however, the effect of lower detection probability on abundance estimates by volunteers was balanced out by the larger number of site visits. Variation in abundance estimates was lower for volunteers than for biologists because volunteers surveyed more sites more frequently.

The number of site visits by volunteers did not affect detection probability during multiple observer surveys, but it may explain why uncorrected high counts and density estimates by volunteers and biologists were not more highly correlated. By surveying sites more often, volunteers captured a higher minimum count of mountain goats at >50% of sites. The larger variation in detection probability among volunteers, however, led to volunteers reporting a larger proportion of counts that were lower than biologist counts. Therefore, N-mixture model estimates by biologists were higher despite higher minimum counts by volunteers.

When large sample sizes can be obtained through large spatial and temporal coverage, citizen science can provide mountain goat population estimates that are statistically equivalent to those of biologists. However, to be useful for long-term monitoring, baseline estimates must have statistical power to detect significant population changes (Field et al. 2005). In our study, statistical power of estimates by volunteers and biologists were not sufficient to detect a 30% decline in mountain goat population size over 10 years. Because I had no measure of variation and was unable to quantify the power of aerial surveys, I cannot comment on the likelihood of detecting such a decline using this method. As a general rule, aerial surveys that do not incorporate some measure of detection probability (e.g., distance sampling, double observer surveys, repeat surveys; Gonzalez-Voyer et al. 2001) have low power to accurately detect changes in population size (Rice et al. 2009).

Optimization of our survey design may increase the power to detect mountain goat population trends. To attain power of 0.8 while maintaining the current level of survey effort, the standard error for volunteer estimates of abundance would need to decrease to 1.502. This may be attainable with a greater number of volunteers across more sites. Recruiting more volunteers may be an option for increasing sample size but would require additional funds for volunteer coordination. Alternatively, greater precision and higher power could be achieved by reducing the number of sites and increasing surveys/site (MacKenzie and Royle 2005). This could be particularly effective when probability of occupancy is high but detection probability low (MacKenzie and Royle 2005), as is the case with common but wide-ranging species.

The capacity for citizen science programs to continue monitoring over a long period of time could be harnessed as an additional means of increasing power. Funding for monitoring of species that are not in imminent danger of extinction is often limited. When resources are scarce and population declines may not be evident immediately, the long-term nature of citizen science is particularly advantageous. If monitoring at the current spatial and temporal scale continues for 30 years, citizen science volunteers will have an 80% likelihood ($\alpha = 0.1$) of detecting a 30% decline in mountain goat populations. For species that do not require management action over the short-term, 30 years may represent a suitable time frame for conservation. When properly designed by incorporating detection probability by volunteers, citizen science programs can contribute to long term population monitoring.

MANAGEMENT IMPLICATIONS

Wildlife managers faced with limited funding to meet their monitoring needs are increasingly turning to the free labor source provided by the public (Silvertown 2009), but establishing and coordinating of a citizen science program requires financial commitment and effort (Yung 2007). Managers must determine which will better meet their conservation objectives: hiring a citizen science project manager to coordinate volunteers to cover a larger sample area, or enlisting a small number of biologists to cover a smaller sample area. Our results suggest that the 2 methods may yield statistically similar population estimates if enough data are collected by volunteers. Further studies comparing citizen science and professional approaches will help to establish the generality of our results.

Citizen science programs involved with long-term monitoring should incorporate some measure of data quality. The cost of employing biologists or using other methods (e.g., mark-recapture) limits comparisons of data on a similar scale to the data that can be collected by volunteers. The use of multiple observer surveys to correct volunteer data may not be viable for citizen science data due to the high variability in detection probability. However, double sampling using data collected by biologists or data from aerial surveys over a smaller subsample offers a useful comparison providing that enough data are collected to measure detection probability. Data quality comparisons will likely be most effective once the program has been established (e.g., ≥1 year after initiation), because managers can then determine the scale of data collection by volunteers.

Because citizen science population estimates from small sample sizes are not comparable to biologist estimates I do not recommend citizen science as a direct substitute for professional monitoring. Citizen science will only produce similar populations estimates to those of biologists when sample sizes are larger those attainable by biologists. I reported that abundance estimates were positively influenced by number of site visits and that variation was negatively influenced by number of sites surveyed. Future research on the number of site visits and number of surveys at each site that maximize precision of citizen science estimates would contribute toward increasing the power of detecting trends.

LITERATURE CITED

Alldredge, M.W., K. H. Pollock, and T. R. Simons. 2006. Estimating detection probabilities from multiple observer point counts. The Auk 123: 1172-1182.

Burnham, K.P., and D.R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach, 2nd ed. Springer-Verlag, New York, New York, USA.

Brandborg, S. M. 1955. Life history and management of the mountain goat in Idaho. Idaho Wildlife Bulletin 2:1–142.

Chadwick, D. H. 1976. Ecological relationships of mountain goats in Glacier National Park. American Institute of Biological Sciences First Conference on Scientific Research in the National Parks, New Orleans, Louisiana, USA.

Chadwick, D. H. 1977. The influence of mountain goat social relationships on population size and distribution. Pages 74- 91 *in* Proceedings of the first international mountain goat symposium. 19 February 1977. Kalispell, Montana, USA.

Cohn, J. P. 2008. Citizen science: can volunteers do real research? BioScience. 58:192-197.

Cooper, C. B.; J. Dickinson, T. Phillips, R. Bonney. 2007. Citizen science as a Tool for Conservation in Residential Ecosystems. Ecology and Society 12:11.

Côté, S. D., M. Festa-Bianchet, and F. Fournier. 1998. Life-history effects of chemical immobilization and radiocollars on mountain goats. Journal of Wildlife Management 62:745–52.

Danielsen, F., N.D. Burgess, A. Balmford. 2005. Monitoring matters: examining the potential of locally-based approaches. Biodiversity and Conservation 14:2507–2542.

Delaney, D.G., C.D. Sperling, C.S. Adams, and B. Leung. 2008. Marine invasive species: validation of citizen science and implications for national monitoring networks. Biological Invasions 10:117–128.

Dunn, E. H., C. M. Francis, P. J. Blancher, S. R. Drennan, M. A. Howe, D. Lepage, C. S. Robbins, K. V. Rosenberg, J. R. Sauer, and K. G. Smith. 2005. Enhancing the value of the Christmas bird count. The Auk 122:338-346.

Ericsson, G., and K. Wallin. 1999. Hunter observations as an index of moose *Alces alces* population parameters. Wildlife Biology 5: 177-185.

Festa-Bianchet. M., and S. D. Côté. 2008. Mountain goats: ecology, behavior and conservation of an alpine ungulate. Island Press, Washington, D.C., USA.

Field, S. A., A. J. Tyre, and H. P. Possingham. 2005. Optimizing allocation of monitoring effort under economic and observation constraints. Journal of Wildlife Management 69:473–482.

Fitzpatrick, M. C., E. L Preisser, A. M Ellison, J. S. Elkington. 2009. Observer bias and the detection of low-density populations. Ecological Applications 19:1673-1679.

Genet, K.S and L.G. Sargent. 2003. Evaluation of methods and data quality from a volunteer-based amphibian call survey. Wildlife Society Bulletin 31: 703-714.

Gonzalez-Voyer, A., M. Festa-Bianchet, M. and K. G. Smith. 2001. Efficiency of aerial surveys of mountain goats. Wildlife Society Bulletin 29: 140-144.

Greenwood, J. J. D. 2007. Citizens, science and bird conservation. Journal of Ornithology 148: 77–124.

Gross, J. E., M. C. Kneeland, D. F. Reed, and R. M. Reich. 2002. GIS-based habitat models for mountain goats. Journal of Mammology 83:218–228.

Hamel, S., and S. D. Côté. 2007. Habitat use patterns in relation to escape terrain: are alpine ungulate females trading off better foraging sites for safety? Canadian Journal of Zoology 85:933-943.

Haynes, L. A. 1992. Mountain goat habitat of Wyoming's Beartooth Plateau: implications for management. Biennial Symposium of the Northern Wild Sheep and Goat Council 8:325–339.

Hochachka, W. M; K. Martin, F. Doyle, and C. J. Krebs. 2000. Monitoring vertebrate populations using observational data. Canadian Journal of Zoology 78: 521–529.

Hop, K., M. Reid, J. Dieck, S. Lubinski, and S. Cooper. 2007. U.S. Geological Survey-National Park Service Vegetation Mapping Program: Waterton-Glacier International Peace Park. U.S. Geological Survey, Upper Midwest Environmental Sciences Center, La Crosse, Wisconsin, USA.

Johnson, D.H. 2008. In defense of indices: the case of bird surveys. Journal of Wildlife Management 72:857–868.

Kindberg, J., G. Ericsson, J. E. Swenson. 2009. Monitoring rare or elusive large mammals using effort-corrected voluntary observers. Biological Conservation 142: 159-165.

Koeth, Chrissy. 2008. Clinging to existence. Montana Outdoors. March-April: 23-27.

Krausman, P. R., L. K. Harris, C. L Blasch, K. K. G. Koenen, and J. Francine. 2004. Effects of military operations on behavior and hearing of endangered Sonoran pronghorn. Wildlife Monographs 157: 1-41.

MacKenzie, D.I., and W.L. Kendall. 2002. How should detection probability be incorporated into estimates of relative abundance? Ecology 83: 2387-2393.

MacKenzie, D.I., J.D. Nichols, G.B. Lachman, S. Droege, J.A. Royle, and C.A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83: 2248–2255.

MacKenzie, D.I., J.D. Nichols, J.E. Hines, M.G. Knutson, and A.B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84: 2200–2207.

MacKenzie, D.I, and J. A. Royle. 2005. Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology 42:1105-1114.

McFetridge, R. J. 1977. Strategy of resource use by mountain goat nursery groups. Pages 169-173 *in* Proceedings of the first international mountain goat symposium. 19 February 1977. Kalispell, Montana, USA.

Martin, S. R. 1997. Specialization and differences in setting preferences among wildlife viewers. Human Dimensions of Wildlife 2: 1-18.

Morisette, J.T., A.D. Richardson, A.K. Knapp, J.I. Fisher, E.A. Graham, J. Abatzoglou, B.E. Wilson, D.D. Breshears, G.M. Henebry, J.M. Hanes, and L. Liang. 2009. Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. Frontiers in Ecology and the Environment 7: *253-260.*

Newman, C., C. D. Buesching, and D. W. Macdonald. 2003. Validating mammal monitoring methods and assessing the performance of volunteers in wildlife conservation—''*Sed quis custodiet ipsos custodies* ?'' Biological Conservation 113: 189-197.

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

Pettorelli, N., F. Pelletier, A. Von Hardenberg, M. Festa-Bianchet, S.D. Côté. 2007. Early onset of vegetation growth vs. rapid green-up: impacts on juvenile mountain ungulates. Ecology 88: 381–390.

Pilz, D., H. Ballard, and E. T. Jones. 2005. Broadening participation in biological monitoring: guidelines for scientists and managers. Institute for Culture and Ecology. Portland, Oregon, USA.

Rice, C. G., K.J. Jenkins, and W. Chang. 2009. A sightability model for mountain goats. Journal of Wildlife Management 73(3): 469-478.

Rideout, C. B. 1977. Mountain goat home ranges in the Sapphire Mountains of Montana. Pages 201-211 *in* Proceedings of the first international mountain goat symposium. 19 February 1977. Kalispell, Montana, USA.

Root, T.L. and P. Alpert. 1994. Volunteers and the NBS. Science 263: 1205.

Royle, J. A. 2004. N-Mixture models for estimating population size from spatially replicated counts. Biometrics 60: 108-115.

Royle, J. A. and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology 84: 777-790.

Scott, D., R. B. Ditton, J. R. Stoll, and T. L Eubanks, Jr. 2005. Measuring specialization among birders: utility of a self-classification measure. Human Dimensions of Wildlife 10:53-74.

Shackleton, D. M. (ed.) and the International Union for Conservation of Nature/ Status Survey and Conservation Caprinae Specialist Group. 1997. Wild Sheep and Goats and their Relatives. Status Survey and Conservation Action Plan for Caprinae, International Union for Conservation of Nature, Gland, Switzerland and Cambridge, United Kingdom.

Silvertown, J. 2009. A new dawn for citizen science. Trends in Ecology and Evolution 24: 467-471.

Singer, F. J., and J. L. Doherty. 1985. Movements and habitat use in an unhunted population of mountain goats, *Oreamnos americanus*. Canadian Field–Naturalist 99: 205–17.

Smith, B.L. Criteria for determining age and sex of American mountain goats in the field. 1988. Journal of Mammalogy 69: 395-402.

Trumbull, D. J., R. Bonney, D. Bascom, and A.Cabral. 2000. Thinking scientifically during participation in a citizen-science project. Science Education 84:265–275.

Varley, N. C. 1994. Summer-fall habitat use and fall diets of mountain goats and bighorn sheep in the Absaroka Range, Montana. Biennial Symposium of the Northern Wild Sheep and Goat Council. 9:131- 138.

Veitch, A., E. Simmons, M. Promislow, D. Tate, M. Sallow, and R. Popko. 2002. The status of mountain goats in Canada's Northwest Territories. Northern Wild Sheep and Goat Council 13: 49-62.

Welch, B. L. 1947. The generalization of "Student's" problem when several different population variances are involved. Biometrika 34: 28–35.

Yoccoz, N.G., J. D. Nichols, and T. Boulinier. 2003 Monitoring of biological diversity—a response to Danielsen et al. Oryx 37: 412.

Yung, L. 2007. Citizen monitoring and restoration: volunteers and community involvement in wilderness stewardship. United States Department of Agriculture Forest Service Proceedings. Rocky Mountain Research Station-P-49.

Table 1: Site visits and surveys conducted by volunteers and biologists in Glacier National Park, Montana for mountain goat population estimates.

Year	Observer	Sites visited	No. surveys	$\bar{x}$ no. site visits	SD	Sites surveyed ≥3 times
2008	Volunteers	30	132	5.95	2.6	22
2008	Biologists	21	33	1.84	0.62	0
2009	Volunteers	32	197	7.24	2.85	31
2009	Biologists	25	76	4.78	2.94	14

Table 2: Logistic regression models for detection probability of volunteers ($\hat{P}_{vol}$) and biologists ($\hat{P}_{biol}$) during multiple observer surveys for mountain goats in Glacier National Park, Montana. Variables included in models are observer experience (ExpRank), size of largest group of mountain goats detected (GroupSize), total number of mountain goats (TotalGoats), temperature (Temp), binocular power (BinocPower), binocular field of view (BinocView), scope power (ScopePower), scope field of view (ScopeView), start time of survey (StartTime), wind speed (WindSpeed), number of visits to site by year and observer (SiteVisits), and weather (SkyCover). All variables are continuous except for SkyCover which is a factor variable with 5 levels.

Model for volunteers	K	AIC_c	Δ_i	w_i
GroupSize, TotalGoats	3	46.99	0.00	0.556
GroupSize, TotalGoats, SiteVisits	4	48.05	1.07	0.326
ExpRank, GroupSize, TotalGoats, SiteVisits	5	50.25	3.26	0.109
ExpRank, GroupSize, BinocPower, BinocView, TotalGoats, SiteVisits	7	55.79	8.80	0.007
ExpRank, GroupSize, BinocPower, BinocView, WindSpeed, TotalGoats, SiteVisits	8	59.46	12.47	0.001
TotalGoats	2	60.97	13.98	0.001
Model for biologists	**K**	**AIC_c**	**Δ_i**	**w_i**
ExpRank, GroupSize,TotalGoats	4	72.75	0.00	0.338
GroupSize,TotalGoats	3	73.13	0.38	0.279
ExpRank, GroupSize,TotalGoats, SiteVisits	5	73.14	0.39	0.277
ExpRank, GroupSize,WindSpeed, TotalGoats, SiteVisits	6	75.76	3.02	0.075
ExpRank, GroupSize,Temp, WindSpeed, TotalGoats, SiteVisits	7	78.06	5.32	0.024
ExpRank, GroupSize,Temp, BinocPower, BinocView, WindSpeed, TotalGoats, SiteVisits	9	80.16	7.42	0.008

Table 3: Parameter estimates, errors and deviance for best model for detection probability of volunteers ($\hat{P}_{vol}$) during multiple observer surveys for mountain goats in Glacier National Park, Montana.

$\hat{P}_{vol}$ ~ GroupSize+ TotalGoats				
Coefficients	Estimate	SE	z value	Pr(>|z|)
Intercept	0.55931	0.2962	1.888	0.059
GroupSize	0.16067	0.05951	2.7	0.0069
TotalGoats	-0.09027	0.03231	-2.793	0.0052
Null deviance: 26.943 on 49 degrees of freedom				
Residual deviance: 15.894 on 47 degrees of freedom				

Table 4: Parameter estimates, errors and deviance for best model for detection probability of biologists ($\hat{P}_{biol}$) during multiple observer surveys for mountain goats in Glacier National Park, Montana.

$\hat{P}_{biol}$ ~ GroupSize+ TotalGoats				
Coefficients	Estimate	SE	z value	Pr(>|z|)
Intercept	0.97	0.26165	3.719	0.0002
GroupSize	0.24	0.06972	3.486	0.0005
TotalGoats	-0.13	0.03817	-3.483	0.0005
Null deviance: 47.764 on 105 degrees of freedom				
Residual deviance: 27.844 on 103 degrees of freedom				

Table 5: Relative density estimates of mountain goats from corrected 2009 high counts from volunteers and biologists and raw counts from aerial surveys in Glacier National Park, Montana.

Site	volunteer relative density estimate	biologist relative density estimate	aerial relative density estimate
Apikuni Falls	very high	very high	--
Autumn Creek	very high	very high	--
Avalanche Lake	moderate	moderate	--
Beaver Woman	high	moderate	high
Boulder Pass	low	--	moderate
Coal Creek	low	high	--
Cobalt Lake	moderate	moderate	--
Cut Bank	very high	moderate	--
Elizabeth Lake	very high	moderate	very high
Fifty Mountain	high	--	--
Firebrand Pass	moderate	no goats	--
Grace Lake	moderate	no goats	--
Gunsight Pass	high	high	very high
Harrison Lake	moderate	no goats	moderate
Haystack Butte	very high	very high	--
Iceberg Lake	very high	very high	--

Janet Lake	moderate	--	low
Numa Lookout	low	no goats	no goats
Ole Creek	low	no goats	--
Otokomi Lake	moderate	moderate	--
Park Creek	no goats	--	--
Pitamakin Pass	high	--	--
Poia Lake	no goats	very high	very high
Preston Park	moderate	moderate	--
Red Eagle	high	moderate	--
Scenic Point	no goats	--	--
Siyeh Pass	moderate	moderate	--
Sperry Chalet	moderate	moderate	moderate
Triple Divide	high	very high	moderate
Trout Lake	no goats	low	low
Upper Kintla	no goats	--	--
Upper Nyack	low	moderate	--

Table 6: N-mixture models for mountain goat abundance from 2009 volunteer survey data and 2009 biologist survey data from Glacier National Park, Montana. Detection probability covariates for N-mixture models were observer experience (ExpRank), size of largest group of mountain goats detected (GroupSize), temperature (Temp), binocular power (BinocPower), binocular field of view (BinocView), scope power (ScopePower), scope field of view (ScopeView), start time of survey (StartTime), wind speed (WindSpeed), weather (SkyCover) , percentage of mountain goats seen that moving when detected (PercentMoving), visibility of mountain goats (Visibility), and landscape feature in which the majority of mountain goats were seen (DomFeat). Abundance covariates with potential to influence λ were viewshed area in km² (Viewshed), area of escape terrain ≥25° within viewshed (ViewshedEscape), area of escape terrain ≥25° survey site (SiteEscape), and number of site visits (SiteVisits). Detection probability of mountain goats is denoted in the column labeled *p*.

Models from 2009 volunteer data	K	AIC	Δ_i	w_i	λ	SE(λ)	*p*
~GroupSize + DomFeat ~ SiteVisits	10	462.20	0.00	0.40	27.52	2.25	0.06
~GroupSize + DomFeat ~ SiteVisits +Viewshed	11	464.18	1.98	0.15	27.54	2.26	0.06
~GroupSize + DomFeat ~ SiteVisits + Viewshed + ViewshedEscape	12	465.90	3.70	0.06	27.39	2.27	0.06
~GroupSize + DomFeat + StartTime + WindSpeed + ExpRank + SkyCover + Visibility + Temp + PercentMov + Visibility ~ SiteVisits + Viewshed + ViewshedEscape + SiteEscape	26	469.92	7.72	0.01	24.48	2.08	0.05
~GroupSize + DomFeat + StartTime + WindSpeed + ExpRank +							

	K	AIC	Δ_i	w_i	λ	SE(λ)	*p*
SkyCover + Visibility + Temp + PercentMov + BinocPower + BinocView + ScopePower + ScopeView ~ SiteVisits + Viewshed + ViewshedEscape + SiteEscape	30	473.79	11.59	0.00	23.64	2.04	0.06
Null	2	1687.85	1225.65	0.00	27.39	2.28	0.12
Models from 2009 biologist data	K	AIC	Δ_i	w_i	λ	SE(λ)	*p*
~GroupSize + DomFeat + Visibility ~ ViewshedEscape	9	264.13	0.20	0.26	32.95	3.89	0.09
~GroupSize + DomFeat + Visibility ~ ViewshedEscape + SiteVisits + SiteEscape + Viewshed	12	267.21	3.28	0.06	31.58	3.78	0.09
~GroupSize+ DomFeat ~ ViewshedEscape	8	272.43	8.50	0.00	34.36	3.99	0.14
~GroupSize + DomFeat + Visibility + Temp + SkyCover+ StartTime +WindSpeed +ExpRank + SkyCover + PercentMov ~ ViewshedEscape + SiteVisits + SiteEscape + Viewshed	24	279.60	15.67	0.00	32.52	4.12	0.09
~GroupSize + DomFeat + Visibility + Temp + SkyCover+ StartTime +WindSpeed +ExpRank + SkyCover + PercentMov + BinocPower + BinocView + ScopePower + ScopeView ~ ViewshedEscape + SiteVisits + SiteEscape + Viewshed	28	284.48	20.55	0.00	31.56	4.09	0.09
Null	2	739.94	476.01	0.00	22.87	2.84	0.26

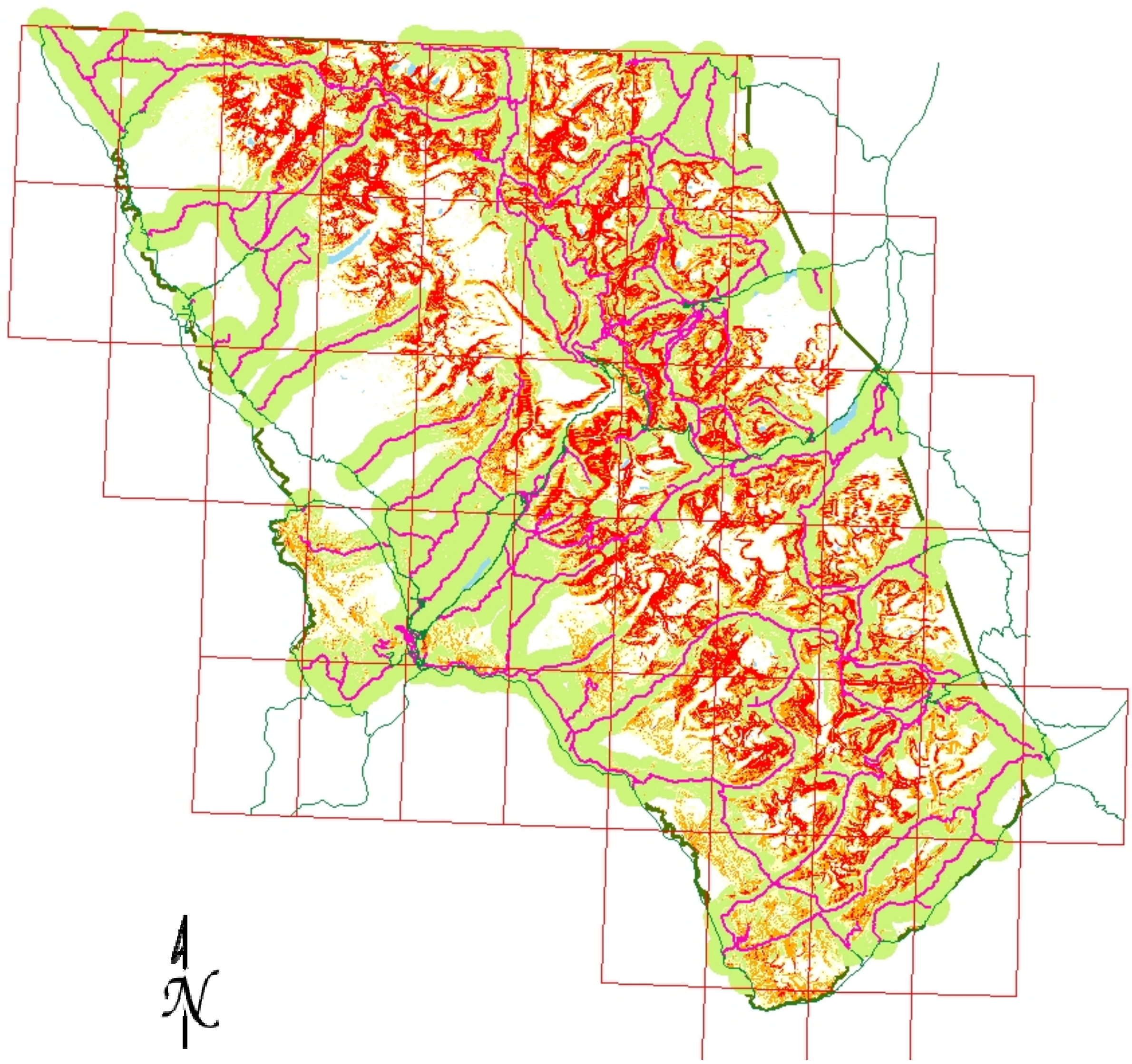

Fig. 1: Escape terrain in GNP with slope angle classifications of 25 to 32 (light grey/ yellow), 33 to 39 (dark grey/ orange), and 40 to 90 º (black/ red) used to select sites in each grid cell. Pink lines are hiking trails and light green areas are 3.2 km buffers around trails.

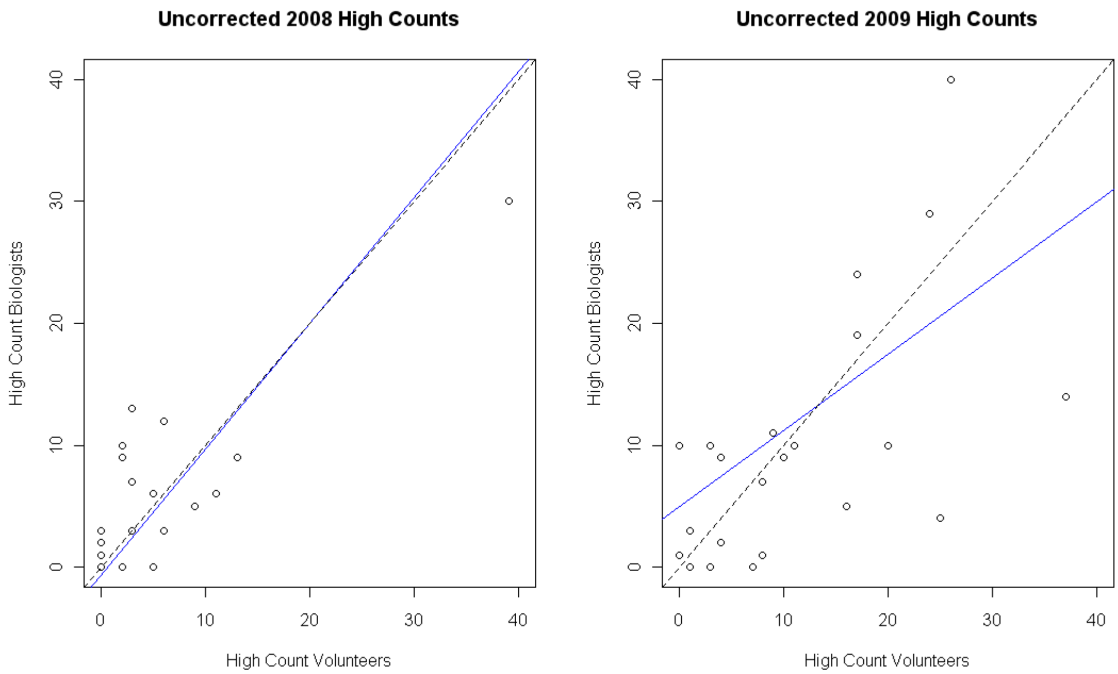

Fig. 2: Regression of high counts of mountain goats by volunteers with high counts by biologists for 2008 (left) and 2009 (right) in Glacier National Park, Montana. The solid blue line is the regression line. The dashed line is 1:1 line.

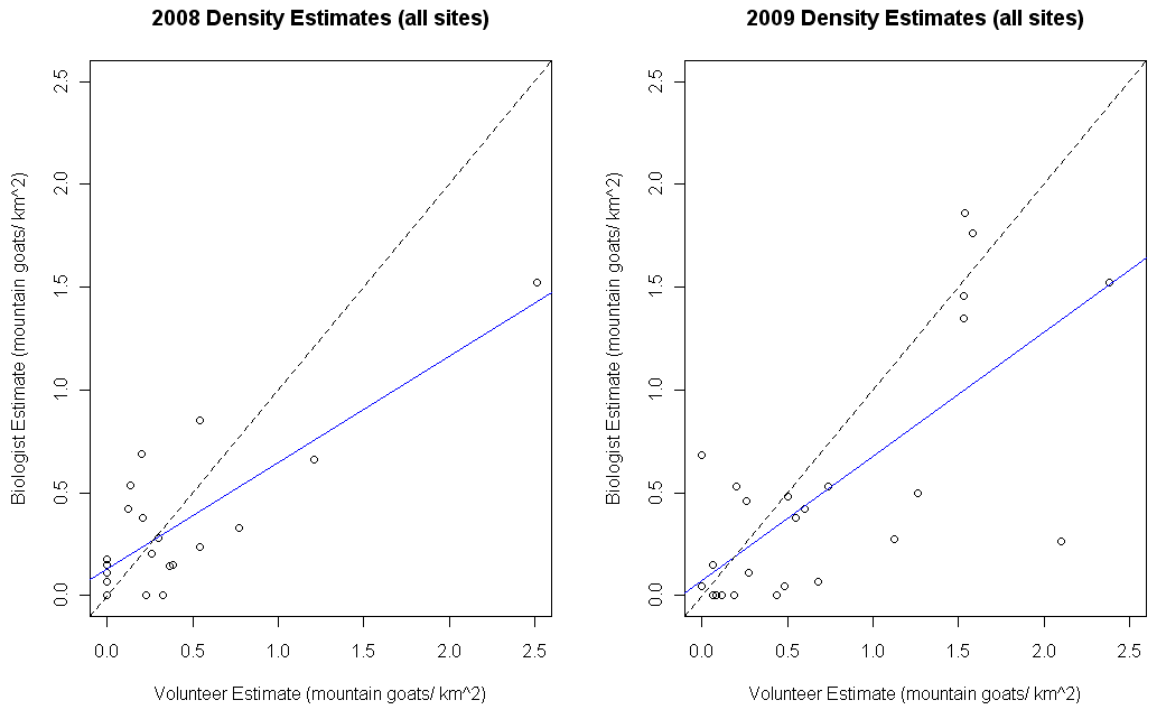

Fig. 3: Regression of density estimates of mountain goats by volunteers with density estimates by biologists at all survey sites for 2008 (left) and 2009 (right) in Glacier National Park, Montana. The solid blue line is the regression line. The dashed line is 1:1 line.

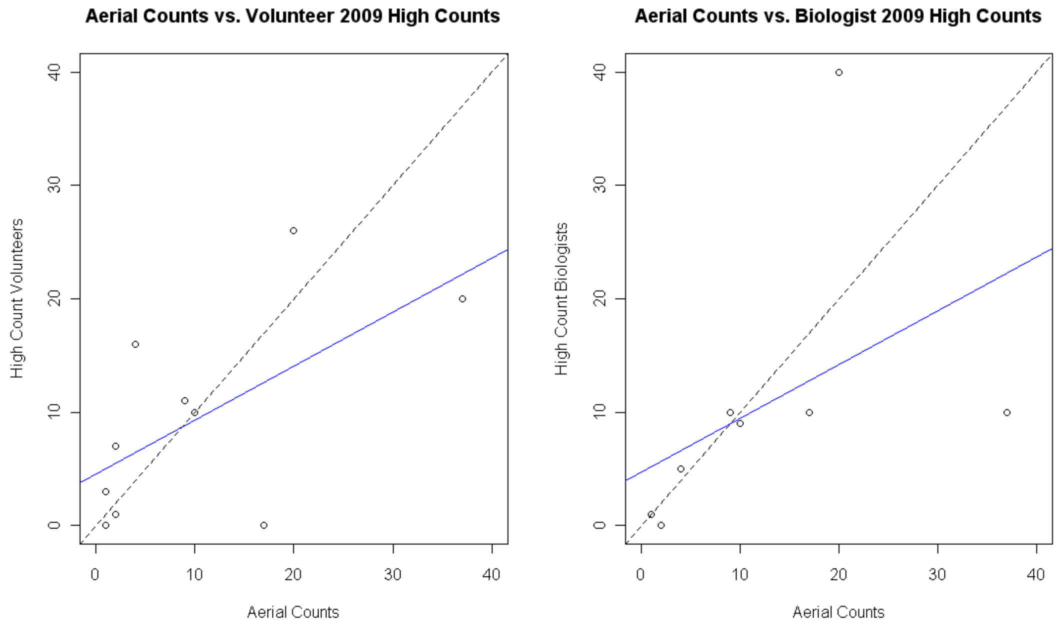

Fig. 4: Regression of 2009 corrected high counts of mountain goats by volunteers (left) and corrected high counts by biologists (right) with raw counts from 2009 helicopter surveys in Glacier National Park, Montana. High counts for volunteers and biologists are corrected by mean detection probability. The solid blue line is the regression line. The dashed line is 1:1 line.

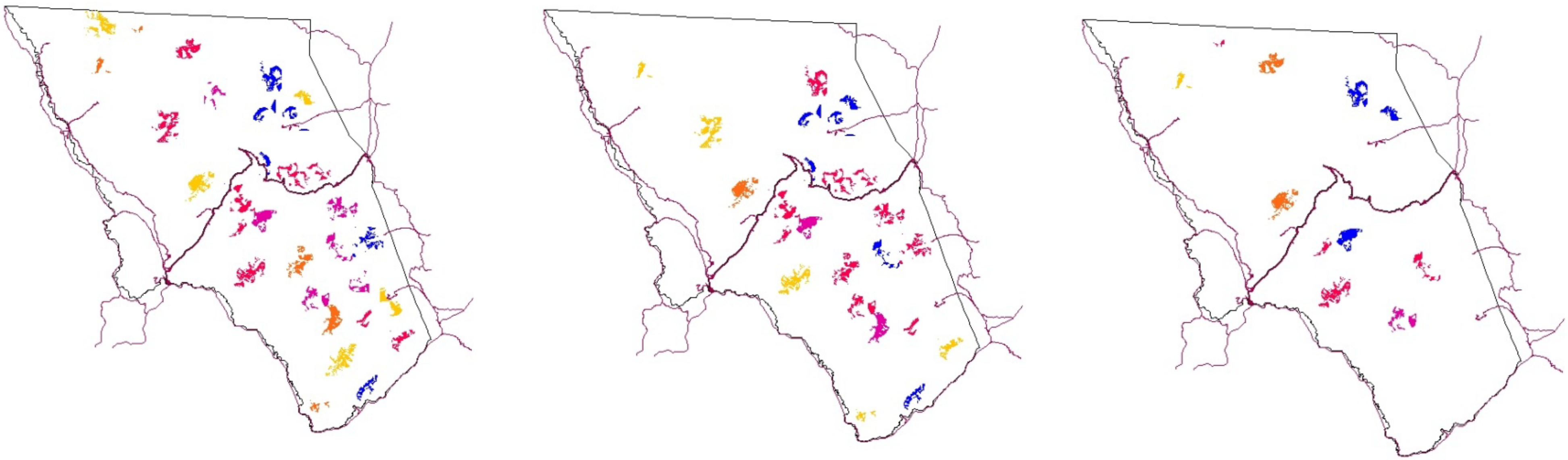

Fig. 5: Maps of relative densities of mountain goats estimated from corrected high counts by volunteers (left) and biologists (center), and raw counts from and helicopter surveys (right) in Glacier National Park, Montana. Legend: yellow= no mountain goats, orange= low density, red= moderate density, pink= high density and blue = very high density.

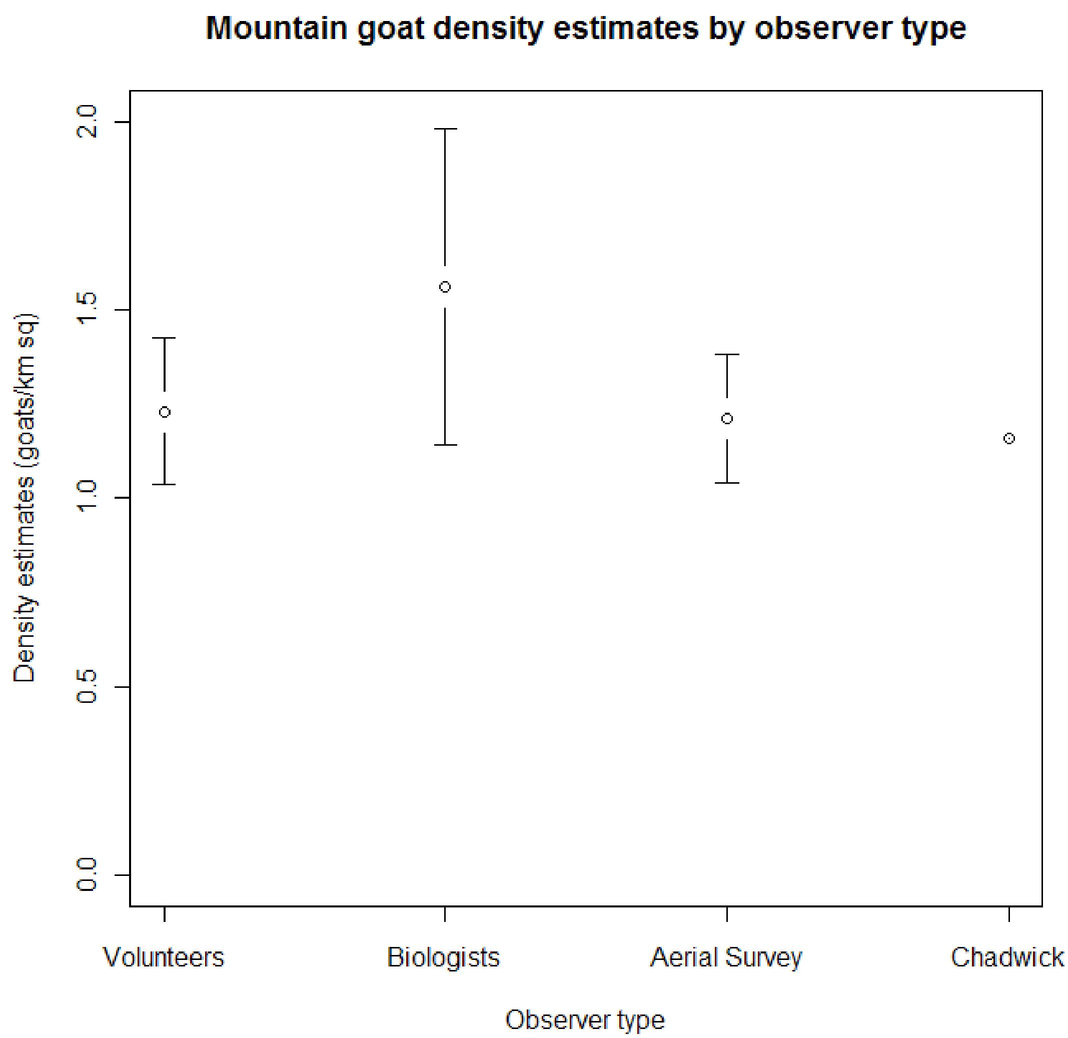

Fig. 6: Density estimates of mountain goats by volunteers, biologists, aerials surveys and previous research (Chadwick 1977) in Glacier National Park, Montana derived from N-mixture models with confidence intervals where available. Confidence intervals for aerial survey estimates are derived using a range of detection probabilities (0.55- 0.91) from other mountain goat aerial surveys (Gonzalez-Voyer et al. 2001; Rice et al. 2009) as a hypothetical correction factor for our aerial survey count.

Appendix 1: Questions from participant information sheet used to assess experience level of volunteers and biologists conducting mountain goat surveys for the High Country Citizen Science program in Glacier National Park, Montana. The range of possible scores for each question in listed in parentheses after the question. Minimum score was 9, maximum score was 42

1. Experience using binoculars to find wildlife at a distance (scores 1-5):

_____ limited use of binoculars _____ 1 to 2 years ____ 2 to 4 years

____ 4 to 8 years _____ over 8 years

2. Experience using a spotting scope (scores 1-6):

_____ have never used a spotting scope _____ 1 month to 1 year

_____ 1 to 2 years ____ 2 to 4 years _____ 4 to 8 years ___over 8 years

3. Which best describes your approach to wildlife watching (scores 1-3):

_____ casual/incidental watching

_____actively search for wildlife during other recreation

_____take trips with specific intention of watching wildlife

4. Experience with wildlife photography (scores 1-3):

_____ casual/incidental photography of wildlife

_____actively search for wildlife to photograph during other recreation

_____take trips with specific intention of photographing wildlife

5. Approximate number of days spent wildlife watching/ wildlife photographing trips in the past year (scores 1-5):

_____0-10 _____11-30 _____31-60 _____61-100 _____101+

6. Wildlife data collection experience (other than GNP citizen science) (scores 1-5):

_____ never collected wildlife data

_____occasional participation in an organized wildlife/ bird count (e.g. Christmas Bird Count)

____ have worked as a wildlife field technician for 1 to 2 year

____ have worked as a wildlife field technician for 2 to 4 years

____ have worked as a wildlife field technician for 4 or more years

7. Do you document the behavior, habitat or other aspects of wildlife you have seen (scores 1-3)?

_____ never _____ occasionally for significant wildlife sightings

_____ regularly document wildlife sightings

8. Ability to spot large mammals at a distance using binoculars or a spotting scope (scores 1-5):

_____1(novice) ______2 ______3 _____ 4 _____5(expert)

9. Ability to identify mountain goats at a distance (scores 1-5):

_____1(novice) ______2 ______3 _____ 4 _____5(expert)