Aerial survey methodology for bison population estimation in Yellowstone National Park
by Steven Craig Hess

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Fish and Wildlife Management
Montana State University

Abstract:
I developed aerial survey methods for statistically rigorous bison population estimation in Yellowstone National Park to support sound resource management decisions and to understand bison ecology. Survey protocols, data recording procedures, a geographic framework, and seasonal stratifications were based on field observations from February 1998-September 2000. The reliability of this framework and strata were tested with long-term data from 1970-1997. I simulated different sample survey designs and compared them to high-effort censuses of well-defined large areas to evaluate effort, precision, and bias. Sample survey designs require much effort and extensive information on the current spatial distribution of bison and therefore do not offer any substantial reduction in time and effort over censuses. I conducted concurrent ground surveys, or 'double sampling' to estimate detection probability during aerial surveys. Group size distribution and habitat strongly affected detection probability. In winter, 75% of the groups and 92% of individual bison were detected on average from aircraft, while in summer, 79% of groups and 97% of individual bison were detected. I also used photography to quantify the bias due to counting large groups of bison accurately and found that undercounting increased with group size and could reach 15%. I compared survey conditions between seasons and identified optimal time windows for conducting surveys in both winter and summer. These windows account for the habitats and total area bison occupy, and group size distribution. Bison became increasingly scattered over the Yellowstone region in smaller groups and more occupied unfavorable habitats as winter progressed. Therefore, the best conditions for winter surveys occur early in the season (Dec-Jan). In summer, bison were most spatially aggregated and occurred in the largest groups by early August. Low variability between surveys and high detection probability provide population estimates with an overall coefficient of variation of approximately 8% and have high power for detecting trends in population change. I demonstrated how population estimates from winter and summer can be integrated into a comprehensive monitoring program to estimate annual growth rates, overall winter mortality, and an index of calf production, requiring about 30 hours of flight per year.

AERIAL SURVEY METHODOLOGY FOR BISON POPULATION ESTIMATION

IN YELLOWSTONE NATIONAL PARK

by

Steven Craig Hess

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Fish and Wildlife Management

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2002

Approval

of a dissertation submitted by

Steven Craig Hess

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Robert A. Garrott (Signature) 4/18/02 (Date)

Approved for the Department of Ecology

Jay J. Rotella (Signature) 4/19/02 (Date)

Approved for the College of Graduate Studies

Bruce R. McLeod (Signature) 4-19-02 (Date)

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road. Ann Arbor, Michigan 48106 to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature Steven C. [illegible]

Date 19 April 2002

ACKNOWLEDGMENTS

I would like to gratefully acknowledge those who have participated or contributed to this project: D. D. Bjornlie, D. M. Fagone, M. J. Ferrari, A. R. Hardy, R. Jaffe, C. R. Kolarich, J. L. MacDonald, T. Olenicki, A. C. Pils conducted ground surveys of bison and provided other logistical support. D. S. Ouren of U. S. Geological Survey provided advice with GIS setup. P. J. Gogan and E. Olexa of U. S. Geological Survey provided information on bison distribution. R. Arment, D. Chapman, S. Monger, and D. Stradley piloted aircraft. W. Clark, G. Kurz, and J. A. Mack of the U. S. National Park Service provided information on recent bison distribution and advice on survey design. J. J. Borkowski and L. L. Eberhardt provided advice on statistical analysis. A. R. Hardy, R. Jaffe, J. L. Sands, and R. C. Krueger provided helpful constructive criticism on this manuscript. R. A. Garrott provided a wealth of advice, constructive criticism, and support, throughout the duration of this study. I would also like to acknowledge my parents for their encouragement, and my dog, for understanding. This research was funded by the U. S. Geological Survey Biological Resources Division with cooperation from the U. S. National Park Service.

TABLE OF CONTENTS

TABLE OF CONTENTS—CONTINUED

LIST OF TABLES

LIST OF TABLES—CONTINUED

LIST OF FIGURES

LIST OF FIGURES—CONTINUED

Abstract

I developed aerial survey methods for statistically rigorous bison population estimation in Yellowstone National Park to support sound resource management decisions and to understand bison ecology. Survey protocols, data recording procedures, a geographic framework, and seasonal stratifications were based on field observations from February 1998–September 2000. The reliability of this framework and strata were tested with long-term data from 1970–1997. I simulated different sample survey designs and compared them to high-effort censuses of well-defined large areas to evaluate effort, precision, and bias. Sample survey designs require much effort and extensive information on the current spatial distribution of bison and therefore do not offer any substantial reduction in time and effort over censuses. I conducted concurrent ground surveys, or 'double sampling' to estimate detection probability during aerial surveys. Group size distribution and habitat strongly affected detection probability. In winter, 75% of the groups and 92% of individual bison were detected on average from aircraft, while in summer, 79% of groups and 97% of individual bison were detected. I also used photography to quantify the bias due to counting large groups of bison accurately and found that undercounting increased with group size and could reach 15%. I compared survey conditions between seasons and identified optimal time windows for conducting surveys in both winter and summer. These windows account for the habitats and total area bison occupy, and group size distribution. Bison became increasingly scattered over the Yellowstone region in smaller groups and more occupied unfavorable habitats as winter progressed. Therefore, the best conditions for winter surveys occur early in the season (Dec–Jan). In summer, bison were most spatially aggregated and occurred in the largest groups by early August. Low variability between surveys and high detection probability provide population estimates with an overall coefficient of variation of approximately 8% and have high power for detecting trends in population change. I demonstrated how population estimates from winter and summer can be integrated into a comprehensive monitoring program to estimate annual growth rates, overall winter mortality, and an index of calf production, requiring about 30 hours of flight per year.

INTRODUCTION

Aerial surveys are important tools for wildlife management that may be used for a variety of purposes such as monitoring spatial distribution, age structure, and estimating density or abundance of large animals. Additionally, time series of reliable abundance estimates and demographic data that span many generations may provide valuable insights into the factors that regulate populations (Dennis and Taper 1994, Murdoch 1994, Sæther 1997, Shenk et al. 1998, Elkington 2000). Therefore, aerial surveys offer efficient means for gaining valuable management information and addressing long-term ecological questions at the population level. Obtaining precise and unbiased population estimates of many organisms, however, is difficult and methodologies for enumerating wildlife populations have advanced considerably over the past 20 years with the development of sophisticated sampling, estimation, and bias correction procedures (Lancia et al. 1996, Thompson et al. 1998). The reliability of population estimates may be evaluated by two key measures: bias and precision. Precision determines the statistical power to detect trends in time series (Gerrodette 1987, Gibbs 2000). It has also long been recognized that aerial surveys may be biased, underestimating the true abundance of large animals because many are not detected (Caughley 1974; Caughley and Goddard 1972).

For many aerial surveys of large mammals, the proportion detected can vary widely depending on conditions, but typically range from 50–60% (Caughley 1974). Surveys that typically do not enumerate an entire population must be considered an index of abundance that is proportional to or correlated with the population size (Caughley 1977). In some types of analyses, such as population growth rates and time-series models, using either indices or absolute population estimates should result in the same conclusions (e.g., Dennis and Taper 1994). However, there may be implicit assumptions that the same proportion of animals is detected during each survey. Temporal variation in detection rates may confound these analyses, and analyses of spatial patterns of abundance may be confounded by spatial variation in detection rates (Slade and Blair 2000). Also, population growth projections that must account for known human-caused reductions may be badly biased or non-sensical, because reductions are usually known absolutely, whereas indices are not measured on an absolute scale (Eberhardt 1987). In these cases, additional methods must be employed in conjunction with aerial surveys to estimate the proportion that were not detected, adjust population indices, and account for this additional source of uncertainty. Few long-term data series meet the criteria of precision and low bias necessary to thoroughly understand the roles that density-dependent and density-independent environmental mechanisms play in regulating wild populations (Murdoch 1994, Sæther 1997).

Estimates of the number of bison (*Bison bison*) in Yellowstone National Park (YNP) are available for a century and represent one of the longest time series for a large mammal population in North America (Fig. 1; Meagher 1973, Dobson and Meagher 1996, Yellowstone National Park 1997, National Park Service 2000, M. L. Taper et al., Montana State University, unpublished data).

Aerial surveys have been employed to census YNP bison since 1949 by attempting to fly over all areas where observers thought bison might exist (Meagher 1973). These data provide an outstanding and unique opportunity for understanding the population dynamics of this large herbivore and for predicting future trend in the population, underscoring the importance of continued population monitoring.

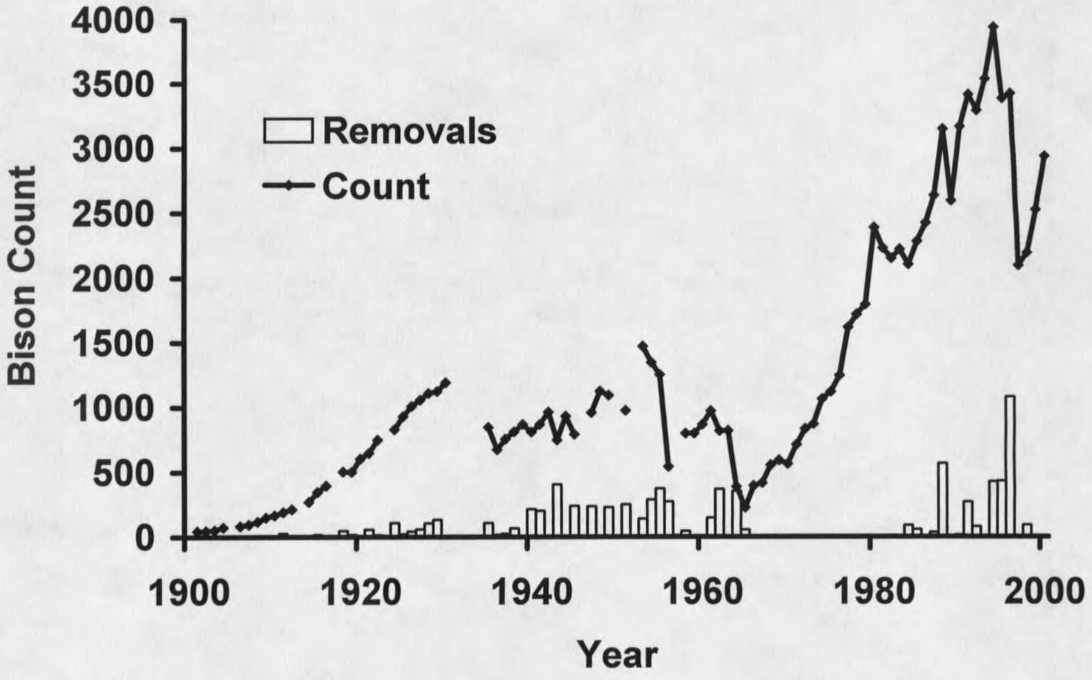

Fig. 1. The number of bison counted during annual censuses and the total number of known annual removals of bison from 1901-2000 in Yellowstone National Park, Wyoming. Data are from Meagher 1973, Dobson and Meagher 1996, Yellowstone National Park 1997, National Park Service 2000, M. L. Taper et al., Montana State University, unpublished data.

During most of the past century counts of the bison population have remained below 2000 animals, and the population was concentrated in several low-elevation valleys, facilitating such censuses (Meagher 1973). Recently, however, counts have increased to > 3000 animals. Consequent with this increase in numbers bison landscape-use patterns have become more dynamic with an apparent increase in movements and an expansion of distributional patterns to include areas that were not historically occupied during the previous century (Meagher 1989, Bjornlie and Garrott 2001). These changes have increased the difficulty of obtaining precise censuses as larger areas must be surveyed, and the probability that a given survey covers all the areas where bison occur decreases. Furthermore, there has never been an empirical attempt to determine the probability of detecting bison from aircraft in YNP or to correct population estimates for the proportion not detected (Meagher 1973). In sparse vegetation on Antelope Island, Utah, nearly 94% of bison were detected from aircraft (Wolfe and Kimball 1989), while detection probability in forested areas of northern Canada, was 75% (Larter et al. 2000), suggesting that detection probability of bison may be considerably higher than for most other large animal species, but varies in different habitats. While the methods employed previously to census the YNP bison population have been adequate for immediate needs, the changes in bison numbers, distribution, and spatial-use patterns combined with advancements in population estimation procedures suggests that new survey methodologies be developed to address long-term ecological and epidemiological management issues of the largest wild, free ranging bison herd in the United States.

There is considerable public interest in YNP bison management activities from diverse social, scientific, and political groups (Keiter 1997). Understanding population regulation in YNP bison (M. L. Taper et al., Montana State University, unpublished data) and effects on ecological processes, such as condition and species composition of winter range grasslands and competition with other ungulates is important for sound management decisions (Yellowstone National Park 1997). Epidemiology is also being addressed as part of an interagency effort to assess and manage the risk of brucellosis transmission between bison and cattle ranging near YNP boundaries (Baskin 1998). Cheville and McCullough (1998) found that the number of bison exiting the park and possibly coming into close association with cattle was related to the total population of YNP bison. Specific target population levels, therefore, have been identified in environmental impact alternatives to reduce the possibility of bison exiting YNP boundaries, associating with cattle, and possibly transmitting brucellosis (National Park Service 2000). These alternatives identify specific threshold population levels above which management reductions will be implemented. Interested parties may disagree about reported population estimates, the survey methods used by agencies, and the true abundance of bison in YNP, especially when estimates may trigger management reductions. Therefore, scientifically valid population estimates with measures of precision (Lancia et al. 1996) are of key importance to the National Park Service (NPS) for managing bison within YNP and to satisfy legal obligations of a finalized environmental impact statement. National parks, however, are mandated not only to monitor and manage wildlife, but also to provide visitors with a safe and enjoyable experience in a natural setting. Low-level overflights by research aircraft may disrupt visitors' experiences and present an unreasonable safety risk as well. Additionally, bison are one of the most popular and highly visible species in YNP and occupy many developed areas where visitors gather. Consequently, aerial survey methodology for YNP bison must meet the simultaneous challenges of safety, minimizing disturbance to both the public and to wildlife, while providing precise and unbiased population estimates.

In considering potential sources of bias and precision in designing aerial survey methodology for bison population estimation, factors that primarily affect bias are: 1) the geographic scale of surveys; 2) detection probability within surveys and; 3) undercounting large aggregations of bison. High levels of precision may be difficult to achieve if: 1) a population is dispersed or large enough to require sampling; 2) the sampling regime is not adequate; 3) the population is gregarious or migratory; 4) detection probability is low, which causes high variability and; 4) observation conditions are variable between surveys. These factors may change within and between seasons as bison make seasonal movements corresponding to the phenology of available forage, depth of accumulating snow, and seasonal aggregations for breeding, which all affect the size of groups, the types of habitats, and amount of area they occupy, and ultimately, the probability of detection.

There have been several different solutions proposed to correct detection biases in aerial surveys, but some state of the art techniques may not be suitable for YNP. Capture-mark-resight (White and Burnham 1999) techniques that rely on highly visible marks or many radio-transmitter collars on animals are objectionable to some visitors who seek a natural experience. These methods, as well as telemetry techniques to establish a sighting curve (Samuel et al. 1987, Steinhorst and Samuel 1989, White and Garrott 1990) would require low-level flights to identify individual animals, posing an unacceptable safety liability, disturbing wildlife, and also the visiting public. Recalibration of a sighting model would require maintenance of instrumented animals that is too costly for monitoring purposes. Both capture-mark-resight and sightability curve methods may be confounded when more than a thousand bison occupy a relatively small valley during summer breeding, making the identification of specific radio signals with individuals difficult (Steinhorst and Samuel 1989, White and Burnham 1999). Line transect methodology (Buckland et al. 1993) offers convenient integrated sampling and bias correction designs for areas with less extreme topography, however, it would be dangerous to fly a straight and level flight path in the steep terrain of YNP. Transect-based designs may also be inefficient where aircraft have to deviate from the flight path frequently to enumerate large groups of animals. Enumerating these large groups of bison also poses special problems. However, photography has been used to establish ratio corrections for counting large groups of animals and holds promise for establishing regression correction coefficients (Watson 1969; Norton-Griffiths 1974, 1978).

Double sampling is a method that allows calibration of indices with intensive subsamples to determine the proportion of animals that were not detected (Cochran 1977, Geibel and Miller 1984, Eberhardt and Simmons 1987). A ratio estimator, which takes the form of regression through the origin, may be used to improve the estimates of the larger sample, assuming sub-samples achieve complete detection (Eberhardt et al. 1979). However, if intensive subsamples fall short of complete detection, the data should be treated differently. In this case, a modified Lincoln-Petersen mark-recapture method may be applied (Anthony at al. 1999). This method was found to be well suited to large animals in open habitats with high sighting probabilities, such as bison. Furthermore, it does not depend on individual identification of animals, per se, that requires low-level flight, posing an unacceptable safety risk to biologists given the remoteness and geographic scale of surveys in YNP.

To cope with the problems of surveying large geographic areas, Siniff and Skoog (1964), and Evans et al. (1966) were among the first to use stratified sampling of quadrats and allocate aerial survey effort in proportion to density for estimating caribou (*Rangifer tarandus*) and moose (*Alces alces*) populations in Alaska. Optimal allocation of aerial survey resources was also employed to develop a comprehensive monitoring program for estimating moose demographic parameters in Alaska (Gasaway et al. 1986). Experiments have also been conducted to test the efficiency of different sampling designs and statistical estimators for aerial surveys of pronghorn antelope (*Antilocarpa americana*) in Colorado (Pojar et al. 1995). Although stratified sampling designs may be optimized based on theoretical formulas (Cochran 1977, Barnett 1991, Thompson 1992), Monte Carlo simulations may also be conducted to directly evaluate precision and bias in sampling designs (Manly 1991).

The overall objective of this research was to develop an efficient monitoring program to provide statistically defensible population estimates and demographic parameters of bison that account for the geographic scale of aerial surveys, the proportion of animals undetected, minimize the disturbance to wildlife and the public, and minimize the safety risk to the public and biologists. The focus of this research was to: 1) conduct aerial surveys to determine the spatial distribution and group size distribution of bison in both winter and summer seasons; 2) design and validate a sampling framework and seasonal stratification; 3) simulate the efficiency of different sampling regimes and replication to evaluate cost, effort, disturbance, precision due to survey design, and bias due to statistical estimation; 4) use simultaneous air and ground survey methods to determine the proportion of bison not detected during aerial surveys; 5) photography to determine the number of bison undercounted in large groups and; 6) to model correction coefficients for groups not detected, corrections for undercounting large groups, and covariates associated with specific survey conditions.

STUDY AREA

I considered all of the area within YNP boundaries (9,018 km^2), and also approximately 500 km^2 outside YNP in the overall scope of this study. Main areas of bison concentration within YNP are the Northern Range from the plateaus and ridges of the upper Lamar Valley downstream to the Northwest entrance, including Swan Lake Flats, and in the central part of YNP, the Madison, Gibbon, and Firehole River drainages to the West entrance, and Hayden and Pelican Valleys (Fig. 2). Areas outside the YNP boundary near the towns of West Yellowstone and Gardiner where bison have recently been regularly congregating during winter periods such as Hebgen Lake and the South Fork of the Madison River were included in the study. Areas within the Park where small numbers of bison are typically found south of the main population near Yellowstone and Shoshone lakes were also included in the study.

The elevational range of the study area extended from about 1,565 m in valleys near the northern boundary of YNP to about 2,700 m on higher plateaus and included steep canyons and mountainous terrain in parts of the Gallatin and Absoroka ranges. Valleys occupied by bison lie at: Lamar, 1,950 m; Firehole, 2194 m; Hayden, 2,346 m; and Pelican 2,377 m. Vegetation cover in valleys ranges from sedges (*Carex* spp.) near streams to short grasses, forbs, and sagebrush (*Artemisia* spp.) in Hayden and Lamar Valleys, and some patchy forest cover. Slopes and plateaus are mostly forested, but some plateau areas in the Northern Range have patchy open meadows with grasses and sedges.

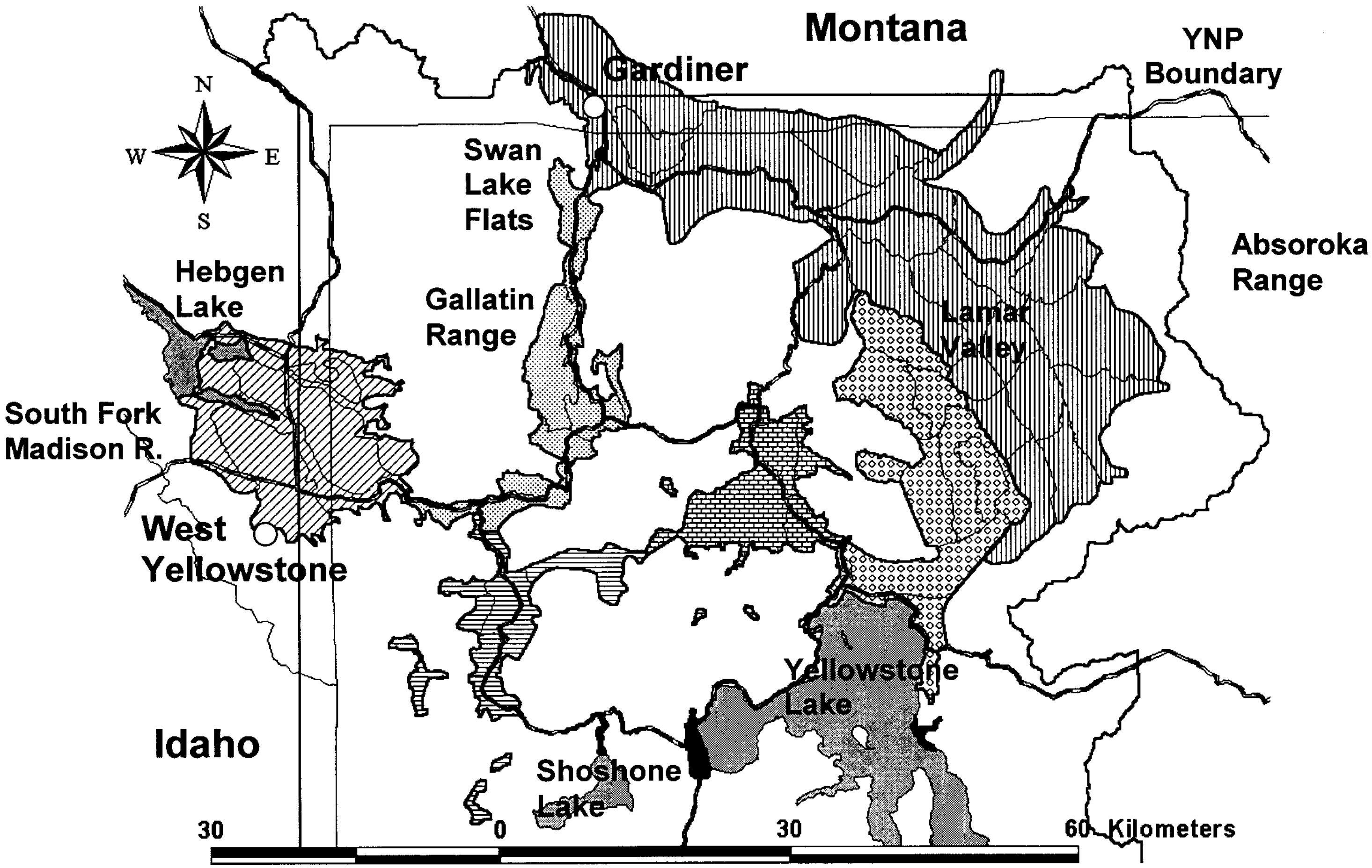

Fig. 2. Aerial survey study area and survey units for bison population estimation in Yellowstone National Park (YNP). Ranges (areas where bison movements are restricted during surveys) occupied by bison include: Firehole River (FH- horizontal lines), Hayden Valley (HV- brick pattern), Madison and Gibbon Rivers (MG- stipled), Northern Range (NR- vertical lines), Pelican Valley (PV- open circles), West Yellowstone area (WY- diagonal lines), and southern areas (SO- solid black). Major roads are indicated with dashed lines.

Major forest cover species include lodgepole pine (*Pinus contorta*), with intermixed climax species such as Engelmann spruce (*Picea engelmannii*), subalpine fir (*Abies lasiocarpa*), and whitebark pine (*Pinus albicaulis*). Forest fires in 1988 are responsible for a mosaic of burned and unburned forests and large areas of forest regeneration.

Major geothermal features, such as geysers, hot springs, and steam vents occur in the Firehole and Gibbon river drainages, Pelican and Hayden valleys, and to a lesser extent in the Northern Range and southern areas of YNP. Geothermal features reduce winter snow cover even during peak periods of accumulation, often exposing soil and vegetation. The Yellowstone area is seasonally covered with continuous snow from about mid-November until mid-April (Meagher 1973), but snow conditions vary at different elevations. In spring, some lower elevation valleys and south facing slopes may be snow-free weeks before higher elevation areas.

METHODS

Factors Affecting Bias and Precision

When considering potential sources of bias and precision (Lancia et al. 1996) in designing aerial survey methodology for bison population estimation, factors that primarily affect bias are: 1) the geographic scale of surveys; 2) detection probability within surveys; and 3) undercounting large aggregations of bison. Precision may be affected by: 1) the sampling regime; 2) gregarious and migratory behavior contributing to patchy spatial distribution; 3) detection probability, which becomes more variable when it is further from unity; and 4) variability among different observers. These factors may also change within and between seasons as bison aggregate annually for breeding and make movements corresponding to snowpack dynamics and the phenology of available forage (Frank et al. 1998). The dynamic nature of these movements determines the habitats and total amount of area they occupy, and the number of animals found in groups. For many ungulate species, detection probability is related to group size distribution, therefore, group size distribution may affect both precision and bias in population estimates (Cook and Martin 1974, Samuel and Pollock 1981). I conducted aerial surveys and addressed these potential sources of bias and precision with: 1) computer simulations of different sampling regimes using survey data with an explicitly defined survey framework, survey units, and seasonal stratification; 2) simultaneous double sampling from the ground and aircraft to determine detection probability; 3) photography to determine counting bias for large groups; and 4) replication to estimate variation between surveys and different observers.

Survey Framework and Stratification

Prior to initiating field research, published historical accounts (Meagher 1973) and information from recent aerial surveys (1996–97) conducted by NPS biologists were used to define the survey framework as the approximately 2,400 km^2 area where bison regularly occur within the Yellowstone region (J. Mack and W. Clark, Yellowstone National Park, unpublished data). I then subdivided the survey framework into 7 major ranges: Firehole River; Hayden Valley; Madison and Gibbon Rivers; Northern Range; Pelican Valley; West Yellowstone; and isolated areas in the southern part of YNP (Fig. 2) I further subdivided these ranges into survey units based on major topographical features, primarily divides, or other features visible from aircraft, such as roads or streams where there were no other prominent topographical features. I determined the size of each survey unit using computer Geographic Information System (GIS) to quantify the total area covered by each survey, and effort (time / area).

I defined high-density strata within the survey framework from recent (1996–97) aerial surveys and quantified the number and density of bison within these survey units during both winter (January–April) and summer periods (June–September; J. Mack and W. Clark, Yellowstone National Park, unpublished data). These high-density strata roughly correspond to the area NPS biologists have been conducting total counts of bison since 1996 and were areas known to have seasonally high densities of bison. In winter, the high-density stratum primarily contained grassland-shrub plant communities associated with valley bottoms near streams and rivers, while in summer, this stratum contained high elevation valleys and large areas of high elevation plateaus with grassland-shrub communities and burned forests. The remainder of the units within the framework constituted the low-density stratum during each season. These areas had smaller, isolated patches of habitat similar to the high-density strata areas or vast continuations of marginally suitable habitat adjacent to high-density strata, usually at higher elevations.

To understand how the spatial scale and stratification of surveys may change across a wider range of population levels and habitat conditions, I used historical bison survey data to determine the total area bison occupied and to identify survey units with predictably high densities of bison over time (M. L. Taper et al., Montana State University, unpublished data). These geo-referenced bison survey data from 1970–1997 became available after my field research was completed. I chose years with population levels of < 1,500 (1970–78), 1,500–3,000 (1979–90), and >3,000 (1991–97), differentiated winter from summer surveys, then ranked my survey units by the total number of bison present in each unit using GIS. I assigned the lowest rank to survey units with the most bison, ties to units with equal numbers of bison, and no rank to units without bison. I averaged ranks across surveys, but excluded units with no rank.

Aerial Data Collection Procedures

I conducted survey flights in 3 types of aircraft: a Cessna 182; an Aviat A-1; and a Piper PA-18 from February 1998, until August 2001. Winter surveys were conducted between January–April and summer surveys in June–September. Aircraft ground speed was 70–110 kph depending on winds, and altitude above ground level was $\geq$ approximately 150 m under good wind conditions and higher in gusty conditions. I attempted to survey all units in the high-density strata within 3 days. To minimize the possibility of bison moving among the 7 major ranges of occupation between flights, I completely surveyed all units in the high-density stratum within a range before proceeding to the next range. Data recorded for each unit surveyed included starting time, the number of bison in each group, the time each group was observed, and the location of observed groups in Universal Transverse Mercator (UTM) or decimal degree coordinates using an onboard Global Positioning System (GPS). I attempted to repeat surveys of all units in high-density strata > 3 times per season. I also surveyed a subset of low-density strata units during each flight such that all low-density units were surveyed >1 time over the course of each season and areas outside the survey framework once in winter and twice in summer. Effort was held as constant as possible to minimize variation between surveys by exhaustively searching small and irregularly shaped units and by using parallel flight lines in large and open units where there were no high canyon walls in the vicinity. Coverage of survey units was checked with flight path tracks recorded on the GPS.

I defined a bison group as the maximum number of individuals that did not influence the observer to detect surrounding bison, such that the probability of sighting groups was independent. The minimum distance between independent groups was approximately an order of magnitude greater than the average distance between bison within a group. The covariate attributes of vegetation type or background (forest, burnt forest, meadow, sagebrush, geothermal, bare soil, aquatic, road, and snow) and the presence of fog were recorded for each group. Forest and snow cover were estimated in 10% increments. During summer surveys, I enumerated large groups of bison by circling, searching for features of the landscape or the group which could be used as reference points to start and end counting. I then counted all adults on 1 or more circling passes, and then counted calves, usually during subsequent circling passes.

I also report and analyze data from aerial surveys conducted by NPS biologists from the same time period (J. Mack and W. Clark, Yellowstone National Park, unpublished data). These surveys were usually conducted during a single flight on a monthly basis. Although there was no explicitly defined framework, each flight surveyed all of the major ranges of bison occupation during an average of 5.2 hours (range 3.0 – 6.9). All surveys were conducted in Piper PA-18 aircraft. Speed and height above ground level were both greater than my surveys. The locations of bison groups were recorded with a 1:250,000 scale map or an onboard GPS unit.

Temporal Windows for Surveys

Several aspects of bison behavior vary temporally and affect survey timing. Group size distribution and habitat use affect detection probability, and therefore, both bias and precision of surveys. The total area bison occupy must also be considered in the design of precise sampling protocols that reflect seasonal distribution. I examined these temporal trends in the spatial distribution and group size distribution of bison within and between seasons to determine how seasonal factors may affect the spatial scale of surveys and probability of detection. Conducting surveys at times when the total area bison occupy is small, group sizes are large, and bison inhabit areas with favorable detection characteristics will provide high detection probability and reduce the variability between surveys.

I pooled my aerial survey data across years with surveys conducted by YNP biologists from a concurrent period to examine change in group size distribution and total area of survey units occupied by bison over the course of winter and summer (J. Mack and W. Clark, Yellowstone National Park, unpublished data). Using data from surveys with complete coverage of high-density strata, I summed the area of survey units occupied by bison and regressed total area occupied by bison on Julian date and tested for year and observer effects, as different pilot/observer combinations may make different judgments about groups sizes and even locations.

To determine how group size distribution changed within seasons and may affect the probability of detection, I regressed maximum group size on Julian date and tested for observer and year effects. I used data where ≥ 2 major ranges of bison occupation were surveyed to insure representative samples of group sizes across each season. I also compared bootstrap estimates of average summer and winter group sizes to determine how seasonal differences in group size may affect detection probability (Efron and Tibshirani 1993). To ensure independence of the data, I used only the first complete survey of high-density strata each season and pooled seasons across years. I used the nonparametric bootstrap percentile approach to determine 95% confidence intervals about mean group size due to the non-normal distribution of the data, and unequal variances and sample sizes. I also examined the cumulative distribution function (CDF) of bison groups sizes pooled across years from all of my surveys.

To determine seasonal patterns of habitat use that may affect the detection probability of bison in winter, I examined patterns of winter habitat use from 10-day synoptic surveys of bison in the Madison, Gibbon, and Firehole (MGF) river drainages (Bjornlie and Garrott 2001). Observers covered the entire bison winter range in these drainages within 2–3 days on snowmobiles, showshoes, wheeled-vehicles and/or on foot and recorded UTM coordinates and habitat association of all bison. I pooled data across the years 1998–2000 and regressed the proportion of bison associated with geothermal habitats on Julian date and a measure of snowpack, snow water equivalent (SWE; the equivalent amount of water contained in a column of snow) at West Yellowstone, Montana (Farnes 1996). I also regressed the proportion of bison associated with forested habitats on average daily temperature at West Yellowstone. I obtained these climate data from an automated weather station operated by the Natural Resource Conservation Service.

Simulations

I conducted Monte Carlo simulations to determine bias and precision in the first stage of sampling with simple random sampling (SRS) and sampling proportional to size (PPS) designs using Matlab version 5.3 software (Barnett 1991, Manly 1991, The Math Works 1999). I created a composite data set by combining complete survey coverage data from the high-density stratum for January 1999 with low-density stratum data from the first time each low-density unit was surveyed over the course of the 1999 winter season. I created a composite data set from the August 1998 survey and other surveys from the summer 1998 season in the same manner. I used data from 62 survey units that were known to be occupied at any point in winter since 1970. For summer simulations, I used 52 survey units that were occupied during summer since 1970, plus 2 units that were surrounded by units with periodically high numbers of bison for a total of 54 survey units. Survey units were sampled without replacement for SRS simulations and sampled proportional to the size of survey units in area with replacement for PPS simulations. The total number of adult bison within each sampled survey unit was the response used in estimation. I first simulated a SRS sampling design with the SRS estimator, sample average ratio (ratio 1) estimator, the ratio of sample averages (ratio 2) estimator, and then PPS sampling with a Horvitz-Thompson PPS estimator (See Appendix A for sample estimator formulas). I varied sample size from 5–60 survey units in increments of 5 without stratification in winter and 5–50 survey units in summer. Each sampling design/estimator/sample size combination was simulated 1,000 times. I used a Satterthwaite (1946) adjustment formula for each estimator to determine the appropriate degrees of freedom and then calculated 1) 95% CI coverage of true density, 2) CI width as the percent of total density (upper CI – lower CI / density X 100), and 3) percent bias in density due to the estimator.

I then stratified and varied the sample allocation of high:low-density strata units from 2:3, 1:1, 2:1, 3:2, and 3:1, simulating stratified SRS with the SRS estimator, the separate ratio estimator, the combined ratio estimator, and stratified PPS sampling with a Horvitz-Thompson PPS estimator 1,000 times for each combination. From the average survey unit ranks since 1990, I designated the 32 highest ranked survey units as the high-density stratum, and 30 units as the low-density stratum for winter simulations, and 17 and 47 units as the high and low-density stratum for summer simulations. Sample sizes were varied, but chosen to be exact combinations of stratum assignments and allocation ratios. To cope with the problem of sampling from a highly skewed population, Hansen et al. (1953) recommended that a larger sampling fraction should be drawn from the stratum consisting of the largest aggregate. Optimum allocation formulas may even indicate that >100% of these units should be sampled, but of course, the maximum sample cannot exceed the total number of units in the stratum. I simulated the case where all of the units within a stratum were included in the sample and identified this design element as a complete coverage block. I treated 8 of the highest ranked, non-contiguous survey units as a complete coverage block in winter, and repeated stratified simulations on the remaining units with only SRS and PPS estimators. In summer, I treated 4 non-contiguous units as a complete coverage block and repeated stratified simulations with only SRS and separate ratio estimators, but also added 1:2 and 4:1 allocation ratios. I evaluated efficiency in different sampling plans by relating the area surveyed to precision (CI width as the percent of density) in each simulation. I divided the area of the survey by average effort (143.0 km^2/hr in winter, and 168.1 km^2/hr in summer) to determine the amount of survey flight time required. I also tested composite data from the following year (January 2000 and August 1999) for winter and summer to simulate the application of an optimized sampling design as it would be employed in practice.

Double Sampling

I used double sampling, or two-phase sampling to estimate the proportion of bison detected during aerial surveys (Cochran 1977, Geibel and Miller 1984, Eberhardt and Simmons 1987). Observers located on the ground primarily in either the Hayden Valley in summer, and MGF areas in winter, recorded all bison in a subsample of units on predetermined routes during simultaneous aerial surveys (Bjornlie and Garrott 2001). I also conducted simultaneous ground surveys during aerial surveys of the Northern Range by NPS biologists on March 13, 1999, and December 10, 1999. Because not all bison were detected by either ground or aerial surveys, I conducted a mark-recapture analysis to calculate unbiased estimates of the total number of bison groups not detected during surveys (Caughley 1974, Cook and Jacobson 1979, Anthony et al. 1999). I classified observations according to the number of bison groups observed from aircraft, but not from the ground (S_1), the number of groups observed from the ground, but not from aircraft (S_2), and the number of groups detected by both survey methods (B). I used a Lincoln-Petersen estimator following Anthony et al. (1999) with log-normal 95% confidence intervals (Chao 1989) to determine the proportion of groups detected by each survey method pooled across all surveys. The unbiased Lincoln-Petersen maximum likelihood estimator is

$$\hat{N} = [(S_1 + B + 1)\ (S_2 + B + 1)/(B + 1)] - 1$$

with estimated variance

$$V\hat{a}r\,(\hat{N}) = \frac{[S_1\ S_2\ (S_1 + B + 1)\ (S_2 + B + 1)]}{[(B + 1)^2(B + 2)].}$$

Lognormal confidence intervals may provide better coverage properties than confidence intervals based on normal theory if the data are positively (right) skewed (Burnham et al. 1987, Chao 1989, Buckland et al. 1993, Nichols et al. 2000). The advantage of asymmetrical lognormal confidence intervals is that the lower bound is not less than the actual observations. Confidence limits were calculated as $\left[\hat{N}/C, \hat{N}C\right]$ where Z = 1.96 for a standard normal distribution and C is

$$C = exp\left(Z\left[ln\left(1+\frac{V\hat{a}r(\hat{N})}{\hat{N}^2}\right)\right]^{1/2}\right).$$

Lincoln-Peterson estimates of detection probabilities from aircraft ($\hat{p}_1$) and from the ground ($\hat{p}_2$) can also be expressed as

$$\hat{p}_1 = B/(B + S_2) \qquad \text{and} \qquad \hat{p}_2 = B/(B + S_1).$$

Details on the derivation of these equations may be found in Magnusson et al. (1978), Estes and Jameson (1988), and Graham and Bell (1989). These quantities are therefore amenable to binomial logistic regression modeling with covariates. I tested 8 *a priori* models of detection probability for ground and aerial surveys separately using maximum likelihood logistic regression (PROC LOGISTIC; SAS Institute 1990). This also allowed a direct examination of factors affecting the probability of detection by ground and aerial surveys separately when overall detection probabilities appeared to be similar for both methods. Each model contained the predictor group size (p_g) in all possible combinations with seasons (p_s; 2 levels: summer and winter), habitats (p_h; 3 levels: open, forest, and geothermal), and ranges (p_r; 3 levels: Hayden Valley, MGF area, and the Northern Range). The most general model (p_{gshr}) allowed differences in detection probability due to all 4 factors simultaneously, and reduced models contained combinations of group size and 1 or 2 other factors at a time. I used Akaike's Information Criteria (AIC) to identify the single most parsimonious model for each method (Burnham and Anderson 1998), Wald χ^2 values to evaluate the overall effects of factors, normal approximations to evaluate the significance of individual coefficients, and the overlap of 95% confidence intervals with zero to evaluate the utility of coefficient estimates. I did not use Akaike weights (Burnham and Anderson 1998) for model averaging due to the difficultly of calculating confidence intervals from multiple variance-covariance matrices of different dimensions. I applied logistic regression coefficients and the variance-covariance matrix from best models to calculate the predicted detection probabilities and 95% confidence intervals from the double survey data (Collett 1999), then used the identity $\hat{p}_o$ = 1- (1 - $\hat{p}_1$)(1 - $\hat{p}_2$) to model the overall estimated probability of detection ($\hat{p}_o$) during double surveys (i.e., $\hat{p}_o$ = 1- the probability of being undetected by both observers; Nichols et al. 2000). I used an SAS jackknife procedure to evaluate the reliability of each model (PROC LOGISTIC; SAS Institute 1990; van Manen et al. 2000, Marcot et al. 1983). This procedure estimated overall correct classification rates, sensitivity (proportion of detections predicted to be detections), specificity (proportion of non-detections predicted to be non-detections), false positive rates, false negative rates, and reliability, where reliability was defined as the percent of model predictions that were empirically correct (sensitivity + specificity)/(sensitivity + specificity + false positive rate + false negative rate).

Photography

To determine the accuracy of counting bison groups from aircraft, I photographed all groups of bison that numbered > 200 and opportunistically sampled groups of < 200. I framed entire groups of bison using a variable focal length lens (28–90 mm), or fixed focal lengths of 50 mm or 105 mm, usually making 4–5 images per group. During photography, aircraft flew at the same altitude and speed as they normally flew during surveys. I projected the photographic slides with a high-quality lens and counted adult bison from the single best image. I regressed the number of bison in the best image on the actual field count to determine if it was necessary to correct for the number counted within groups. I then modeled the relationship of accuracy to group size and established a correction coefficient for counting large groups of bison.

Replicate total counts

To provide a method of analysis comparable to census attempts from the past century (Dobson and Meagher 1996, National Park Service 2000), I treated the sum of bison detected during repeated winter surveys as replicate total counts. I combined these data with surveys conducted by NPS biologists during the same period (J. Mack and W. Clark, Yellowstone National Park, unpublished data) and calculated bootstrap estimates of means, coefficients of variation (CV), and variances (Efron and Tibshirani 1993). I also corrected mean total count estimates for the proportion of bison not detected and summed the bootstrap estimates of variance with the additional binomial variance due to detection probability using the delta method to estimate overall 95% confidence intervals (Mood et al. 1974, Seber 1982).

This simple method of population estimation is based on the variability between repeated total counts x_i, the mean of repeated counts $\overline{x}$, and the mean sighting probability $\hat{p}$. A corrected point estimate $\hat{N}$ for the population is found by dividing the mean total count $\overline{x}$ by $\hat{p}$, the estimated detection probability from simultaneous air/ground double surveys

$$\hat{N} = \frac{\overline{x}}{\hat{p}}.$$

The mean detection probability $\hat{p}$ is a simple binomial correction determined by dividing the total number of bison observed from aircraft during double surveys by the total number predicted by logistic regression. I calculated bootstrap estimates of the mean $\overline{x}$ and variance of counts within years (Efron and Tibshirani 1993). Bootstrapping may be used to provide precise estimates of the mean and variance from the n counts within each year if the number of counts is $n > 4$. When $n = 5$ counts, there are 3,125 (5^5) possible recombinations of the data, whereas when $n = 4$ counts, there are only 256 (4^4) different recombinations.

The delta method may be used to estimate the overall variance of $\hat{N}$ (Mood et al. 1974, Seber 1982, Eberhardt 1982). The delta method is a special technique for variance estimation that is applicable to statistics that are functions of observed averages when there is more than one source of variation. In this case, the sources of variability come from different statistical distributions

$$\hat{var}\left[\frac{x_i}{p_i}\right] = \left(\frac{\overline{x}}{\hat{p}}\right)^2 \left(\frac{s^2}{\overline{x}^2} + \frac{\hat{p}(1-\hat{p})/n}{\hat{p}^2}\right)$$

$$\hat{var}(\hat{N}) = \left(\frac{\overline{x}}{\hat{p}}\right)^2 \left(CV^2(x_i) + CV^2(\hat{p})\right).$$

The squared coefficient of variation of the mean proportion missed is added to the coefficient of variation of the mean annual count to determine the overall variance of bias-corrected counts. Bounds of 95% confidence intervals are determined by calculating

$$\hat{N} \pm Z\left(\sqrt{\hat{var}(\hat{N})}/\sqrt{n}\right)$$

where $Z = 1.96$ for a standard normal distribution and n is the number of replicate counts. Alternatively, Z may be replaced by the appropriate t-value corresponding to n replicate total counts for more conservative confidence intervals.

Detection of Statistical Trends

I used program TRENDS (Gerrodette 1993) to estimate the statistical power to detect trends in time series of 3 bison population estimates. Power may be defined here as the probability of detecting a trend given that a trend exists in the data, or, as 1- β, where β is the probability of wrongly accepting the null hypothesis of no trend when a trend in fact exists. I estimated power over a range of values for rate of change (2% – 30%) with $\alpha = 0.05$ and 0.1 for 2-tailed tests following a *Z*-distribution. I used an initial value of CV (8%), which was comparable to the observed average CV for winter population estimates from replicate total counts. I assumed an exponential rate of increase as opposed to a linear rate of increase, and CV to be proportional to the square root of abundance. There was limited evidence to support the assumption of CV proportional to the 1/square root of abundance ($r^2 = 0.32$; n = 8) over CV proportional to square root of abundance ($r^2 = 0.29$), or CV proportional to abundance ($r^2 = 0.27$).

Validation of an Index

Caughley (1977) defined an index as 'any measurable correlative of density'. A key test of an index, therefore, is whether a correlation exists between counts and abundance. I tested the validity of total counts of bison as an index of population size by regressing the total number of bison observed during single surveys by NPS biologists on my corrected population estimates (Gibbs 2000). There were 8 such surveys conducted within 16 days (mean interval = 8 days) of each other for direct comparison. I tested the regression for effects of year, the number of days between my surveys and NPS surveys, and season because group size distribution, a major determinant of detection probability, was quite different between seasons.

Demographic Trends

I conducted simple analyses of population growth rates and apparent mortality rates to demonstrate the value and application of estimated population time series data for YNP bison. I estimated annual growth rates of YNP bison from 1998–2000 using simple population models described by Eberhardt (1987). These models are appropriate for assessing population dynamics in the presence of human-caused removals. Models that do not account for removals may be badly biased or result in biologically impossible rates of population change. During the period of this study, there was only one major removal event of 90 bison in the winter of 1998–99 (Yellowstone National Park, unpubl. data). These animals were detained and essentially unavailable to surveys that began in late January 1999, therefore, I consider them removed from the previous time step, and subtracted this number from each of my corrected 1997–98 survey estimates of adult bison from high-density strata (Table 23). I also excluded the obviously biased population estimate from the 4–5 of March 2000. I used simple linear regression to estimate annual increment (Cheville and McCullough 1998), and log-linear regression to estimate annual percent increase for the winter and summer series separately (Eberhardt 1987).

To determine if apparent mortality was related to winter conditions, I estimated an index of overall winter mortality by calculating the difference between mean winter and summer population estimates within each of the 3 years of study. I considered these differences to be representative of overall winter mortality despite the apparent negative bias in summer population estimates. I then regressed these differences on accumulated snow water equivalent (SWE_{acc}) from West Yellowstone for each year of the study (Farnes 1996).

RESULTS

Survey Framework and Stratification

I defined a total of 76 survey units within the 2,339 km^2 survey framework area based primarily on ridges that delineate drainages (Table 1). The total area of the survey framework comprised approximately 26% of the entire area of YNP (Fig. 2). Using long-term aerial survey observations from 1970–1997, 65 of these survey units were occupied by bison, and the locations of 99% of all groups (n = 21,450) and individuals (n = 318,651) occurred within my designated survey framework. The same framework and survey units were used in both winter and in summer, but survey units had different strata designations for each season. The winter high-density stratum comprised 52% of the total survey framework (Fig. 3), while 41% of the total survey framework was in the summer high-density stratum (Fig. 4).

In winter months of December–March, 93% of all bison group locations (n = 11,098) and 95% of all individual bison (n = 108,873) observed from 1970–1997 occurred within my designated winter high-density stratum. During the summer months of June–September, 88% of all bison group locations (n = 4,833) and 93% of all observed individual bison (n = 134,939) occurred within my designated summer high-density stratum. Ranking survey units according to the number of bison showed areas with consistently high use despite a trend of increasing range expansion and more than doubling in population counts over a

Table 1. Number and area (km^2) of aerial survey units delineated to provide complete coverage of 5 bison ranges (areas where bison movements are restricted during surveys) in and adjacent to Yellowstone National Park, Wyoming.

Range	Number of Survey Units	Total Area	Mean Area	Maximum Area	Minimum Area
Northern Range	23	1,185	51.5	110.1	8.5
Firehole-West	28	640	22.9	87.7	2.1
Pelican Valley	13	341	26.3	68.7	2.3
Hayden Valley	7	155	22.2	75.6	1.8
Southern Range	5	18	3.6	10.5	1.0
Total	76	2,339	30.8	110.1	1.0

27-year time series. I divided the time series into 3 periods corresponding to different population counts. Years 1970–1978 corresponded to population counts < 1,500 bison, 1979–1990 to years with 1,500–3,000 bison, and 1991–1997 to years with counts > 3,000 bison. Over this time series, in both winter and summer, bison occupied an increasing area, while in winter there was also a westward shift in distribution. The highest ranking unit in winter shifted from Hayden Valley to the Firehole drainage, while small numbers of bison increasingly occupied more units near the western boundary of YNP. In summer, bison consistently occupied less area than in winter, and Hayden Valley was exclusively the area of highest use. There was also more consistency in the

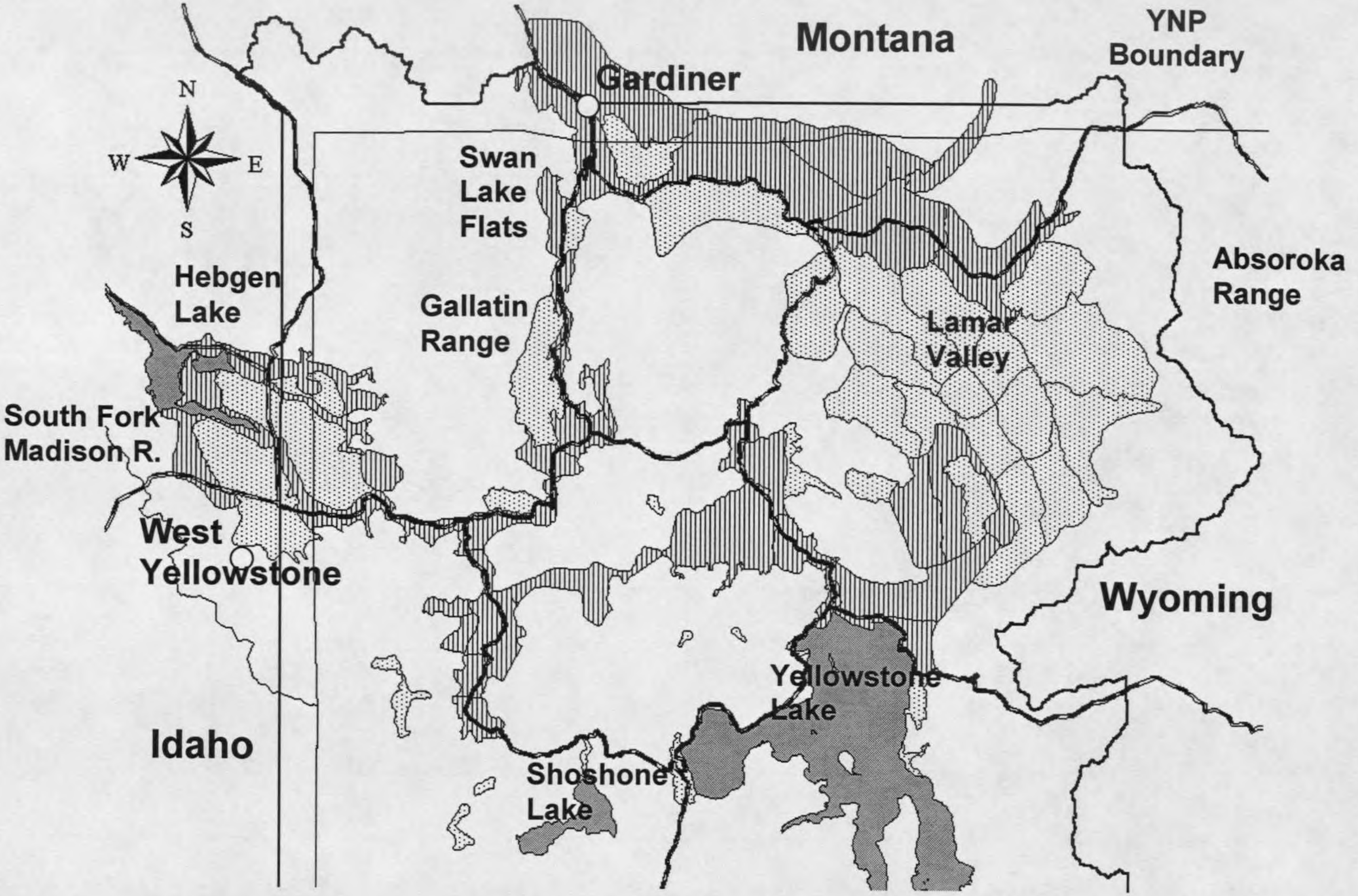

Fig. 3. Winter survey unit stratification for aerial surveys of bison in Yellowstone National Park, Wyoming. The 40 survey units in the high density stratum (vertical Lines) had an area of 1,219 km^2. The 36 survey units in the low density stratum (stipled) had an area of 1,120 km^2.

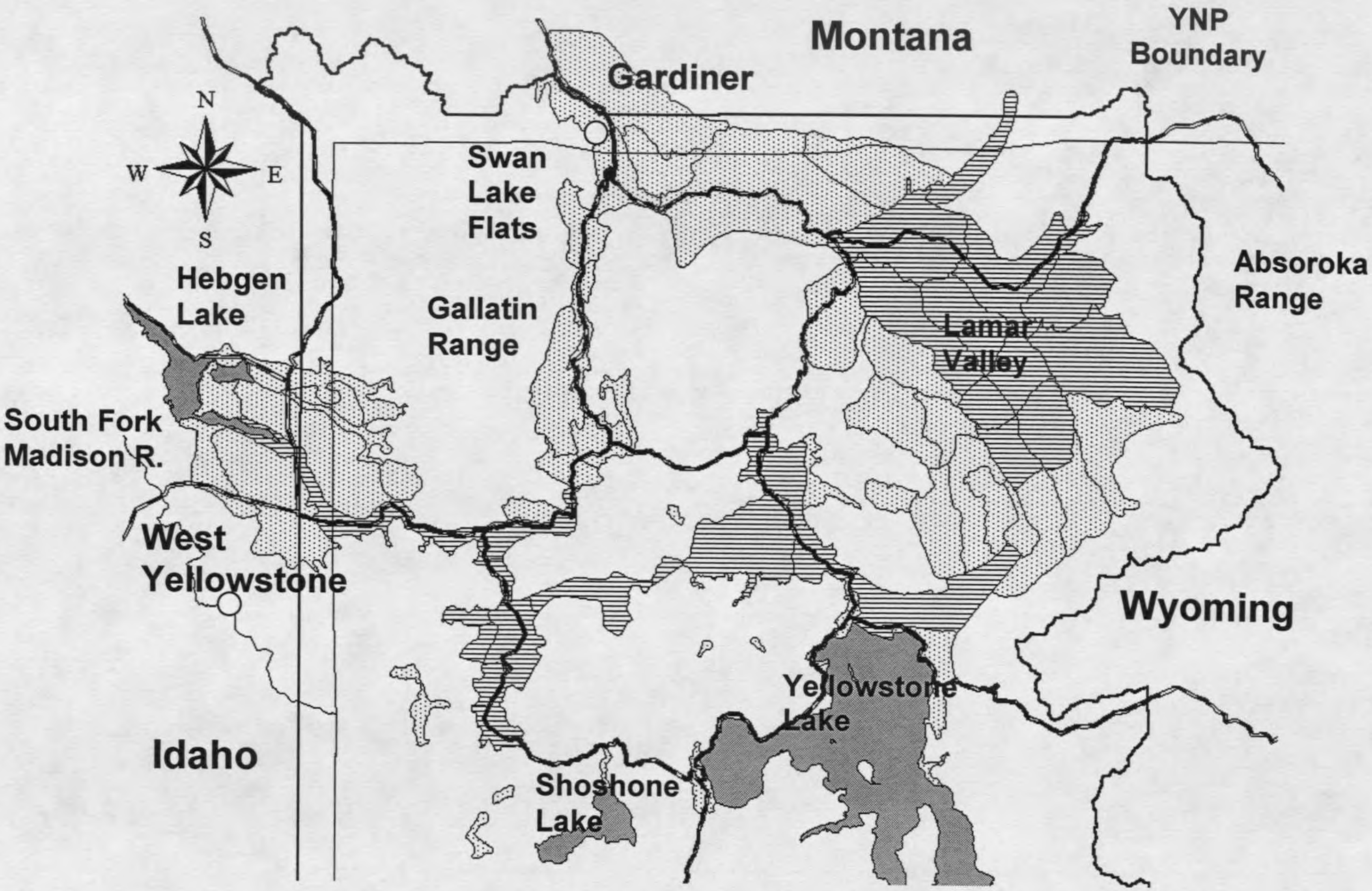

Fig. 4. Summer survey unit stratification for aerial surveys of bison in Yellowstone National Park, Wyoming. The 29 survey units in the high density stratum (horizontal Lines) had an area of 947 km^2. The 47 survey units in the low density stratum (stipled) had an area of 1,392 km^2.

rankings of summer survey units, in contrast to winter, when no single unit always held the greatest number of bison within any of the 3 time periods.

In the winters of 1970–78, 34 of 76 survey units were occupied by bison. A mean of 217 bison per survey occurred in the largest unit in Hayden Valley, representing 6–40% of the population count. This unit had the highest mean number of bison per survey, although it did not always contain the greatest number of bison during this time, or any other winter time period (Fig. 5). Single survey units in both Pelican Valley and in Lamar Valley also had consistently high numbers of bison, with ranks averaging 3–6, while surrounding units in these valleys, and in the Firehole drainage had units with lower average ranks. During the years 1979–90, 41 survey units were occupied and a similar general pattern held. The same Hayden Valley unit remained the highest average ranking unit with a mean of 338 bison per survey, representing 5–22% of the population count, while a survey unit in the Lower Geyser Basin of the Firehole drainage had the second greatest average rank with a mean of 278 bison per survey, representing 5–20% of the population count (Fig. 6). During the winters of 1991–97, 57 survey units were occupied. An increased number of survey units near the western boundary of YNP became occupied during this time and the Lower Geyser Basin of the Firehole drainage became the highest ranking survey unit with an average of 291 bison per survey, representing 3–10% of the population count, while the Hayden Valley unit decreased in rank with an average of 220 bison per survey representing 1–12% of the population count (Fig. 7).

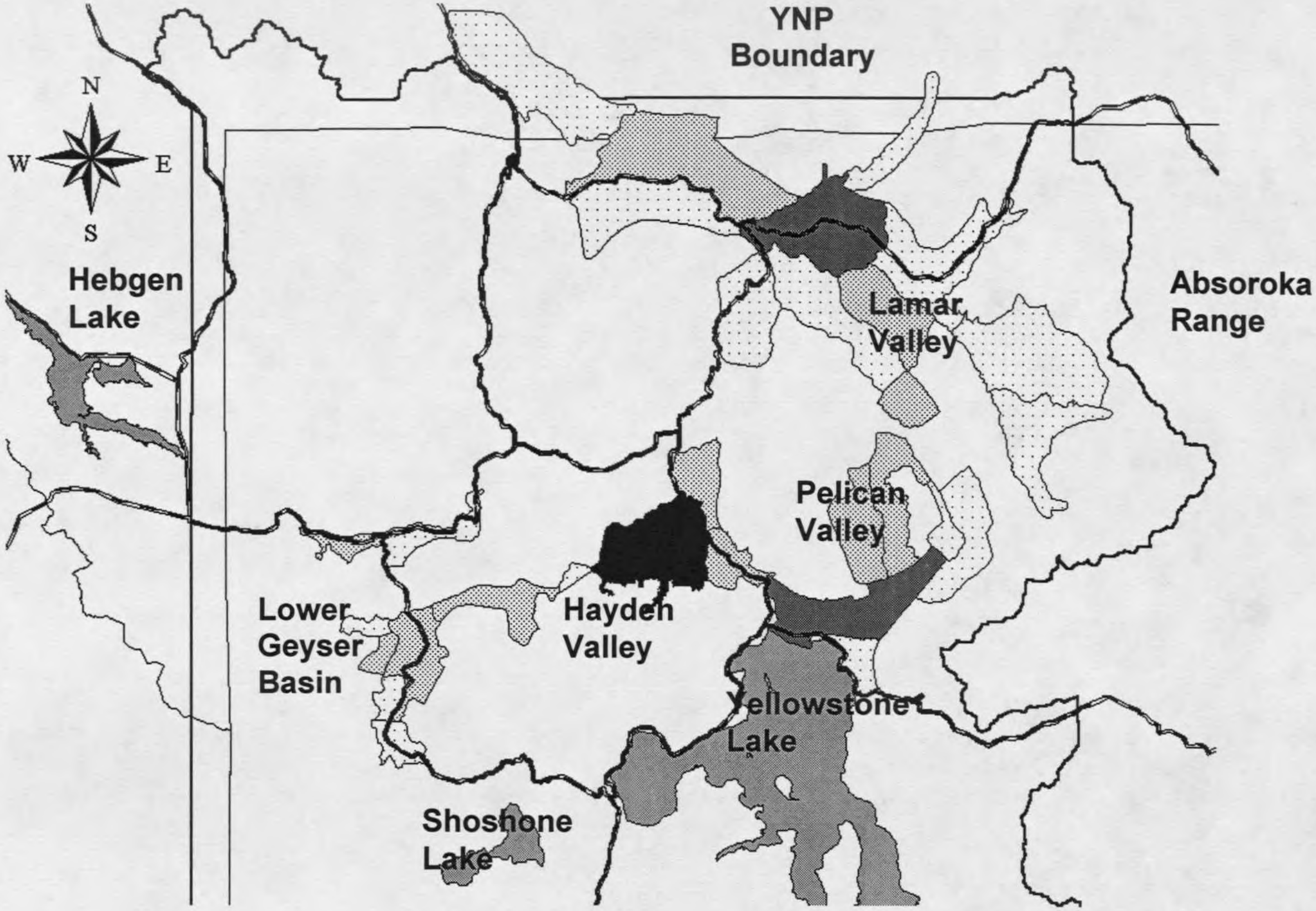

Fig. 5. Ranks of survey units according to number of bison observed during winter (January-April) aerial surveys of bison (n = 16) in Yellowstone National Park, Wyoming when population level was < 1,500. (M.L.Taper et al., Montana State University, unpubl. data). Solid black color indicates rank 1-3 (highest number of bison), darkest stipling- rank 3-6, intermediate stipling- rank 6-9, and lightest stipling- rank >9.

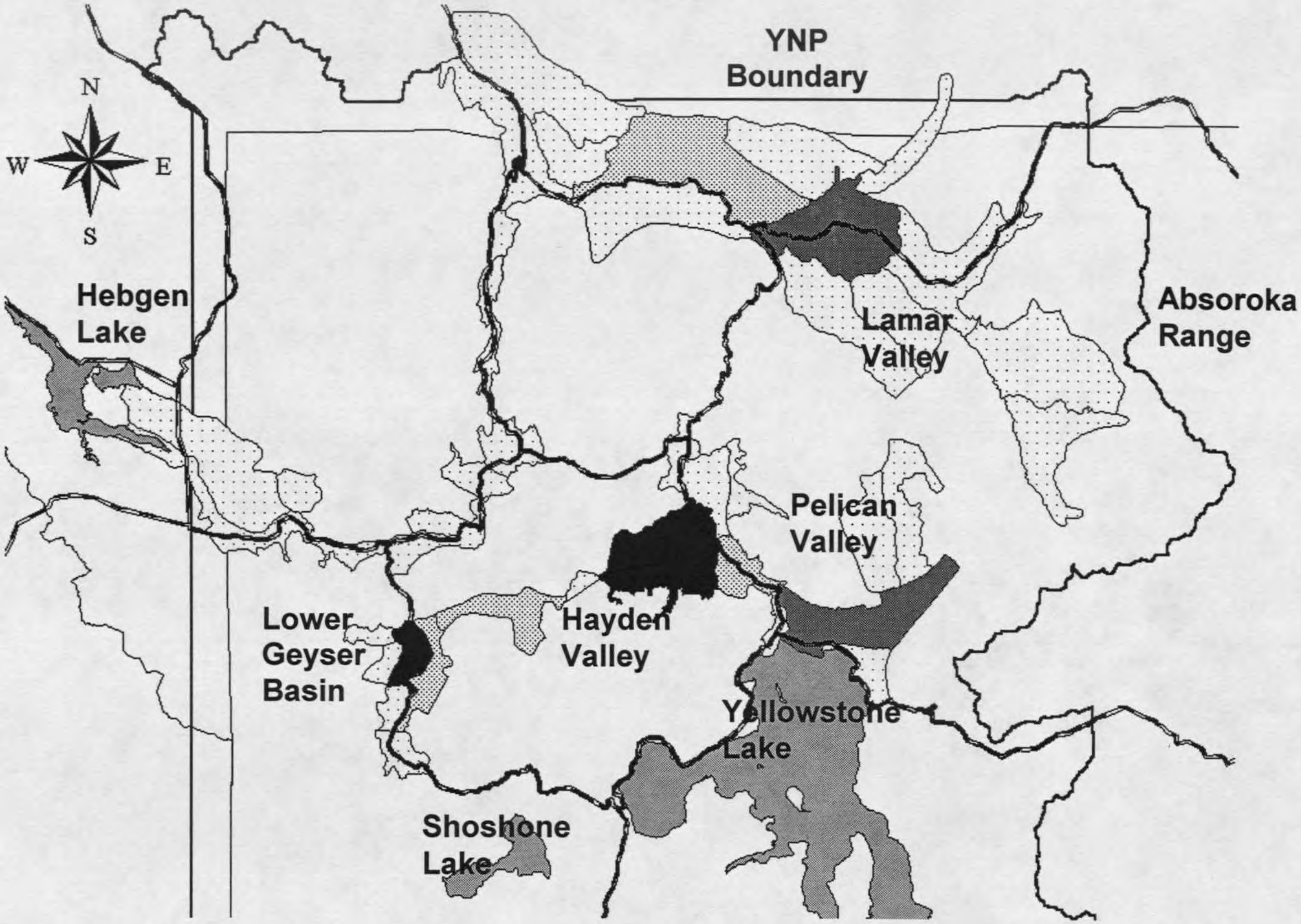

Fig. 6. Ranks of survey units according to number of bison observed during winter (January-April) aerial surveys of bison (n = 20) in Yellowstone National Park, Wyoming when population level was 1,500 - 3,000. (M.L.Taper et al., Montana State University, unpubl. data). Solid black color indicates rank 1-3 (highest number of bison), darkest stipling- rank 3-6, intermediate stipling- rank 6-9, and lightest stipling- rank >9.

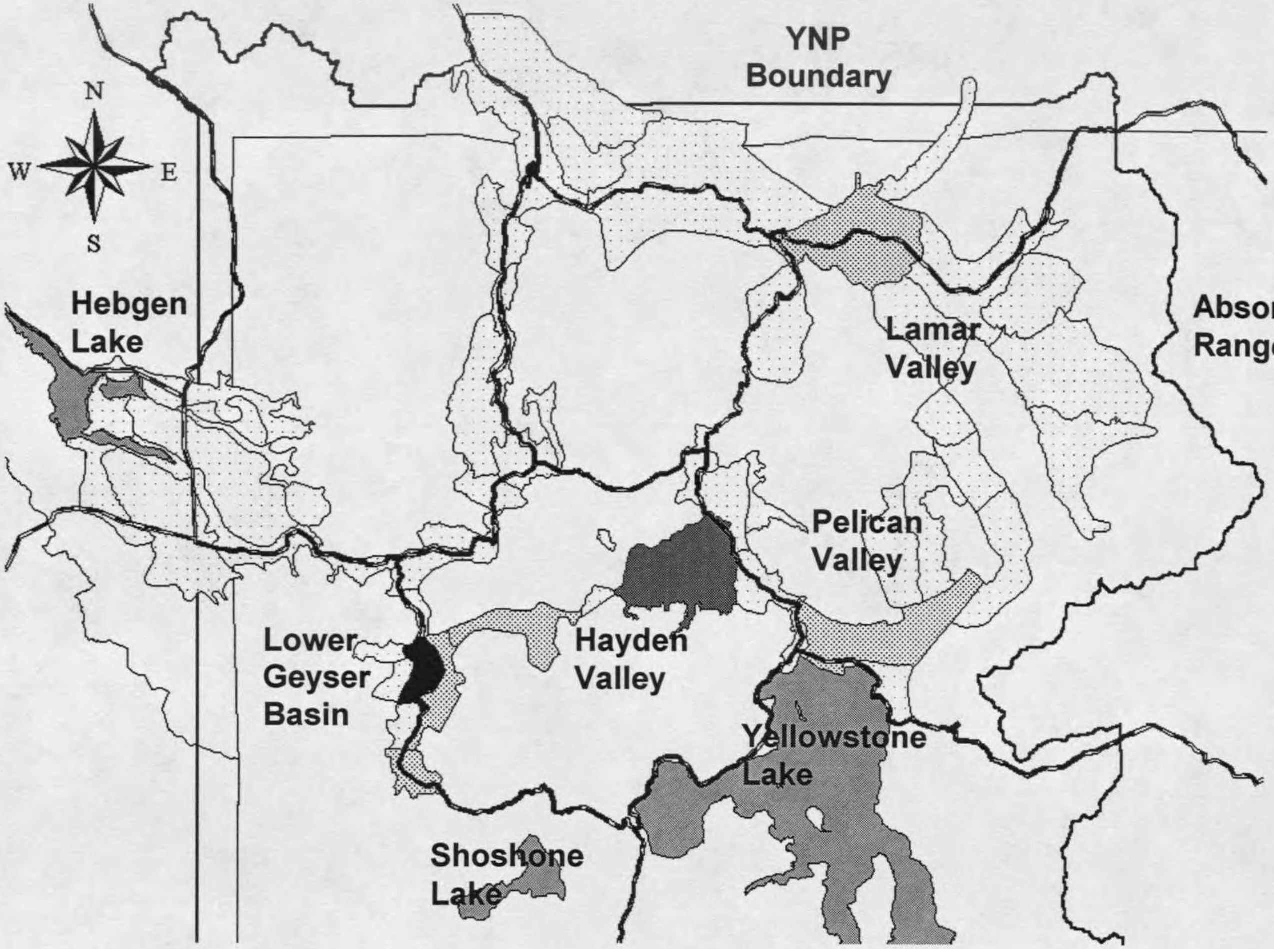

Fig. 7. Ranks of survey units according to number of bison observed during winter (January-April) aerial surveys of bison (n = 23) in Yellowstone National Park, Wyoming when population level was > 3,000. (M.L.Taper et al., Montana State University, unpubl. data). Solid black color indicates rank 1-3 (highest number of bison), darkest stipling- rank 3-6, intermediate stipling- rank 6-9, and lightest stipling- rank >9.

In the summers of 1970–1978, 30 of 76 survey units were occupied by bison. An average of 623 bison per survey occurred in the largest unit in Hayden Valley, representing 51–73% of the population count (Fig. 8). During the summers of 1979–1990, 37 survey units were occupied and the same Hayden Valley unit had a mean of 1,417 bison per survey, representing 42–67% of the population count (Fig. 9). During the summers of 1991–1997, 52 survey units were occupied. A mean of 1,395 bison per survey were observed in the Hayden Valley unit, representing 21–52% of the population count (Fig. 10). The largest survey unit in Hayden Valley had the highest mean number of bison per survey and the corresponding first rank each time it was surveyed across the entire summer time series, except for 3 surveys during the 1991–1997 period when higher counts occurred in immediately adjacent survey units within Hayden Valley. Other areas that contained consistently high numbers of bison in summer were the Lamar Valley and adjacent slopes of the Absoroka Range, Pelican Valley, and the Lower Geyser Basin of the Firehole River drainage.

<u>Aerial Survey Data</u>

From February 1998 through September 2000, I conducted 53 aerial survey flights totaling 180 hours in both summer and winter periods. I began my first winter and summer aerial survey field seasons in 1998 (Table 2). In 1999, winter surveys were complicated by unusually severe weather that caused some

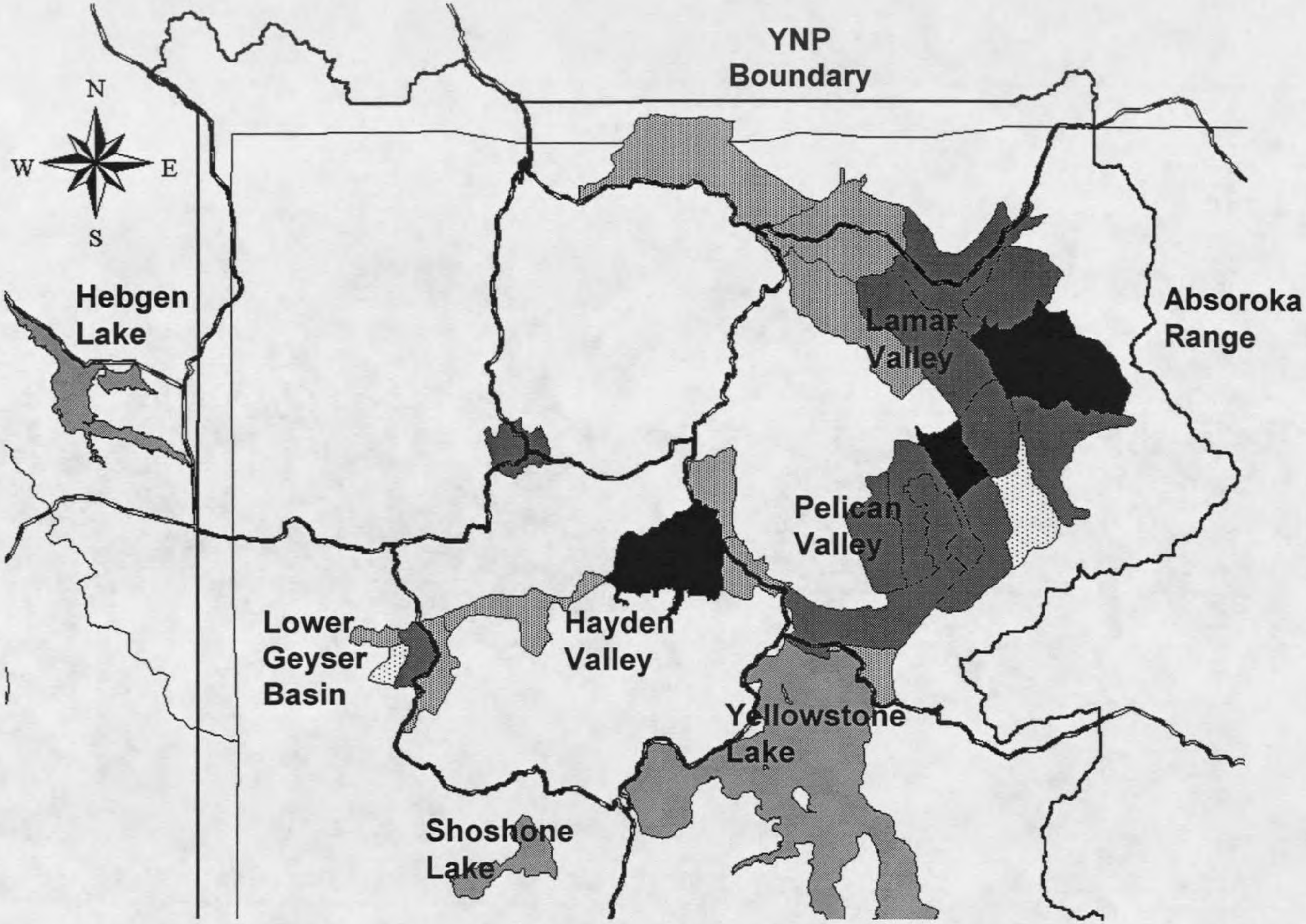

Fig. 8. Ranks of survey units according to number of bison observed during summer (June-September) aerial surveys of bison (n = 16) in Yellowstone National Park, Wyoming when population level was < 1,500. (M.L.Taper et al., Montana State University, unpubl. data). Solid black color indicates rank 1-3 (highest number of bison), darkest stipling- rank 3-6, intermediate stipling- rank 6-9, and lightest stipling- rank >9.

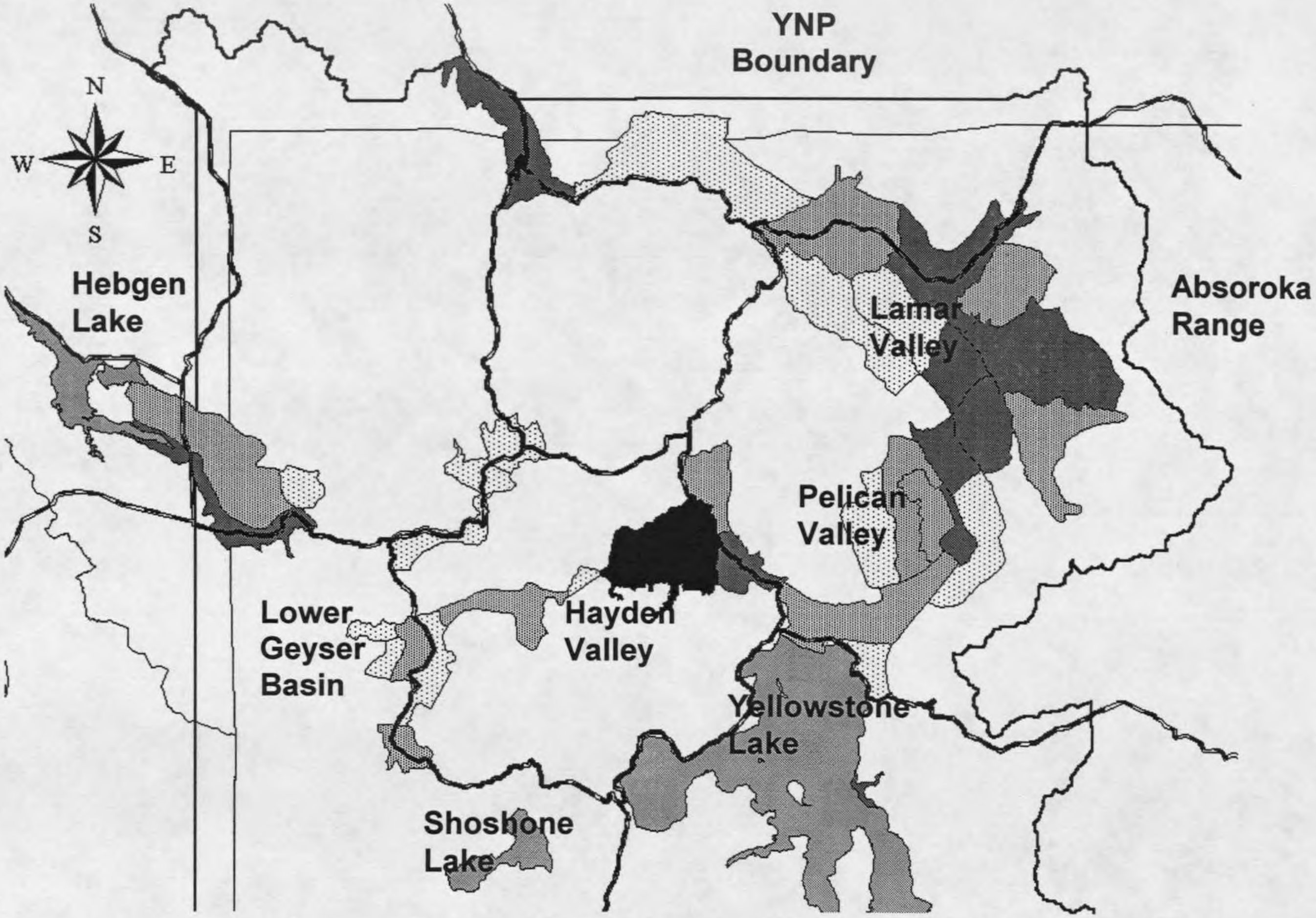

Fig. 9. Ranks of survey units according to number of bison observed during summer (June-September) aerial surveys of bison (n = 20) in Yellowstone National Park, Wyoming when population level was 1,500 - 3,000. (M.L.Taper et al., Montana State University, unpubl. data). Solid black color indicates rank 1-3 (highest number of bison), darkest stipling- rank 3-6, intermediate stipling- rank 6-9, and lightest stipling- rank >9.

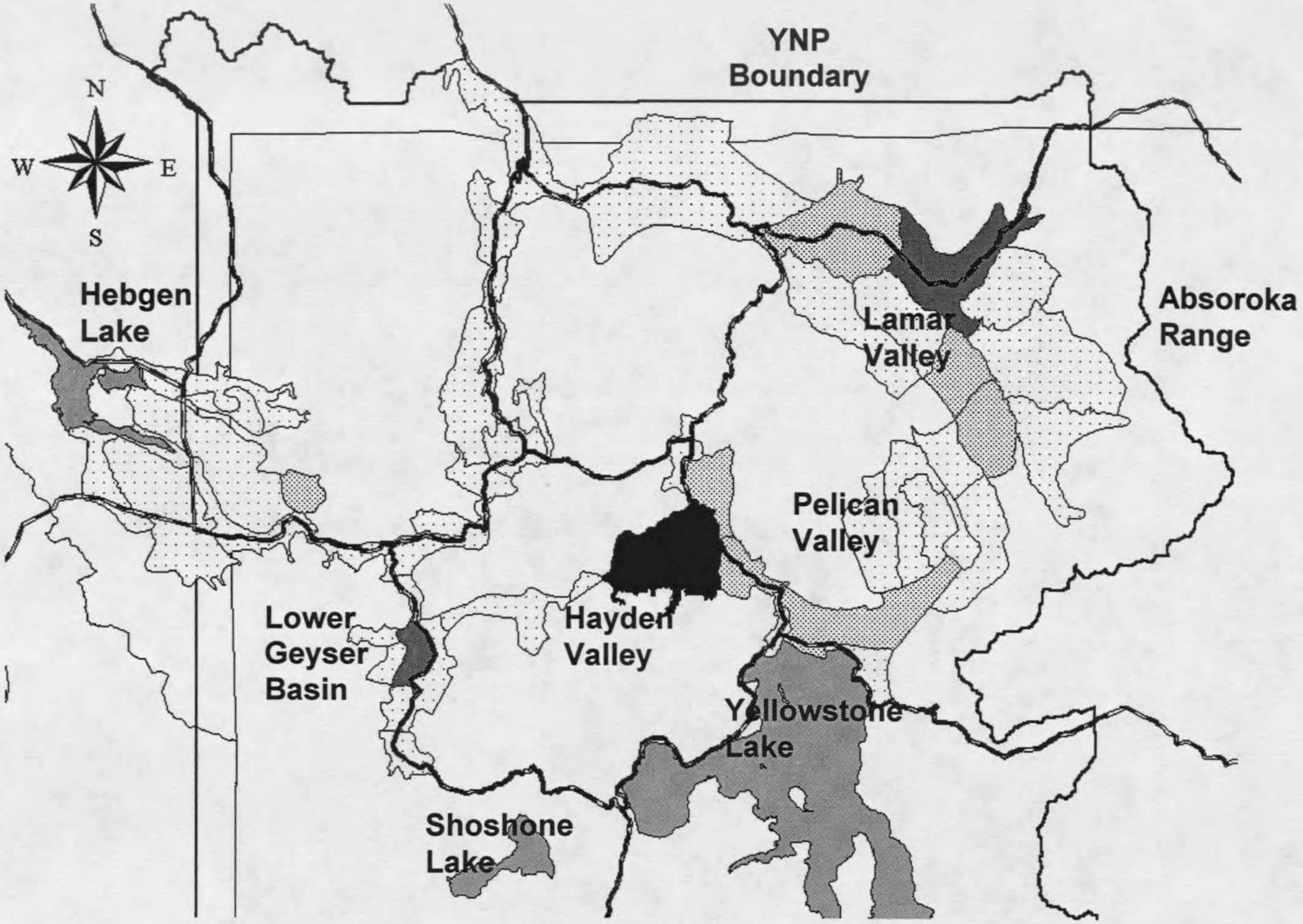

Fig. 10. Ranks of survey units according to number of bison observed during summer (June-September) aerial surveys of bison (n = 23) in Yellowstone National Park, Wyoming when population level was > 3,000. (M.L.Taper et al., Montana State University, unpubl. data). Solid black color indicates rank 1-3 (highest number of bison), darkest stipling- rank 3-6, intermediate stipling- rank 6-9, and lightest stipling- rank >9.

Table 2. Date, survey time in hours, type of aircraft, and number of bison detected in the Northern Range (NR), Pelican Valley (PV), Hayden Valley (HV), Firehole, Madison, Gibbon Rivers and West Yellowstone (F-W), and Southern areas (S) during aerial surveys in Yellowstone National Park, Wyoming, 1998.

Date	Time	Aircraft	Number of Bison Detected					
			NR	PV	HV	F-W	S	Total
17 February	4.3	Piper PA-18	396	208	586	708	--	1,898
28 February	5.9	Cessna 182	361	202	647	689	--	1,899
14 March	2.3	Cessna 182	--	--	--	642[a,b]	2	---
22-23 April	6.6	Cessna 182	410	78	61	1,265[b]	--	1,814
25 May	2.1	Cessna 182	--	--	110	618[a,b]	--	---
13 June	2.0	Cessna 182	--	53	717[b]	--	--	---
3-4 July	5.8	Cessna 182	472	14	1,130[b]	305[b]	--	1,921[c]
17 July	2.0	Aviat A-1	--	--	1,507[b]	--	--	---
27 July	2.1	Aviat A-1	--	--	1,518[b]	--	--	---
31 July	2.7	Aviat A-1	142[a]	0[a]	1,529[b]	--	--	---
1 August	2.7	Cessna 182	514	0	--	--	--	---
2 August	2.1	Aviat A-1	--	--	1,472[b]	--	--	1,986[d]
4 August	4.2	Aviat A-1	0[a]	--	--	0	--	---
5 August	3.3	Cessna 182	513	--	--	--	--	---
6 August	2.8	Aviat A-1	--	0[a]	1,530[a]	--	--	2,043[e]
8 August	3.6	Aviat A-1	572	--	--	--	0	---
26 August	1.9	Aviat A-1	--	--	1,615	3[a]	--	---

[a] Partial surveys.
[b] Aerial survey conducted with simultaneous ground survey.
[c] Sum of July 3-4
[d] Sum of August 1-4
[e] Sum of August 6-8

flights to be terminated prior to completion. In the 1999 summer season, however, flights were not affected by weather (Table 3). In my third summer field season in 2000, pilot and aircraft availability were limited due to wildfires (Table 4). I completed 3 surveys of high-density strata and one complete survey of low-density strata over the course of each season except during the summer of 2000, and surveyed areas outside of the framework 3 times over the course of the study. Overall survey effort was 143.0 km^2/hr (SE 18.3) in winter, and 168.1 km^2/hr (SE 11.3) in summer. In addition to the 65 survey units that were occupied by bison from 1970–1997, I found 3 other occupied units during my surveys, totaling 68 of the 76 units in the entire survey framework. I did not observe bison in areas outside of the survey framework in the southeastern part of YNP during winter, 1998, or on plateaus in the southwestern part of YNP during the summers of 1998 and 1999. The overall average CV for replicate surveys was 5.4% (Table 5), however, the CV for winter 2000 was 12.2%. Because the Northern Range herd and the Central herd had limited exchange, I also calculated the CV for each subpopulation. The overall average CV for the Northern Range was 9.0% while the overall average CV for the Central Herd was 5.8% (Table 6). Again, in year 2000, the CV for each subpopulation was much higher than the mean CV across years. This may be due, in part, to some mixing between subpopulations primarily in winter. The non-independent nature of exchange between these subpopulations may lead to a higher mean CV for each subpopulation than the mean CV of the entire YNP population.

Table 3. Date, survey time in hours, type of aircraft, and number of bison detected in the Northern Range (NR), Pelican Valley (PV), Hayden Valley (HV), Firehole, Madison, Gibbon, West Yellowstone (F-W), and Southern areas (S) during aerial surveys in Yellowstone National Park, Wyoming, 1999.

			Number of Bison Detected					
Date	Time	Aircraft	NR	PV	HV	F-W	S	Total
30 January	4.5	Aviat A-1	--	--	351	890[a]	--	---
31 January	4.3	Aviat A-1	383	353	--	--	--	---
1 February	1.5	Aviat A-1	--	--	--	114	--	2,091[c]
8 February	2.7	Aviat A-1	--	--	--	916[c]	--	---
15 February	3.4	Aviat A-1	392	--	--	--	--	---
20 February	3.4	Aviat A-1	--	296	243	--	--	---
27 February	1.2	Piper PA-18	--	--	--	587[a,b]	--	---
6 March	3.5	Aviat A-1	--	203	361	--	4	---
11 March	2.9	Piper PA-18	--	--	--	902	--	1,862[d]
12 March	1.2	Piper PA-18	99[a]	--	--	--	--	---
14 March	4.5	Aviat A-1	382	--	--	--	--	---
17 March	2.6	Cessna 182	--	179	346	--	1	---
18 March	2.6	Cessna 182	--	--	--	945	--	---
19 March	2.2	Cessna 182	--	--	--	132	0	1,984[e]
20 March	1.4	Aviat A-1	--	--	--	176[a]	--	---
21 March	0.7	Aviat A-1	--	--	--	205[a,b]	--	---
15 June	1.6	Aviat A-1	--	--	900	--	--	---
24 July	3.6	Aviat A-1	545	9	--	--	0	---
26 July	4.6	Aviat A-1	--	--	1,679	9	--	2,242[f]
28 July	4.7	Aviat A-1	509	--	1,576	--	--	2,085[g]
13 August	3.5	Aviat A-1	548	--	--	--	--	---
14 August	2.6	Aviat A-1	--	--	1,687	--	--	2,235[h]
13 September	2.6	Aviat A-1	--	--	1,533 [a]	--	--	---

[a] Partial aerial survey of area conducted with simultaneous ground survey.
[b] Survey terminated early due to poor weather conditions.
[c] Sum of 30 January - 1 February, 1999.
[d] Sum of 15 February, 6 March & 11 March, 1999.
[e] Sum of 14 March - 19 March, 1999.
[f] Sum of 24 July - 26 July, 1999.
[g] Sum of 28 July, 1999.
[h] Sum of 13 August - 14 August, 1999.

Table 4. Date, survey time in hours, type of aircraft, and number of bison detected in the Northern Range (NR), Pelican Valley (PV), Hayden Valley (HV), Firehole, Madison, Gibbon, West Yellowstone (F-W), and Southern areas (S) during aerial surveys in Yellowstone National Park, Wyoming, 2000.

			Number of Bison Detected					
Date	Time	Aircraft	NR	PV	HV	F-W	S	Total
29 January	4.8	Aviat A-1	417	341	--	--	--	---
30 January	5.8	Aviat A-1	--	--	582	1,196[a]	--	2,536[b]
31 January	3.3	Aviat A-1	--	--	--	1,031[c]	0	---
18 February	3.5	Piper PA-18	440	371	--	--	2	---
19 February	4.8	Piper PA-18	--	--	273	1,341	--	2,427[d]
4 March	5.4	Piper PA-18	272	309	351	--	--	---
5 March	3.5	Piper PA-18	--	--	--	1,070	--	2,002[e]
25 March	4.2	Piper PA-18	293	162	--	--	--	---
26 March	2.3	Piper PA-18	--	--	--	1,220[a]	--	---
16 July	2.1	Piper PA-18	--	--	1,841[a]	--	--	---
30 July	2.8	Piper PA-18	634	--	--	--	--	---
4 August	5.5	Piper PA-18	89	4	2,139[a]	1	--	2,867[f]
26 September	5.5	Piper PA-18	--	372	1,020[a]	904	--	---

[a] Aerial survey of area conducted with simultaneous ground survey.
[b] Sum of 29 January - 31 January, 2000.
[c] Partial aerial survey of area conducted with simultaneous ground survey.
[d] Sum of 18 February - 19 February, 2000.
[e] Sum of 4 March - 5 March, 2000.
[f] Sum of 30 July - 4 August, 2000.

Table 5. Mean and variation in total number of bison observed during aerial surveys in Yellowstone National Park, Wyoming, 1998-2000. CV indicates percent coefficient of variation.

Season	N	Mean Count	Maximum Count	Minimum Count	Standard Error	CV (%)
Winter 1998	3	1870	1899	1814	28.1	2.1
Summer 1998	3	1983	2043	1921	35.2	3.1
Winter 1999	3	1983	2091	1874	62.6	5.5
Summer 1999	3	2187	2242	2085	51.2	4.1
Winter 2000	3	2321	2536	2002	162.9	12.2
Summer 2000	1	2867	---	---	---	---

Table 6. Mean and variation in total number of bison observed during aerial surveys of the Northern Range and Central Herds in Yellowstone National Park, Wyoming, 1998-2000.

	Northern Range					Central Herd				
Season	N	Mean	Max	Min	CV (%)	N	Mean	Max	Min	CV (%)
Winter 1998	3	389	410	361	6.5	3	1481	1538	1404	4.7
Summer 1998	4	518	572	472	7.9	7	1434	1620	1449	3.8
Winter 1999	3	389	402	382	2.9	3	1594	1708	1476	7.4
Summer 1999	3	534	548	509	4.1	3	1653	1697	1576	4.1
Winter 2000	4	356	440	272	23.7	3	1945	2119	1730	10.2
Summer 2000	1	634	---	---	---	2	2220	2296	2144	4.8

During surveys conducted between July and the end of August, I was able to classify bison as either adults that were older than 1 year or as calves of the year. Although the overall variability between counts of adult bison was low with < 6% CV, in year 2000 there was a much higher CV of 14.1% during counts of the Central Herd, primarily due the inability to complete one of my surveys (Table 7). In 1998 and 1999, variability was slightly greater in counts of the Central Herd (mean CV = 5.1%) and the Northern Range (mean CV = 4.5%) than the overall variability in counts (mean CV = 4.0%). Due to the great distance between the aggregations of bison I observed, I believe there was limited exchange between the Central herd and Northern Range subpopulations during summer, and that the differences in variability was due to error in counts. The variability in counts of adult bison was less than the variability of counts in calves of the year in most cases.

Bison calves were located in only 3 areas of YNP during summer surveys after early July. The majority of all calves were located in Hayden Valley, but on the Northern Range, they were located both in Lamar Valley and on the Mirror Plateau (Table 7). When I examined variation in total counts of bison calves, variation in Hayden Valley (mean CV = 10.0%) and the Northern Range (mean CV = 10.1%) was greater than the overall variation from both areas combined (mean CV = 8.4%). There are 3 reasons for greater variation in calf counts: 1) new calves continue to be born in the months of July and August, but 2) there is also high mortality associated with these neonates, and 3) calves tend to be closely attended by dams, and many are obscured from the aircraft vantage point. Additionally, observer experience may be more important in accurately enumerating calves than adult bison. These counts probably underestimate the true number of calves during any given survey and are difficult to reconcile, however, may serve as an index of calf production if they are conducted at the same time each year.

Table 7. Mean and variation in summer counts of adult and calf bison in Yellowstone National Park, Wyoming, 1998-2000.

			Adults				Calves			
Location	Year	N	Mean	Min	Max	CV (%)	Mean	Min	Max	CV (%)
Central Herd	1998	7	1230	1158	1308	4.4	286	234	349	13.5
	1999	3	1370	1300	1454	5.7	260	233	276	9.0
	2000	2	1683	1515	1851	14.1	310	293	326	7.5
Northern Range	1998	3	417	398	427	4.0	87	86	88	1.3
	1999	3	469	446	493	5.0	54	43	63	18.8
	2000	1	543	--	--	---	91	--	--	---
Overall	1998	3	1626	1598	1672	2.5	354	322	371	8.6
	1999	3	1840	1746	1947	5.5	313	288	339	8.1
	2000	1	2226	---	---	---	384	---	---	---

Six of my surveys of high-density areas that were within 11 days of surveys conducted by NPS showed agreement of 98.7%–105.6% (W. Clark and

J. Mack, unpubl. data) although some of my surveys were conducted in different aircraft than the Piper PA-18 employed by NPS, with different pilot/observer combinations and levels of experience and different survey methodologies (Table 8).

Table 8. Date, total number of bison observed, aircraft platform, and percent of NPS counts, in aerial surveys conducted by Montana State University research and the National Park Service in Yellowstone National Park, Wyoming, 1998 - 2000.

	NPS Group[a]		MSU Group			
Date	Aircraft	Total	Date	Aircraft	Total	% of NPS Count
11 Mar 98	Piper PA-18	1,799	28 Feb 98	Cessna 182	1,899	105.6
22 Apr 98	Piper PA-18	1,281[b]	22-23 Apr 98	Cessna 182	1,326[b]	103.5
7 Jul 98	Piper PA-18	1,946	3-4 Jul 98	Cessna 182	1,921	98.7
4 Aug 98	Piper PA-18	2,016	4-6 Aug 98	Aviat A-1/ Cessna 182	2,043	101.3
13 Mar 99	Piper PA-18	1,683	6-14 Mar 99	Aviat A-1/ Cessna 182	1,858	110.4
30 Jan 00	Piper PA-18	2,410	29-31 Jan 00	Aviat A-1	2,536	105.2

[a] J. Mack and W. Clark, National Park Service, unpubl. data.
[b] Partial survey by NPS compared to same areas surveyed by MSU group.

Temporal Windows for Surveys

I regressed the area occupied by bison and maximum group size from aerial survey data in both winter and summer on Julian date to determine trends in these variables and define optimal temporal windows for conducting surveys. The total area occupied by bison over the course of winter increased ($P < 0.04$; $r^2 = 0.51$; Fig. 11) after controlling for marginally significant observer effects ($P < 0.11$). The resulting regression equation for the predicted area occupied by bison in winter is:

$$y = 1038 + (1.46 \times \text{Julian Date}) + (75.5 \times \text{Observer})$$

where observer is an indicator variable taking a value of 1 for MSU, and zero for YNP. This increase in area occupied by bison corresponded to a major distributional shift from high valleys of the interior of YNP to lower elevation areas on the north and west boundaries of the park. The observed maximum group size also simultaneously declined in a log-linear manner over the course of winter ($P < 0.001$; $r^2 = 0.81$). After controlling for observer effects ($P < 0.005$), the winters of 1998 and 2000 had larger predicted maximum group sizes than in 1999 ($P < 0.009$), a year with above average snow accumulation (Fig. 11). The resulting regression equation for predicted maximum group size in winter is:

$$\text{Ln(Max Group Size)} = 5.66 - (0.018 \times \text{Julian Date}) - (0.403 \times \text{Year}) - (0.518 \times \text{Observer})$$

where year and observer are indicator variables taking on values of 1 for 1999 and MSU, and zero otherwise. Bison were most spatially concentrated in large

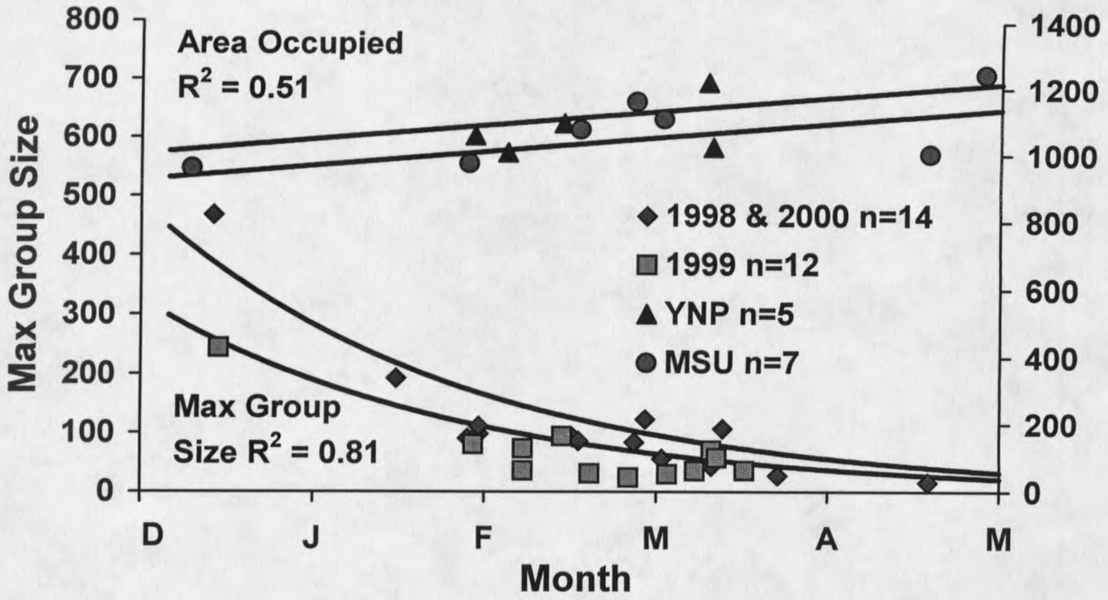

Fig. 11. Change in area occupied and maximum group size of bison during winter aerial surveys in Yellowstone National Park, Wyoming, 1998-2000. Predicted area occupied was controlled for observer effects (MSU vs YNP), while maximum group size was controlled for both observer and year effects (1999 vs 1998 and 2000).

groups in December and January and then became more widely scattered across the landscape in smaller groups as winter progressed. From aerial surveys in summer, the annual low point of area occupied by bison occurred in early August, at approximately Julian day 215 ($P < 0.017$; $r^2 = 0.74$; Fig. 12). At that time, bison occupied about 500 sq km, which was roughly half the area that bison occupied at any time during winter. The resulting regression equation for predicted minimum area occupied in summer is:

$$y = 18922 - (172.1 \text{ x Julian Date}) + (0.40 \text{ x Julian Date}^2).$$

Peak maximum group size also occurred in August (Julian day 227; $P < 0.001$; $r^2 = 0.51$; Fig. 12) after controlling for observer effects ($P < 0.02$), although slightly later than when bison occupied minimum area, at approximately the second week of August. The resulting regression equation for predicted maximum summer group size is:

$$\text{Ln(Max Group Size)} = -1.42 + (0.066 \times \text{Julian Date}) - (0.0002 \times \text{Julian Date}^2) - (0.63\ \text{Observer})$$

where observer is an indicator variables taking on values of 1 for MSU, and zero otherwise. Consequently, within seasons, survey conditions were closest to optimal when maximum group size was greatest and the area occupied by bison was lowest in winter during December and January, and then again in summer during the first 2 weeks of August.

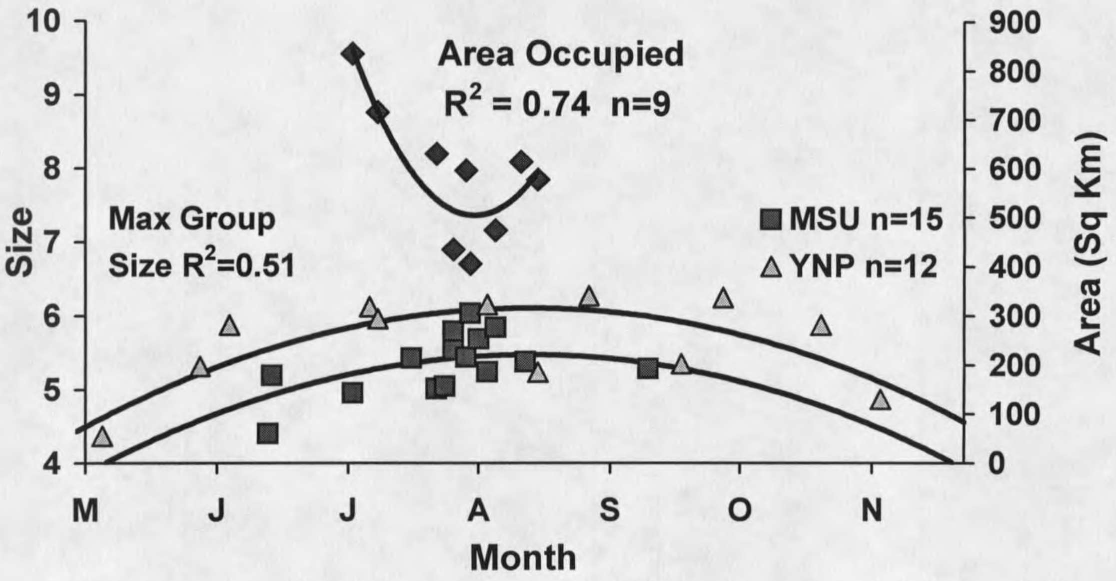

Fig. 12. Change in natural log of maximum group size and area occupied by bison during summer aerial surveys in Yellowstone National Park, Wyoming, 1998-2000. Predicted maximum group size was controlled for observer effects (MSU vs YNP).

I compared mean group size and cumulative group size distribution between seasons to determine if group size bias may affect surveys during one season more than another. Bison group sizes were significantly larger during the first complete survey in summer of each year (mean = 17.6; 95% CI 13.6–22.3; n=307) than during the first complete survey in winter (mean = 9.5; 95% CI 8.5–10.7; n=685; $P < 0.001$; 2-sample bootstrap test; Efron and Tibshirani 1993). The CDF of bison group sizes provides a graphic comparison of summer and winter distributions (Fig. 13). During summer surveys, I observed a high proportion of bison groups numbering in the hundreds while in winter I did not observe maximum group size >100 bison. Consequently, > 33% of groups in winter numbered $\leq$ 10 bison while in summer, < 9% of groups numbered $\leq$ 10 bison. Approximately half of all bison observed during winter were in groups of $\geq$ 17, while in summer, half of all bison were observed in groups $\geq$ 95. Because bison were much more highly concentrated in large groups in the open valleys of Hayden and Lamar during summer breeding period than in winter, the effect of group size bias on detection is minimized at this time.

I used bison ground survey data from the Madison, Gibbon, and Firehole River drainages to determine the potential effects of geothermal and forest habitat use on timing of aerial surveys. The proportion of bison found in geothermal areas was related to Julian date ($y = 0.08 + (0.003 \times \text{Julian Date}) + (0.00003 \times \text{Julian Date}^2)$; $r^2 = 0.45$; $P < 0.001$; Fig. 14), however, SWE was not a significant predictor of geothermal use in this model ($P > 0.61$). Although SWE and Julian date were related, the peak in geothermal use occurred before the

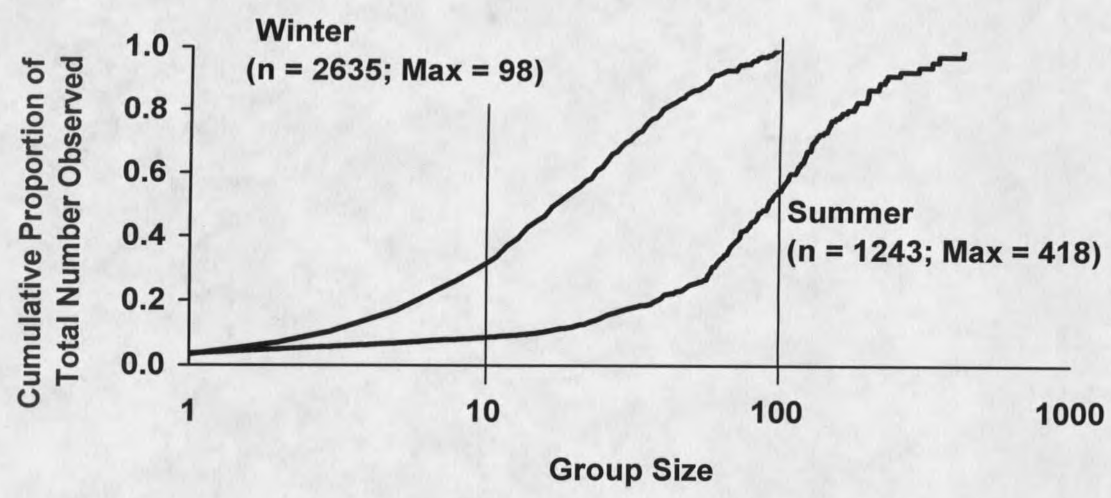

Fig. 13. Cumulative distribution function of Log 10 bison group size during winter (January-April) and summer (June-September) aerial surveys in Yellowstone National Park, Wyoming, 1998-2000.

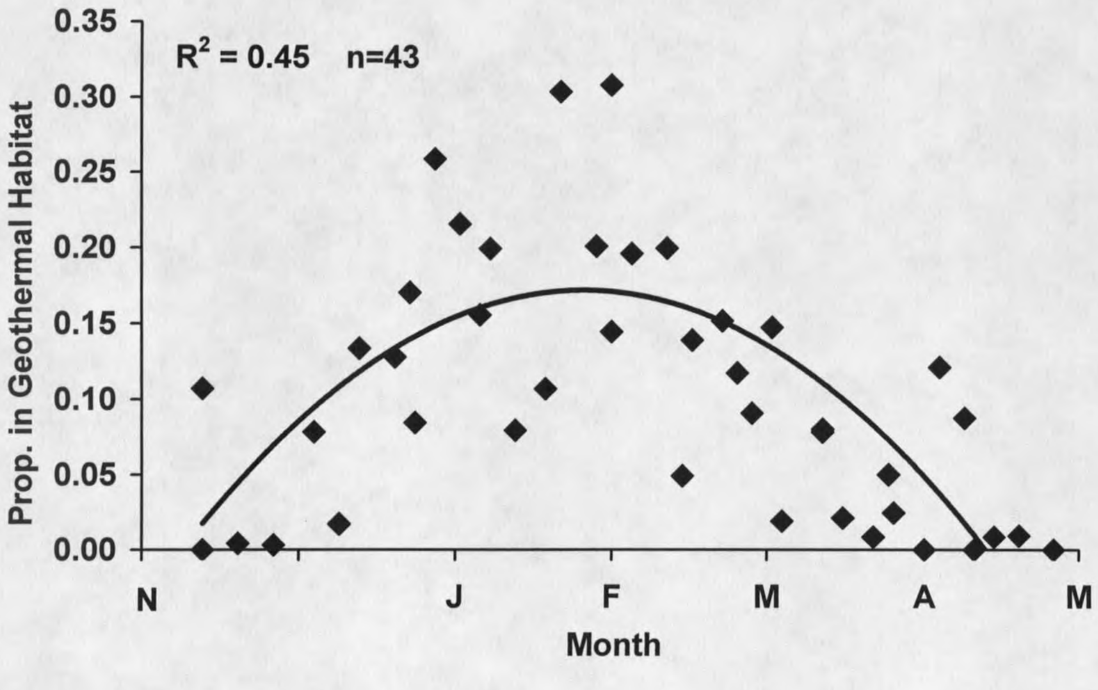

Fig. 14. Change in winter geothermal habitat use of bison during ground surveys in the Madison, Gibbon and Firehole drainages of Yellowstone National Park, Wyoming, 1998-2000.

end of February, long before the April peak in SWE. The proportion of bison found in forested areas in winter was related to average daily temperature ($y = 0.46 + (0.032 \times Temp) + (0.0006 \times Temp^2)$; $r^2 = 0.52$; $P < 0.001$; Fig. 15). A lower proportion of bison occupied forested habitats on days with colder average temperature than on warmer days. Conducting winter surveys prior to February should reduce the proportion of bison associated with unfavorable geothermal background conditions. This timing also coincides with larger group sizes and less total area occupied by bison. Winter surveys should also be conducted during periods of colder temperatures, thereby reducing the proportion of bison associated with forest habitats. These conditions may vary daily, but unusually warm winter conditions should be avoided during surveys if possible.

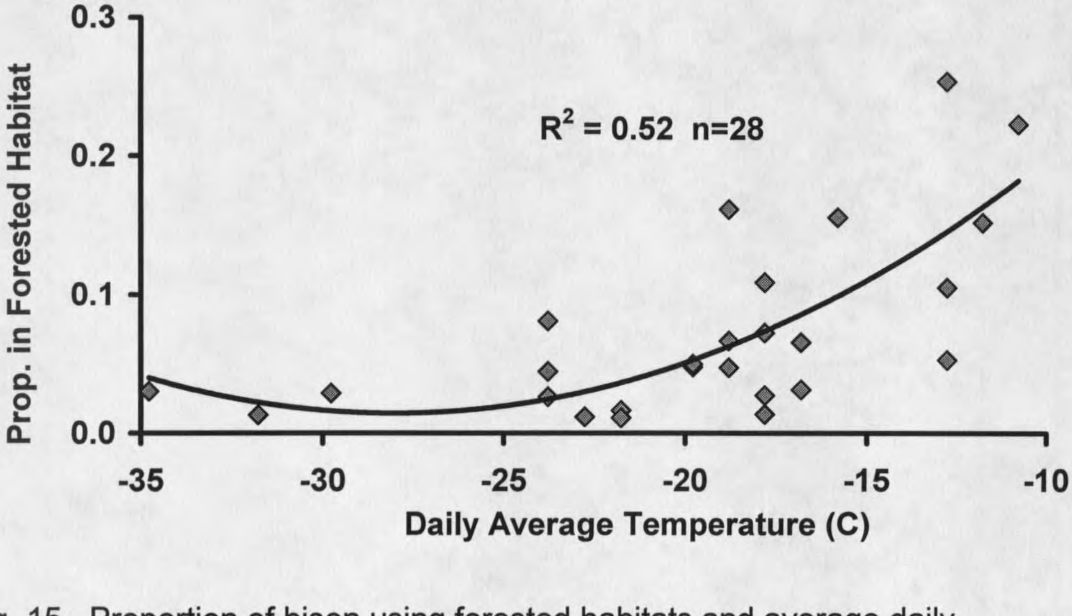

Fig. 15. Proportion of bison using forested habitats and average daily temperature (C) during ground surveys in the Madison, Gibbon, and Firehole River drainages of Yellowstone National Park, Wyoming, 1998-1999.

Simulations

In simulations of 5–60 sample survey units without stratification from the January 1999 data each replicated 1,000 times, SRS provided less than nominal CI coverage, but performed better with samples > 35 in most cases (Table 9). SRS was also the least biased overall, and had the second lowest CI width. The ratio 1 estimator was the only estimator to achieve nominal CI coverage on average. It was the second least biased but had the highest CI width in every case. The ratio 2 estimator had less than nominal CI coverage overall, the greatest bias, but also had the smallest CI width in every simulation. The PPS estimator had the lowest CI coverage for sample sizes 25–55, and the lowest coverage overall. PPS had less than nominal CI coverage and the second highest CI width across all sample sizes, and CI width did not decrease with sample size as rapidly as the other estimators. PPS may perform poorly if the survey unit area is not highly correlated with the number of bison within them. Most of the CI coverage failures were of right (upper) limits (i.e., the upper confidence limits were less than true density). On average, 7.3% of SRS upper confidence limits were less than true density, while only 0.4% of SRS lower confidence limits were greater than true density. In comparison, the ratio 1 estimator failed only 1.2% of upper, and 0.2% of lower limits, while the ratio 2 estimator failed 5.6% of upper and 1.7% of lower limits, and PPS failed 9.2% of upper and 0.2% of lower limits.

Stratification improved the overall precision of all of the estimators, while CI coverage improved for all but the separate ratio estimator, and average bias increased for both of the ratio estimators (Table 10). From simulations of 10–40 stratified sample survey units allocated 1:1 high:low-density from the January 1999 data, stratified SRS showed the second highest precision, close to nominal CI coverage, and bias of < 1%. The separate ratio estimator had the highest precision, but had the lowest nominal CI coverage for samples of 20–30, and had the greatest average bias. The combined ratio estimator was the only one to achieve nominal CI coverage on average, but showed negative bias, and had the second lowest precision. PPS had the lowest nominal CI coverage in every case, the lowest precision, and the second lowest bias after stratified SRS.

When I imposed 2:1 high:low-density allocation, the efficiency of both PPS and SRS estimators improved over 1:1 allocation with similar sample sizes. SRS performed second best in terms of CI coverage and precision, and best in terms of bias. The separate ratio estimator had close to nominal CI coverage, the lowest CI width, but also showed negative bias by as much as 2.4% with smaller samples, while the combined ratio estimator simultaneously had the second highest CI width on average and highest nominal CI coverage, but was negatively biased by > 3%. PPS had the lowest CI coverage and precision, with < 1.8% bias.

Increasing the allocation ratio to 3:1 high:low-density units resulted in equal precision for all but the combined ratio estimator. SRS had the second highest precision, nominal CI coverage in most cases, and negative bias of

Table 9. Sample size, percent confidence interval (CI) coverage, CI width, and percent bias from 1,000 simulated un-stratified samples of 4 different estimators for bison in Yellowstone National Park, Wyoming, January, 1999.

	SRS			Ratio 1			Ratio 2			PPS		
N	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias
5	76.6	365.4	-0.80	83.1	545.2	0.91	76.9	344.7	-6.09	80.8	376.6	-2.79
10	82.9	221.7	1.65	93.0	324.3	2.20	81.6	215.4	-0.50	87.1	229.3	-0.05
15	88.4	172.4	-0.84	95.9	250.8	0.16	88.4	167.3	-1.67	90.2	182.7	-2.37
20	88.3	137.0	0.07	97.6	212.1	-1.07	89.7	133.9	-1.88	88.7	152.2	-0.29
25	91.1	113.8	2.37	97.6	181.8	2.47	91.0	111.4	1.70	90.0	135.8	-0.64
30	92.4	97.6	0.42	98.1	158.5	-1.13	92.5	95.5	-1.42	89.7	120.4	0.58
35	94.8	83.9	-1.90	99.1	144.7	-1.92	94.5	82.1	-2.32	91.1	110.2	-0.03
40	94.7	70.3	-0.35	99.3	130.1	-0.51	94.7	68.8	-0.75	89.9	100.2	0.98
45	94.1	58.0	0.05	99.3	118.3	0.05	93.4	56.7	-0.15	91.6	93.6	0.59
50	94.2	46.3	0.14	99.4	107.6	0.39	93.6	45.2	0.24	91.7	87.7	0.15
55	94.6	33.6	0.00	100.0	101.5	0.18	95.0	32.9	0.09	91.0	82.1	0.18
60	91.1	17.2	0.12	100.0	94.4	0.15	93.6	16.8	0.13	91.8	77.8	-0.15
Mean	90.3	118.1	0.08	96.9	197.4	0.16	90.4	114.2	-1.05	89.5	145.7	-0.32

Table 10. Stratification ratio, sample size, percent confidence interval (CI) coverage, CI width, and percent bias from 1,000 simulated samples of 4 different estimators for bison in Yellowstone National Park, Wyoming, January, 1999.

		SRS			Separate Ratio			Combined Ratio			PPS		
Hi:Lo Ratio	N	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias
1:1	20	93.2	132.2	-0.36	91.6	125.3	-2.25	94.7	146.1	-1.89	91.6	156.6	-0.25
	30	92.8	90.4	0.99	91.3	85.8	-0.88	95.7	100.3	-0.33	92.6	122.2	-1.43
	40	95.8	65.6	-0.88	93.6	62.6	-0.71	96.7	72.5	-0.89	91.4	101.6	0.72
	50	95.4	43.8	0.20	94.3	42.2	0.04	97.5	48.0	0.06	93.6	89.5	-0.27
2:1	15	92.0	131.7	0.58	90.6	124.1	-2.42	94.0	174.8	-3.15	91.0	158.4	-1.74
	36	95.3	48.2	-0.62	95.0	62.1	-0.45	97.5	75.9	-1.39	91.6	90.0	1.79
	42	94.6	31.2	0.10	94.6	30.1	0.21	98.5	60.1	-0.13	93.2	83.2	-0.78
	45	94.9	21.3	0.14	94.2	20.4	0.01	98.9	53.0	0.07	92.4	79.8	-0.10
3:1	20	94.1	90.8	-0.25	93.4	87.0	-1.98	96.3	153.9	-4.60	91.3	122.4	-1.37
	24	95.2	74.8	-0.04	94.1	71.4	-1.01	96.0	128.9	-3.25	92.1	108.7	-0.22
	28	95.8	60.5	0.41	95.2	58.0	-0.37	97.6	112.2	-1.56	91.7	99.3	-0.47
	32	94.5	48.0	0.31	93.8	45.9	-0.50	97.5	98.1	-1.50	92.2	92.4	-0.54
	36	96.8	35.7	-0.68	94.1	34.2	-0.66	98.0	87.3	-1.25	93.0	86.0	-0.75
	40	94.0	21.3	0.01	93.6	20.4	-0.12	97.1	76.0	-1.53	91.4	78.7	1.04
Mean		94.6	64.0	-0.01	93.5	62.1	-0.79	96.9	99.1	-1.52	92.1	104.9	-0.31

< 0.4% on average. The separate ratio estimator had the lowest CI width in every case, slightly lower than nominal CI coverage, but the second highest bias on average. The combined ratio estimator had the highest CI width and coverage, but was negatively biased, by as much as 4.6%. PPS had the lowest CI coverage, < 1.4% bias, but nearly twice the CI width as SRS. Stratification resulted in a lower overall CI coverage failure rate than sampling without stratification, but most of the CI coverage failures were still from right limits, except for the ratio estimators. On average, 5.0% of SRS upper confidence limits were less than true density, while only 0.3% of SRS lower confidence limits were greater than true density. The separate ratio estimator failed 0.3% of both upper and lower limits, the combined ratio estimator failed 1.5% of both limits, and PPS failed 7.7% of upper and 0.3% of lower limits. The ratio estimators exhibited no major advantages over SRS, and the most bias on average, hence I excluded them from further winter simulations.

When I created a complete coverage block from 8 survey units from the winter 1999 data and repeated stratified simulations using only SRS and PPS for the remaining survey units, there was a similar gain in efficiency with greater high:low stratification ratios, but in each case, the highest precision required nearly complete coverage of the high-density stratum (Table 11). Overall CI coverage was above nominal levels for SRS, but only 90.2% for PPS, while average CI width for PPS was more than twice SRS. Most of the CI coverage failures were, again, from right limits. For SRS, 4.2% of upper limits were less

Table 11. Stratification ratio, sample size, percent confidence interval (CI) coverage, CI width, and percent bias from 1,000 simulated stratified samples with a complete coverage block of 8 survey units using simple random sampling (SRS) and sampling proportional to size (PPS) estimators for bison in Yellowstone National Park, Wyoming, January, 1999.

		SRS			PPS		
Hi:Lo Ratio	N	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias
2:3	20	92.4	64.2	0.31	84.1	104.6	0.62
	25	94.7	52.0	0.49	85.8	92.2	-0.05
	30	94.9	43.3	-0.14	88.9	81.7	0.12
	35	93.4	35.4	0.55	89.5	74.4	0.01
	40	95.4	29.7	0.00	92.0	69.5	-0.60
	45	95.9	24.1	-0.19	91.1	64.1	-0.17
1:1	20	94.5	52.0	-0.05	87.6	92.8	-0.14
	24	93.9	43.2	1.27	88.0	81.6	0.13
	28	94.9	36.0	-0.30	88.8	73.6	0.27
	32	96.0	29.7	-0.14	91.9	69.6	-0.33
	36	95.7	24.1	-0.03	91.0	63.2	0.14
	40	95.6	18.4	0.12	91.9	60.1	-0.09
	44	96.5	12.4	0.07	94.3	57.2	-0.52
3:2	20	93.9	42.9	0.01	87.7	81.7	0.46
	25	96.0	32.9	-0.34	89.3	70.5	-0.03
	30	96.3	24.0	0.09	90.8	63.5	-0.22
	35	95.8	15.6	0.17	90.9	58.0	-0.09
2:1	21	94.7	35.9	0.13	90.5	74.7	-0.06
	24	96.6	29.7	0.11	91.7	68.4	0.12
	27	96.1	24.0	-0.01	90.3	63.1	0.40
	30	96.5	18.5	-0.03	91.5	59.9	-0.01
	33	96.3	12.4	-0.02	92.0	56.6	-0.11
3:1	20	95.3	32.8	0.05	91.1	71.9	-0.13
	24	95.8	24.0	0.05	93.5	65.8	-0.98
	28	96.2	15.6	-0.13	91.5	58.1	-0.25
Mean		95.3	30.9	0.08	90.2	71.1	-0.06

than true density, and 0.5% of lower limits were greater than true density, while 9.6% of upper limits, and 0.1% of lower limits failed with PPS on average.

When I related the total amount of area surveyed in winter to precision, including the area of the complete coverage block, SRS had more gain in precision (Fig. 16) from stratification than PPS (Fig. 17), but when I increased high:low-density allocation from 2:1 to 3:1, it did not result in substantial increases in efficiency for either sampling plan. Furthermore, when a complete coverage block was included in the sampling framework, there was very little difference between strata allocations, and precision was determined by sample combinations that allowed the greatest coverage of high-density strata. The differences in mean density and its associated variance between strata were not great enough to realize substantial gains in efficiency by allocating much greater effort to a high-density stratum for the winter 1999 data.

In simulations of sample surveys without stratification from the August 1998 data, overall CI coverage tended to be far less than nominal, and CI width and bias were much greater than the January 1999 simulations primarily because bison were highly aggregated (Table 12). SRS achieved only 70.9% CI coverage on average despite the second highest CI width, but had the second lowest bias overall. The ratio 1 estimator showed higher CI coverage for samples of 30–40 than SRS, lower CI width with small samples, but much higher CI width with large samples, and low bias on average despite some large negative values. The ratio 2 estimator mirrored SRS in most respects, having similar CI coverage, but lower CI width, and a positive overall bias due to the

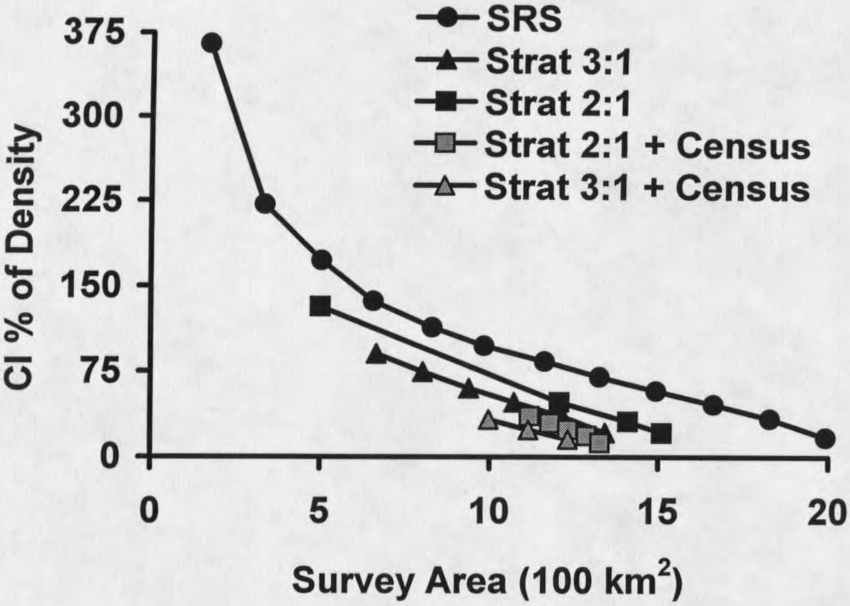

Figure 16. Efficiency of simple random sampling (SRS) and stratified SRS for winter aerial surveys of bison in Yellowstone National Park, Wyoming. Each point represents 1,000 simulations of the average CI width as the percent of density and the survey area with stratification ratios and with a census block in addition to stratification. See Appendix B for details on survey units used in strata and census blocks.

largest bias value (5.7%) with samples of 5. When this value was excluded, however, the average bias was 0.8%. PPS performed unlike other estimators, having the highest CI coverage except for samples of 15 and 30, the lowest CI width for samples of 10–35, and also showed large negative values in bias. All of the CI failures were due to the right (upper) confidence limit being less than true density, except for the ratio 2 estimator with small samples, in which < 0.1% of lower confidence limits were greater than true density. Upper limit failures were 27.6% for SRS, 27.6% for ratio 1, 28.9% for ratio 2, and 22.1% for PPS on average.

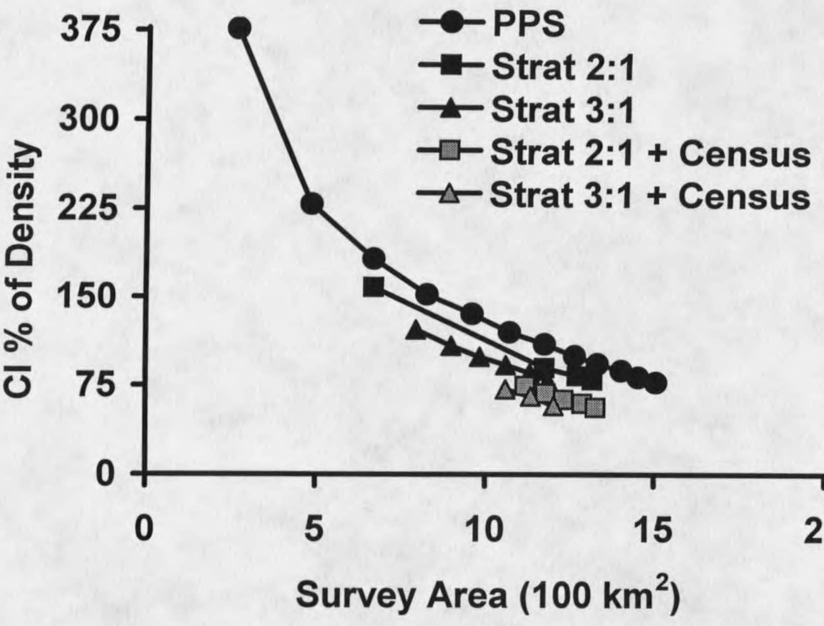

Figure 17. Efficiency of sampling proportional to size (PPS) and stratified PPS for winter aerial surveys of bison in Yellowstone National Park, Wyoming. Each point represents 1,000 simulations of the average CI width as the percent of density and the survey area with stratification ratios and with a census block in addition to stratification. See Appendix B for details on survey units used in strata and census blocks.

Stratification improved the overall CI coverage and reduced CI width of all of the estimators in comparison to similar sample sizes for the summer data without stratification, however, mean bias increased for both of the ratio estimators (Table 13). CI coverage was lower, and both CI width and bias were greater than the comparable simulations for winter due to the aggregated distribution of bison during summer. Overall CI coverage for SRS was less than nominal, but highest with the largest sample from each allocation ratio. The separate ratio provided slightly better CI coverage than stratified SRS, and had the lowest CI width in each case with only slightly greater bias overall. The

Table 12. Sample size, percent confidence interval (CI) coverage, CI width, and percent bias from 1,000 simulated un-stratified samples of 4 different estimators for bison in Yellowstone National Park, Wyoming, August, 1998.

	SRS			Ratio 1			Ratio 2			PPS		
N	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias
5	40.8	501.9	-1.66	40.8	456.0	-0.64	40.6	426.2	5.65	58.1	499.9	-3.71
10	55.8	355.9	-1.21	55.4	329.0	0.12	56.7	328.7	2.50	66.1	316.5	3.70
15	63.9	294.9	1.38	64.4	277.9	1.56	64.3	278.5	2.53	63.2	271.3	-3.86
20	71.4	264.8	-2.94	70.3	252.5	-3.01	71.3	251.4	-1.85	72.1	232.2	-3.06
25	74.8	222.0	1.29	74.8	219.4	1.27	72.5	213.0	1.85	75.5	199.1	-0.02
30	77.4	196.8	-2.93	82.3	205.6	-2.73	78.8	189.2	-2.14	81.5	172.8	0.97
35	78.5	161.6	0.44	83.5	181.7	0.59	78.8	155.8	0.94	88.1	161.3	-3.18
40	74.2	133.8	0.08	80.8	166.2	0.34	75.3	129.3	0.65	89.3	139.8	0.52
45	80.2	100.8	1.86	80.2	147.3	1.85	80.2	97.7	2.04	91.3	126.1	0.83
50	91.7	65.7	0.80	91.7	138.0	0.75	91.7	63.7	0.79	93.7	116.4	-0.13
Mean	70.9	229.8	-0.29	72.4	237.3	0.01	71.0	213.3	1.30	77.9	223.6	-0.79

Table 13. Stratification ratio, sample size, percent confidence interval (CI) coverage, CI width, and percent bias from 1,000 simulated samples of 4 different estimators for bison in Yellowstone National Park, Wyoming, August, 1998.

		SRS			Separate Ratio			Combined Ratio			PPS		
Hi:Lo Ratio	N	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias
2:3	30	80.3	163.9	0.45	80.9	154.4	0.85	81.4	170.5	0.33	79.9	196.8	0.07
	35	78.3	125.8	-0.34	78.6	118.8	-0.37	79.7	132.0	-0.84	84.9	173.0	1.16
	40	87.8	83.2	0.52	87.8	78.8	0.64	87.8	89.5	0.15	88.2	155.3	0.46
1:1	20	78.7	204.7	0.73	79.6	191.3	-0.04	79.3	217.5	-1.19	74.8	227.3	0.42
	30	85.9	107.3	-1.45	85.9	101.4	-1.51	86.0	121.9	-2.17	86.5	163.8	0.19
	32	88.7	83.6	0.20	88.7	79.2	0.38	88.4	99.6	-1.06	88.4	155.7	0.36
2:1	18	81.7	151.0	0.34	80.3	143.1	0.33	88.7	206.4	4.48	80.9	192.5	-1.07
	21	77.2	108.5	-1.00	81.5	103.0	3.04	87.7	166.9	2.53	87.6	173.4	1.58
	24	90.0	59.6	-0.49	94.2	56.8	-0.43	92.0	135.1	3.36	88.6	153.2	0.15
3:1	16	79.9	161.0	1.70	80.8	151.4	2.00	84.5	234.3	-2.96	80.6	193.8	3.00
	20	81.7	104.2	0.99	81.7	98.6	1.17	90.1	182.8	-2.17	87.4	164.5	0.44
4:1	15	81.1	161.7	1.25	81.5	152.2	1.36	90.9	303.7	-4.78	80.8	196.1	0.89
	20	88.6	127.2	0.29	88.6	79.3	0.10	95.1	221.5	-5.67	90.1	158.8	-2.07
Mean		83.1	127.2	0.29	83.4	119.9	0.36	86.5	174.9	-2.22	84.3	178.0	0.29

combined ratio estimator had the highest CI coverage overall due to high CI width, but also had the largest negative bias on average. PPS had the second highest CI coverage and highest CI width overall, and low bias on average, equal to SRS. Almost all of the CI failures were due to the right (upper) confidence limit being less than true density. SRS had a right CI failure rate of 16.9%, the separate ratio estimator was 16.6%, PPS was 15.7%, while the combined ratio estimator was 12.8%, and was also the only estimator with left failures, which averaged 0.7%.

Efficiency improved when allocation ratios of high:low-density units were increased up to 3:1, but CI width deteriorated for stratified SRS and combined ratio estimators with 4:1 allocation. For the other estimators, the gain in efficiency for 4:1 over 3:1 allocation was much less than the gain from 2:1 to 3:1 allocation. Furthermore, large samples of the high-density stratum could not be obtained with 3:1 and 4:1 allocation given the strata assignments I used for the simulation. SRS and the separate ratio estimators appear to be the best suited for stratified sampling during summer. The combined ratio estimator has the dual disadvantages of wide CI intervals and high bias, while PPS does not appear to offer any gains in efficiency for the highly aggregated distribution of bison in summer, most likely because the size of survey units is poorly correlated with the number of bison in them at that time. Based on the high bias of the combined ratio estimator and the poor CI coverage and precision of PPS, I excluded these estimators from further simulation.

When I created a complete coverage block from 4 survey units for the summer 1998 data and repeated stratified simulations for the remaining survey units, only 6 of which were occupied, the highly aggregated distribution of bison in summer required that at least 9 of the 14 high-density units be included for an adequate sample, thus only limited sample combinations could be simulated. Increased ratios of high:low allocation resulted in improved CI coverage and lower CI width for comparable sample sizes (Table 14). For both estimators, sample sizes of 16 with 4:1 allocation ratio performed nearly as well as sample size 36 with 1:2 allocation ratio. CI coverage never reached nominal levels, which were the same in each case for both estimators. Stratified SRS provided better precision, and substantially lower bias than the separate ratio estimator on average. All of the CI failures were due to the right (upper) confidence interval being less than true density, which averaged 20.5%.

When I related the amount of area surveyed in summer to precision, SRS had more gain in precision (Fig. 18) from stratification than PPS (Fig. 19), but when I increased high:low-density allocation from 2:1 to 3:1, it resulted in a greater increase in efficiency for both sampling plans than for the winter simulations. When a complete coverage block was created in the sampling framework, increased high:low-density strata allocations improved efficiency, as well as overall precision, and reduced bias for stratified SRS. Although there was only 1 combination of sample sizes for 3:1 and 4:1 allocation, they had almost as high CI coverage and precision, and as low bias as any other allocation that required more survey area. The differences in mean density and

Table 14. Stratification ratio, sample size, percent confidence interval (CI) coverage, CI width, and percent bias from 1,000 simulated stratified samples with a complete coverage block of 4 survey units using simple random sampling (SRS) and sampling proportional to size (PPS) estimators for bison in Yellowstone National Park, Wyoming, August, 1998.

		SRS			Separate Ratio		
Hi:Lo Ratio	N	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias
1:2	30	69.0	59.7	0.93	69.0	63.9	0.33
	36	87.3	41.3	-0.24	87.3	44.0	-0.34
2:3	25	70.4	60.9	0.24	70.4	65.1	-0.37
	30	85.5	40.8	0.06	85.5	43.3	-0.32
1:1	20	70.6	61.1	0.23	70.6	65.3	-0.33
	24	85.1	40.6	0.15	85.1	43.2	-0.26
3:2	15	68.4	74.9	-1.34	68.4	80.5	-2.06
	20	87.1	41.2	-0.22	87.1	43.8	-0.49
2:1	15	74.5	63.3	-0.69	74.5	67.7	-1.42
	18	86.4	41.1	-0.20	86.4	43.7	-0.38
3:1	16	83.5	40.1	0.52	83.5	42.6	0.25
4:1	15	86.7	41.2	-0.30	86.7	43.8	-0.50
Mean		79.5	50.5	-0.07	79.5	53.9	-0.49

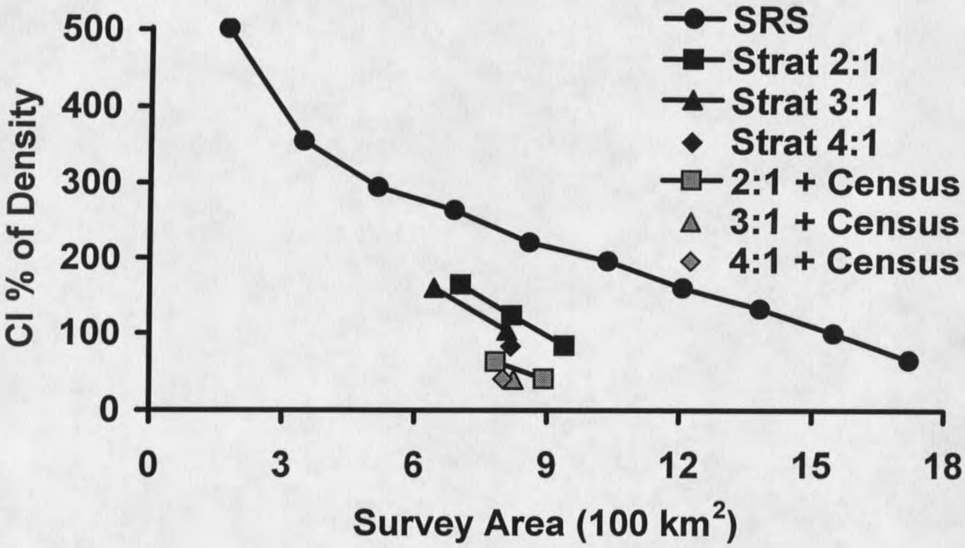

Figure 18. Efficiency of simple random sampling (SRS) and stratified SRS for summer aerial surveys of bison in Yellowstone National Park, Wyoming. Each point represents 1,000 simulations of the average CI width as the percent of density and the survey area with stratification ratios and with a census block in addition to stratification. See Appendix B for details on survey units used in strata and census blocks.

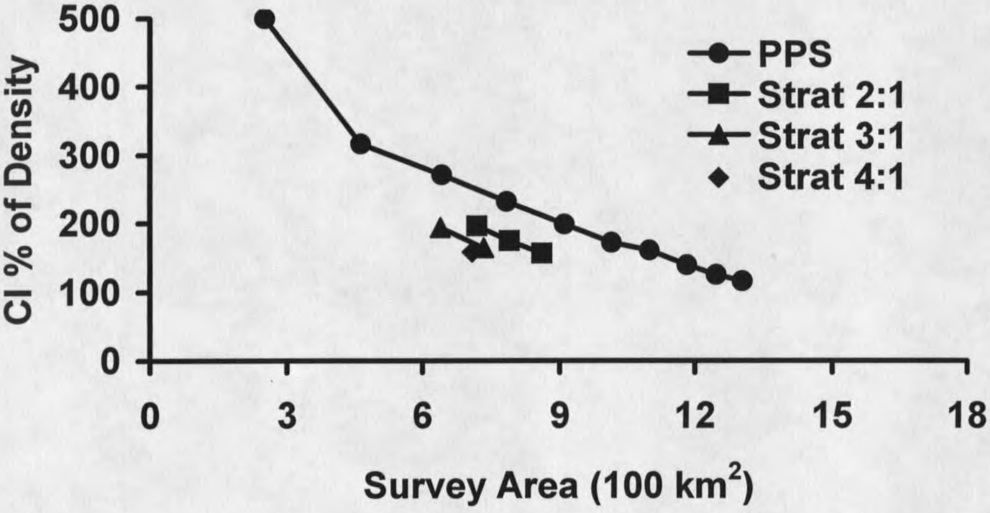

Figure 19. Efficiency of sampling proportional to size (PPS) and stratified PPS for summer aerial surveys of bison in Yellowstone National Park, Wyoming. Each point represents 1,000 simulations of the average CI width as the percent of density and the survey area required with different stratification ratios. See Appendix B for details on survey units used in strata and census blocks.

its associated variance between strata appear to be much greater in summer than in winter, allowing equivalent precision by allocating as much as 4:1 effort to a high-density stratum.

When I tested the stratified sampling plan with a complete coverage block on the winter 2000 data (Table 15), SRS had lower CI coverage and precision, but slightly less bias overall than for winter 1999 (Table 11). PPS, however, had slightly higher CI coverage, better precision, and similar overall bias. Higher allocation ratios of high:low-density resulted in similar CI coverage, precision, and overall bias with much reduced overall sample sizes, similar to winter 1999. When I tested the stratified sampling plan with a complete coverage block on the summer 1999 data (Table 16), SRS had lower CI coverage, but much higher precision, and less bias overall than for the 1998 data (Table 14). The separate ratio estimator, however, had similar CI coverage, much better precision, and much lower overall bias than for the 1998 data. Although CI coverage never reached nominal levels for summer 1999, the highest precision was achieved with little bias for any set of simulations. The poor precision of the summer 1998 simulations was caused by a single survey unit in the high-density stratum that contained 338 bison and created an extreme density value. Assigning this unit to the complete coverage block would improve estimates dramatically, but would require prior knowledge of the distribution, which could only be accomplished by a reconnaissance flight.

I examined the best stratified SRS estimates with complete coverage blocks and estimated the amount of time it would take to conduct surveys under

Table 15. Stratification ratio, sample size, percent confidence interval (CI) coverage, CI width, and percent bias from 1,000 simulated stratified samples with a complete coverage block of 8 survey units using simple random sampling (SRS) and sampling proportional to size (PPS) estimators for bison in Yellowstone National Park, Wyoming, January, 2000.

		SRS			PPS		
Hi:Lo Ratio	N	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias
2:3	20	87.1	74.3	-0.30	86.3	92.8	-0.12
	25	87.1	60.5	0.06	87.0	78.5	0.31
	30	90.1	50.8	-0.11	89.3	69.5	-0.19
	35	88.6	43.0	-0.09	91.4	62.9	0.00
	40	88.4	35.4	0.39	92.5	57.0	-0.18
	45	89.9	29.1	-0.12	92.5	52.3	0.18
1:1	20	88.7	61.0	-0.30	86.5	77.5	0.48
	24	88.3	50.4	0.39	89.5	70.0	0.02
	28	89.7	42.7	-0.06	90.8	63.1	-0.42
	32	90.5	35.7	-0.15	90.5	56.8	0.27
	36	90.3	29.2	-0.10	92.9	53.0	-0.46
	40	89.8	22.5	0.22	94.2	48.7	0.39
	44	90.7	15.3	0.08	92.9	45.7	-0.07
3:2	20	87.8	51.8	-0.33	89.2	70.1	0.05
	25	88.8	39.1	-0.07	93.4	60.9	-0.30
	30	90.0	29.5	-0.19	94.3	53.5	-0.42
	35	91.0	19.4	0.04	94.1	47.7	-0.24
2:1	21	89.3	43.6	-0.22	91.5	63.9	-0.02
	24	90.3	36.6	-0.50	92.7	58.1	0.06
	27	89.6	29.6	0.09	93.5	53.6	-0.03
	30	90.1	23.0	0.12	93.3	49.6	-0.01
	33	93.2	15.9	-0.15	93.6	46.4	0.02
3:1	20	90.4	39.4	0.18	91.1	61.3	0.40
	24	89.9	30.0	-0.02	94.6	54.8	-0.55
	28	91.5	20.2	-0.19	94.2	49.1	-0.36
Mean		89.6	37.1	-0.05	91.7	59.9	-0.05

Table 16. Stratification ratio, sample size, percent confidence interval (CI) coverage, CI width, and percent bias from 1,000 simulated stratified samples with a complete coverage block of 4 survey units using simple random sampling (SRS) and sampling proportional to size (PPS) estimators for bison in Yellowstone National Park, Wyoming, August, 1999.

		SRS			Separate Ratio		
Hi:Lo Ratio	N	% CI Cov.	% CI Width	% Bias	% CI Cov.	% CI Width	% Bias
1:2	30	68.4	10.0	0.14	68.4	10.3	0.08
	36	83.7	6.9	0.08	83.7	7.1	0.07
2:3	25	72.0	10.5	-0.04	72.0	10.8	-0.11
	30	85.7	7.0	0.00	85.7	7.2	-0.05
1:1	20	71.5	10.5	-0.02	71.5	10.8	-0.10
	24	85.8	7.1	0.00	85.8	7.3	-0.07
3:2	15	64.8	12.3	-0.04	65.0	12.8	-0.14
	20	84.1	7.0	0.06	84.1	7.3	0.01
2:1	15	69.5	10.4	0.12	69.6	10.8	-0.03
	18	86.3	7.2	-0.03	86.4	7.5	-0.09
3:1	16	87.6	7.4	-0.02	88.6	7.7	-0.13
4:1	15	88.2	7.6	-0.04	87.8	8.0	-0.15
Mean		79.0	8.7	0.02	79.1	9.0	-0.06

different allocation ratios (Table 17). When strata were well specified, CI width was essentially equivalent across allocation ratios, and higher allocation ratios provided much reduction in the survey area required for equivalent precision. When strata were not well specified, as in the summer of 1998, even a moderate level of precision could not be achieved. Poor specification resulted from an extreme density value that was unlike other survey units within the stratum and caused increased variance estimates. In practice, poorly specified strata can be avoided by using reconnaissance immediately before a sample survey to modify strata assignments.

Overall, stratified SRS with 3:1 allocation and a complete coverage block worked well in winter, providing much increase in precision for comparable sample sizes, although CI coverage was less than nominal for 2000 (Fig. 20). Stratified PPS with a complete coverage block had lower CI coverage in both winters than SRS, was never as precise, but had similar bias properties. In summer, extreme density values can make estimation difficult. When survey units contain large numbers of bison and are not included in samples, underestimates will occur. If however, such units are included in samples, overestimates will occur. Stratified SRS and separate ratio estimators with a complete coverage block performed better than other sampling plan/estimator combinations with comparable sample sizes during summer simulations, but CI coverage never reached nominal levels (Fig. 21). Estimates remained stable with allocation ratios as high as 4:1, which offered some reduction in effort. The

Table 17. Season, stratification ratio, CI width as the percent of density, sample area, and estimated survey time required from 1,000 simulated stratified simple random (SRS) samples with a complete coverage block of 8 survey units in winter and 4 survey units in summer for bison in Yellowstone National Park, Wyoming, 1998-2000.

Season	Hi:Lo Ratio	% CI Width 1999	% CI Width 2000	Area km^2	Time Hours
Winter	2:3	24.1	29.1	1800	12.6
	1:1	12.4	15.5	1750	12.2
	3:2	15.6	19.4	1460	10.2
	2:1	12.4	15.9	1390	9.7
	3:1	15.6	20.2	1230	8.6
		1998	1999		
Summer	1:2	41.3	6.9	1427	8.5
	2:3	40.8	7.0	1243	7.4
	1:1	40.6	7.1	1065	6.4
	3:2	41.2	7.0	951	5.7
	2:1	41.1	7.2	893	5.4
	3:1	40.1	7.4	825	5.1
	4:1	41.2	7.6	801	4.8

conditions for precise PPS estimation were not met in these simulations. Because these estimators require fundamentally different sample selection procedures, I recommend PPS selection not be employed in sample surveys of YNP bison.

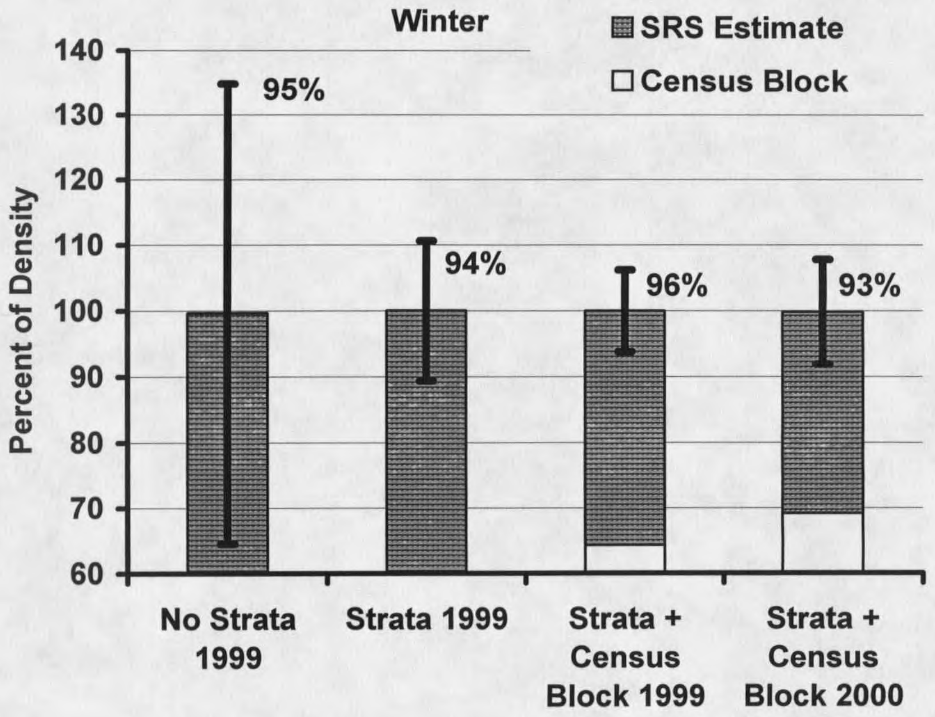

Fig. 20. Estimates as percent of true density (cross-hatched bars), confidence interval (CI) width (error bars), CI coverage (numerals), and percent of estimate from census blocks (open bars) for 1,000 simulations of simple random sampling (SRS) for bison in Yellowstone National Park during summer. Sample sizes are n = 40 survey units with no strata and with stratified samples for 1998, and n = 38 with stratified samples + 8 census blocks for 1998 and 1999. See Appendix C for details on area and number of bison in strata.

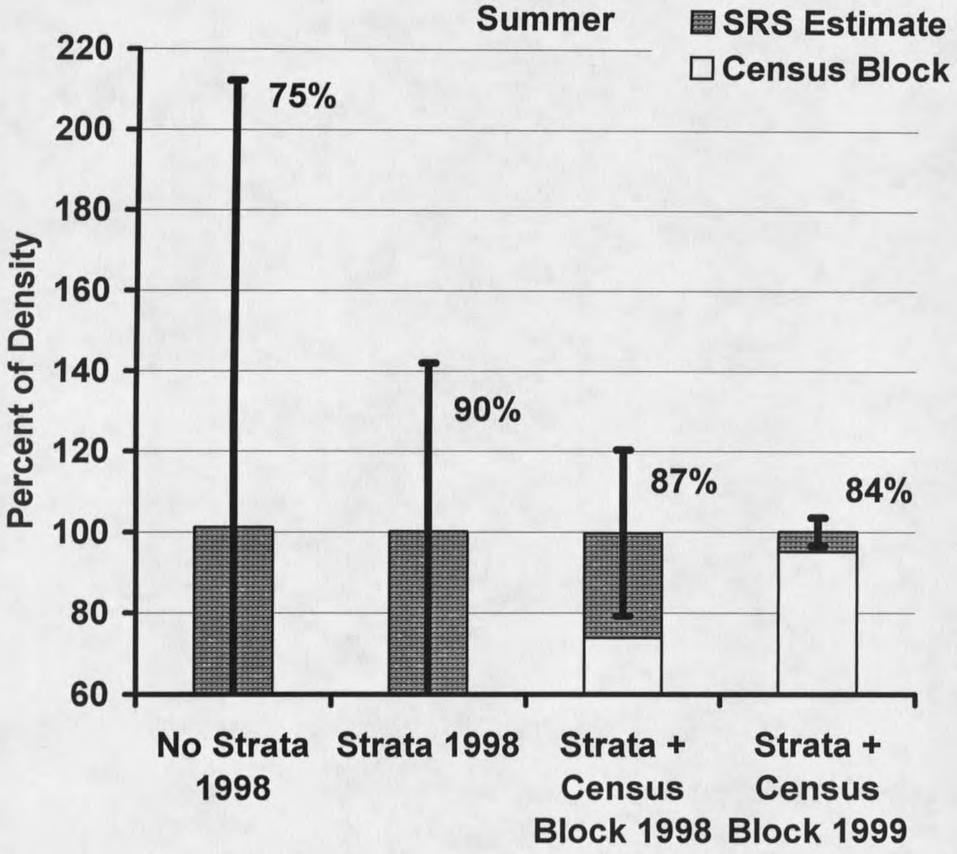

Fig. 21. Estimates as percent of true density (cross-hatched bars), CI coverage (numerals), and percent of estimate from census blocks for 1,000 simulations of simple random sampling (SRS) for bison in Yellowstone National Park during summer. Sample sizes are n = 25 survey units with no strata for 1998, n = 24 with stratified sample for 1998, and n = 20 with stratified samples + 4 census blocks for 1998 and 1999. See Appendix C for details on area and number of bison in strata.

Double Sampling

I classified 3,116 bison in 298 groups from 3 different ranges (MGF, Hayden Valley, Northern Range) in YNP according to which groups were seen from the ground only, from aircraft only, or seen from both ground and aircraft (Table 18). The Lincoln-Petersen maximum likelihood estimates of group

Table 18. Date, location (Firehole; FH, Hayden Valley; HV, Northern Range; NR), and number of bison groups detected by aircraft, but not from ground (S_1), from ground, but not from aircraft (S_2), and by both methods (B) during simultaneous surveys in Yellowstone National Park, Wyoming, 1998-2000.

Date	Range	Number of Bison Detected		
		S_1	S_2	B
22 Apr 98	FH	3	8	13
25 May 98	FH	0	1	8
13 Jun 98	HV	5	6	20
3 Jul 98	FH	3	1	8
3 Jul 98	HV	10	3	28
27 Jul 98	HV	8	7	20
31 Jul 98	HV	5	7	11
2 Aug 98	HV	1	1	6
6 Aug 98	HV	4	3	13
10 Dec 98	NR	2	1	10
30 Jan 99	FH	3	2	8
13 Mar 99	FH	5	4	18
20 Mar 99	FH	3	8	15
21 Mar 99	FH	0	2	6
30 Jan 00	FH	1	2	5
Total		53	56	189

detection probability from aircraft ($\hat{p}_1$ = 0.77; 95% CI 0.67–0.80; n = 245) and ground detection probability ($\hat{p}_2$ = 0.78; 95% CI 0.67–0.81; n = 242) were not different ($P > 0.8$; test for two proportions), and the combined overall detection probability for double surveys was high ($\hat{p}_o$ = 0.95; 95% CI 0.86–0.96; n = 298). No groups of $\geq$ 11 bison were undetected from aircraft and no groups of $\geq$ 7 bison were undetected from ground. Detection probabilities also did not differ between aircraft and ground within seasons. The proportion of all groups detected in winter from aircraft ($\hat{p}_1$ = 0.748; 95% CI 0.655–0.773; n = 109) was not different from ground ($P > 0.14$; $\hat{p}_2$ = 0.838; 95% CI 0.727–0.859; n = 100). In summer, the proportion of all groups detected from aircraft ($\hat{p}_1$ = 0.791; 95% CI 0.671–0.825; n = 142) was also not different from ground ($P > 0.37$; $\hat{p}_2$ = 0.747; 95% CI 0.633–0.779; n =134). The proportions of groups detected by both methods were very similar between winter ($\hat{p}_o$ = 0.959; 95% CI 0.906–0.968; n = 128) and summer ($P > 0.56$; $\hat{p}_o$ = 0.947; 95% CI 0.879–0.961; n = 170).

The 8 logistic regression models for both ground and aerial surveys were closely ranked (Table 19); all were within 3.4 AIC units of one another and 9.2 AIC units for ground models. Hosmer-Lemeshow tests failed to reject goodness-of-fit for general models (p_{gshr}) of either data set (χ^2_6 = 6.3; $P > 0.388$ for aircraft, χ^2_6 = 5.3; $P > 0.512$ for ground). Group size was a significant factor for both ground (χ^2_1 = 10.9, $P < 0.001$) and aerial surveys (χ^2_1 = 7.2, $P < 0.008$), but showed different forms for each survey method (Fig. 22). The significance of

Table 19. Logistic regression detection probability models ranked by AIC for aerial and ground surveys of bison in Yellowstone National Park, Wyoming, 1998 - 2000. Factors affecting detection probability included group size (p_g) in all possible combinations with seasons (p_s; 2 levels), habitats (p_h; 3 levels, p_{h2}; 2 levels), and range (p_r; 3 levels). The number of model parameters (K), Delta (Δ) AIC, and jackknife estimates of reliability are reported for each model.

Aerial Surveys				Ground Surveys			
Model	K	Δ AIC	Reliability	Model	K	Δ AIC	Reliability
*p_{gh2}	3	-2.0	70.0%	--	--	---	---
p_{gh}	4	0.0	69.8%	p_g	2	0.0	75.4%
p_g	2	1.6	73.7%	p_{gs}	3	1.6	75.4%
p_{gsh}	5	1.7	66.4%	p_{gr}	4	3.5	74.6%
p_{gs}	3	2.4	73.8%	p_{gh}	4	4.0	71.8%
p_{gsr}	5	2.7	72.2%	p_{gsr}	5	5.4	74.6%
p_{gshr}	7	2.8	65.4%	p_{gsh}	5	5.5	71.7%
p_{ghr}	6	3.4	68.4%	p_{ghr}	6	7.2	71.6%
p_{gr}	4	3.4	72.4%	p_{gshr}	7	9.2	71.3%

* *A posteriori* model

these predictors demonstrates strong effects of group size bias on detection probability (Cook and Martin 1974, Samuel and Pollock 1981). While the highest ranked *a priori* model for ground surveys contained only group size as a predictor, the highest ranked aerial survey model also included habitat as a predictor. The overall effect of habitat on detection probability was significant

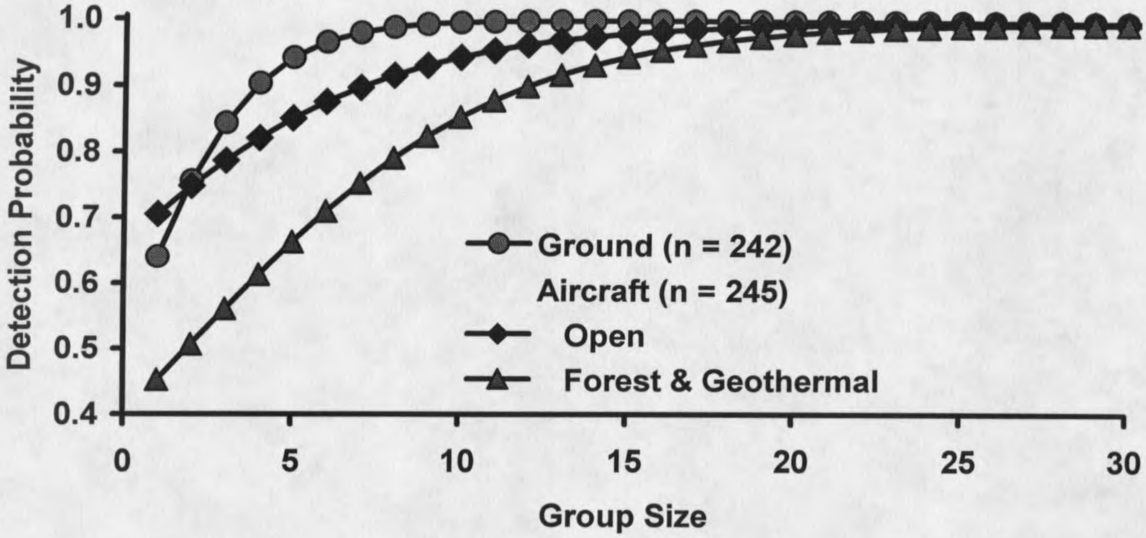

Fig. 22. Logistic regression estimates for detection probability of bison groups during ground and aerial surveys in Yellowstone National Park, 1998-2000. Detection probability in forest and geothermal habitats was lower than in open habitats during aerial surveys. See Appendix D for details on coefficients and variance/covariance estimates for aerial survey model.

($\chi^2_2 = 6.1$, $P < 0.048$), although the coefficients were only marginally significant for forested ($P < 0.055$) and geothermal ($P < 0.114$) areas in reference to all open habitats. Both coefficients were negative, indicating lower detection probabilities from aircraft in these habitats compared to open habitats. The coefficient for detecting bison in forested habitats was -1.011 (Table 20; 95% CI 2.040–0.020), while the coefficient for detecting bison in geothermal habitats was even lower, but had a wider confidence interval and overlap with zero (-1.167; 95% CI -2.670–0.285). Although Δ AIC for 4 aerial survey models was < 3.4, the second ranked model (p_g) was a special reduced case of the highest ranked model (p_{gh}), while the season effect for the third and fourth ranked models did not approach significance (p_{gsh} $\chi^2_1 = 0.3$, $P > 0.59$; p_{gs} $\chi^2_1 = 1.1$, $P > 0.28$). Given

Table 20. Logistic regression parameters, approximate *P*-values, estimated coefficients, standard errors (SE), and confidence intervals (CI) for highest ranked models of bison detection probability during aerial surveys in Yellowstone National Park, Wyoming, 1998 – 2000. Factors affecting detection probability include group size (p_g), seasons (p_s; winter in comparison to summer), and habitats (p_h; forest and geothermal separately compared to open habitats, p_{h2}; forest and geothermal combined in comparison to open habitats).

Model	Parameters	*P*-Value	Estimated Coefficients	SE	Lower 95% CI	Upper 95% CI
*p_{gh2}	Intercept	0.007	0.654	0.241	---	---
	Group Size	0.008	0.217	0.082	0.058	0.378
	Forest & Geothermal	0.017	-1.061	0.445	-1.966	-0.186
p_{gh}	Intercept	0.007	0.654	0.242	---	---
	Group Size	0.008	0.218	0.082	0.058	0.378
	Forest	0.055	-1.011	0.527	-2.040	0.020
	Geothermal	0.114	-1.167	0.739	-2.670	0.285
p_g	Intercept	0.028	0.503	0.229	---	---
	Group Size	0.008	0.217	0.081	0.058	0.378
p_{gsh}	Intercept	0.008	0.713	0.269	---	---
	Group Size	0.008	0.225	0.085	0.058	0.392
	Forest	0.072	-0.962	0.535	-2.040	0.086
	Geothermal	0.149	-1.088	0.754	-2.256	0.392
	Winter	0.599	-0.178	0.334	-0.844	0.489

* *A posteriori* model

that the coefficients for detection probability in forest and geothermal habitats were very similar to each other relative to open habitats, I hypothesized an *a posteriori* model (p_{gh2}) where forest and geothermal habitats were combined in comparison to open habitats. This model was ranked higher than the 3 habitat model, having one less estimable parameter (AIC $p_{gh2} - p_{gh} = -2.0$; Table 19), and no loss in reliability. The coefficient for forest and geothermal habitats was significant ($P < 0.017$), similar to the coefficient for forest in the 3 habitat model, and confidence intervals did not overlap with zero (-1.061; 95% CI -1.966–-0.186), hence I chose this as the best approximating model to correct aerial surveys.

The second highest ranked model for ground detection probability contained season as a predictor, and, although the coefficient for winter relative to summer was not near significance ($\chi^2_1 = 0.4$, $P > 0.52$), model reliability was the same (Table 19). This effect, however, may be confounded by significantly smaller average group sizes in winter (mean = 6.7; 2-sample bootstrap test; Efron and Tibshirani 1993; $P < 0.027$) than in summer (mean = 13.3) within these data. The overall effect of habitat on detection probability was not significant ($\chi^2_2 = 0.02$, $P > 0.99$), indicating that habitat had little effect on detection probability during ground surveys. For both methods, range did not appear among the top ranked models, but Δ AIC for these models was < 3.5. Range was, however, confounded with season, as these 2 variables were highly associated with each other ($\chi^2_2 = 203.2$, $P < 0.001$).

The overall group detection probability for both survey methods combined was high, ranging from 0.89 for solitary bison to unity for groups sizes $\geq$ 27 (Fig. 22). Using logistic regression coefficients from model p_{gh2} for aircraft surveys and model p_g for ground surveys, I calculated the proportion of individual bison detected from the observed groups size distribution and habitat associations (for aircraft surveys) in the double-survey data. The logistic regression models predicted 3,147 bison were present during the aircraft surveys in which 3,010 were observed, and 3,146 bison were present during the ground surveys in which 3,043 bison were observed. The proportions of individuals detected then were 0.956 (95% CI 0.907–0.974) from aircraft and 0.967 (95% CI 0.944–0.979) from ground, resulting in a combined proportion of 0.999 (95% CI 0.994–0.999) for both survey methods combined. When these predictions were restricted to the smaller group sizes observed in winter, however, detection probabilities were 0.919 (95% CI 0.819–0.955) from aircraft and 0.940 (95% CI 0.895–0.961) for ground surveys. In summer, detection probabilities were 0.971 (95% CI 0.943–0.982) from aircraft and 0.975 (95% CI 0.959–0.983) for ground surveys. The overall CV from the logistic regression model applied to the number of adult bison detected during surveys was 3.4%. Point estimates with 95% confidence intervals of adults detected during winter surveys had a mean CV of 5.4%, while summer surveys had a mean CV of 1.3% (Table 21).

Table 21. Number of bison detected, point estimates, 95% confidence intervals (CI), and coefficient of variation (CV) corrected by logistic regression model during surveys in Yellowstone National Park in winters of 1998-2000.

Date	Number Detected	Point Estimate	95% CI	CV (%)
February 28, 1998	1,899	1,963	(1,934, 2,114)	2.3
April 22-23, 1998	1,814	1,909	(1,867, 2,078)	2.8
July 3-4, 1998	1,598	1,670	(1,642, 1,692)	1.6
August 1-4, 1998	1,607	1,641	(1,627, 1,692)	1.0
August 5-6, 1998	1,672	1,717	(1,699, 1,794)	1.2
January 30- February 1, 1999	2,091	2,258	(2,180, 2,533)	4.0
February 15- 11 March, 1999	1,862	2,212	(2,022, 2,821)	9.2
March 14-19, 1999	1,984	2,295	(2,117, 2,929)	9.0
July 24-26, 1999	1,823	1,894	(1,865, 1,963)	1.2
July 28, 1999	1,746	1,800	(1,778, 1,869)	1.3
August 13-14, 1999	1,947	1,990	(1,972, 2,063)	1.3
January 29-31, 2000	2,536	2,723	(2,631, 3,068)	4.1
February 18-19, 2000	2,427	2,656	(2,533, 3,138)	5.8
March 4-5, 2000	2,002	2,204	(2,101, 2,584)	5.6
July 30- August 4, 2000	2,474	2,530	(2,504, 2,630)	1.3

Photography

I made 453 images of 96 different bison groups to determine if correction for counting bias was necessary. There was a close one-to-one correspondence between visual counts from aircraft and photographs of groups with less than 100 bison (Fig. 23). However, for groups greater than 100, there was an increasing tendency to undercount, described by a marginally significant interaction term in the regression ($P < 0.096$), such that slopes differed between groups $\leq$ 100 and > 100 ($y = 2.06 + 0.98x + [5.42 + 0.15x \text{ (for } x > 100)]$; $P < 0.001$; $r^2 = 0.96$). As a consequence, all groups greater than 100 should be photographed and counted afterwards to provide a better enumeration of large bison groups, while groups of less than 100 should not need correction. Because groups greater than 100 bison were rarely observed in winter, this correction will only be necessary for summer surveys. This regression coefficient was specific to a single observer, and will have to be established for each different observer.

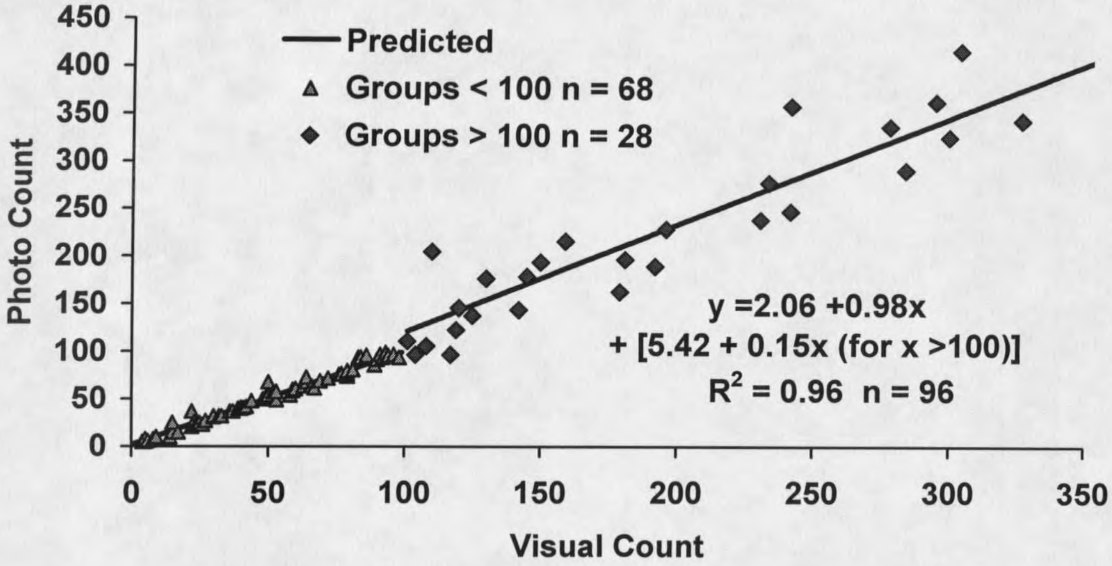

Fig. 23. Comparison of photographic images and visual counts of bison from aircraft during aerial surveys in Yellowstone National Park, 1998-2000. Regression shows a 1-to-1 correspondence for groups <100, but increasing tendency to undercount groups > 100.

Replicate total counts

I combined all total winter counts conducted by NPS biologists (J. Mack and W. Clark, unpubl. data) with my total winter counts (total n = 19) for each of 3 winters (1998–2000). I calculated bootstrap estimates of the mean $\bar{x}$ and variance of counts within seasons (Efron and Tibshirani 1993). I then determined bias-corrected population estimates $\hat{N}$ using a value of $\hat{p}$ from 3 combined winters of simultaneous ground and aerial surveys. The value of $\hat{p}$ I used was the number of bison observed during winter double surveys divided by the total number predicted by logistic regression ($\hat{p}$ = 803/874 = 0.919). I used the delta method to estimate the overall variance of bias corrected population estimates and then calculated normal confidence intervals using *t*-values corresponding to *n* replicate total counts.

Overall population estimates (Table 22) have a combined coefficient of variation (CV) of 8.30%. Estimates were also further broken into the 2 major bison herds that have limited interchange, the Northern Range herd (NR), and the Central herd (CH). The overall CV for NR was 11.7% while CV of the CH was 8.9%, the average CV of the 2 herds combined was 10.3% (Table 23).

Table 22. Overall population estimates, 95% confidence intervals (CI), and coefficient of variation (CV) from replicate total counts of bison in Yellowstone National Park in winters of 1998-2000.

Year	N	Point Estimate	95% CI	CV (%)
1998	8	2090	(1,986, 2,194)	7.2
1999	6	2151	(1,993, 2,310)	9.2
2000	5	2583	(2,391, 2,776)	8.5

Table 23. Population estimates, 95% confidence intervals (CI), and coefficient of variation (CV) from replicate total counts of bison in the Northern Range and central herd of Yellowstone National Park in winters of 1998-2000.

	Northern Range				Central Herd			
Year	N	Point Est.	95% CI	CV (%)	N	Point Est.	95% CI	CV (%)
1998	8	412	(383, 441)	8.7	8	1,678	(1,566, 1,791)	8.2
1999	6	434	(394, 474)	9.3	6	1,718	(1,542, 1,894)	10.3
2000	5	416	(334, 499)	17.2	5	2,167	(1,961, 2,374)	8.3

Detection of Statistical Trends

Estimated statistical power was high for detecting trends in time series data with an 8.0% CV, reaching 0.75 for 0.152 rate of change with $\alpha = 0.10$, and a 0.175 rate of change with $\alpha = 0.05$ for 2-sided tests (Fig. 24). Because these tests estimate the power for 2-sided tests where alternative hypotheses are either an increasing or decreasing population trend, smaller rates of change may be detected, if only 1-sided hypotheses are of interest. For example, if an investigator suspects that the population has decreased because of a management action, and formulates an appropriate 1-sided *a priori* hypothesis, the rates of change that can be detected with power 0.75 should be 0.073 with α = 0.10, and 0.102 with α = 0.05.

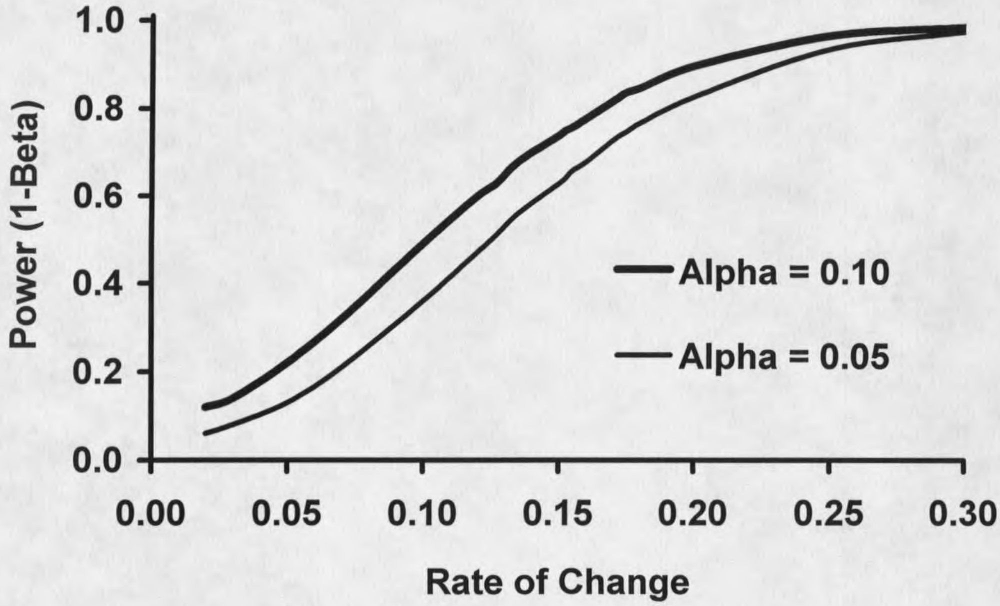

Fig. 24. Estimated statistical power to detect trends in time series of 3 population estimates with 2-sided hypotheses and CV = 8.0% for bison in Yellowstone National Park, Wyoming. Power was calculated with program TRENDS (Gerrodette 1993).

<u>Validation of an Index</u>

Winter surveys conducted by NPS biologists were highly correlated with my corrected estimates of abundance. There were no significant effects of season ($P > 0.41$) or days between surveys ($P > 0.79$), however, index surveys were a strong predictor of population estimates (Fig. 25; $P < 0.003$) when the other effects were dropped from the model. A single influential point, however, had a strong negative effect on the strength of the regression. When this point

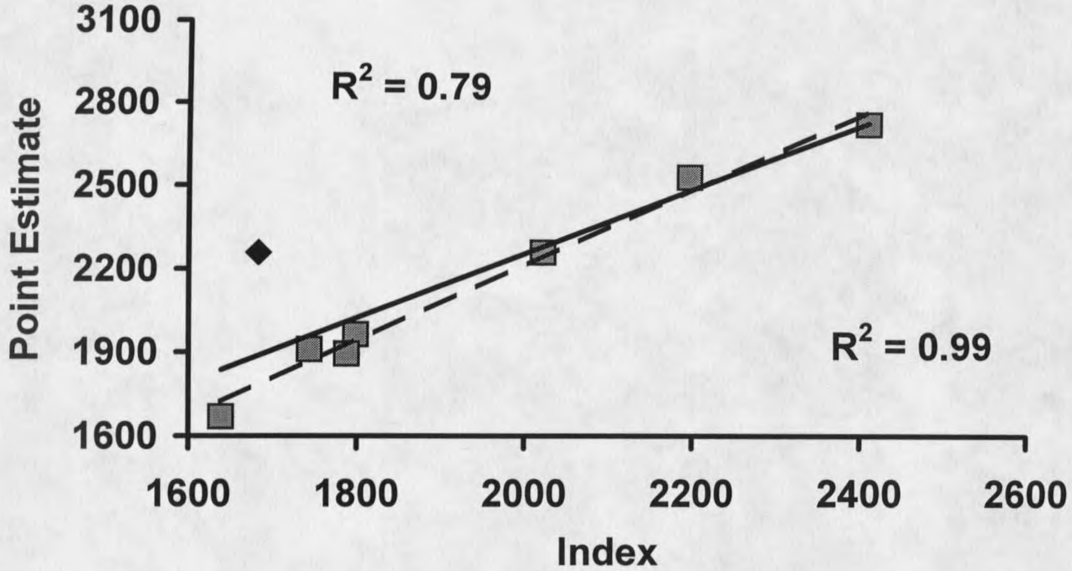

Fig. 25. An index of bison population from aerial surveys conducted by Yellowstone National Park biologists (Unpubl. data) validated with population estimates from aerial surveys conducted by Montana State University, 1998-2000. Linear regression of the full data set shows a strong relationship between methods. Deletion of a single influential point results in a much stronger relationship, but also results in an extreme negative intercept.

was excluded from the regression, the r^2 value changed from 0.75 to 0.99 and the significance of index surveys as a predictor increased ($P > 0.001$). The model for the reduced data set, however, predicted an unrealistic intercept of -504, and should have been closer to zero, (i.e., the true population should be close to zero when index counts are close to zero). In comparison, the full data set had an intercept of -54, and its 95% CI included zero (-1204.5, 1096.3), whereas the intercept for the reduced data set did not (-856.3, -158.9). Therefore, the regression equation from the full data set may be used as a crude correction method for the range of index values observed during this study

(1,639–2,410), even though it contains considerably more uncertainty than the model for the reduced data set. The regression equation is:

$$\hat{N} = -54.1 + (1.154x_i)$$

where x_i is the observed total count (index) value. During replicated surveys, the observer may recognize such a survey as a 'bad count' and choose to disregard it. If surveys are not replicated, however, there would be no way to recognize bad surveys.

Demographic Trends

Using simple linear regression to estimate annual increment for 2 time steps between 1997–98 and 1999–2000, growth models had slope coefficients of 422 (95% CI 372, 471) for winter ($P < 0.0001$; $r^2 = 0.99$; $n = 7$) and 431 (95% CI 285, 592) for summer ($P < .001$; $r^2 = 0.90$; $n = 7$). When I log-transformed population estimates and regressed these values on year, the coefficients were 0.188 (95% CI 0.165, 0.212) for winter ($P < .0001$; $r^2 = 0.99$; $n = 7$) and 0.220 (95% CI 0.158, 0.288) for summer ($P < .001$; $r^2 = 0.93$; $n = 7$). These slope values are indicative of the annual percent increase, although they may be overestimated due to the nature of the method (Eberhardt 1987).

The difference between mean winter and summer population estimates within each of the 3 years were 260.0 in 1997–98, 360.3 in 1998–99, and 159.5 in 1999–2000. SWE_{acc} values from West Yellowstone were 1714 cm in 1997–98, 2950 cm in 1998–99, and 1468 cm in 1999–2000. From these limited data, regression shows a linear trend ($P < 0.235$; $r^2 = 0.74$), suggesting that overall annual mortality of YNP bison may have been related to snow accumulation over the entire winter season during the years of this study. I lacked an adequate series to test hypotheses of non-linear effects, however, I was able to approximate this by log-transformation of the SWE_{acc} values. Although this model had a slightly better fit ($P < 0.198$; $r^2 = 0.81$), it should not be regarded as conclusive. Linear and non-linear effects of snow accumulation on mortality should both be considered as hypotheses for future testing with a longer time series.

DISCUSSION

Aerial surveys are valuable tools for obtaining population estimates and demographic parameters of wildlife when survey methods are specifically designed for the target species within a particular region of interest. There are clearly a number of limitations that the species, physical geography, or local environment can have on a survey methodology, and one that is suited to a species in one region may not be well suited to the same species in another region. There are also inherent limitations to research and monitoring programs in any National Park, further restricting the types of survey methodology that may be used. Yellowstone National Park presents some unique problems when designing an effective aerial survey methodology for bison. In addition to biological and geographical considerations, survey methodology must respect the fact that YNP has millions of visitors every year, many coming primarily to view wildlife. Visitors deserve a safe experience first and foremost, but also one with minimal low-level flight disturbance, and without obvious markers on wildlife.

Given these practical constraints, there are 3 potential sources of statistical error that must be addressed in the design of aerial survey methodology for Yellowstone bison; the first is the geographic scope of surveys. YNP is a vast region of mountains, plateaus, canyons, and valleys over 9,000 km^2 in area. Since it is practically difficult to survey all areas of YNP that may contain bison, I identified regions that contain the majority of YNP bison during both winter and summer seasons and a sample survey design to eliminate bias due to incomplete spatial coverage. Because bison are highly migratory, I also identified how the area bison occupy changes within seasons, and optimal temporal windows to conduct surveys. Secondly, not all groups within survey units will be detected. YNP has a heterogeneous mix of grasslands and a mosaic of forest habitats ranging from mature to early seral stages, but the most unique aspect of YNP, and the primary reason for establishing the first national park of its kind, is a large number of geothermal features. Because snow does not accumulate near geothermal features in winter, these areas provide year-round access to forage, influencing the distribution of bison, but also making for poorly contrasting background conditions for detection from aircraft. I developed a detection probability model that may be used to correct for the proportion of bison obscured by forest or poor background conditions. Thirdly, not all bison within groups will be counted accurately. Bison are highly gregarious, occurring in groups of up to 500 individuals. Photography may be effectively employed to account for this bias that arises as a result of enumerating large groups.

I found these factors to be important considerations for designing aerial survey methodology for bison in YNP because they all affect safety, biases due to detection and enumeration, the spatial scale of surveys, which varies due to seasonal movements, and consequently also affects the timing of surveys. The least biased estimate of the number of bison in YNP will simultaneously account for all of these sources of error, however, managers may choose not to correct for all sources of error because they each contribute additional uncertainty to estimates. The minimum design requirement for obtaining a statistically valid population estimate from a single survey should at least account for the amount of area searched based on a spatially explicit survey framework.

An explicit survey framework has never before been defined for bison in YNP, however, it is important for several reasons: first, any survey methodology must be able to account for the total area that was searched. The amount of time spent searching this area also provides an essential measure of effort. A framework of sampling units is also a basic necessity for any sampling protocol. Finally, standardization of the spatial extent is essential to make any comparisons between surveys. I found that the total area that bison inhabit and must be considered in the survey framework is less than 25% of the entire area of YNP, and that a very high proportion of all bison occur within a small amount of this area during periods of time in both summer and winter. I used long-term survey data to determine the reliability of my framework and found that only a small number of bison, on the order of 1–2% of the population, regularly occurred outside of this framework. This pattern of restricted distribution held even when bison were at their highest population level during the past century (>4,000), in 1994. Bison are not likely to occupy more area than this because target population levels and management protocols have been identified to limit numbers at <3,000 (National Park Service 2000). These management actions will occur near the northern and western boundaries of YNP, effectively preventing bison from annexing adjacent areas outside of their currently established range. The survey framework I developed should therefore be applicable for the indefinite future.

Bison exhibit strong tendencies to repeatedly use the same areas of YNP year after year, however, primary areas of occupation change dramatically between seasons. I also used long-term survey data to test predictability of the seasonal distribution. In winter, bison used Hayden Valley in high numbers when population levels were <3,000. When population levels grew to levels greater than this, however, bison began using the Lower Geyser Basin of the Firehole drainage in higher numbers, and occupied much more area on the western boundary of YNP. These areas ranked consistently high in use, but did not always contain the highest number of bison. This pattern may be explained by movements within the winter season from the higher valleys of Pelican and Hayden to the lower Firehole drainage. A high proportion of the Central Herd followed this corridor between Hayden Valley in early winter and the Firehole drainage area, where they aggregated in relatively high-density later in winter (Bjornlie and Garrott 2001). This resulted in shifting areas of abundance, causing survey units with the highest number of bison to change over the season and increase the average rank values of several units through which these bison traveled.

In summer, the greatest number of bison to be found in any area of YNP occurred in Hayden Valley, an area that is traditionally used for breeding aggregation. This area was ranked with the highest number of bison every time it was surveyed for 27 years, and also throughout this study. Other areas with historically high numbers of bison in summer also include the Lamar Valley and adjacent slopes of the Absoroka Range, Pelican Valley, and the Lower Geyser

Basin of the Firehole River drainage. Bison may be extremely concentrated during a few weeks at the end of July and in early August. During one survey in the summer of 1999, only four survey units were occupied by bison, concentrating heavily in the Hayden Valley, and high plateaus between the Northern Range and Pelican Valley. This predictable distribution of bison in summer, while extremely patchy, provided a basis for establishing the strata that I used for this research and also serves as a basis for aerial survey protocols.

From these rankings, and from my own aerial survey data, I found a set of 8 centrally located survey units that predictably contained large numbers of bison in winter. These units include 2 in the Hellroaring and Junction Butte area of the Northern Range, and in the central part of YNP, the largest unit in the Pelican Valley, the 2 largest units in Hayden Valley, 2 units in the Firehole drainage and 1 unit in the Nez Perce Creek area between Hayden Valley and the Firehole Drainage. In summer, I also found a set of 8 units that predictably contained large numbers of bison. These units are the 3 largest units in Hayden Valley, 2 units on Mirror Plateau near Mirror Lake, and in the Lamar Valley, Soda Butte Creek, the ridge between Cache and Calfee Creeks, and the Junction Butte area. These units have predictably high numbers of bison that account for more than 3/4 of the entire population and create problems for statistical estimation because their values lie in the extreme tail of the sampling distribution. Conducting censuses of these areas will reduce the overall variation in population estimates by eliminating a major component of variation due to sampling. The remaining units should comprise high and low-density strata, both to be sampled. To effectively reduce the overall variation due to sampling, it will be important to time surveys so that a large proportion of the population occur in the complete coverage blocks. A thorough understanding of temporal patterns of distribution is therefore important not only for an effective sample survey design, but also for higher effort surveys that census a large proportion of traditionally occupied areas, and to provide the best conditions for detecting bison.

The practical benefits of identifying optimal temporal windows for conducting surveys are a reduction of the total area required for high effort surveys or sample surveys and a simultaneous reduction in the variability between surveys. When I examined temporal patterns in the total area occupied by bison, I found that in winter, although they were never as aggregated as in summer, bison were both spatially concentrated and aggregated in December and January. As winter progressed, however, bison became more dispersed across the landscape in smaller groups. The effect of this change in distribution over the course of a winter is that it increases the total amount of area that must be considered in surveys and it decreases the probability of detection because maximum group size also decreases. I also found that in late February, the proportion of bison occupying geothermal habitats peaked. Snow does not accumulate in these areas, and therefore, they had unfavorable background color conditions for detecting bison. Throughout winter, a lower proportion of bison occupied forested habitats on days with colder average temperature than on warmer days. Surveys that are planned for early in the winter season should provide conditions where bison are spatially aggregated, with a low proportion in geothermal habitats, and predictably colder temperatures in December and January where a higher proportion of bison occupy non-forested areas. Later in winter, however, bison may be dispersed over a much larger total area in smaller groups, and a high proportion of bison may occupy geothermal habitats in late February, and forested habitats on warmer winter days. In a year when the onset of winter occurred earlier, and was more severe, as in 1999, maximum group size was lower and decreased earlier than during years with milder winter conditions. It is therefore important to plan surveys as early in the winter season as possible during years with more severe winter conditions.

Early winter conditions offer a combination of advantages over later winter surveys, but never approach the advantageous conditions found in summer. The total area occupied by bison reached an annual low point in early August, and group size also simultaneously reached an annual peak with many groups numbering in the hundreds. These large groups represented a high proportion of the entire population and were mostly located in the open valleys of Hayden and Lamar, offering the best conditions possible for surveys. By taking advantage of these conditions, it is possible to reduce the total amount of effort put into surveys, benefiting from large aggregations that facilitate detection and a relatively small area that must be surveyed. Using this information, standardizing the temporal windows for both high effort surveys and for sample surveys should provide a substantial reduction in the variability between surveys and in overall effort.

Sample survey designs differ fundamentally in how survey units are chosen to be included in a sample and this affects the overall effort that must be put into a survey. In any sample survey, the first stage of sampling determines where the survey will be conducted and how effort will be distributed among those units or units within different strata. I simulated this first stage of SRS and PPS sampling plans first without stratification, then with stratification, and finally with a complete coverage block in addition to stratification. I estimated the nominal CI coverage, precision, bias, and the effort needed to provide statistically reliable estimates with the actual distribution of bison in Yellowstone in both winter and summer. Mountainous topography dictates a sampling scheme based on survey units of unequal size bounded by ridges and other observable features, however, bison are highly aggregated within these areas, especially in summer, making precise statistical estimation difficult. If bison were randomly distributed within survey units, then the number of bison in a survey unit should be related to the size of the unit. In this case, PPS should be the most reliable selection method for the first stage of sampling because larger units should have a higher probability of being included in the sample, and contain a large number of bison within them. If the number of bison is not randomly distributed within survey units, then SRS may be a more reliable selection method. Because bison exhibit this aggregated spatial distribution, sampling plans without stratification did not provide high precision with any reasonable sample sizes, but allowed an evaluation of some properties of estimators. In the absence of stratification, PPS and its associated estimator were not substantially better in any measure than SRS with the SRS estimator or the ratio estimators in winter or summer.

Stratification may substantially reduce the amount of flight time necessary to survey a population by concentrating effort in strata that have high mean density and associated variance. Stratification improved the efficiency of SRS more than PPS in both winter and summer, but overall precision was still relatively low due to some survey units that contained extremely high numbers of bison. In winter, Stratified PPS provided estimates with little bias relative to ratio estimators, but precision was still poor and offered no real advantages over SRS. For winter simulations, strata were well specified, and reliable estimates could be obtained with effort allocated as high as 3:1 high:low-density. Stratification had a dramatic effect on efficiency and provided estimates that were equally precise, but required surveying only 33% in winter, and 52% in summer of the area surveyed without stratification, allowing further reduction in flight time. The highest precision that could be obtained, however, required nearly complete coverage of high-density strata. Estimates for summer 1998 lacked any reasonable precision because of a single survey unit in the high-density stratum that contained extremely high numbers of bison relative to other survey units.

When I treated previously identified units as complete coverage blocks and simulated a stratified sample survey for the remaining units, there was another substantial increase in efficiency for both winter and summer surveys, as well as a reduction in bias. This is due to the fact that a large proportion of the total population may be predictably concentrated in a few survey units.

Removing these units from the sampling design and treating them as a complete coverage block removed these extreme density values from sample estimation and reduced the large sampling variance associated with them, improving the precision in overall estimates. This design reduced the survey area required in winter 38%, and 53% in summer over SRS without stratification. Winter estimates with CI width less than 20% of the overall density could only be achieved when a complete coverage block of 8 units were included in the sample survey design. For winter 1999, 12 of the 24 survey units in the high-density stratum had to be sampled in order for mean CI widths to be less than the density of the entire sample region, while in 2000, and even larger sample of 16 units were required. More precise estimates could only be obtained with nearly complete coverage of the high-density stratum, which was not much more efficient than a census of these areas. When I applied complete coverage blocks in summer, estimates for 1998 did not provide reasonable precision because a survey unit with extremely high-density was not among the 4 units identified to be in a complete coverage block from the past decade of surveys. In 1999, these same 4 survey units in the complete coverage block contained 95% of the entire population, and overall estimates were precise. For both summers, however, mean CI widths less than the density of the entire sample region could not be obtained even with samples of 12 out of 14 survey units in the high-density stratum. Therefore, this design offered no advantage over a census of these areas during the summer season. To effectively employ complete coverage blocks in winter stratified sample survey plans, survey units with extremely high numbers of bison must be identified prior to the sample survey. If this can be accomplished, it may be used to provide unbiased population estimates, but will still require a great amount of overall survey effort.

Any sample survey design based on survey units of specific geographic areas will be subject to the same shortcoming due to the nature of a gregarious and nomadic species inhabiting a dissected mountainous region. Given the non-random distribution of bison and high variability in density between survey units, reliable estimation from a sample survey design may be extremely difficult to achieve, resulting in imprecise and potentially highly biased estimates. Sample surveys of species that are highly aggregated and nomadic, such as walrus (*Odobenus rosmarus*), may be overwhelmed by sampling variability resulting in very low power for detection of statistical trends (Hills and Gilbert 1994). In contrast, repeated censuses of bison have a low CV relative to population estimates of many other large mammal species (Gibbs 2000). Furthermore, it is possible for bison to move *en masse* between areas necessitating the designation of a large and highly inclusive high-density stratum, and then nearly complete coverage of this stratum will be required for precise estimation. Although I developed objective criteria for strata designation, I cannot recommend generic sample survey strata because moving aggregations of bison may result in high variance and imprecise estimates. Therefore, if a sample survey is necessary, survey strata should be based on reconnaissance conducted immediately prior to the sample survey. These points must be carefully considered when assigning survey units to strata: 1) relatively large samples are needed from each stratum, 2) better sampling efficiency will result from greater differences in mean density and its associated variance, but no strata should be vacant, 3) more reliable estimates will be achieved when overall sample sizes are larger, 4) pre-survey reconnaissance should be used to identify survey units with high numbers of bison, and these units should be included in a complete coverage block to improve estimates. Having small sample sizes from a stratum will result in fewer degrees of freedom for estimation, and therefore result in low precision. If the differences in mean density between strata are small, stratification will have little effect on the efficiency of surveys, and higher ratios of high:low-density allocation will result in poor precision. If there are no bison in a stratum (the stratum is vacant), however, a valid population estimate cannot be calculated. Lower ratios of high:low-density allocation will have larger overall sample sizes requiring more flight time in surveys, but will always be as precise as higher ratios of high:low-density allocation. For these reasons, it is important to make strata assignments with the most current information on distribution. Ideally, this information should come from a recent reconnaissance flight, or from a recent high effort survey or census of high-density strata.

When population estimates from a sample survey are interpreted, it is important to realize that the actual confidence level may be less than the stated nominal level, and there may also be some bias due to sample selection. I have shown the conditions under which these properties fail through simulations. The true CI coverage and bias in estimation may be approximated by examining sampling plan, sample size, and strata allocation to those found in comparable simulations. Additionally, while the bias reported in these simulations was due only to sampling and statistical estimation, there is an additional bias that must be considered due to incomplete detection. This type of bias is a result of the second stage of sampling, and relates to the detection and enumeration of groups of animals within the first stage of sampling, a geographic area.

The first type of bias in the second stage of sampling is due to incomplete detection of bison groups, and may be corrected by an estimate of detection probability (Cochran 1977, Steinhorst and Samuel 1989). The simultaneous ground and air double-survey method I employed to estimate detection probability did not rely on low-level flights that disturb visitors and wildlife, or endanger the biologists conducting these surveys, unlike sightability model and mark-resight methods that require individual identification of marked animals from the air (Samuel et al. 1987, White and Burnham 1999). While these other methods may be considered state-of-the-art wildlife science, the disturbance and liability risks involved in obtaining these data can be inappropriate for a high-profile wildlife species in a National Park, especially when many of these animals occur near developed areas. The double-survey method may also be used again in the future to recalibrate or validate the detection probability model without relying on costly radio-telemetry instrumentation or individual markers on animals. Therefore, the method offers a suitable alternative for large animals with high detection probabilities (Anthony et al. 1999) that is easily repeatable, and also less risky and disruptive for areas with much public visitation.

In this case, because Lincoln-Peterson estimates of detection probabilities were similar and consistently high between ground and air methods, using logistic regression to model detection probabilities for air and ground surveys separately was justified. I found that detection probability of bison in Yellowstone was higher than many other species of ungulates (Caughley 1974), but, as in other studies, detection probability from both aircraft and the ground was most strongly affected by group size, a phenomenon known as group size bias (Cook and Martin 1974, Samuel and Pollock 1981). Habitat was also highly influential; from aircraft, bison in forested and geothermal areas had much lower detection probabilities than those in open habitats. The air and ground methods tended to be complimentary in this respect because habitat was an important factor for aerial surveys, but it was not for ground surveys. This is probably because observers on the ground walked directly through all of these areas, paying particular attention to large, brown animals that may have been obscured by vegetation, but may not have been as vigilant in open habitats where animals could have been concealed behind small topographic rises. This had the effect of making detection probabilities of groups observed from the ground similar across different habitats. Observers in aircraft are not always afforded the luxury of examining particular areas more closely than others, but have unrestricted views from a higher angle in open habitats.

I found detection probability from aircraft to be high for Yellowstone bison due to their use of open habitats, large body size, dark coloration, and highly gregarious behavior. In winter, 74.8 % of bison groups and 91.9 % of individual bison were detected on average from aircraft, while in summer, 79.1% of groups and 96.7 % of individuals were detected. These estimates are not, however, unusually high in comparison to species with similar characteristics. In shrub-steppe habitats in Utah, Wolf and Kimball (1989) also reported that 93.7% of bison were detected during aerial surveys from January–July. In forested habitats of Northwest Territories, Canada, Larter et al. (2000) estimated that 75% of bison were detected during spring aerial surveys, but assumed that all were detected in open habitats. In various areas of Alaska, detection ranged from 83–97% of Moose (*Alces alces*), which have similar body size and coloration as bison, but do not occur in large groups (Gasaway et al. 1986). Lower detection probabilities were found in smaller and more cryptically colored ungulates that occur in less favorable habitats and in smaller groups, such as elk (*Cervus elaphus*) and mountain sheep (*Ovis canadensis*). On the Northern Range of YNP, 50–83% of elk were detected by aircraft (Cougenhour and Singer 1996), while 67% of mountain sheep in canyon habitats of Idaho (Bodie et al. 1995), and 58% of mountain sheep in Colorado were detected on average (Neal et al. 1993). These other species may also attempt to evade aircraft during surveys, accounting for a portion of their lower detection probabilities relative to Yellowstone bison.

Beyond interspecific differences in detection probability, the model I developed during this research also has limitations in generalization to less rigorous aerial survey procedures. Application of this detection probability model requires substantial adherence to the methods I designed in this research and should not be applied to *ad hoc* aerial surveys or those conducted in the past (i.e., Taper et al., unpublished data). There are 3 primary reasons this detection probability is not applicable to other survey methodology: 1) survey effort must be rigorously controlled within areas and not based on prior knowledge or a cursory inspection of an area, 2) surveys should be conducted during the temporal windows I identified, and 3) the criteria for group size definition and habitat categories must be rigorously standardized. Previous aerial surveys in Yellowstone did not employ an explicitly defined survey framework and therefore may not have been as consistent in the total amount of area surveyed and the effort expended per unit area, resulting in more variability between surveys. Surveys conducted outside of optimal temporal windows may result in a mixture of conditions that we did not encounter during this research. For example, fall surveys may be intermediate between winter and summer for group size distribution and total area occupied by bison, but not have consistent background contrast found in those seasons. The detection probability model is sensitive to small groups, solitary animals, and particularly those deemed to be in forest or geothermal habitats. Differences in observer categorization of these factors may therefore introduce some bias to final point estimates and confidence intervals, and therefore requires rigorous standardization.

There may also be small biases in the application of the detection probability due to the double-survey methodology employed in this research, but this should be consistent across surveys that are rigorously standardized. First, I did not account for bison outside of the delineated survey framework, and because animals may occur outside of this area, a slight underestimate may happen. I estimated, however, that this accounted for < 2% of the entire population, and surveying these vast areas was not warranted. I also acknowledge that the double-survey method and analysis I employed may have contained a small bias caused by groups of animals that were not observed from either ground observers or aircraft. Because both ground and aerial survey methods tended to miss solitary bison at higher rates than larger groups, small groups had the highest probability of being simultaneously undetected by both methods and could have caused some minor biases in detection probability. I estimated this percent of all bison groups that were undetected by both air and ground survey methods combined to be 5%, ranging from 11% for solitary bison to 0% for groups $\geq$ 27 in open habitats. It is in the smallest group size categories where the potential for bias exists, however, I found a high proportion of bison in groups $\leq$ 98 in winter, and $\leq$ 418 in summer. Other methods may eliminate this bias, but the increased disturbance and risk associated with these methods are not warranted.

One consequence of lower detection probability in smaller groups for a species that is extremely gregarious like bison, and may be found in aggregations of several hundred animals, is that it tends to have a small effect on the proportion of the total number of animals detected. When these corrections are applied to the observed groups size and habitat distributions observed during the winter aerial surveys, detection probability was approximately 5% higher in summer, primarily because bison were aggregated in much larger groups.

Another consequence when the detection probability model is used to correct population estimates is that there will be less error when bison are aggregated in larger groups than when bison are dispersed in smaller groups. I found that the overall coefficient of variation from the detection probability model applied to winter aerial surveys was 5.4%, but only 1.3% during summer surveys. A large proportion of the population may be concentrated in a few very large aggregations, particularly during summer, and enumerating these large groups may contribute more error to population estimates than accounting for detection of small groups and solitary animals.

Bison are one of the most gregarious large mammals in North America, with the notable exception of caribou (*Rangifer tarandus*). Enumerating these large groups accurately can be challenging, especially for inexperienced observers. Photogrammetric methods have been applied to estimate extremely large aggregations of caribou in Alaska (Fancy et al. 1994), however, this type of approach is unnecessarily effort-intensive for bison in Yellowstone. The oblique 35mm handheld photography I employed was very useful for determining the accuracy of my field counts and how counting error increased with larger groups. I recommend that all groups over 100 animals be photographed 3–5 times during aerial surveys. The cost in additional flight time for photography, film, and processing is trivial relative to the overall expense of a survey, especially considering that the potential for counting error increases with group size and can reach 15% among the large groups of bison.

The regression coefficient I estimated to correct counts of large groups is likely to be highly observer-specific, therefore, I recommend that each new observer verify field counts by photograph counts for all bison groups greater than 100 animals and directly replace field counts with photograph counts in all cases of discrepancy. The process of photographing large groups and counting animals on images is also the best way to practice counting bison in flight, which can be particularly daunting for inexperienced observers. A number of distractions can occur while counting animals in flight, including radio noise, fog, glare, turbulence, other aircraft, and nausea. It is therefore important to be able to reconcile field counts of large groups observed during surveys with a visual record. In cases where a roll of film has failed or images are of poor quality for particular groups, the regression coefficient and its associated error may have to be applied as a correction. This will require any new data from photography to be added to the existing data set so the regression model can be periodically updated. Updating will also allow the model to account for differences between observers.

The application of corrections for counting large groups and detection probability introduce discontinuities with previous methodology used to census the Yellowstone bison population. I provided an analytical method based on replicate total counts to allow continuity in the time series of abundance estimates, which is now more than a century in length. Replicate total counts technically have 3 sources of error, of which only 2 can be accounted for. Total counts cannot account for the entire area inhabited by bison, therefore, an unknown number of animals were outside of the area surveyed each time a total count was conducted. The count-to-count error was treated as sampling error in this case, so that the mean number of bison per survey and associated variance could be estimated by bootstrapping (Efron and Tibshirani 1993). A simple binomial correction can also be applied to adjust for detection probability. The method can be used to determine the average total count within a defined period of time, usually a season, corrected by average detection probability, and the average CV of this method was approximately 8%. Nonetheless, this approach contains an inherent negative bias due to incomplete spatial coverage, and should not be considered a population estimate per se, but an index that may be compared to previous abundance estimates to evaluate trends in population growth.

The statistical detection of trends requires time series data that are both precise and unbiased. High detection probability in bison facilitates precise estimation of abundance and also the detection of demographic change over time. I found that a change of 17.5% could be detected with 75% power under a 2-sided hypothesis and α= 0.05 with a series of 3 years. While the power to detect a trend with a 2-sided hypothesis is high relative to other species, the power is even greater with a 1-sided hypothesis. In practice, managers are more likely to be interested in a 1-sided alternative hypothesis, for example, that there may be a measurable reduction in the population due to particularly severe winter conditions, or, the eradication of diseased animals may result in higher reproduction and population growth rates. In these cases, a 10.2% rate change can be detected if all other parameters remain the same. There is a chance, however, of obtaining a 'bad' estimate that does not reflect the true direction of a trend in some circumstances. If surveys are not replicated, there is no way to know if a particular survey is out of line with the others. Such a survey may lead to a wrong conclusion about a trend.

To determine if index counts are sufficiently repeatable to make correct conclusions about population trends, I attempted to validate an index of abundance estimates. I found that there was a strong relationship between index counts and surveys of much higher effort that were corrected by the detection probability model. There is, however, also a strong chance that an index count may be badly biased. Out of 8 index counts conducted by NPS biologists, one was clearly not comparable to the higher effort surveys. Rigorous control of effort will help reduce the likelihood of a 'bad' index count or survey, however, out of 15 of my surveys with higher effort, one of these was also clearly out of line with the others. Both were conducted late in the winter season of different years and both were negatively biased. If these surveys had been the only available data for a particular season within respective years, inferences regarding trends would have also been badly biased, resulting in incorrect conclusions that the population had remained at the same level when it had in fact increased. It is therefore important to conduct replicate surveys or at least a duplicate survey each season to be certain that anomalous conditions during one of the surveys do not lead to spurious conclusions regarding trends.

When I analyzed the 1998-2000 series of population estimates for trends, I found that winter and summer estimates returned nearly identical values for strong linear population increases despite an apparent negative bias in summer estimates. Although Cheville and McCullough (1998) also found a strong linear pattern of annual increment of 145 bison/year from the 1972–1997 YNP time series after accounting for removals, the annual increment I estimated (422–427) was nearly threefold greater during the 2 time steps from 1998–2000. I also found from my limited data over the course of 3 years that an index of overall annual mortality may be related to accumulated winter SWE. Although 2 of these years were characterized by mild winters, 1998–99 SWE_{acc} in West Yellowstone was 17% greater than the average of 2519 cm from 1968–2000. Despite the relatively high snow accumulation of 1998–99, snow crusting layers associated with the extreme winter of 1996–97 were not present and bison did not make mass movements beyond YNP boundaries. Other studies have also found that the number of elk (*Cervus elaphus*) and bison carcasses in Yellowstone was related to variability in annual climate and winter severity (Green et al. 1997, Podruzny and Gunther 1999). Extreme winters may be associated with high mortality for the entire YNP bison population (Meagher 1973), but winters with greater than average snow accumulation and less severe crusting condition may actually confer increased survival and recruitment of younger age classes during the growing season if total annual precipitation limits availability of forage (Merrill and Boyce 1996). Because I lacked an adequate series to rigorously test non-linear effects of snow accumulation or any alternative hypotheses on apparent mortality, it is premature to conclude that density-independent effect of snow accumulation is a determinant of population limitation in YNP bison. It would, however, be valuable to continue to estimate the population in both summer and winter so that a longer series of this mortality index may be obtained and a more rigorous analysis may be conducted to test the hypothesis of winter climate as a limiting factor against alternatives, such as density-dependent effects of forage limitation during the growing season, or other factors.

RECOMMENDED SURVEY METHODS

There are several major considerations for a long-term monitoring protocol of bison in Yellowstone. Among these are the effort, frequency, and timing of surveys, which are all closely tied to design. The first consideration is how much time will be spent conducting aerial surveys and how surveys will be distributed between seasons, if at all. While the overall effort will be dictated by available funding, more time spent in conducting surveys will provide more information for management, but also increase the amount of time biologists are at risk. There will be a trade-off between the quality of single surveys with high effort and multiple surveys of lower effort. Another key consideration in design is the statistical rigor of the survey. Technically, surveys which are based on a probability sampling scheme have the highest statistical validity (Thompson et al. 1998). Options for different survey designs include 1) replicate total counts, 2) single or duplicate high-effort censuses of rigorously defined areas traditionally occupied by bison, and 3) a sample survey design. Each of these designs has advantages and disadvantages in effort, safety, statistical rigor, and continuity with previous bison survey methodology in Yellowstone.

Conducting replicate total counts would require the most effort and still not provide a statistically rigorous method based on a true probability sampling scheme. Estimates of abundance would be based on the mean of repeated counts in a defined area and error would be based on count-to-count variance, requiring at least 5 repeated counts to obtain robust estimates of this variance. I estimate that each replicate survey of the high-density stratum in winter will require approximately 8.5 hours of flight to complete, and 5.7 hours in summer. Winter surveys of this type should be accomplished in 2 flights on consecutive days, due to fuel limitations and observer fatigue. Summer surveys, however, may be accomplished in one day, although they may require refueling. While I was not able to assess replicate total counts during summer, it would probably be difficult to conduct 5 replicate surveys within a period of time while bison maintain a stable distribution and aggregation. Therefore, this is probably not a practical strategy during summer. Such high effort is not likely to be continued for any long-term monitoring program, but may be implemented periodically when there are management actions that trigger public concern. The method may also violate assumptions of demographic closure if some of the variation between surveys may come from mortality or movements in or out of the survey framework. This type of estimate, however, does provide a link to the long time series of abundance estimates for bison over the last century. A simple correction may also be applied by adjusting the mean of repeated counts by the average detection probability to reduce the negative bias associated with those bison that are missed under average survey conditions. The resulting population estimates would therefore represent the average bison count within a defined area over a period of time, corrected by average detection probability conditions. It is also important to realize that it would be nearly impossible to achieve complete spatial coverage during replicate total counts, therefore, estimates would contain some unknown negative bias. The method would, however, allow surveys that are clearly out of line with others from a similar time period to be recognized as ‘bad’ surveys, and also offer the most continuity with previous bison survey methodology.

A single or duplicated high-effort census of rigorously defined areas traditionally occupied by bison is another survey method that is comparable to previous methodology. Much effort may be put into a single census survey each season to provide complete spatial coverage of these well-defined areas, and while it would still contain some negative bias due to incomplete spatial coverage of all areas bison might inhabit, this negative bias may be corrected by implementing a sample survey of a low-density stratum in conjunction with a census of a high-density stratum. The detection probability model may also be used to correct for the unseen proportion of bison during the survey. Each census survey of this type in winter will require 2 flights on consecutive days with approximately 8.5 hours of total flight time to complete, while in summer, each survey may be accomplished in one day, requiring approximately 5.7 hours. Additional flight time will also be required to implement sample surveys of low-density strata. There is also a possibility for anomalous conditions that result in a ‘bad’ survey. In this case, a ‘bad’ survey would not be recognizable unless comparable data from the same time period exist. Therefore, a duplicate survey may be implemented as a reliability check. For this method, there would be no statistical error associated with a probability sampling scheme because it is a census of high-density stratum. Therefore, the total number of bison observed would provide an absolute lower boundary, and a 95% confidence interval could be estimated with the detection probability model to account for specific factors during the survey. This approach would have lower cost and much less time at risk than replicate total counts, and the overall effort would be comparable to a sample survey. Although incomplete spatial coverage may result in underestimation of a small part of the population, conducting surveys at times that reduce the variability associated with detection probability would control for this type of bias, providing a reliably consistent and relatively precise index of the population from year to year.

Single or multiple surveys based on a sample survey scheme should offer the most statistically rigorous method of determining population size while simultaneously decreasing the effort and risk for biologists. Sample surveys uphold the highest standard in wildlife survey designs (Lancia et al. 1996, Thompson et al. 1998), are potentially the least biased, however, they are also the least precise of the methods I considered. By sampling the entire framework, these types of surveys can account for areas not sampled, using mean density and variability between survey units to estimate the population, and thereby eliminating the negative bias due to incomplete spatial coverage. There is, however, the potential for a great deal of sample variability for species that are highly aggregated and nomadic that may result in very low power for detection of statistical trends if sample surveys are not properly planned. In summer, the method may result in estimates that are imprecise and/or biased beyond usefulness due to the extremely aggregated distribution. I therefore do not recommend sample surveys be conducted in summer. Anomalous conditions may also lead to a 'bad' survey, which would not be recognizable as such in the absence of replicate surveys. I estimate that under the best sample survey conditions where strata are carefully specified from reconnaissance immediately prior to the survey, that the amount of time required would be approximately 8.6 hours in winter, requiring 2 days to complete. This method requires approximately the same amount of effort in survey time in comparison to a census of the entire area traditionally used by bison, but it is also the least comparable to methods employed in the past, thereby introducing a break in continuity to the long-term time series. If sample surveys are corrected for the proportion of bison that are undetected, the additional uncertainty simultaneously associated with these 2 sources of error may make detection of population trends more difficult. Instead of correcting sample surveys for detection probability, I recommend that sample surveys be conducted at times that reduce the variability associated with detection probability, thereby rigorously standardizing the comparison of estimates between years.

Regardless of which design is employed, there are seasonal windows of opportunity which reduce the spatial extent of area occupied by bison, and the proportion of bison that are not detected, thereby minimizing the variability in population estimates. I recommend conducting winter surveys in December and January and summer surveys in late July and early August. Bison are most spatially aggregated during these periods, providing the highest detection probability and lowest variability between surveys. I also recommend avoiding unusually warm winter conditions during surveys, as a higher proportion of bison may inhabit forested areas under these conditions. Timing surveys to occur during specified windows of opportunity will also standardize comparisons between annual population estimates, reducing the variability associated with over-winter mortality, seasonal movements, habitat use, gregarious behavior, and the likelihood of anomalous conditions that result in a ‘bad’ survey. Conditions during summer may reduce the likelihood of such a ‘bad’ survey, may be optimal for reducing variability due to detection, and offer savings in both effort and expense, but may also be complicated by extremely large groups that are difficult to enumerate. It is therefore valuable to have population estimates from both seasons as a type of replication to cross-check between seasons.

Conducting surveys during both summer and winter also provides useful demographic information beyond a time series from a single season, and may be integrated with a comprehensive population monitoring scheme (Fig. 26). From a series of either winter or summer estimates, annual increment or growth rates may be estimated. Having population estimates of adults in both summer and winter, however, may provide an index of the overall herd mortality that occurs primarily in late winter (Meagher 1973). This index of mortality may be used to investigate density-dependent and density-independent population regulation mechanisms, but differences between winter and summer estimates may also be checked for consistency and considered a type of replication. Another high-priority parameter in bison demography is reproductive rate. While I found that my aerial counts of bison calves underestimated the numbers required to account for the observed annual increment in population growth, they may still

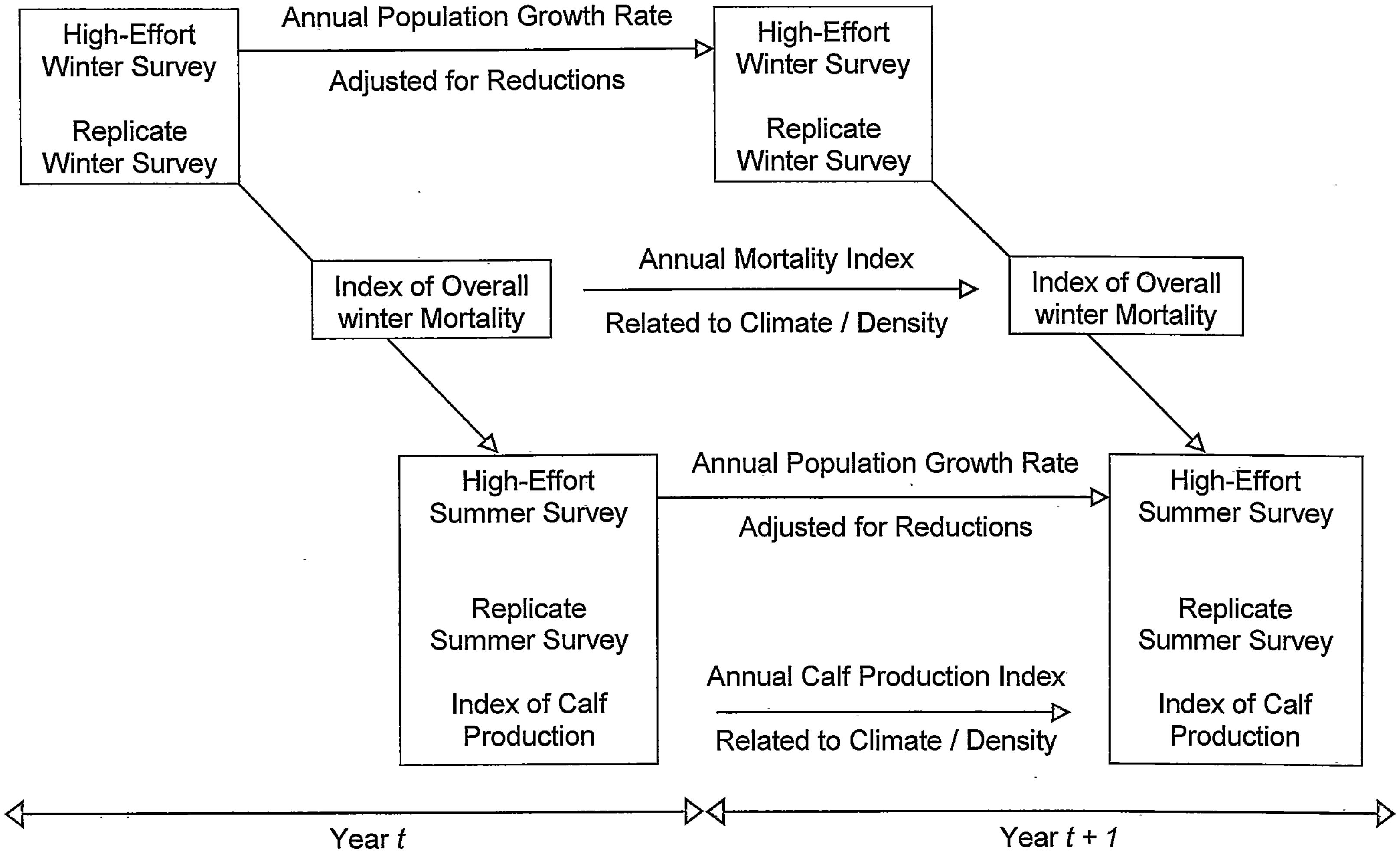

Fig. 26. A comprehensive population monitoring scheme for bison in Yellowstone National Park.

serve as a useful index of calf production. Methods other than aerial surveys may also be employed to estimate rates of reproduction. Calves were located in only 3 areas of YNP during summer, and 2 of these areas were easily accessible on the ground, which may facilitate ground-based surveys. Wolfe and Kimball (1989) determined that ground-based surveys of bison calves were also biased, however Larter et al. (2000) employed ground-based surveys of calves in demographic analyses of bison in Canada. If such surveys of calves can be well designed and implemented, they may provide a useful index of annual reproduction that relates to environmental variation, density dependent growth responses, and the effectiveness of management actions. If, for example, brucellosis is eradicated from YNP bison (Cheville and McCullough 1998), age-specific reproduction rates may be altered, and younger age classes of females may begin to contribute to higher annual overall reproductive rates. It is therefore important to monitor this aspect of demography to understand natural population regulation mechanisms and the effects of management on bison in YNP. Finally, removals are generally well documented and can be easily incorporated to adjust population growth rates, but it would also be useful for at least 2 purposes to estimate the rates and causes of natural mortality. First, if overall herd mortality rates were estimated and incorporated with removals, index-removal methods could be used to determine detection probability without intensive field efforts (Eberhardt 1982). Secondly, mortality estimates would allow an investigation of predator-prey relationships and provide insights into potential population regulation mechanisms such as compensatory versus additive mortality. Therefore, distributing surveys between both summer and winter offers much more demographic information than single-season surveys, but also incurs additional expense, effort, and associated risks, as would attempts to estimate reproduction and mortality rates.

I therefore recommend at least one high effort survey be conducted at the beginning of both the winter and the summer seasons, and then duplicate surveys be conducted within a short period to assess the agreement between estimates. The results of the first winter high effort survey may also be used to modify strata assignments in preparation for a sample survey, if necessary. The 2 surveys each season may be compared to determine their consistency, and in the case of discrepancy, another survey may be required for resolution. In terms of an overall monitoring program, useful demographic data may be obtained for a modest overall expense and effort of 2 surveys per season each year. The total amount of flight time required to implement this monitoring program would be less than 30 hours/year entailing duplicated surveys of approximately 8.5 hours in winter and 5.7 hours in summer.

Several physical and behavioral attributes make the Yellowstone bison well suited for population monitoring with aerial surveys. The combination of large body size, dark coloration, gregariousness, and predominant use of open habitats are all highly advantageous during aerial surveys. Additionally, nearly ideal survey conditions can be achieved during the temporal windows identified in this study. These factors all contribute to high detection probability, facilitating more precise and repeatable population estimates than for many other large mammals in YNP while requiring only a relatively modest overall effort in a monitoring framework to estimate the demographic parameters of unbiased population growth rates, and indices of overall annual mortality and calf production. If this comprehensive population monitoring scheme is implemented over an extended period, there is a very high potential for understanding fine-scale population dynamics in bison. While understanding the relative contribution of density-dependent and density-independent mechanisms in population regulation is valuable in itself, it may also be useful for interpreting the existing century-long time series for bison in Yellowstone, and these general findings for bison may provide insights into the population ecology of other species that experience the Yellowstone environment in a fundamentally similar way.

LITERATURE CITED

Anthony, R. G., M. G. Garrett, and F. B. Isaacs. 1999. Double-survey estimates of bald eagle populations in Oregon. Journal of Wildlife Management 63:794-802.

Barnett, V. 1991. Sample survey: principles & methods. Arnold Press, London, UK.

Baskin, Y. 1998. Home on the range. Bioscience 48:245-251.

Bjornlie, D. D., and R. A. Garrott. 2001. Ecological effects of winter road grooming on bison in Yellowstone National Park. Journal of Wildlife Management 65:423-435.

Bodie, W. L., E. O. Garton, E. R. Taylor, and M. McCody. 1995. A sightability model for bighorn sheep in canyon habitats. Journal of Wildlife Management 59:832-840.

Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 1993. Distance Sampling. Chapman and Hall, New York, New York, USA.

Burnham, K. P., D. R. Anderson, G.C. White, C. Brownie, and K. H. Pollack. 1987. Design and analysis methods for fish survival experiments based on release-recapture. American Fisheries Society Monograph 5.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: A practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

Caughley, G. 1974. Bias in aerial survey. Journal of Wildlife Management 38:921-933.

Caughley, G. 1977. Analysis of vertebrate populations. John Wiley and Sons, New York, New York, USA.

Caughley, G., and J. Goddard. 1972. Improving the estimates from inaccurate censuses. Journal of Wildlife Management 36:135-140.

Chao, A. 1989. Estimating population size for sparse data in capture-recapture experiments. Biometrics 45:427-438.

Cheville, N. F., and D. R. McCullough 1998. Brucellosis in the Greater Yellowstone Area. National Academy Press, Washington, D. C., USA.

Cochran, W.G. 1977. Sampling techniques. Third edition, John Wiley and Sons, New York, New York, USA.

Collett, D. 1999. Modelling Binary Data. Chapman and Hall/CRC, Boca Raton, Florida, USA.

Cook, R. D., and F. B. Martin. 1974. A model for quadrat sampling with "visibility bias". Journal of the American Statistical Association 69:345-349.

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

Cougenhour, M. B., and F. J. Singer. 1996. Elk population processes in Yellowstone National Park under the policy of natural regulation. Ecological Applications 6:573-593.

Dennis, B., and M. L. Taper. 1994. Density dependence in time series observations of natural populations: estimation and testing. Ecological Monographs 64:205-224.

Dobson, A., and M. Meagher. 1996. The population dynamics of brucellosis in Yellowstone National Park. Ecology 77:1026-1036.

Eberhardt, L. L. 1982. Calibrating an index with removals. Journal of Wildlife Management 46:734-740.

Eberhardt, L. L. 1987. Population projections from simple models. Journal of Applied Ecology 24:103-118.

Eberhardt, L. L., D. G. Chapman, and J. R. Gilbert. 1979. A review of marine mammal census methods. Wildlife Monographs 63.

Eberhardt, L. L., and M. A. Simmons. 1987. Calibrating population indices by double sampling. Journal of Wildlife Management 51:665-675.

Efron, B., and R. J. Tibshirani. 1993. An Introduction to the Bootstrap. Chapman and Hall/CRC, New York, New York, USA.

Elkington, J. S. 2000. Detecting stability and causes of change in population density. Pages 191-212, *in* L. Boitani, and T. K. Fuller, editors, Research techniques in animal ecology: controversies and consequences. Columbia University Press, New York, New York, USA.

Estes, J. A., and R. J. Jameson. 1988. A double-survey estimate for sighting probability of sea otters in California. Journal of Wildlife Management 52:70-76.

Evans, C. D., W. A. Troyer, and C. J. Lensink. 1966. Aerial census of moose by quadrat sampling units. Journal of Wildlife Management 30:767-776.

Fancy, S. G., K. R. Whitten, and D. E. Russell. 1994. Demography of the Porcupine caribou herd, 1983-1992. Canadian Journal of Zoology 72:840-846.

Farnes, P. E. 1996. An index of winter severity for elk. Pages 303-306, in F. J. Singer, editor, Effects of grazing by wild ungulates in Yellowstone National Park. Technical Report NPS/NRYELL/96-01, NPS, Natural Resource Information Division, Denver, Colorado, USA.

Frank, F. A., S. J. McNaughton, and B. F. Tracy. 1998. The ecology of the Earth's grazing ecosystems. BioScience 48:513-521.

Gasaway, W. C., S. D. DuBois, D. J. Reed, and S. J. Harbo. 1986. Estimating moose population parameters from aerial surveys. Biological Papers of the University of Alaska Number 22.

Geibel, J. J., and D. J. Miller. 1984. Estimation of sea otter, *Enhydra lutris*, population, with confidence bounds, from air and ground counts. California Fish and Game 70:225-233.

Gerrodette, T. 1987. A power analysis for detecting trends. Ecology 68:1364-1372.

Gerrodette, T. 1993. TRENDS:software for a power analysis of linear regression. Wildlife Society Bulletin 21:515-516.

Gibbs, J. P. 2000. Monitoring populations. Pages 212-252, *in* L. Boitani, and T. K. Fuller, editors, Research techniques in animal ecology: controversies and consequences. Columbia University Press, New York, New York, USA.

Graham, A., and R. Bell. 1989. Investigating observer bias in aerial survey by simultaneous double-counts. Journal of Wildlife Management 53:1009-1016.

Green, G. I., D. J. Mattson, and J. M. Peek. 1997. Spring feeding on ungulate carcasses by grizzly bears in Yellowstone National Park. Journal of Wildlife Management 61:1040-1055.

Hansen, M. H., Hurvitz, W. N., and W. G. Madow. 1953. Sample survey methods and theory. Volume 1: Methods and applications. John Wiley and Sons, New York, New York, USA.

Hills, S., and J.R. Gilbert. 1994. Detecting Pacific walrus population trends with aerial survey: a review. Transactions of the North American Wildlife and Natural Resource Conference 59. Wildlife Management Institute, Washington, DC, USA.

Keiter, R. B. 1997. Greater Yellowstone's bison: unraveling of an early American wildlife conservation achievement. Journal of Wildlife Management 61:1-11.

Larter, N. C., A. R. E. Sinclair, T. Ellsworth, J. Nishi, and C. C. Gates. 2000. Dynamics of reintroduction in an indigenous large ungulate: the wood bison of northern Canada. Animal Conservation 4:299-309.

Lancia, R. A., J. D. Nichols, and K. H. Pollack. 1996. Estimating the number of animals in wildlife populations. Pages 215-253, *in* T. A. Bookhout, editor, Research and management techniques for wildlife and habitats. The Wildlife Society, Bethesda, Maryland, USA.

Magnusson, W. E., G. J. Caughley, and G. C. Gregg. 1978. A double-survey estimate of population size from incomplete counts. Journal of Wildlife Management 42:174-176.

Manly, B. F. J. 1991. Randomization and Monte Carlo experiments in biology. Chapman and Hall, New York, New York, USA.

Marcot, B. G., M. G. Raphael, and K. H. Berry. 1983. Monitoring wildlife habitat and validation of wildlife-habitat relationships models. Transactions of the North American Wildlife and Natural Resources Conference 48:315-329.

Meagher, M. M. 1973. The bison of Yellowstone National Park. National Park Service Scientific Monograph Series Number 1. National Park Service, Washington, D. C., USA

Meagher, M. M. 1989. Range extension by bison of Yellowstone National Park. Journal of Mammalogy 70:670-675.

Merrill, E. and M. S. Boyce. 1996. Summer range and elk population dynamics in Yellowstone National Park. Pages 263-274 *in* R. B. Keiter and M. S. Boyce, eds. The Greater Yellowstone ecosystem; redefining America's wilderness heritage. Yale University Press, New Haven, Connecticut, USA.

Mood, A. M., F. A. Graybill, and D. C. Boes. 1974. Introduction to the theory of statistics. McGraw-Hill series in probability and statistics. New York, New York, USA.

Murdoch, W.W. 1994. Population regulation in theory and practice. Ecology 75:271-287.

National Park Service (NPS). 2000. Bison management plan for the State of Montana and Yellowstone National Park. Executive Summary and Final Environmental Impact Statement. NPS D-655a/August 2000. U.S. Government Printing Office 2000-592-691/12007, Denver, Colorado, USA.

Neal, A. K., G. C. White, R. B. Gill, D. F. Reed, and J. H. Olterman. 1993. Evaluation of mark-resight model assumptions for estimating mountain sheep numbers. Journal of Wildlife Management 57:436-450.

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

Norton-Griffiths, M. 1974. Reducing counting bias in aerial surveys by photography. East African Wildlife Journal 12:245-248.

Norton-Griffiths, M. 1978. Counting animals. African Wildlife Leadership Foundation Handbook Number 1. Nairobi, Kenya.

Podruzny, S., and K. Gunther. 1999. Spring ungulate availability and use by grizzly bears in Yellowstone National Park. Pages 25-28 *in* C. C. Schwartz and M. A. Haroldson, eds. Yellowstone grizzly bear investigations: annual report of the Interagency Grizzly Bear Study Team, 1999. U. S. Geological Survey, Bozeman, Montana, USA.

Pojar, T. M., D. C. Bowden, and R. B. Gill. 1995. Aerial counting experiments to estimate pronghorn density and herd structure. Journal of Wildlife Management 59:117-128.

Sæther, B. E. 1997. Environmental stochasticity and population dynamics of large herbivores: a search for mechanisms. Trends in Ecology and Evolution 12:143-149.

Samuel, M. D., and K. H. Pollock. 1981. Corrections of visibility bias in aerial surveys where animals occur in groups. Journal of Wildlife Management 45:993-997.

Samuel, M. D., E. O. Garton, M. W. Schlegel, and R. W. Carson. 1987. Visibility bias in aerial surveys of elk in northcentral Idaho. Journal of Wildlife Management 51:622-630.

SAS Institute. 1990. SAS/STAT user's guide, 6.04 edition. SAS Institute, Cary, North Carolina, USA.

Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. Biometrics Bulletin 2:110-114.

Seber, G. A. F. 1982. The estimation of animal abundance. Macmillan, New York, New York, USA.

Shenk, T. M., G. C. White, and K. P. Burnham. 1998. Sampling-variance effects on detecting density dependence from temporal trends in natural populations. Ecological Monographs 68:445-463.

Siniff, D. B., and R. O. Skoog. 1964. Aerial censusing of caribou using stratified random sampling. Journal of Wildlife Management 28:391-401.

Slade, N. A., and S. M. Blair. 2000. An empirical test of using counts of individuals captured as indices of population size. Journal of Mammalogy 81:1035-1045.

Steinhorst, R. K., and M. D. Samuel. 1989. Sightability adjustment method for aerial survey of wildlife populations. Biometrics 45:415-425.

The Math Works. 1999. Matlab student version 5.3, release 11, Natick, Massachusetts, USA.

Thompson, S. K. 1992. Sampling. John Wiley and Sons, New York, New York, USA.

Thompson, W. L., G. C. White, and C. Gowan. 1998. Monitoring Vertebrate Populations. Academic Press, San Diego, California, USA.

van Manen, F. T., B. A. Crawford, and J. D. Clark. 2000. Predicting red wolf success in the southeastern United States. Journal of Wildlife Management 64:895-902.

Watson, R. M. 1969. Aerial photographic methods in censuses of animals. East African Agricultural and Forestry Journal 34:32-37.

White, G. C., and R. A. Garrott. 1990. Analysis of wildlife radio-tracking data. Academic Press, San Diego, California, USA.

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

Wolfe, M. L., and J. F. Kimball. 1989. Comparison of bison population estimates with a total count. Journal of Wildlife Management 53:593-596.

Yellowstone National Park. 1997. Yellowstone's northern range: complexity and change in a wildland system. National Park Service, Mammoth Hot Springs, Wyoming, USA.

APPENDICES

APPENDIX A

SAMPLING ESTIMATORS AND ERROR FORMULAS

Sampling Estimator and Error Formulas

The following sampling estimator and error formulas were used in simulations of aerial surveys of bison in Yellowstone National Park. For each estimator, we calculated $\hat{d}$, the estimate of density. In these formulas, y represents the response, which is the total number of bison within survey units, x represents the area of survey units, μ represents the population mean of area, n represents the number of survey units, S^2 represents the sample variance, and π represents Horvitz-Thompson inclusion probabilities. Strata are indicated by subscripted h, the response of individual units by i, and totals by T.

The simple random sampling estimator of density (SRS; Barnett 1991) is:

$$\hat{d} = N\bar{y} / X_T$$

where

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

with standard error

$$SE(\hat{d}) = \sqrt{\frac{N(N-n)}{X_T^2} S^2} .$$

The stratified simple random sampling estimator is:

$$\hat{d} = \left(\sum_{i=1}^{h} X_h \bar{y}_h \right) \Big/ X_T$$

with standard error

$$SE(\hat{d}) = \sqrt{\frac{\sum_{h=1}^{2} X_h (X_h - x_h) S^2}{X_T^2}}$$

where S^2 is the sample variance of the y_i values.

The ratio I estimator (The sample average ratio estimator; Barnett 1991) is:

$$\bar{r}_1 = \frac{1}{n} \sum_{i=1}^{n} (y_i / x_i).$$

The ratio I estimator with the bias correction term (Cochran 1977) is:

$$\bar{r}_{1c} = \bar{r}_1 + \frac{(N-1)n}{n-1} [\bar{y} - \bar{r}_1 \bar{x}]$$

with root mean square error

$$\sqrt{MSE(\hat{r}_{1c})} = \sqrt{(1 - \frac{n}{N}) \frac{S_r^2}{n} + \frac{(N-1)n}{n-1} [\bar{y} - \bar{r}_1 \bar{x}]}$$

where $r_i = \frac{y_i}{x_i}$ for i = 1, ..., n, and S_r^2 is the sample variance of the r_i values.

The ratio II estimator (Ratio of the sample averages estimator; Barnett 1991) is:

$$\hat{r}_2 = (\bar{y} / \bar{x}) = (Y_T / X_T)$$

with standard error

$$SE(\hat{r}_2) = \sqrt{\left(\frac{N-n}{N \mu_x^2} \right) \frac{S_r^2}{n}}$$

where

$$S_r^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{r}_2 x_i)^2 .$$

The separate ratio estimator of the total for each stratum added together (Cochran 1977) is:

$$\hat{Y}_{Rs} = \sum_{h=1}^{2} \frac{y_h}{x_h} X_h = \sum_{h=1}^{2} \frac{\bar{y}_h}{\bar{x}_h} X_h = \sum_{h=1}^{2} \hat{r}_{2h} X_h$$

$$\hat{d}_{Rs} = \left(\sum_{h=1}^{2} \frac{\bar{y}_h}{\bar{x}_h} X_h \right) \Big/ X_T = \frac{\sum_{h=1}^{2} \hat{r}_{2h} X_h}{X_T}$$

with standard error

$$SE(\hat{d}_{Rs}) = \sqrt{\frac{\frac{\sum_{h=1}^{2} N_h (N_h - n_h)/n_h (n_h - 1)}{(N_h - 1) n_h} \left(\sum_{i=1}^{n_h} y_{ih}^2 + \hat{r}_{2h}^2 \sum_{i=1}^{n_h} x_{ih}^2 - 2\hat{r}_{2h} \sum_{i=1}^{n_h} y_{ih} x_{ih} \right)}{X_T^2}}.$$

The combined ratio estimator where the overall estimate is derived from a single combined ratio, $\hat{Y}_{Rc}$ (Cochran 1977) is:

$$\hat{Y}_{Rc} = \frac{\hat{Y}_{st}}{\hat{X}_{st}} X = \frac{\bar{y}_{st}}{\bar{x}_{st}} X$$

where

$$\hat{Y}_{st} = \sum_h N_h y_h \qquad \hat{X}_{st} = \sum_h N_h x_h$$

$$\hat{Y}_{Rc} = \frac{\sum_{h=1}^{2} m_h \bar{y}}{\sum_{h=1}^{2} m_h \bar{x}}$$

with standard error

$$SE(\hat{Y}_{Rc}) = \sqrt{\frac{\dfrac{\sum_{h=1}^{2} N_h(N_h - n_h)/n_h(n_h - 1)}{(N_h - 1)n_h}\left(\sum_{i=1}^{n_h} y_{ih}^2 + \hat{r}_{Rc}^2 \sum_{i=1}^{n_h} x_{ih}^2 - 2\hat{r}_{Rc} \sum_{i=1}^{n_h} y_{ih} x_{ih}\right)}{X_T^2}}.$$

The estimator for sampling proportional to size (PPS; Barnett 1991) is:

$$\tilde{y}_T = \frac{X_T}{n} \sum_{j=1}^{n} \frac{y_j}{x_j}$$

$$\hat{d} = \frac{1}{n} \sum \frac{y_i}{\pi_i} X_T$$

with standard error

$$SE(\hat{d}) = \sqrt{\frac{\sum_{i=1}^{n} \left(\frac{1}{\pi_i^2} - \frac{1}{\pi_i}\right)^2 y_i^2 + 2 \sum_{i=1}^{n} \sum_{j>i}^{n} \left(\frac{1}{\pi_i \pi_j} - \frac{1}{\pi_{ij}}\right) y_i y_j}{X_T^2}}.$$

The estimator for stratified sampling proportional to size is:

$$\hat{d} = \sum_{i=1}^{h} \left(\frac{1}{n} \sum \frac{y_i}{\pi_i}\right) X_T$$

with standard error

$$SE(\hat{d}) = \sqrt{\frac{\sum_{h=1}^{2} \left(\frac{X_h}{X_T}\right)^2 \left(\frac{1}{\pi_{ih}^2} - \frac{1}{\pi_{ih}}\right)^2 y_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} \left(\frac{1}{\pi_{ih}\pi_{jh}} - \frac{1}{\pi_{ijh}}\right) y_{ih} y_{jh}}{X_T^2}}.$$

APPENDIX B

SURVEY UNIT RANKS FOR SIMULATIONS

Table 24. Ranks of survey units and rank orders used for simulations of sampling for aerial surveys of bison in Yellowstone National Park. Data are from winter periods corresponding to population levels of <1,500 (1970-78), 1,500-3,000 (1979-1990), >3,000 (1991-1997) (M.L.Taper et al., Montana State University, unpubl. data), and during recent years of study (1998-2000).

			Ranks				Rank Order	
			1970-	1979-	1991-	1998-	1970-	1990-
Stratum	Range	Unit	1978	1990	1997	2000	2000	2000
Census	FH	LB	6.3	2.7	2.8	3.2	2	1
Block	HV	HV	1.7	2.4	5.7	3.2	1	2
	FH	NP	6.6	6.6	7.0	4.4	4	3
	FH	HL	6.2	8.2	6.5	6.2	5	4
	PV	PV	3.3	3.4	7.6	6.0	3	5
	NR	HM	7.0	8.5	12.5	5.3	7	6
	NR	JB	4.9	4.2	9.0	9.4	6	7
	FH	OF	11.1	10.2	8.8	9.8	10	8
High	FH	TB	7.2	11.6	10.4	9.2	9	9
Density	HV	EA	7.5	7.2	12.5	10.4	8	10
	WY	MW		16.0	11.4	13.8	14	11
	MG	MJ	9.0	14.2	14.2	11.0	11	12
	PV	TL	9.4	15.6	11.4	14.1	12	13

Table 24. Continued.

			Ranks				Rank Order	
Stratum	Range	Unit	1970-1978	1979-1990	1991-1997	1998-2000	1970-2000	1990-2000
	MG	NG		19.7	14.0	12.9	22	14
	NR	MT		13.3	10.8	17.6	15	15
	NR	DF	17.0	11.3	9.7	20.6	18	16
	PV	MK	7.3	13.4	14.3	16.1	13	17
	MG	GR	15.0	16.4	18.2	13.1	23	18
	FH	SM	10.1	14.4	16.9	14.4	16	19
	WY	HB			18.8	12.6	24	20
	WY	SF			21.5	10.0	25	21
	FH	GP				16.0	26	22
	FH	MG	12.3	15.4	19.3	13.2	20	23
	MG	GM		20.8	17.7	16.5	34	24
	WY	MB		19.7	19.1	15.5	32	25
	PV	FL	8.8	13.0	16.4	18.2	17	26
	NR	FF	7.5	23.3	17.5		28	27
	HV	HS	14.3	17.8	24.3	11.0	29	28
	NR	SB	9.9	16.4	19.6	15.8	21	29
	WY	DC			18.0	18.3	33	30

Table 24. Continued.

			Ranks				Rank Order	
Stratum	Range	Unit	1970-1978	1979-1990	1991-1997	1998-2000	1970-2000	1990-2000
	SO	WT				19.0	35	31
	HV	AP	8.4	13.4	21.1	17.3	19	32
Low	NR	MN			20.0		38	33
Density	SO	PP				20.6	39	34
	WY	WY			21.5	22.0	44	35
	WY	CR			22.1		46	36
	PV	LL		19.1	24.2	20.0	42	37
	PV	RC	10.0			22.2	27	38
	FH	FC			26.6	18.8	47	39
	WY	CF		13.0	22.8		31	40
	MG	SL			22.9	24.7	50	41
	MG	OC		25.0	23.7	24.0	51	42
	NR	BD	13.3	18.7	21.1	26.7	37	43
	NR	LP	15.7		24.0		36	44
	MG	SV		25.0	24.0		53	45
	WY	CC			26.0	23.5	55	46
	NR	MR	9.0		25.0		30	47

Table 24. Continued.

Stratum	Range	Unit	Ranks 1970-1978	1979-1990	1991-1997	1998-2000	Rank Order 1970-2000	1990-2000
	NR	ME		23.3	25.7		52	48
	MG	RM			28.5	23.2	57	49
	NR	UP			26.0		58	50
	MG	RF		27.0	26.3		60	51
	HV	CL			24.0	29.0	59	52
	HV	CV		20.0	26.5	27.0	54	53
	HV	WL		21.0	27.5	26.0	56	54
	NR	BP		20.5		27.0	49	55
	NR	SC	15.8	21.8	28.6	26.0	48	56
	NR	LS	10.5	23.3	28.1		40	57
	NR	WS	15.5	19.0	28.8		41	58
	PV	BB				29.0	62	59
	MG	GL			29.0		61	60
	NR	DD	15.0		29.0		45	61
	NR	AM	11.8	21.8		30.0	43	62

Table 25. Ranks of survey units and rank orders used for simulations of sampling for aerial surveys of bison in Yellowstone National Park. Data are from summer periods corresponding to population levels of <1,500 (1970-78), 1,500-3,000 (1979-1990), >3,000 (1991-1997) (M.L.Taper et al., Montana State University, unpubl. data), and during recent years of study (1998-2000). Bison were never observed in two survey units, but large numbers were often observed in surrounding units.

			Ranks				Rank Order	
Stratum	Range	Unit	1970-1978	1979-1990	1991-1997	1998-2000	1970-2000	1990-2000
Census	HV	HV	1.0	1.0	1.2	1.0	1	1
Block	NR	SB	4.6	4.3	5.2	4.4	2	2
	FH	LB	5.7	7.5	5.2		7	3
	NR	JB	9.0	6.1	8.2	2.8	11	4
High	NR	ML	4.3	5.3	7.0	5.0	4	5
Density	HV	EA	7.5	5.1	8.3	3.8	8	6
	HV	AP	9.0	6.8	8.2	4.0	12	7
	NR	MP	5.0	5.6	7.0	5.7	6	8
	PV	PV	4.5	7.8	7.3	6.0	10	9
	NR	WS	2.5	4.1	9.7	3.8	3	10
	WY	MB		13.5	7.6		24	11
	PV	PC	4.5	8.4	10.3	5.0	13	12

Table 25. Continued.

			Ranks				Rank Order	
Stratum	Range	Unit	1970-1978	1979-1990	1991-1997	1998-2000	1970-2000	1990-2000
	NR	UP	3.0	3.8	9.2	7.0	5	13
	NR	AM	6.2	9.9	10.7	6.3	16	14
	PV	RC	3.5	5.7	9.7		9	15
	NR	FF	5.0	12.3	12.3	8.0	20	16
	FH	NP	7.0	8.4	10.5		17	17
	NR	MN	3.6	7.4	14.9	6.0	15	18
Low	MG	MJ			10.7		26	19
Density	FH	TB	14.0	11.0	11.6		35	20
	WY	MW		4.0	11.7		14	21
	PV	LL			12.0		30	22
	HV	CV			16.7	7.4	32	23
	MG	GR		12.3	12.1		34	24
	NR	BF	10.0				21	25
	PV	MK	5.7	7.4	13.2		18	26
	FH	SM	9.0	12.3	14.0		28	27
	PV	TL	8.0	9.9	14.0		25	28

Table 25. Continued.

Stratum	Range	Unit	Ranks 1970-1978	Ranks 1979-1990	Ranks 1991-1997	Ranks 1998-2000	Rank Order 1970-2000	Rank Order 1990-2000
	FH	OF		9.0	14.8		29	29
	NR	LP	5.0	17.0	15.5		37	30
	HV	HS	6.2	9.8	15.7		23	31
	NR	LS	3.6	7.5	16.0		19	32
	WY	CF		8.0	16.2		33	33
	FH	HL	8.3	9.7	16.4		27	34
	MG	GM		14.0	16.6		39	35
	PV	FL	4.0	10.0	17.0		22	36
	WY	SF			17.0		40	37
	WY	DC			17.3		41	38
	MG	NG	6.0	12.0	18.0		31	39
	NR	HM	8.0	11.0	18.2		36	40
	MG	RF			18.7		42	41
	FH	MG			18.8		43	42
	WY	CR			19.7		44	43
	FH	FC			20.0		45	44
	MG	SV			20.0		46	45

Table 25. Continued.

			Ranks				Rank Order	
Stratum	Range	Unit	1970-1978	1979-1990	1991-1997	1998-2000	1970-2000	1990-2000
	WY	WY			20.0		47	46
	MG	RM			20.7		48	47
	MG	SL			23.0		49	48
	MG	GL			24.0		50	49
	NR	MT		5.0	24.0		38	50
	MG	OC			25.0		51	51
	NR	BD			25.0		52	52
	NR	MR					--	--
	WY	CC					--	--

APPENDIX C

DATA USED FOR SIMULATIONS

Table 26. Season, strata, number of bison, sample area, and number of survey units used in simulations of sampling for bison in Yellowstone National Park. Stratified simulations were first conducted with the census blocks treated as part of high-density strata, then census blocks were removed from sample strata.

		Number of bison		Area	Number
Season	Stratum	1999	2000	km^2	of Units
Winter	Low Density	4	40	969	30
	High Density	742	749	669	24
	Census Block	1345	1750	421	8
	Total	2091	2539	2058	62
		1998	1999		
Summer	Low Density	1	5	1070	36
	High Density	437	90	542	14
	Census Block	1234	1868	247	4
	Total	1672	1963	1859	54

APPENDIX D

LOGISTIC REGRESSION COEFFICIENTS

Table 27. Logistic regression coefficients and variance/covariance estimates for a detection probability model for bison during aerial surveys in Yellowstone National Park. Parameters include the model intercept, effect of group size, and forest and geothermal habitats relative to open habitats.

		Variance/Covariance		
Parameters	Coefficients	Intercept	Group size	Forest/ Geothermal
Intercept	0.65443	0.058187	----	----
Group size	0.21708	-0.013552	0.006716	----
Forest/ Geothermal	-1.0614	-0.028876	-0.000974	0.197901