Biases in estimating population size from an aerial census:

A case study in the Kruger National Park, South Africa

Dedicated to A. M. ("Tony") Starfield on the occasion of his 60th birthday.

J. V. Redfern¹,², P. C. Viljoen³, J. M. Kruger⁴, & W. M. Getz¹

¹Department of Environmental Science, Policy, & Management, 201 Wellman Hall #3112, University of California, Berkeley, CA 94720, USA.

²Corresponding Author (jredfern@nature.berkeley.edu; phone: 510-642-1227; fax: 510-642-7428).

³Environmental & Wildlife Consulting, PO Box 1099, Phalaborwa 1390, South Africa.

⁴Scientific Services, Kruger National Park, PO Box 178, SKUKUZA, 1350, South Africa.

Resubmitted to

South African Journal of Science

October 4, 2002
ABSTRACT

In the Kruger National Park (KNP), South Africa, aerial census data for approximately 15 herbivore species was collected from 1981–1993 using a total area count, strip transect methodology. No estimates of bias or precision error were obtained for the KNP census data. Visibility bias, however, has been shown to be a primary source of error in aerial census data collected using methodologies similar to the KNP’s. Starfield argues that a pragmatic modelling approach can be used to understand the importance of uncollected data and improve data collection strategies. Following this pragmatic approach, we develop a simple, deterministic model to estimate the potential range of bias in the KNP census data. Sources of visibility bias considered in our model include undercounting detected herds and failing to detect small herds. We apply the model to data collected for impala, zebra, wildebeest, and waterbuck, because these species represent a range of potential censusing challenges. The model suggests that visibility bias represents a major source of error in the KNP census data. In particular, the model suggests that visibility bias may confound comparisons among species or comparisons of a particular species’ abundance among years, under different environmental conditions, or in different habitat types.

Key words: aerial census; visibility bias; pragmatic modelling; South Africa; Kruger National Park.
INTRODUCTION

The Kruger National Park (KNP – see Figure 1) is an officially designated long-term ecological research area and has one of the most sophisticated long-term ecological research programs in Africa. One component of this program is an annual, dry season aerial census. The goal of this census is to provide insight into herbivore population trends and the relationship between these trends and environmental variables. From 1981 to 1993, the census used a total area count, strip transect methodology and surveys were conducted throughout most of the park. The census recorded the location and herd size of approximately 15 herbivore species as well as of a range of environmental variables. Although it was acknowledged that the data contained bias and precision errors, no studies were conducted to estimate the size of these errors (pers. obs. PCV).

Visibility bias, resulting from the failure of observers to detect all individuals, may be a primary source of error in an aerial census conducted using a total area count, strip transect methodology. In particular, it has been estimated that a census of wild ungulates may miss 12% to 77% of populations (as summarized by Caughley). Four primary sources of visibility bias have been identified: census methodology, observer effects, environmental variables, and species-specific characteristics.

Controllable sources of bias, such as census methodology and observer effects, were minimized in the KNP aerial censuses through the development of a consistent census methodology including specific rules for environmental conditions under which censuses could be flown, observer training, and the use of multiple observers. We expect, however, that the census data is also subject to uncontrollable sources of visibility bias resulting from environmental variables, such as rainfall, and species-specific characteristics, such as habitat...
Specific problems arising from uncontrollable sources of bias that we expect in the KNP data include missing small groups, undercounting large groups, and missing portions of a group due to the species' habitat preference and/or herd cohesion. KNP censuses conducted from 1999–2002 using a distance sampling technique suggest that these sources of error may bias the census data collected from 1981–1993.

The 1999–2002 KNP censuses followed procedures similar to those used in the 1981–1993 censuses to minimize controllable sources of error. Hence evidence of uncontrollable sources of bias in the 1999–2002 census data suggests that these sources of bias are also present in the 1981–1993 census data. Specifically, the 1999–2002 censuses provide information about how herd size and distance from the observer influence detection probabilities. For many species, data from these recent censuses reveal that detection probabilities decrease for smaller herds and at the edges of the transect (which is the same size as that used in the censuses conducted from 1981–1993). Additionally, the data reveal that these detection probabilities vary between years. Although the 1999–2002 census data suggests that visibility bias is present in the 1981–1993 census data, estimates of bias obtained from the recent data cannot be used to estimate bias in the 1981–1993 census data because distance from the observer was not recorded.

The 1981–1993 KNP census data is assumed to represent trends in population dynamics and has been useful in exploring landscape scale relationships between herbivore distributions, forage condition, and surface water availability. Census data containing unmeasured biased, however, cannot be used to estimate absolute animal abundance. In theory, this unmeasured bias limits the applicability of the census data. For example, any questions regarding the specific number of individuals to harvest or cull cannot be directly addressed because absolute population
sizes are unknown. Starfield\textsuperscript{1} argues that a pragmatic modelling approach can be used to understand the importance of uncollected data and improve data collection strategies. Additionally, Starfield and colleagues have demonstrated the value of using a pragmatic modelling approach, particularly when data are relatively crude or even unavailable, to address critical wildlife management issues\textsuperscript{17,18,19}.

Following Starfield’s pragmatic approach\textsuperscript{1}, we develop a simple, deterministic model to understand how uncontrollable sources of visibility bias affect the applicability of the KNP aerial census data to management and research questions. Our model specifically explores two sources of bias: undercounting detected herds and failure to detect small herds. We apply our model to census data for impala (*Aepyceros melampus*), zebra (*Equus burchelli*), wildebeest (*Connochaetes taurinus*), and waterbuck (*Kobus ellipsiprymnus*). These species were selected to represent a range of potential censusing challenges. The model is used to understand the potential range of bias contained in the census data and how this bias range is affected by factors such as a species' habitat preference, herd cohesion, coloration, and herd size distribution.

\textbf{METHODS}

\textit{Species Characteristics}

We expect sources of bias to affect species differently depending on their habitat preference, behaviour (specifically herd cohesion), coloration, and herd size distribution. Previous research on the habitat preferences and behaviour of impala, zebra, wildebeest, and waterbuck suggests that populations of these species may represent a range of censusing challenges. Impala populations may be composed of large herds and typically occur on the boundary between woodlands and grasslands\textsuperscript{20}. Zebra and wildebeest both occur chiefly in open
Biases in Aerial Census Data

Redfern et al. 2002

grasslands\(^2\), where the wildebeest's dark colouring should increase detection probabilities.

Waterbuck typically occur in riparian habitat where canopy cover may be dense; they also occur in loosely aggregated groups\(^2\).

Differences in relative herd sizes among species were verified for the KNP populations by organizing the census data for the impala, zebra, wildebeest, and waterbuck into a histogram portraying the distribution of herd sizes observed over the study period (1981–1993). Although this histogram is derived from the biased census data, one may assume that the herd size patterns are characteristic of the differences between species. A 5 x 2 contingency test\(^2\) conducted by summing the number of herds detected during the 1981–1993 censuses in size categories 1, 2, 3–5, 6–10, >10 was used to determine whether these herd size distributions were significantly different among species having the same mode herd size, specifically impala, wildebeest, waterbuck. We also used the census data to calculate the mean herd size over the study period for each species.

Statistical Models

The KNP census data is collected as a series of herd size measurements. Analysis of this data, to answer management questions of interest, relies on some form of aggregation of the herd size variable \(y\) including by species, by year, by groups of years organized according to rainfall, etc. Two kinds of bias exist in generating such aggregations: bias in the actual measurements \(x_i\) of the size of the \(i\)-th herd, \(i = 1, \ldots, n\), and omission of herds of various sizes (i.e. we count \(n\) herds but in reality \(n+m\) herds exist so that implicitly we have to correct the values \(x_i = 0, i=n+1, \ldots, n+m\)).
We reconstruct an estimate $y_i$ of the unbiased herd size using an appropriate transform $y_i = f(x_i)$, which accounts for sources of bias that lead to incorrect observations of a herd's size. Several systematic biases may arise in the measurement of $x_i$, $i=1,...,n$. When species reside in areas of high vegetation biomass or cover, it is likely that observers will fail to detect some members of a herd. The same may be true for species that occur in loosely aggregated groups. For these species a simple count of the individuals seen is not adequate to accurately estimate the herd's size; the nearby area must be scanned to ensure detection of the entire herd. Because of the difficulty of detecting single individuals, some members of a diffuse herd may be missed. Additional difficulties may arise in counting large herds. In particular, a stronger than linear increase in the number of individuals missed may occur at larger herd sizes due to factors such as the increased time necessary to ensure an accurate count and the confusion of trying to count a large number of moving individuals.

In general, sources of visibility bias result in an undercount of populations\(^5\,13\). For example, Caughley\(^5\) found that undercounting was a major problem in aerial surveys compared to sources of bias that would tend to inflate the population size estimate and Cogan and Diefenbach\(^13\) found no instances of overcounting in their study. Our model, therefore, only addresses undercounting bias. In particular, we propose that bias in the size of detected herds can be modelling using the transformation:

$$f(x) = x^{1/a}$$

(1)

where $0 < a \leq 1$. This transformation was chosen to model a stronger than linear increase in the number of individuals missed at larger herd sizes. Smaller values of $a$ indicate that higher undercounting bias is expected, particularly at larger herd sizes, for a species.
As mentioned, omission of herds can also bias aggregations of aerial census data and a method must be derived to estimate the number and size of undetected herds. This source of bias is herd size dependent with small herds having a higher probability of omission. Samuel and Pollock\textsuperscript{22} developed a negative exponential detection function (i.e. the relationship between probability of omission and herd size). More recent studies have continued to use an exponential detection function of the form \( D(y) = 1/(1 + \exp(U(y))) \) where \( U(y) \) is some function of the herd size \( y \).\textsuperscript{11,13,14}

The detection function used in this paper generalizes the shape of these previous detection functions by incorporating a shoulder in the omission probabilities for small herds (Fig. 2). This shoulder may be important for species where single individuals and small groups have a similar probability of omission. Specifically, we used the following detection function to calculate the probability of missing a herd of size \( y \):

\[
D(y) = \frac{s}{1 + (\frac{y}{h})^\gamma}
\]

where \( s \) determines the \( y \)-intercept, \( h \) sets the point at which \((1 -(s/2))\) individuals are missed (i.e. the midpoint of the curve), and \( \gamma \) governs the steepness of the curve around this midpoint.

We used equations (1) and (2) to explore the effects of different values of the parameters--\( a, s, h, \) and \( \gamma \)--on population size estimates for impala, zebra, wildebeest, and waterbuck. First, for selected values of \( a \) we used equation (1) to obtain \( y_i \), representing corrected size estimates for all detected herds. The corrected herd sizes were aggregated to form a population histogram. We then applied the detection function, equation (2), to estimate the number of herds missed in each size category. This number was rounded to the nearest integer.
and added to the population histogram of corrected herd sizes. Population size estimates for each species were then calculated from the corrected histograms for each year.

**Parameter Estimates**

Species-specific estimates of the model parameters, $a$, $s$, $h$, and $\gamma$, were derived from discussions with KNP scientists having at least ten years of census experience. Because no measurements of bias were collected during the censuses, this expert knowledge currently provides the only information available to estimate the potential range of bias contained in the census data. For each species, KNP scientists were asked to evaluate how size estimates of detected herds are effected by two sources of visibility bias: 1) bias arising from a species’ habitat preference and behaviour (specifically herd cohesion); and 2) bias arising from the difficulty of counting large herds. KNP scientists expect visibility bias arising from species’ habitat preference and behaviour to have little effect on herd size estimates for species occurring in open habitats, such as zebra and wildebeest. The effect is expected to increase as species, such as impala, occur more frequently in dense cover and reach its greatest impact for species, such as waterbuck, which occur in dense cover and loosely aggregated groups.

KNP scientists expect that visibility bias arising from the difficulty in counting large herds will increase for species in which individuals are small and/or of a light colour. Hence, they expect that this source of bias will have the greatest effect on impala population size estimates, the least impact on wildebeest estimates, with waterbuck and zebra falling in-between. The expectations of KNP scientists, regarding the relative influence of these two sources of visibility bias on size estimates for detected herds, were combined to estimate high and low $a$ values (see Equation 1) for each species (Table 1).
The parameters $s$, $h$, and $\gamma$ determine the shape of the detection function (Equation 2) and hence govern the probability of omission for herds of various sizes. The range of values associated with the parameters $s$ and $h$ were based on each species’ habitat preference and coloration. Higher values were assigned to waterbuck and impala because these species typically occur in dense cover (Table 1). Although zebra and wildebeest both occur in more open habitat, zebra were associated with a slightly higher range of values for the parameter $s$ (Table 1) because of their lighter colouring. The parameter $\gamma$ cannot be automatically described in terms of low and high bias values. A value of $\gamma = 1$ results in a negative exponential detection function while a value of $\gamma = 4$ results in a reverse logistic-like detection function (Fig. 2). Whether a particular curve shape results in high or low bias depends in part on the species’ herd size distribution.

Model simulations used four combinations of the species-specific parameter sets to determine the potential range of bias in the KNP census data for impala, zebra, wildebeest, and waterbuck. Setting all parameters equal to their low bias value (Table 1) simulated a best-case censusing scenario, and provided an estimate of the lower limit of bias contained in the census data for each species. Setting all parameters equal to their high bias value (Table 1) simulated a worst-case censusing scenario and provided an estimate of the upper limit of bias. In both scenarios, values of $\gamma = 1$ and $\gamma = 4$ were tested to determine the effect of the shape of the detection function on these bias limits.

Model simulations also used four additional parameter sets in which parameter estimates were the same for all species (Table 1). This application of the model assumes that the sources of bias have the same effect on all species. This assumption allowed us to derive and compare a non-specific estimate of bias for all species. The lower limit of bias, or a best-case censusing scenario, for this non-specific bias range was determined by setting each parameter equal to the
parameter estimate yielding the lowest bias for any one of the four species (i.e. values of $a$ range from 0.82 – 0.98 for the four species, therefore we used $a = 0.98$ as the non-specific, low bias value) (Table 1). Likewise, the upper limit of bias, or worst-case censusing scenario, for this non-specific bias range was determined by setting each parameter equal to the parameter estimate yielding the highest bias for any one of the four species (Table 1). In both scenarios, values of $\gamma = 1$ and $\gamma = 4$ were tested.

For each parameter set, we applied the model (as explained above) to obtain an estimate of unbiased population sizes for impala, zebra, wildebeest, and waterbuck. Using the model estimate of population size, we also calculated the percentage of the population detected in censuses. Specifically, the model assumes constant bias between years, hence the mean and standard deviation of the percentage of the population detected during the censuses was found using the results from the 13 years of census data. We use the mean percentage of the population detected during the censuses as a diagnostic tool for understanding the magnitude of potential bias error resulting from the different parameter sets.

RESULTS

Species Characteristics

Each species' herd size distribution, calculated by aggregating the census data over the study period (1981–1993), is illustrated in Figure 3. Immediately distinguishable is a marked tendency in the case of impala to estimate herd size to the nearest five individuals (Fig. 3A). Other evidence of this pattern in the herd size distributions, although not as pronounced, occurs at larger herd sizes in the zebra and wildebeest census data (Fig. 3, B and C). Comparison of the herd size distribution curves (Fig. 3) and mean herd sizes (Table 2) indicate that impala are the
only species for which a substantial number of herds are larger than 30. While the mode herd size for impala, wildebeest, and waterbuck is one (Fig. 3, A, C, and D), the 5 x 2 contingency test\textsuperscript{21} conducted by summing the number of herds detected during the 1981–1993 censuses in size categories 1, 2, 3–5, 6–10, >10 reveals that the herd size distributions of these three species are significantly different (P < 0.01). Mode herd size for zebra clusters around 2–4 individuals (Fig. 3B), although the mean herd size is not much larger than that for wildebeest (Table 2).

**Statistical Models**

The results of the model simulations using the non-specific parameter values (Table 1), suggest that during the censuses 45% to 95% of impala, zebra, wildebeest, and waterbuck populations in the KNP may have been counted (Table 3). This estimate of the potential range of bias in the KNP census data falls within the range of bias observed for censuses conducted using total area count, strip transect methodologies\textsuperscript{5}. In general, the estimates of the percent of the population counted were similar between the four species (Table 3). Within the best-case and worst-case censusing scenarios, the shape of the detection function had a small effect on the corrected population size estimates and hence the percentages of populations detected during the censuses (Table 3). In particular, the shape of the detection function had a greater influence on zebra and impala population size estimates compared to wildebeest and waterbuck population size estimates.

Species-specific parameter estimates (Table 1), which take into account species habitat preference, herd cohesion, coloration, and expected herd size distribution, were also used in the model to understand differences in the potential range of bias in the KNP census data for impala, zebra, wildebeest, and waterbuck. Overall, the results of the model simulations using the species-
specific parameters, suggest that during the censuses 49% to 93% of KNP populations may have been counted (Table 3). The results reveal that population size estimates for waterbuck may be the most biased, followed by impala, zebra, and wildebeest (Table 3). The effect of the shape of the detection function was similar to that observed for the model simulations using the non-specific parameter estimates – generally, greater differences were observed for zebra and impala compared to wildebeest and waterbuck.

Corrected population size estimates, derived using the species-specific parameters, are shown for impala, zebra, wildebeest, and waterbuck in Figure 4. For all species the population size distribution each year is skewed with the lower bound of the distribution determined by the observed census data. The model results generally follow the trends in the census data (Fig. 4) as a direct result of the model assumption that bias is constant between years.

DISCUSSION

Our results suggest that visibility bias represents a major source of error in the KNP census data. For example, observers may detect between 51% and 78% of impala populations (Table 3). These percentages translate into large differences in the actual size of the impala population. Specifically, the average population size derived from the 1981–1993 census data is 119 000 while the average model population size varies between 152 000 and 232 000 depending on whether high or low bias estimates of the parameters are used. Similar differences between the population sizes estimated during the census and the unbiased population sizes predicted by the model occur for all species (Fig. 4).

The potential range of bias in the KNP aerial census data affects the applicability of the data to management and research questions. Model simulations using the species-specific
parameter estimates suggest that caution must be used when relying on census data containing unmeasured bias to make comparisons among species. In particular, the differences in bias among species indicated by our model suggest that comparisons of species’ densities, population growth characteristics, and viability may be confounded.

Furthermore, comparisons of a particular species’ abundance under different environmental conditions or in different habitat types may also be confounded by unmeasured sources of visibility bias. While it is possible that observed differences in a species’ abundance are caused by ecological factors, these differences may also be the result of non-constant bias in the data. For example, visibility bias may vary if certain environmental conditions are correlated with the aggregation of individuals of a particular species or of multiple species around important resources (e.g. water sources). Large congregations may increase detection probabilities but decrease counting accuracy. Additionally, increased vegetation coverage during years with above average rainfall or in particular habitat types may correspond to a worst-case censusing scenario while decreased vegetation coverage corresponds to a best-case censusing scenario. Our models suggest large differences in the percentage of the impala, zebra, wildebeest, and waterbuck populations detected during best-case and worst-case censusing scenarios. Specifically a minimum difference of 13% is predicted for the four species using the species-specific parameter estimates (Table 3) and 41% is predicted using the non-specific parameter estimates (Table 3). Hence, unmeasured differences in visibility bias may mask large changes in population sizes from year to year or large differences in population sizes among habitat types.

The applicability of aerial census data to research and management questions will be limited until we are able to objectively estimate the amount of bias contained in the data, and
hence derive more accurate population size estimates for the species counted. Considerable progress in dealing with sources of visibility bias has been made using line transect methodologies. Line transect methods do not assume that all individuals or groups are detected and statistical models for understanding the detection function and its relation to herd size have been proposed\textsuperscript{15,23,24}. A key assumption in line transect analyses, however, is that all individuals are detected on the line. While it can be expected that biases are minimized on the line, uncontrollable sources of bias may cause deviations from this key assumption. For example, the KNP censuses conducted from 1999–2002\textsuperscript{15} using a distance sampling technique reveal that the highest detection probabilities for certain species are not achieved on the line (i.e. 0–50 m), but rather 50–100 m from the line. While the data can be truncated so that estimation of the detection function does not include the lower detection probabilities observed on the line, it is important to determine whether the truncated data are biased if management actions rely on census data.

Regardless of whether line or strip transect methodologies are used in an aerial census, a further investment of resources is needed to collect additional data that can be used to obtain objective estimates of bias error. Several studies have used the detection of herds containing individuals fitted with radio-transmitters to develop bias corrections for undercounting detected herds and failing to detect small herds\textsuperscript{10,11,13,14}. In general, a second set of observers locates these herds aerially or from the ground and obtains an accurate count of the herd sizes just prior to the start of the aerial census. Statistical analyses can then be used to determine the relationship between the aerial census estimates of the size of detected herds containing individuals fitted with radio-transmitters and the actual size of each herd. These analyses could be used to parameterise the transformation proposed in this paper (Equation 1) for modelling bias in the size of detected herds.
Likewise, statistical analyses comparing the actual number and size of herds containing individuals fitted with radio-transmitters to detection of these herds in the aerial census can be used to determine the relationship between herd size and the probability of omission. For example, Samuel et al. 11 used logistic regression to assess the relationship between the detection of elk herds and visibility factors including herd size and vegetation cover. Statistical analyses could also be used to parameterise the detection function proposed in this paper (Equation 2). Unbiased population size estimates can then be obtained from Monte Carlo simulations that correct for undercounting detected herds and failing to detect small herds using the observed statistical relationships.

Obtaining objective estimates of bias error may require a significant investment of additional resources. Eberhardt et al. 14 point out that a sizable sample of herds containing individuals with radio transmitters is needed to estimate potential biases. Additionally, objective estimates of bias error must be collected for the total range of habitat types and environmental conditions that potentially influence the population size estimates derived from the census data to account for possible variations in visibility bias. However, investment in this data collection will be justified by the increased utility of current and historic census data. Specifically, census data in which bias has been estimated can be used to address questions concerning absolute animal abundance, such as the number of individuals to harvest or cull. Estimates of bias error obtained under conditions similar to those encountered during historic censuses may be used to derive unbiased population size estimates and hence extend the data sets available to address management and research questions.

The best investment of resources can be obtained by understanding the relationship between visibility bias and species characteristics such as habitat preference, herd cohesion,
coloration, and herd size distribution. Results from our models, derived from the species-specific parameter ranges suggested by KNP scientists with at least ten years of census experience, suggest that the greatest return on an investment in additional data collection will be achieved by obtaining estimates of bias error for species that prefer dense vegetation cover, occur in loosely aggregated herds, and/or have light coloration. Additionally, our results illustrate the importance of designing data collection methods to determine bias in undercounting detected herds for species that typically occur in large herds. For species in which the expected mode herd size is 1–2 individuals, data collection should be designed to determine omission probabilities for small herds. Finally, our results demonstrate the increasing importance of estimating the shape of the detection function for species whose expected herd size distribution is characterized by a larger mean herd size and a flatter distribution. In particular, determining the shape of the detection function is important because a negative exponential curve implies a higher probability of missing larger herds than a reverse logistic-shaped curve.

CONCLUSION

McNaughton and Campbell\(^2\) identify three important uses of long-term ecological research. They suggest that this research documents subtle, chronic ecosystem changes, signals the onset of acute changes, and can be of use in designing management strategies. In African savanna ecosystems, large investments have been made in long-term ecological monitoring. As early as the 1950's, programs were designed to collect baseline ecosystem data and determine geographical distributions and population sizes of large mammal species\(^2\). Monitoring programs conducted in National Parks are financed in part by African governments and usually provide a foundation for management programs\(^2\).
The Kruger National Park (KNP), located in South Africa, has one of the most sophisticated long-term ecological research programs in Africa. Our model, however, suggests that the unmeasured, systematic error contained in the aerial census data for KNP herbivores limits the applicability of the data to research and management questions. To obtain the data necessary to correct historic and current population size estimates, further investments in calibration censuses are required. Wise investment in calibration censuses can be achieved by understanding the important sources of bias to measure for the species of interest. The pragmatic modelling approach applied in this paper illustrates a relationship between the relative importance of different sources of bias error and a species' characteristics, such as habitat preference, behaviour (specifically herd cohesion), coloration, and expected herd size distribution. Incorporating an understanding of this relationship in survey design can help researchers balance study objectives with the resources allocated for data collection. In the long-term, current investments in calibration censuses will be justified by the increased utility of census data. In particular, the development of calibration censuses that allow us to estimate the bias error contained in historic census data will significantly extend the time series of data available for ecological research.
Biases in Aerial Census Data

Redfern et al. 2002

ACKNOWLEDGEMENTS: We wish to thank Chris Redfern and anonymous reviewers for helpful comments on the manuscript. We are sincerely grateful to Andre Potgieter for useful discussions about the species-specific parameter values. We have also benefited from discussing KNP census techniques with Harry Biggs, Rina Grant, Danie Pienaar, Angela Gaylard, Theresa Davidson, and Naledi Mare. NSF Grant INT-9731202 to WMG supported J.V. Redfern’s research; NSF Grant DEB-0090323 to WMG partially supported W.M. Getz’s research.

LITERATURE CITED


Table 1. Species-specific and non-specific parameter values for worst-case (W) and best-case (B) censusing scenarios. Species-specific parameter values were derived through discussions with KNP scientists. Parameter values reflect the scientists’ understanding of differences in expected census bias resulting from a species' coloration, habitat preference, herd size distribution, and herd cohesion. The non-specific parameters were used in simulations for all four species to derive and compare a non-specific bias estimate. Non-specific high and low bias values for $a$, $s$, and $h$ were defined as the maximum or minimum values of the species-specific parameter values assigned to any one of the four species.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Impala</th>
<th>Zebra</th>
<th>Wildebeest</th>
<th>Waterbuck</th>
<th>Non-specific</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W</td>
<td>B</td>
<td>W</td>
<td>B</td>
<td>W</td>
</tr>
<tr>
<td>$s$</td>
<td>0.75</td>
<td>0.4</td>
<td>0.5</td>
<td>0.3</td>
<td>0.45</td>
</tr>
<tr>
<td>$h$</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$a$</td>
<td>0.85</td>
<td>0.93</td>
<td>0.92</td>
<td>0.97</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Table 2. Species-specific herd size mean and standard deviation (SD) for the 1981–1993 KNP census data.

<table>
<thead>
<tr>
<th>Species</th>
<th>Mean Herd Size ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impala</td>
<td>12 ± 16.4</td>
</tr>
<tr>
<td>Zebra</td>
<td>8 ± 10.4</td>
</tr>
<tr>
<td>Wildebeest</td>
<td>7 ± 9.6</td>
</tr>
<tr>
<td>Waterbuck</td>
<td>4 ± 4.4</td>
</tr>
</tbody>
</table>
Table 3. Estimates of the percent of the population detected under best-case and worst-case censusing scenarios for $\gamma = 4$ and $\gamma = 1$ are shown for simulations using the species-specific and non-specific parameter values. The model was used to estimate the unbiased population size for each year of census data. We then calculated the percentage of the population that was detected in the census. The model assumes constant bias between years, hence the mean and standard deviation (SD) of the percentage of the population detected were found using the results from the 13 years of census data.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$\gamma$</th>
<th>Impala Mean (SD)</th>
<th>Zebra Mean (SD)</th>
<th>Wildebeest Mean (SD)</th>
<th>Waterbuck Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst-case</td>
<td>4</td>
<td>54 (1.1)</td>
<td>76 (0.6)</td>
<td>79 (0.5)</td>
<td>49 (1.2)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>51 (0.7)</td>
<td>72 (0.2)</td>
<td>77 (0.1)</td>
<td>49 (0.3)</td>
</tr>
<tr>
<td>Best-case</td>
<td>4</td>
<td>78 (0.7)</td>
<td>92 (0.3)</td>
<td>93 (0.4)</td>
<td>77 (0.4)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>75 (0.5)</td>
<td>89 (0.1)</td>
<td>92 (0.1)</td>
<td>74 (0.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Impala Mean (SD)</td>
<td>Zebra Mean (SD)</td>
<td>Wildebeest Mean (SD)</td>
<td>Waterbuck Mean (SD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48 (1.3)</td>
<td>51 (1.1)</td>
<td>48 (0.8)</td>
<td>49 (1.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45 (1.0)</td>
<td>48 (0.8)</td>
<td>47 (0.4)</td>
<td>49 (0.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>94 (0.3)</td>
<td>95 (0.2)</td>
<td>93 (0.4)</td>
<td>94 (0.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>92 (0.1)</td>
<td>92 (0.1)</td>
<td>92 (0.1)</td>
<td>92 (0.3)</td>
</tr>
</tbody>
</table>

* Worst-case (W) and best-case (B) values of the parameters ($s, h, a$) are given in Table 1.
Figure 1. A map of southern Africa showing the study site, Kruger National Park.

Figure 2. An example of the detection function (Equation 2) used in this paper is plotted for $s = 0.6$, $h = 4$, and three values of $\gamma$. The assumption that single individuals and small groups have similar omission probabilities while large groups have negligible omission probabilities can be modelled using larger values of $\gamma$, which produce a reverse logistic-shaped curve.

Figure 3. Herd size distributions are plotted as histograms for each species by aggregating the KNP aerial census data collected from 1981–1993. The last category in the histogram (denoted by *) represents the number of herds containing 100 or more individuals.

Figure 4. Population size estimates, derived from the aerial census data (dashed black line) and from the results of the model simulations using the species-specific parameter values for best-case (solid grey line) and worst-case (solid black line) censusing scenarios (Table 1), are plotted annually during the study period. For all species, best-case censusing scenarios occur when $\gamma = 4$. Worst-case scenarios, however, for impala, zebra, and wildebeest occur when $\gamma = 1$, while worst-case scenarios for waterbuck occur when $\gamma = 4$. 

24
Study Site: Kruger National Park

Biases in Aerial Census Data

Redfern et al. 2002

Figure 1
Figure 2

- Graph showing the probability of omission against herd size.
- The graph includes three lines, each labeled with different values of $\gamma$: $\gamma = 1$, $\gamma = 8$, and $\gamma = 4$.
- The x-axis represents herd size, ranging from 1 to 15.
- The y-axis represents the probability of omission, ranging from 0 to 0.6.

The graph illustrates how the probability of omission decreases as herd size increases, with different slopes corresponding to the different values of $\gamma$. The curves for $\gamma = 1$, $\gamma = 8$, and $\gamma = 4$ show distinct trends, with $\gamma = 1$ having the steepest slope and $\gamma = 8$ having the shallowest.
Biases in Aerial Census Data

Figure 3

A) Impala

B) Zebra

C) Wildebeest

D) Waterbuck

Log(Herd Size + 1)
Figure 4

A) Impala

B) Zebra

C) Wildebeest

D) Waterbuck

Population Size Estimate (Thousands)

Year

1981 1993