

Using species traits to predict detectability of animals on aerial surveys

S. SCHLOSSBERG,^{1,3} M. J. CHASE,¹ AND C. R. GRIFFIN²

¹*Elephants Without Borders, P.O. Box 682 Kasane, Botswana*

²*Department of Environmental Conservation, University of Massachusetts Amherst, 160 Holdsworth Way, Amherst, Massachusetts 01003 USA*

Abstract. In animal surveys, detectability can vary widely across species. We hypothesized that detectability of animals should be a function of species traits such as mass, color, and mean herd size. We also hypothesized that models of detectability based on species traits can be used to predict detectability for new species not in the original data set, leading to substantial benefits for ecology and conservation. We tested these hypotheses with double-observer aerial surveys of 10 mammal species in northern Botswana. We combined all 10 species and modeled their detectability with species traits (mass, mean herd size, color) as predictors while controlling for observer effects, vegetation, and herd size. We found support for effects of mass and an interaction between herd size and mean herd size on detectability. This model accurately predicted the ratio of herds detected by two observers vs. one observer for 8 of 10 species. To test whether a model based on species traits could be applied to a new species, we serially deleted each species from the data set, fit a trait-based model to the remaining nine species, and used this model to predict detectability for the deleted species. The model was able to reproduce the species-trait model for seven species and accurately predicted the ratio of detections by one or two observers for a different set of seven species; the model was successful by both measures for five species. To our knowledge, this represents the first time that a mechanistic model for detectability of animals has been used to predict detectability for new species. Prediction failed for species with extreme values of traits, suggesting that predicting detectability is not possible near or beyond the boundaries of one's data set. The approach taken in this paper can potentially be used with a variety of taxa and may provide new opportunities to apply detectability corrections where they have not been possible before.

Key words: aerial survey; body size; Botswana; coloration; detectability; detection probability; double observer; group size; habitat; herd size; large mammals; mass.

INTRODUCTION

Accurate estimates of population size are crucial for ecology and conservation (Gu and Swihart 2004, Rodrigues et al. 2006). Because observers often fail to detect animals on surveys, detectability corrections are widely employed in ecology today (Buckland et al. 2001, MacKenzie et al. 2002, Royle 2004, Guillera-Aroita 2017). Multi-species studies have shown that detectability varies greatly from species to species on animal surveys (Kéry and Schmid 2006, Kéry and Plattner 2007, Johnston et al. 2014). Little, however, is known about underlying causes of interspecific variation in detectability, primarily because the vast majority of studies seeking to account for missed animals treat detectability as a nuisance parameter rather than a subject of interest.

Given the limitations of human perception, detectability should be a function of measurable species traits such as

body size, color, group size, loudness, habitat, or calling frequency (Aldredge et al. 2007, Anderson et al. 2015). A few studies have examined correlations between species' detectabilities and traits such as body size or taxonomic group (Fletcher et al. 2006, Johnston et al. 2014). These studies have generally estimated detectability separately for each species and then examined relationships with species traits or taxonomic groups. Here, we suggest an alternative approach. Rather than estimating detectability separately for each species, we suggest that the detectability of species assemblages should be modeled explicitly as a function of their traits. Accordingly, detectability should be predictable with knowledge of species traits. A recent study by Garrard et al. (2013) showed that interspecific variation in detectability of plants was largely explained by species traits such as abundance and whether or not they were flowering. The success of this approach with plants led us to believe that a similar approach could work in animals.

Developing trait-based models of detectability for animals could have two key benefits. First, explicitly modeling mechanisms underlying detectability could reveal new information about the detection process. This could improve animal surveys by revealing factors that limit

Manuscript received 16 November 2016; revised 2 September 2017; accepted 13 September 2017. Corresponding Editor: Dr. Aaron Wirsing.

³E-mail: xanthocephalus@gmail.com

detectability and focusing observers' training on those factors. For instance, if observers frequently miss light-colored species, training could focus on such species. Second, modelling detectability as a function of species traits could allow us to predict detectability for new species or locations. Such predictive models of detectability could increase the efficiency of surveys in terms of time, money, sample sizes, or logistics (Garrard et al. 2013). N-mixture and occupancy models, for instance, generally require multiple samples at each location within seasons as well as relatively large numbers of observations per species (MacKenzie and Royle 2005, Yamaura et al. 2016). These requirements may preclude the use of detectability corrections for rare species or those that are hard to detect or capture (Banks-Leite et al. 2014). If researchers can predict detectability, we could make sampling more efficient by incorporating previous research and applying corrections from well-validated models based on large sample sizes rather than generating correction factors de novo in each study.

Here, we use African mammals to test whether detectability can be predicted as a function of species traits. We used double-observer aerial surveys to model detectability of 10 mammal species. African mammals provide an excellent study system for this question because they include a diverse assemblage of species that varies widely in traits such as size, color, and group size. Also, past studies suggest that detectability varies substantially across African mammal species (Caughley 1974). Our goals were (1) to develop simple models to predict the detectability of mammals based on easily measured traits and (2) to determine if such models can be applied to new species not in the original data set. To our knowledge, no study of animals has tried to predict detectability purely as a function of species traits or apply such a model to new species; Garrard et al. (2013) showed that this is possible for plants. Thus, our study was intended to be a proof of concept, not an exhaustive exploration of all factors affecting detectability.

METHODS

Study area

We conducted our study in the Okavango Delta of northern Botswana. This area has a diversity of habitats ranging from open water and emergent marsh to woodlands and shrublands. Most of the area is protected as game reserves or wildlife management areas, and large populations of many medium-sized and large mammals reside here. We conducted surveys for mammals on five study areas (range 236–545 km²), each either all or part of a concession used for ecotourism.

Survey methods

We used double-observer aerial surveys to estimate detectability of African mammals. This method involves

two observers on the same side of the aircraft making independent observations on the same survey strip (Magnusson et al. 1978). By analyzing the frequency of observations made by the front, rear, and both observers, one can estimate detectability and model effects of covariates on it.

We flew double-observer surveys in a GippsAero GA8 Airvan configured with four rows of two seats. The pilot and the front recorder sat in the first row, followed by two "front" observers, two "rear" observers, and the rear recorder in the final row. We hung opaque cloth behind the front observers to prevent rear observers from cueing on head movements of front observers. The front observers and recorder had a separate headphone/intercom system from the rear observers and recorder to prevent front and rear observers from hearing one another speaking.

We flew surveys over four mornings, from 28 to 31 July 2014. This was during the dry season in Botswana when deciduous trees are leafless, allowing good visibility on aerial surveys. The pilot flew along parallel transects 2 km apart and was instructed to maintain groundspeed of 180 km/h. To control for positional effects, observers rotated seat positions between days, so each observer sat in each of the four seats for one day of surveys. Three of the four observers had >100 h previous experience conducting aerial surveys in southern Africa. The fourth observer was highly experienced in mammal identification but had not conducted aerial surveys before the study. This observer received 5 h of airborne training in survey techniques before data collection began.

Observers were instructed to count all medium and large mammals, including any species at least the size of a warthog (*Phacochoerus africanus*), seen in a strip demarcated by two metal wands attached to each wing strut. We calibrated each strip to be 200 m wide at the target flight altitude of 91 m. Observers called out the species and number of animals for each herd observed, and the recorder transcribed the data and used a GPS receiver to mark the location and time. We used digital voice recorders, connected to the intercom system, as a backup to the data sheets. For each observer, we mounted a 35-mm digital camera in the aircraft window and aimed it at the observer's survey strip. Cameras were equipped with a remote shutter, and we instructed observers to take photographs of each herd observed.

ANALYSIS

Data processing

We first aligned the independent front and rear observations to determine whether one or both observers detected each herd. Photographs were available for 85% of herds observed. For these herds, we simply used the photographs to align observations. Because photographs were necessary to identify vegetation types, we discarded from the analysis herds that were not photographed by at least one observer.

Observers in a fast-moving aircraft often err in estimating the number of animals in a herd (Norton-Griffiths 1978). Thus, we used the photographs taken by observers to correct herd-size estimates. We first used each observer's photograph(s) to correct the observer's own estimates. We counted the number of animals visible in the photograph(s) and used that number as the herd size with two exceptions. First, for recorded herd sizes of six or fewer, we assumed that observers were accurate and did not reduce estimates below the observers' estimate. Second, we did not reduce an observer's herd-size estimate for any herd unless we were confident that all animals were in view in the photograph and none had been left out of the photo or occluded by vegetation or other animals. For herds seen by both observers, the final estimate of herd size depended on the availability of photographs. If a photograph was available for only one observer, we used that observer's corrected estimate. If both observers had photographed a herd, we examined both photos to identify the number of unique individuals visible. Because front and rear observers' photographs were often taken from disparate angles, the total number of individuals in the photographs was sometimes greater than either individual estimate.

Modeling strategy

To model detectability, we used the conditional likelihood model developed by Huggins (1989, 1991), as implemented in Program MARK (White and Burnham 1999). This model uses maximum likelihood to estimate detectability by conditioning on the number of herds observed. The model treats detection probability as a linear-logistic function of covariates, as discussed below. Our unit of analysis was the herd rather than the individual animal because group-forming animals are detected as a herd rather than individually (Buckland et al. 2001). We arbitrarily included only species with at least 20 herd observations in our analyses.

Our models included three species traits: (1) color; (2) mass, and (3) mean herd size for the species. Color may be important if animals with darker coloration are easier to see against the light-colored soils in northern Botswana. Based on Jachmann (2001), we assigned each species in our data set a score from 1 (pale sandy) to 14 (black). Mass was included as a covariate because, all else being equal, larger animals should be more visible to an aerial observer than smaller animals (Fletcher et al. 2006, Berthiaume et al. 2009). We took species' masses from Jachmann (2001) and Skinner and Chimimba (2006). Because masses varied over three orders of magnitude, we log-transformed values for analysis. We used mean herd size as a covariate because we expected that the effect of herd size may be different for species that form small or large herds. For instance, detectability of species that tend to form large herds may be insensitive to herd size, but detectability of species occurring in small herds may be more responsive to herd size. Previous research by Jachmann (2002) showed

that mean herd size was a good predictor of species' detectabilities. In our study, herd sizes differed substantially among species, as noted in other studies of African mammals (Burger and Gochfeld 1994, Jachmann 2002), so mean herd size is a meaningful species trait. To calculate this covariate, we simply took the mean of the photo-corrected herd sizes for each species in the data set.

Though our goal was to model species traits, past research suggested that two herd-level traits, the number of animals in a herd and the type of vegetation around the herd, would also affect detectability. Herd size is consistently the most important factor affecting detectability of group-forming species on aerial surveys (Samuel et al. 1987, Graham and Bell 1989, Ransom 2012, Schlossberg et al. 2016). Research has also shown that detectability of mammals decreases as vegetation density increases (Bayliss and Yeomans 1989, Anderson and Lindzey 1996, Cogan and Diefenbach 1998). All species encountered on our surveys in Botswana occur in herds, and vegetation in the Okavango Delta is highly heterogeneous, so including herd size and vegetation in our models was necessary to accurately estimate detectability. In our models, therefore, detectability for an individual herd was a function of both species-level and herd-level traits. We treated herd size as a continuous covariate in our models. Our models also included interaction terms between herd size and species traits because we predicted that trait effects may differ for small and large herds. For instance, small herds of light-colored animals may be harder to detect than small herds of dark animals, but the two colors may be equally detectable when herds are large. Vegetation was a categorical covariate in our models. We created a simple classification system for vegetation in the Okavango Delta and assigned each herd photographed in the study to one of eight categories (Table 1). To do this, we first assigned each individual animal in the photograph to the dominant vegetation type within one body length of the animal; this small scale seemed most relevant for detectability on aerial surveys. The vegetation category for the herd, then, was the modal vegetation category for the individual animals. In cases of ties between vegetation categories, we used the vegetation

TABLE 1. Vegetation categories on double-observer aerial surveys, Botswana, 2014.

Category	Definition
Water	open water
Bare ground	no vegetation
Low grass	leaving elephant legs at least partially exposed
Open shrub	woody plants up to height of adult; canopy cover <50%
Open tree	woody plants taller than adult; canopy cover <50%
Tall grass	completely covering the legs or taller
Closed shrub	woody plants up to height of adult; canopy cover >50%
Closed tree	woody plants taller than adult; canopy cover >50%

category deemed more open (nearer the top of Table 1). Undoubtedly, many alternative schemes could be developed to classify vegetation around herds, but we were interested in a simple and reproducible method. Because the sample size for the “closed tree” category was only 10 herds, we combined the closed tree and closed shrub categories for analysis. Preliminary analyses suggested that detectabilities for these two categories were similar (S. Schlossberg, *unpublished data*). Photographs and, therefore, vegetation classifications, were unavailable for some herds observed in the study. Because the Huggins model cannot account for missing data, our dataset included only herds that were photographed.

Besides species and herd traits, characteristics of observers or survey aircraft can also affect detectability on aerial surveys (Conroy et al. 2008, Koneff et al. 2008). Observer effects were not the focus of our study, but excluding them could lead to inaccurate predictions. A detailed study of elephants (Schlossberg et al. 2016) and exploratory analyses of the multi-species data set used in this study revealed that three covariates related to observers influenced detectability. First, one observer had a higher detection probability than the other three observers. Second, detection probabilities were higher in the rear row of seats than in the front row of seats in the Airvan (see Schlossberg et al. 2016). Third, detectability increased with a species’ abundance, likely due to observers’ search image improving as they repeatedly observe a species. The measure of abundance used in our study was the density of each species (in herds/km²) estimated via helicopter total counts conducted on each study area a week prior to the double-observer samples. We expected that the quality of an observer’s search image would be a function of their recent experience with a species as well as the number on the current study area. Consequently, we used a cumulative measure of abundance rather than separate measures for each study area. For each area, the abundance measure used in the model was the species’ mean density on that study area and all study areas previously surveyed in the double-observer study. Thus, if a species’ density was 1 herd/km² in the first study area surveyed, the covariate value was 1 for that study area. If the same species had density of 2 herds/km² in the second study area, the covariate value for that study area was 1.5.

To simplify model selection, we included these three observer effects in each model we tested. For the set of species-trait models described below, the top model lacking observer identity, row, and herd density had $\Delta\text{AIC}_c = 36$ (where ΔAIC_c is the difference in the Akaike information criterion corrected for sample size between the model and the top model). Thus, we felt justified in including observer effects as a baseline for all models.

Species-trait model

To determine how well species-level traits predict detectability, we analyzed the full data set with all 10 species to create what we term the “species-trait model.”

Variables for predicting detectability were the three species-level covariates (color, mean herd size, and mass), the vegetation category and herd size for each observation, interactions between herd size and each of the three species-level covariates, and the three observer effects. The full model, including all species-level and herd-level covariates took the form

$$\begin{aligned} \text{logit}(p_{ijkl}) = & \alpha + \beta_1(\text{color score})_i + \beta_2(\text{mean herd size})_i \\ & + \beta_3(\log[\text{mass}])_i + n_j\gamma_1(\text{color score})_i \\ & + n_j\gamma_2(\text{mean herd size})_i + n_j\gamma_3(\log[\text{mass}])_i \\ & + n_j\delta + \mathbf{v}_j\boldsymbol{\tau} + \omega_1 O_{jl} + \omega_2 R_{jl} + \omega_3 \rho_{ik} \end{aligned} \quad (1)$$

where p is detection probability, i indexes species, j indexes individual herds, k indexes study areas, l indexes observer, α is a global intercept, β indicates main effects of species traits, γ indicates interactions between species traits and herd size, n , and δ indicates the global intercept for the herd-size effect. For vegetation effects, \mathbf{v}_j is a 1×6 row vector of indicator variables equaling 1 for the observed vegetation category for herd j and 0 for all others, and $\boldsymbol{\tau}$ is a 6×1 column vector of coefficients for the vegetation category effects. “Bare ground” was the reference category for the vegetation coding. For observer effects, O_{jl} is an indicator variable for whether or not observer l was an observer for herd j , R is an indicator variable for whether or not observer l was in the rear row of seats, ρ_{ik} is the cumulative density of herds of species i for surveys up to and including study area k , and the ω terms are coefficients for those effects. When traits for species i are substituted into the equation, the parameters describe seven curves, one for each vegetation type, relating detectability and herd size, with intercepts = $\alpha + \beta$ terms + $\boldsymbol{\tau}$ and slope = $\delta + \boldsymbol{\gamma}$ terms, which are linear on a logit scale.

We used MARK to run models with all possible subsets of the eight covariates of interest, three species traits, three interactions between species traits and herd size, herd size alone, and vegetation, for a total of 256 models. As mentioned above, each model also contained the three observer effects. We ranked models by AIC_c and then removed models with uninformative parameters. Specifically, we removed any model with $\Delta\text{AIC}_c > 0$, for which the top model was a nested subset of that model; the additional parameters in that model did not improve model likelihood sufficiently to be considered informative (Arnold 2010). Because there was substantial model-selection uncertainty, we used Akaike weights to restrict models to a 90% confidence set and then used model-averaging on predicted detectabilities, not model parameters, to make inferences. This method avoids pitfalls that can occur when model-averaging parameters (Cade 2015). All analyses besides MARK models were run with R version 3.3.2 (R Core Team 2016); see Data S1 for data and Data S2 for all R code.

We used two methods to test how well the species-trait model fit the data. First, we used a parametric bootstrap

to compare the deviance of the species-trait model against that of 1,000 data sets simulated from model parameters (see MacKenzie and Bailey 2004). Second, for each species, we computed what we call the “detection ratio”: the number of herds detected by two observers divided by the number detected by one observer. We then compared the actual detection ratio against the prediction from the species-trait model. We used this method because directly comparing modeled detectability against raw data was not possible. The raw proportion of herds detected overestimates detectability because some herds are missed by both observers. Instead, we took advantage of the fact that the relative number of herds detected by one vs. two observers should be predictable based on a multinomial distribution. If a herd is detected with probability p_1 by the front observer and p_2 by the rear observer, the probabilities of 0, 1, and 2 detections are $(1 - p_1)(1 - p_2)$, $p_1(1 - p_2) + p_2(1 - p_1)$, and p_1p_2 , respectively. The expected ratio of herds detected by two observers to herds detected by just one observer is, therefore, $p_1p_2 / (p_1(1 - p_2) + p_2(1 - p_1))$. To calculate the detection ratio predicted by the species-trait model, we first estimated front and rear detectability for each observed herd by plugging observation-specific covariates (herd size, species traits, observer, etc.) into the species-trait model. We then calculated the mean of the numerator and denominator of the above expression for each species. The quotient of those two means was the predicted detection ratio for that species. We calculated standard errors of model-predicted detection ratios with the delta method. We considered the predicted detection ratio for a species to be accurate if the actual ratio calculated from raw data was within the 95% confidence interval of the predicted ratio.

Deleted model

The second goal of our study was to determine if models based on species traits can predict detectability for a new species, as might happen when a model is applied to a rare or cryptic species with insufficient observations for analysis. We used a type of training/testing procedure to determine whether species traits can predict detectability for new species. We serially deleted each species from the data set and then repeated model fitting with the truncated data set as described above for the species-trait model. We used the resulting models to predict detectability for the deleted species based on its traits. As above, we made inferences based on the 90% confidence set of models. Accordingly, for each model in the confidence set, we used the traits of the deleted species to predict its detectability in each vegetation type over that species' observed range of herd sizes. We then model averaged the predictions by vegetation type and herd size to produce a single detectability estimate for each combination of vegetation type and herd size. Finally, for each herd size, we combined the estimates for the seven vegetation types by taking a weighted mean of the estimates, with weights equal to the proportion of observed herds in each

vegetation type. We repeated this process for all 10 species and then aggregated the predictions in what we term the “deleted model.” To validate the deleted model, we used the two methods described above: parametric bootstrap to measure goodness of fit and comparison of predicted and actual detection ratios. Because the deleted model was fit separately for each species, we ran the parametric bootstrap separately for each species.

As an indicator of how well the deleted model reproduced the species-trait model, we compared the parameters of the two models. This comparison was complicated by the fact that we did not compute model-averaged parameters for either model. As described above, Eq. 1 describes a series of lines, on a logit scale, relating herd size and detectability for each vegetation type. Thus, to compare parameters between the deleted and species-trait models, we computed the logit-scale intercepts and slopes for the relationship between herd size and detectability for each species in each model. We then compared these parameters between the species-trait and deleted models. To illustrate this procedure, we use the case of African buffalo in the species-trait model as an example. The 90% confidence set for the species-trait model included 30 models. For each of the 30 models, we predicted the detectability of buffalo for herd sizes from 1 to 118, the maximum herd size observed, for each of the seven vegetation categories, for a total of 24,780 predictions. For observer effects, we averaged detectabilities between the front and rear rows, used the values for observers 2, 3, and 4 (not observer 1), and used the mean density of buffalo across all strata as the covariate for the abundance effect. We made predictions on the scale of the linear predictor, with logit-transformed detectability values, as in Eq. 1. For each combination of herd size and vegetation category, we then model-averaged the predicted detectabilities, which gave us 826 predicted values, one for each possible combination of herd size and vegetation type. This was equivalent to one line relating herd size and logit-transformed detectability for each vegetation category. Finally, we used linear regression to calculate the slope and intercept of the line for each vegetation category. Because vegetation did not affect slope in our models, slopes were nearly identical between vegetation categories, and we averaged slopes over the seven vegetation categories. The final parameters for buffalo for the species-trait model were a single slope for the herd size effect and seven intercept values, one for each of the seven vegetation types. We repeated these calculations for the other nine species for the species-trait model and for all ten species with the deleted model, and we plotted the results by species to determine how well the deleted model matched the species-trait model.

RESULTS

Over 4 days of double-observer sampling, we recorded a total of 973 herds for the 10 species with at least 20 observations. Vegetation usage was available for

TABLE 2. Variables used in species-trait models and number of herds observed on double-observer aerial surveys, Botswana, 2014.

Species	Scientific name	Mean herd size	Mass (kg)	Color score	No. herds observed
Buffalo	<i>Syncerus caffer</i>	12.6	486	14	32
Elephant	<i>Loxodonta africana</i>	5.5	2338	12	215
Giraffe	<i>Giraffa camelopardalis</i>	3.3	1045	8	24
Hippo	<i>Hippopotamus amphibius</i>	2.4	1465	12	62
Impala	<i>Aepyceros melampus</i>	14.1	49	7	82
Kudu	<i>Tragelaphus strepsiceros</i>	4.0	205	4	26
Lechwe	<i>Kobus leche</i>	7.6	103	9	263
Tsessebe	<i>Damaliscus lunatus</i>	4.2	132	10	23
Warthog	<i>Phacochoerus africanus</i>	2.7	67	13	34
Zebra	<i>Equus quagga</i>	10.2	316	11	38

Note: The number of herds observed includes only herds photographed by the observers. Color score is based on Jachmann (2001), ranging from 1 (pale sandy) to 14 (black).

799 herds, which was the sample used in our analyses (Table 2). For the species-trait model, after removing models with uninformative parameters, the top model had 49% of the Akaike weight in the full set of models. Thus, we made inferences by model-averaging predicted detectabilities for models in the 90% confidence set (Appendix S1). Model parameters were generally consistent across models (Data S3).

Model-averaged predictions showed that detectability increased with a species' mass, though detectability

approached 1 for herd sizes >30 regardless of mass (Fig. 1a). The effect of mean herd size interacted with herd size such that species with larger average herd sizes showed less sensitivity to herd size than species with smaller average herd sizes (Fig. 1b). Color had little effect on detectability (Fig. 1c), with estimated detectabilities nearly identical across the observed range of color categories. Color also had little support in the confidence set of models; the top model including any color effect had $\Delta AIC_c = 5.7$, and nearly all parameter estimates for color

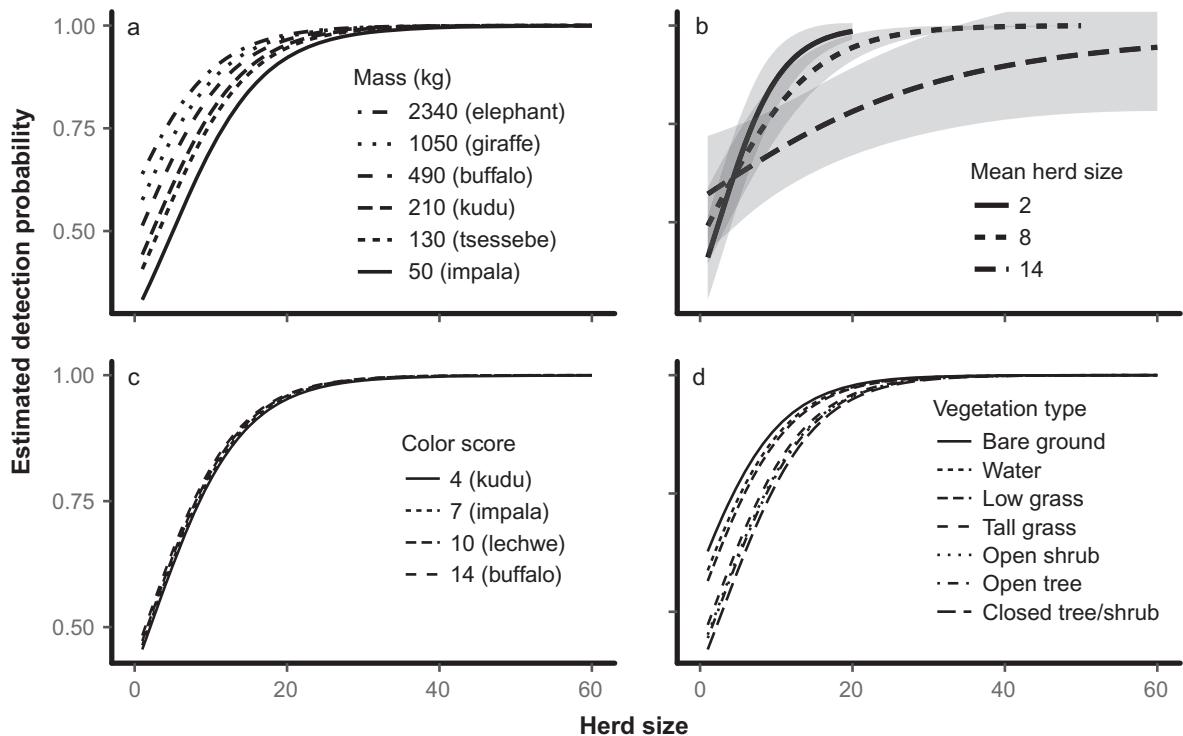


FIG. 1. Predicted effects of species traits on detectability for (a) mass, shown for selected values, (b) mean herd size (including interaction with herd size), shown for selected values, (c) color, shown for selected values, and (d) vegetation type. All graphs are based on model-averaged predictions from the species-trait model. Estimates were calculated at the mean values for all other traits in the data set. Shaded areas indicate \pm SE. SEs not shown in a, c, and d because of extensive overlap between levels.

had confidence intervals that overlapped 0 (Data S3). We found support for both herd-level covariates in the models. Detectability clearly increased with herd size (Fig. 1a–d). Vegetation type also affected detectability, with vegetation categories clustering into two groups (Fig. 1d). Detectability was highest in bare ground, water, and low grass; detectability was lower in tall grass and vegetation categories with trees or shrubs. As with mass, effects of vegetation were primarily evident at smaller herd sizes; for herds of >30 , predicted detectability was near 1 in all vegetation types.

The goodness-of-fit test for the species-trait model was not significant ($P = 0.21$), suggesting that the model fit the data. The species-trait model was a good predictor of the “detection ratio” of herds with two vs. one detections in the raw data set (linear regression: $r^2 = 0.80$, $P < 0.001$; Fig. 2). For 8 of the 10 species, the actual ratio was within the 95% confidence interval of the predicted ratio; actual detection ratios for buffalo and tsessebe were significantly greater than predicted values.

We tested the power of species traits to predict detectability for new species by serially deleting each species from the data set and then predicting that species’ detectability with data from the other nine species. For most species, the deleted model showed a good correspondence to the species-trait model fit with data from all 10 species (Fig. 3). For impala and warthog, however, the two models deviated substantially. The goodness-of-fit test for the deleted models produced similar and non-significant results for all 10 species (range, $P = 0.21$ – 0.27); the narrow range of these values is due to extensive overlap of the

input data for each model. We note that the goodness-of-fit tests reveal only the fit of the model to the input data, not the predictive power of the model with the deleted species. The deleted model had modest ability to predict detection ratios (linear regression, $r^2 = 0.54$, $P = 0.02$; Fig. 2). Actual ratios were within 95% confidence intervals of model predictions for seven species; actual ratios for buffalo, tsessebe, and hippopotamus were outside the confidence limits of the model. Confidence intervals on predicted ratios were quite wide for some species.

As a final method of comparing the species-trait and deleted models, we compared the parameters for the relationship between herd size and detectability between the two models. Slope parameters differed little between the two models for most species (Fig. 4). For buffalo and impala, however, the two models produced very different slope estimates. For intercept parameters, we compared the intercept separately for each vegetation type, resulting in seven estimates per species. For all species, correlations between the two models were positive and >0.9 , suggesting that the species-trait and deleted models ranked vegetation categories similarly in detectability (Fig. 5). For impala, warthog, and buffalo, however, the predicted intercepts from the deleted models were biased relative to the intercepts from the species-trait model.

DISCUSSION

Our results show that detectability of 10 African mammals on aerial surveys is largely predictable as a function of species traits. A model using species’ mass and mean

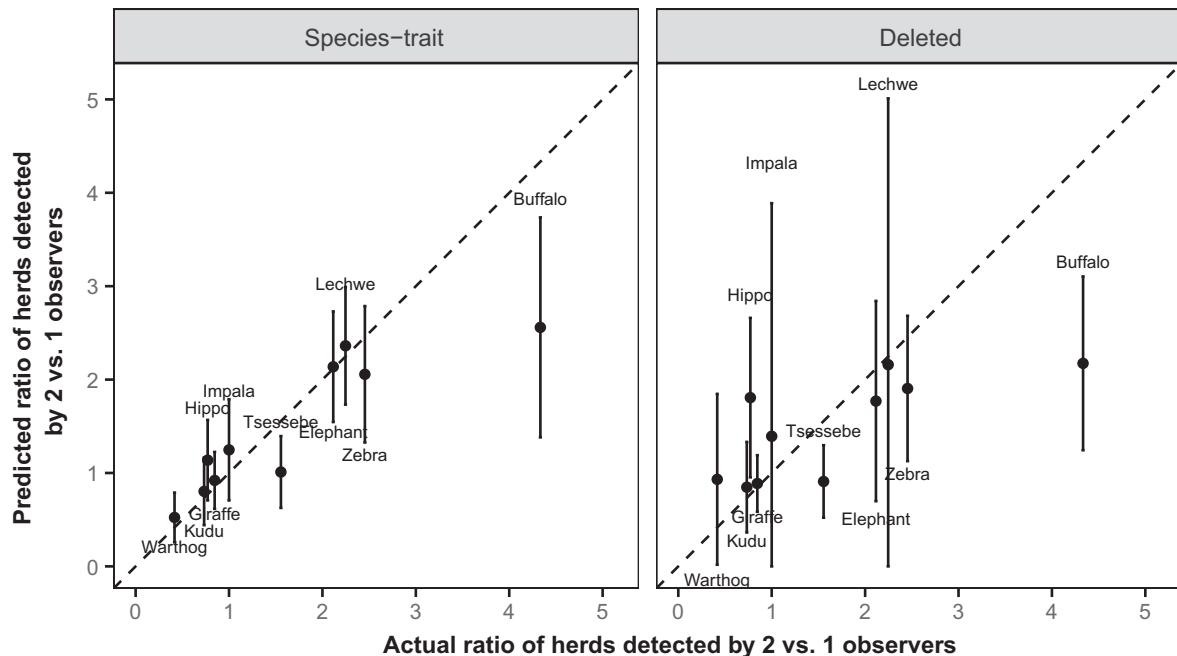


FIG. 2. Actual vs. predicted detection ratios for the species-trait and deleted models for 10 mammal species in Botswana. Error bars indicate 95% confidence intervals. Dashed line indicates equality of predicted and actual ratios.

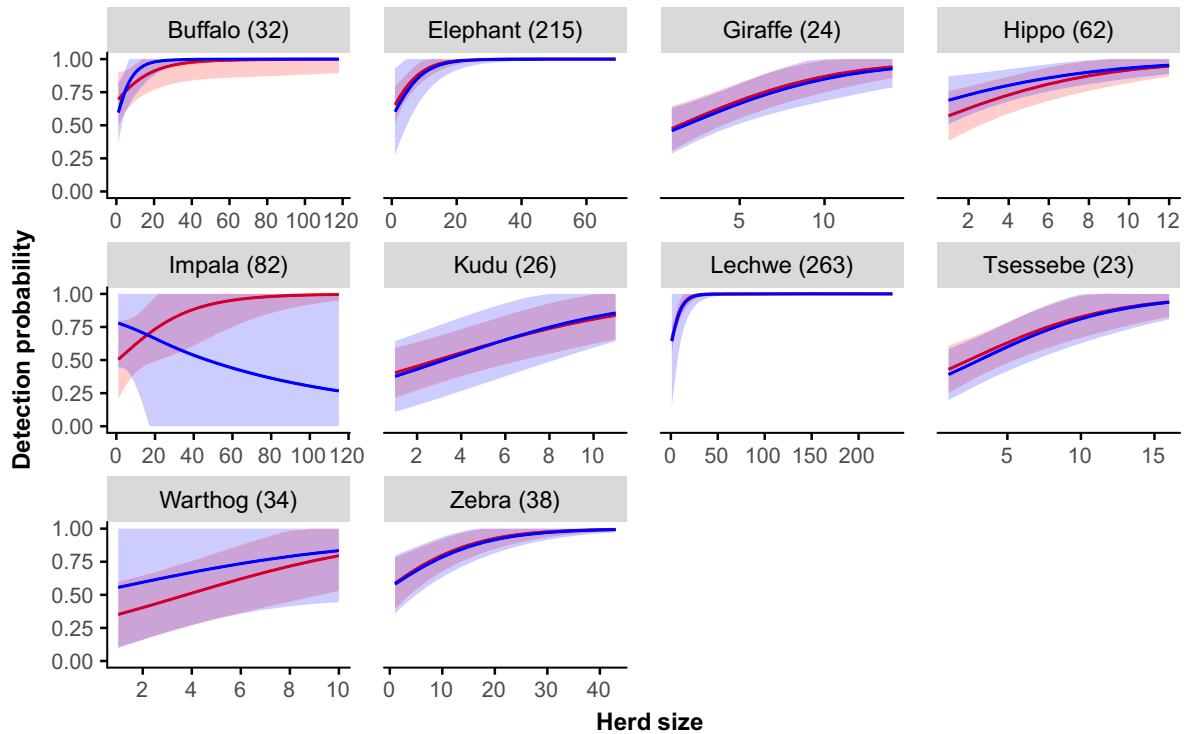


FIG. 3. Predicted detectability from the deleted model (blue) and the species-trait model (red) by species. Shaded areas indicate \pm SE for each model. The total number of herds observed for each species is in parentheses. [Color figure can be viewed at wileyonlinelibrary.com]

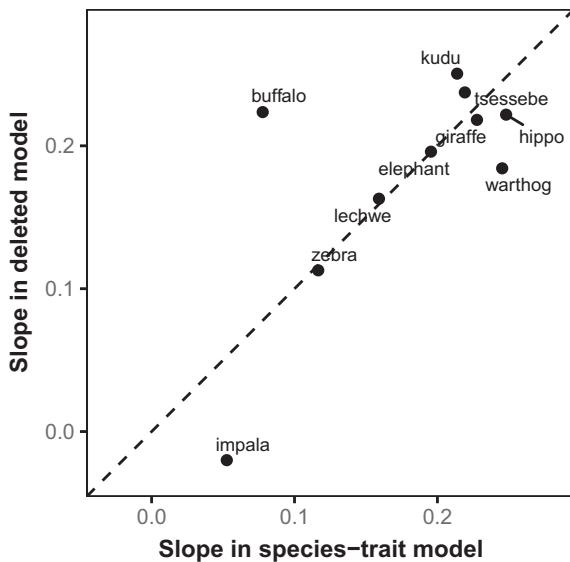


FIG. 4. Estimated slopes for the predicted relationship between herd size and detectability for the species-trait and deleted models. Slopes are on a logit scale. Dashed line indicates equal slopes between the two models.

herd size along with herd-level covariates and observer effects accurately predicted the actual ratio of detections by one vs. two observers for 8 of 10 species. This finding

should not be surprising because detectability is a function of human perceptual abilities, and such perceptions are surely influenced by species' characteristics. Our findings add to previous research by Garrard et al. (2013) showing that plant detectability was also a predictable function of species traits. We hope that the success of our endeavor will encourage researchers in other animal systems to examine traits affecting detectability across species.

During the review process, one reviewer brought up the possibility that our models may be pseudoreplicated because our effective sample size was 10 species, and a species trait such as mass could show a correlation with species' detectabilities by chance. Garrard et al. (2013) dealt with this issue by including a random effect of species in their models. We explored this possibility by running our species-trait model as a Bayesian data augmentation model. Results showed that including a random intercept term for species had little effect on mean estimates of model parameters, but the variance in parameter estimates increased with the addition of the random effect (S. Schlossberg, *unpublished data*). Thus, the lack of a random effect in our frequentist models may have led us to overestimate the precision of our species-trait models. The question of how to deal with unmodeled species-level heterogeneity in species-trait models is an interesting one, and we plan to explore it via simulation in a future paper. For future attempts to model detectability with species traits, however, researchers may want to include a random

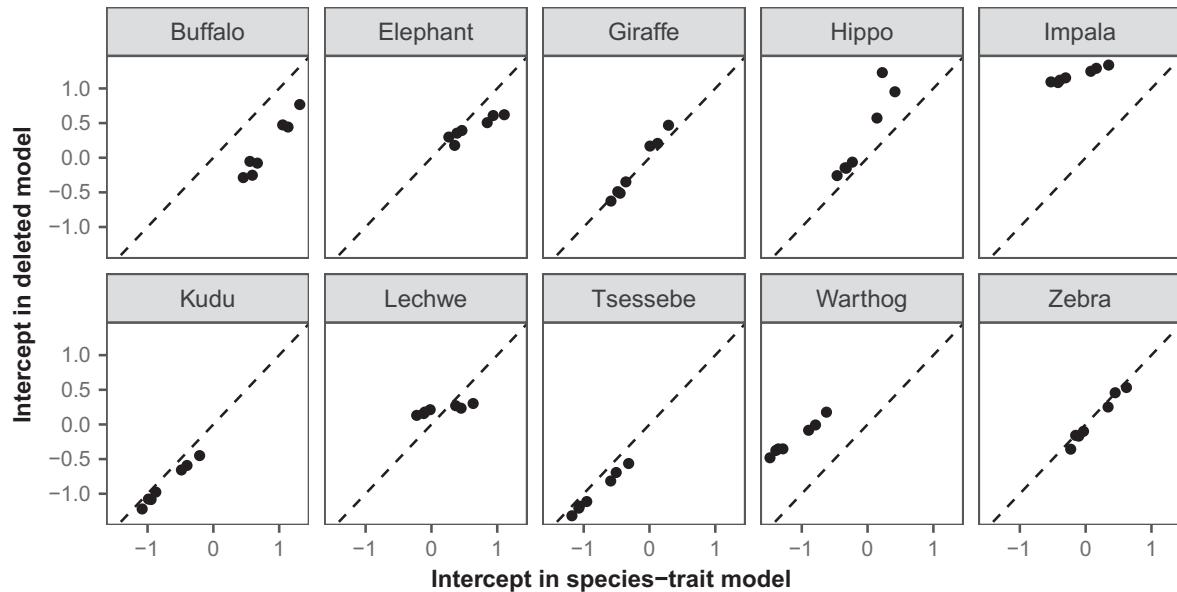


FIG. 5. Estimated intercepts for the predicted relationship between herd size and detectability for the species-trait and deleted models. Each point on a graph indicates a separate vegetation type. All intercepts are on a logit scale. Dashed line indicates equal intercepts between the two models.

effect of species or at least test such effects to ensure that pseudoreplication is not affecting results.

Factors affecting detectability

Herd size was the most important factor affecting detectability in our study. In the species-trait model, detectability increased with herd size for all 10 species, and the top-ranked model without herd size had $\Delta AIC_c = 7.3$. These results echo findings from other systems that larger animal groups are easier to detect (Samuel et al. 1987, Graham and Bell 1989, Cogan and Diefenbach 1998, Ransom 2012). Moreover, the effects of other species traits on detectability depended on herd size. Mass and vegetation affected detectability only at smaller herd sizes; for larger herds, predicted detectability was high regardless of covariate values (Fig. 1).

For animal species that typically occur singly, detectability can be summarized with a single parameter for each species (ignoring effects of vegetation or other categorical covariates such as observer). For herd-forming animals like the mammals we studied, detectability is best thought of as a function of herd size rather than a single value. Thus, the goal of modeling detectability for herd-forming animals should be to determine the relationship between herd size and detectability. As a result, modeling detectability will be more complex for group-forming species than in species that occur singly.

In the species-trait model, herd size also showed an important interaction with mean herd size (Fig. 1b). For species that typically occur in larger groups, predicted detectability varied relatively little with herd size. For species that typically occur in smaller groups, herd size

had larger effects. This effect has not been reported previously in the literature and suggests that effects of herd size on detectability are likely to be different for species with different aggregation patterns.

Mass was another important predictor of detectability in the species-trait model, with heavier species having higher predicted detectability. An obvious explanation for this result is that the human eye notices larger animals more easily on aerial surveys. Studies of birds have also reported that detectability increases with species' mass (Fletcher et al. 2006, Berthiaume et al. 2009, Johnston et al. 2014). Because masses in our study varied over three orders of magnitude, we log-transformed masses for analysis. During exploratory analyses, we also tested a $2/3$ power transformation, as surface area should be proportional to mass^{2/3}. We found, however, that the log-transformed values fit the data better (S. Schlossberg, *unpublished data*). Though mass clearly influenced detectability in our sample, other measures of body size such as height, length, and body surface area could all potentially be predictors of detectability and should be explored in future studies.

In the species-trait model, color had very little effect on detectability. We included color as a predictor because we expected that lighter-colored species, such as impala or lechwe, would be harder to detect than darker species. In reality, the background colors against which animals were viewed varied greatly in our study area, from whitish to dark brown (S. Schlossberg, *personal observation*). Also, observers may cue on factors such as movement or shadows that are independent of color. Because the effect of color on detectability of animals has not been tested frequently, more research is on this subject is needed.

As expected, detectability decreased with increasing vegetative cover around herds. In particular, detectability was higher in grassy or bare areas than in habitats with woody plants. These results reinforce earlier studies reporting that detectability on aerial surveys decreased with canopy cover or vegetation density (Bayliss and Yeomans 1989, Anderson and Lindzey 1996, Cogan and Diefenbach 1998). Unfortunately, including vegetation in detectability models is more onerous than accounting for animal mass or color because of the need to describe the habitat around each herd. In areas with heterogeneous vegetation, however, accounting for habitat usage is likely to be important for accurate prediction of detectability.

Despite the general success of the species-trait model in predicting detectability, model predictions of detection ratios deviated from actual results for tsessebe and buffalo. For both of these species, detections by two observers were more common than the model predicted, with the difference substantial for buffalo. One possible reason why the model fit these two species poorly is that buffalo ($n = 32$) and tsessebe ($n = 23$) had two of the smallest sample sizes in our data set. Thus, the model may have been accurate, but sampling error may have influenced the observed detection ratios for these two species. Alternately, the model may have been missing traits that affect detectability of tsessebe and buffalo.

Applying the species-trait model to new species

The success of our attempt to predict detectability for new species with the deleted model depends on one's measure of model accuracy. Theoretically, if the deleted model is accurate, its predictions should match those of the species-trait model. When we compared model parameters between the deleted and species-trait models, the deleted model fit the species-trait model poorly for impala, warthog, and buffalo but appeared to fit the other seven species more closely (Figs. 4, 5). On the other hand, one can also ask how well the deleted model fit the observed data. When we compared predicted detection ratios against actual values, the deleted model underestimated detectability for tsessebe and buffalo, just as the species-trait model did. The deleted model also overestimated detection ratios for hippopotamus. By either standard, the deleted model appeared to be accurate in predicting the detectability of elephant, giraffe, kudu, lechwe, and zebra. As discussed above, these results come with the caveat that frequentist models that do not include a random effect of species may overestimate the precision of predicted detectabilities.

To our knowledge, Garrard et al. (2013) is the only previous study that attempted to predict detectability for a new species with a trait-based model. Both Garrard's study and ours found similar levels of predictive accuracy when applying a model to new species. Garrard's top predictive model had $r^2 = 0.63$ when comparing predicted and actual times to detection, and our deleted model had $r^2 = 0.54$ when comparing predicted and

actual detection ratios. One question that neither Garrard's study nor ours addressed is what level of accuracy is sufficient for a predictive model of detectability to be useful. Because the idea of applying detectability models to new species is in its infancy, this issue has not been addressed to date. Garrard's suggestion of using predictive detectability models to determine sampling effort may require less accuracy than attempting to correct population estimates for detectability.

In our models, one problem with attempting to predict detectability for new species was that predicted detectabilities had low precision for some species. For impala and warthog, this is likely due to extrapolation beyond the limits of the input data, as we will discuss. For the other eight species' deleted models, SEs for predicted detectabilities for herd sizes of 1–30 averaged 65% greater than SEs from the species trait model; we limited this calculation to smaller herd sizes because, for larger herd sizes, predicted detectabilities were near 1 and SEs were generally small. The larger error in the deleted models is largely explained by the smaller sample size for that model vs. the species-trait model. In fact, for the eight species excluding impala and warthog, the percent increase in SE between the species-trait model and the deleted model was largely explained by the number of herds observed ($r = 0.94$).

If researchers are comfortable with the error rates of predictive models, our findings suggest that it may be possible to apply detectability models developed with data-rich species to data-poor species. Previously, detectability for data-poor species has been modeled by "borrowing," grouping data-poor and data-rich species based on perceived similarity in detectability (MacKenzie et al. 2005). For instance, Alldredge et al. (2007) grouped bird species based on maximum detection distance and expert opinion of singing rate to increase the effective sample size for uncommon species. This strategy is based on the untested assumption that the species being grouped have similar detectabilities. Our procedure, which explicitly models detectability as a function of measurable traits, is likely to be more robust because opinion is not involved, and effects of traits used to predict detectability can be directly tested. Another advantage of modeling detectability as a function of species traits is that trait-based models from different studies can potentially be combined by averaging parameters for specific traits. This would result in large effective sample sizes for detectability corrections. Nonetheless, applying models across studies will only be possible if different studies produce similar results for effects of traits on detectability, and observer or aircraft-specific effects can be controlled. Clearly, more studies along the lines of ours will be necessary before we can know if combining results from multiple studies will be feasible.

Examining cases where the deleted model failed reveals a key limitation to our ability to predict detectability. For impala, the species-trait model predicted that detectability would increase with herd size. When impala was removed from the data set, however, the resulting model predicted

that impala detectability would decrease with herd size (Fig. 3). Impala had the lowest mass and largest mean herd size in our data set. Thus, the deleted model was extrapolating beyond the boundaries of the data, and our attempt to predict detectability for impala failed as a result. Likewise, the deleted model for warthog showed a poor fit to the species-trait model and wide confidence intervals. Warthog had the second lowest mean herd size and the second lowest mass in the data set. In regression models, predictions far from the mean values for independent variables produce imprecise estimates. Thus, attempting to predict detectability for warthog led to detectability estimates with wide confidence intervals. These two examples show that for species with extreme values of traits affecting detectability, predicting detectability with observations from other species may not be possible. An alternative hypothesis for the relatively poor fit of models for warthog and impala is that the relationship between log-transformed mass and detectability may be nonlinear at the lower end of the range of masses observed, and our linear-logistic models did not capture this effect.

The deleted model's predicted detection ratios for buffalo, tsessebe, and hippopotamus were inaccurate. For buffalo and tsessebe, predictions from the deleted model were similar to those from the species-trait model. This suggests that, as mentioned above, traits that affect detectability in these species may have been left out of the model, or sampling error may have affected the observed detection ratios. For hippopotamus, the deleted model overestimated detectability compared to both the species-trait model (Fig. 3) and the actual detections (Fig. 2). Of the 62 hippopotamus herds used in the study, 44% ($n = 27$) were observed in water, and this species made up over one-half of all observations of any species in water. Consequently, the parameter estimate for the vegetation category of water was based on a much smaller sample size in the deleted model than in the species-trait model, and the resulting intercepts differed greatly (species-trait model, estimate = 0.22, 95% CI = -0.46 to 0.93; deleted model for hippopotamus, estimate = 1.22, CI = 0.28 to 2.29). This finding reinforces the idea that predicting detectability for species with atypical trait values may be problematic.

Other factors affecting detectability

We found strong effects of observer, seating position in the plane, and cumulative abundance on detectability of African mammals. These effects were supported in every member of the species-trait model confidence set (Data S3). Several past studies have shown that observer and seating position can influence detectability on aerial surveys (Koneff et al. 2008, Ransom 2012, Schlossberg et al. 2016). In this study, one of the two most experienced observers had a higher detection probability than the other three observers. Our observers also exhibited higher detection probabilities while sitting in the rear row of

seats than the front row, likely because the rear seats were further from the wing struts, potentially allowing better visibility than the front seats. To our knowledge, our finding that detectability increased with a species' cumulative abundance has not been reported previously for animals. This finding is consistent with the idea that observers have a search image for each species, and that search image may improve during the survey as an observer repeatedly sees a species. This subject clearly needs additional attention, as correcting aerial surveys for abundance effects on detectability is not standard practice.

Implications for aerial surveys

For decades, scientists have acknowledged that observers on aerial surveys miss some animals (Caughley 1974). For African mammals, however, little effort has been made to quantify detectability on aerial surveys or correct results for missed herds. Our results suggest that ignoring detectability can cause serious bias. Because of missed herds, observers would underestimate population sizes by an estimated mean of 29% for the 10 species we studied (S. Schlossberg, *unpublished data*). By species, the undercount ranged from 7% to 65%. Thus, some sort of detectability analysis should be employed for all aerial surveys of large mammals. If one's goal is to accurately estimate population size, corrections will be especially important for the species that are hardest to detect: the smallest, those that use the densest vegetation, and those that tend to have smaller herd sizes. Fortunately, a variety of methods, including double-observer counts, are available to correct aerial surveys for detectability (Magnusson et al. 1978, Steinhorst and Samuel 1989, Koneff et al. 2008, Riddle et al. 2010).

Our findings also suggest areas where observers on aerial surveys could potentially focus their training to improve detection probabilities. Because observers are generally missing smaller species, smaller herds, and animals that use denser vegetation, observers should pay greater attention to these types of herds. One possible training technique could be to practice locating cryptic herds using photographs from previous surveys. Whether or not such training can actually improve detection probabilities has not been studied. Because, however, observers miss many herds, attempting such a training exercise could be useful.

CONCLUSIONS

In recent years, our ability to measure and control for detectability on animal surveys has advanced greatly. Modeling detectability as a function of species traits could advance this field further by elucidating factors hindering detections and allowing researchers to apply robust models to rare species for which detectability correction has been impossible in the past. Our study was simply a proof of concept, an initial attempt to model the factors determining species' detectabilities. Despite modest sample

sizes, we were able to predict detectability for most species. Additional research on this subject is needed. Fortunately, the tools to conduct this research are readily available, and many existing data sets may be suitable for addressing these questions without collecting new data (e.g., Alldredge et al. 2007, Kéry and Plattner 2007).

ACKNOWLEDGMENTS

This study was funded by Paul G. Allen, as part of the Great Elephant Census. We are thankful to the staff at Vulcan Inc.: Dave Stewart, Dune Ives, and Lauren Kickham. We appreciate the support provided by Jody Allen and Ferina Keshavjee. This study was endorsed by the Botswana Government, and we thank the Department of Wildlife and National Parks for allowing us to fly these surveys. Honorable Minister Khama, Michael Flyman, Cyril Taolo, Amo Keitsile, Elford Seonyatseng, Robert Sutcliffe, Tempe Adams, Kelly Landen, Marcus Hancock, and the staff of Abu Camp and Wilderness Safaris are acknowledged for their assistance. Thanks to A. Wirsing, A. Johnston, and two anonymous reviewers for their outstanding comments on the manuscript.

LITERATURE CITED

- Allredge, M. W., K. H. Pollock, T. R. Simons, and S. A. Shriener. 2007. Multiple-species analysis of point count data: a more parsimonious modelling framework. *Journal of Applied Ecology* 44:281–290.
- Anderson Jr., C. R. and F. G. Lindzey. 1996. Moose sightability model developed from helicopter surveys. *Wildlife Society Bulletin* 24:247–259.
- Anderson, A. S., T. A. Marques, L. P. Shoo, and S. E. Williams. 2015. Detectability in audio-visual surveys of tropical rainforest birds: the influence of species, weather and habitat characteristics. *PLoS ONE* 10:e0128464.
- Arnold, T. W. 2010. Uninformative parameters and model selection using Akaike's Information Criterion. *Journal of Wildlife Management* 74:1175–1178.
- Banks-Leite, C., R. Pardini, D. Boscolo, C. R. Cassano, T. Püttker, C. S. Barros, and J. Barlow. 2014. Assessing the utility of statistical adjustments for imperfect detection in tropical conservation science. *Journal of Applied Ecology* 51:849–859.
- Bayliss, P., and K. M. Yeomans. 1989. Correcting bias in aerial survey population estimates of feral livestock in northern Australia using the double-count technique. *Journal of Applied Ecology* 26:925–933.
- Berthiaume, É., M. Bélisle, and J.-P. Savard. 2009. Incorporating detectability into analyses of population trends based on hawk counts: a double-observer approach. *Condor* 111:43–58.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to distance sampling: estimating abundance of biological populations*. Oxford University Press, Oxford, UK.
- Burger, J., and M. Gochfeld. 1994. Vigilance in African mammals: differences among mothers, other females, and males. *Behaviour* 131:153–169.
- Cade, B. S. 2015. Model averaging and muddled multimodel inferences. *Ecology* 96:2370–2382.
- Caughley, G. 1974. Bias in aerial survey. *Journal of Wildlife Management* 38:921–933.
- Cogan, R. D., and D. R. Diefenbach. 1998. Effect of undercounting and model selection on a sightability-adjustment estimator for elk. *Journal of Wildlife Management* 62:269.
- Conroy, M. J., J. T. Peterson, O. L. Bass, C. J. Fonnesebeck, J. E. Howell, C. T. Moore, and J. P. Runge. 2008. Sources of variation in detection of wading birds from aerial surveys in the Florida Everglades. *Auk* 125:731–743.
- Fletcher Jr., R. J., R. L. Hutto, and D. A. Haukos. 2006. Estimating detection probabilities of river birds using double surveys. *Auk* 123:695–707.
- Garrard, G. E., M. A. McCarthy, N. S. G. Williams, S. A. Bekessy, and B. A. Wintle. 2013. A general model of detectability using species traits. *Methods in Ecology and Evolution* 4:45–52.
- Graham, A., and R. Bell. 1989. Investigating observer bias in aerial survey by simultaneous double-counts. *Journal of Wildlife Management* 53:1009–1016.
- Gu, W., and R. K. Swihart. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation* 116:195–203.
- Guillera-Aroita, G. 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* 40:281–295.
- Huggins, R. M. 1989. On the statistical analysis of capture experiments. *Biometrika* 76:133–140.
- Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics* 47:725–732.
- Jachmann, H. 2001. *Estimating abundance of African wildlife: an aid to adaptive management*. Kluwer Academic Publishers, Boston, Massachusetts, USA.
- Jachmann, H. 2002. Comparison of aerial counts with ground counts for large African herbivores. *Journal of Applied Ecology* 39:841–852.
- Johnston, A., S. E. Newson, K. Risely, A. J. Musgrove, D. Massimino, S. R. Baillie, and J. W. Pearce-Higgins. 2014. Species traits explain variation in detectability of UK birds. *Bird Study* 61:340–350.
- Kéry, M., and M. Plattner. 2007. Species richness estimation and determinants of species detectability in butterfly monitoring programmes. *Ecological Entomology* 32:53–61.
- Kéry, M., and H. Schmid. 2006. Estimating species richness: calibrating a large avian monitoring programme. *Journal of Applied Ecology* 43:101–110.
- Koneff, M. D., J. A. Royle, M. C. Otto, J. S. Wortham, and J. K. Bidwell. 2008. A double-observer method to estimate detection rate during aerial waterfowl surveys. *Journal of Wildlife Management* 72:1641–1649.
- MacKenzie, D. I., and L. L. Bailey. 2004. Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics* 9:300–318.
- MacKenzie, D. I., and J. A. Royle. 2005. Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology* 42:1105–1114.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255.
- MacKenzie, D. I., J. D. Nichols, N. Sutton, K. Kawanishi, and L. L. Bailey. 2005. Improving inferences in population studies of rare species that are detected imperfectly. *Ecology* 86:1101–1113.
- Magnusson, W. E., G. J. Caughley, and G. C. Grigg. 1978. A double-survey estimate of population size from incomplete counts. *Journal of Wildlife Management* 42:174.
- Norton-Griffiths, M. 1978. *Counting animals*. African Wildlife Foundation, Nairobi, Kenya.
- R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Ransom, J. I. 2012. Detection probability in aerial surveys of feral horses. *Journal of Wildlife Management* 76:299–307.

- Riddle, J. D., K. H. Pollock, and T. R. Simons. 2010. An unrecoupled double-observer method for estimating detection probability and abundance. *Auk* 127:841–849.
- Rodrigues, A. S. L., J. D. Pilgrim, J. F. Lamoreux, M. Hoffmann, and T. M. Brooks. 2006. The value of the IUCN Red List for conservation. *Trends in Ecology & Evolution* 21:71–76.
- Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108–115.
- Samuel, M. D., E. O. Garton, M. W. Schlegel, and R. G. Carson. 1987. Visibility bias during aerial surveys of elk in northcentral Idaho. *Journal of Wildlife Management* 51:622–630.
- Schlossberg, S., M. J. Chase, and C. R. Griffin. 2016. Testing the accuracy of aerial surveys for large mammals: an experiment with African savanna elephants (*Loxodonta africana*). *PLoS ONE* 11:e0164904.
- Skinner, J. D., and C. T. Chimimba. 2006. *The mammals of the southern African sub-region*. Cambridge University Press, Cambridge, UK.
- Steinhorst, R. K., and M. D. Samuel. 1989. Sightability adjustment methods for aerial surveys of wildlife populations. *Biometrics* 45:415–425.
- White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study* 46:120–138.
- Yamaura, Y., M. Kéry, and J. A. Royle. 2016. Study of biological communities subject to imperfect detection: bias and precision of community N-mixture abundance models in small-sample situations. *Ecological Research* 31:289–305.

SUPPORTING INFORMATION

Additional supporting information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/eap.1632/full>