SHORT COMMUNICATION

Do you see what I see? Variation in detection, identification and enumeration of mammals during transect surveys

[**Running title**: Variation in mammal survey data]

Anne E. Goodenough[1*], Danielle L. Berry[1], William S. Carpenter[1], Melissa Dawson[2], Natasha Furlong[1], Rachel J. Lamb[1], Lynne MacTavish[2], Niall O'Reilly[1], Hannah Toms[1], Lauren H. Whitehead[1], Adam G. Hart[1]

1. Department of Natural and Social Science, University of Gloucestershire, Francis Close Hall, Cheltenham, GL50 4AZ

2. Mankwe Wildlife Reserve, Mogwase, NW Province, South Africa

* Corresponding author email = aegoodenough@glos.ac.uk

Abstract

Effective monitoring, management and conservation of wildlife axiomatically depend on accurate data but causes of variation, including inter-observer variation, are rarely explicitly quantified. Here, under controlled conditions, we demonstrate considerable variation in detection, identification and enumeration of (theoretically) readily-identifiable African mammals at a reserve with a known species assemblage. Detection: 97.8% of sightings were missed by ≥1 observer; frequency of detection was affected by observer ID, detection distance, visibility, and animal group size. Identification: just 3/14 species were identified consistently at all sightings. Enumeration: lack of consensus for 60.5% of sightings; consensus likelihood was affected by visibility and group size.

Keywords

Inter-observer variation, ecological data reliability, mammal detection, species misidentification, estimating group size.

Introduction

Accurate surveys of species' presence and abundance are essential for developing effective monitoring strategies and conservation management of wildlife (Spellerberg, 2005; Thomas et al., 2010; Goodenough and Hart, 2017). This requires individuals to be initially detected, then identified correctly, and finally enumerated with precision and accuracy. However, variation can arise during detection (Diefenbach et al., 2003; Shirley et al., 2012), identification (Hobbs and Waite, 2010; Hoekman, 2021), and enumeration (Frederick et al., 2003). Variability can be caused by differences in: (1) survey conditions, especially detection distance and the risk of observers being partly unsighted; (2) species conspicuousness, which is affected by size, abundance, camouflage, and behaviour; and (3) inter-observer variation (IOV), which can be innate and/or due to differences in experience or training.

Variation linked to survey conditions – especially in relation to availability for detection, visibility, and distance – has been studied in a variety of contexts. This includes visual surveys on land (e.g. Kissling and Garton, 2006) and water (e.g. Ronconi and Burger, 2009; Schwarz et al., 2010), auditory surveys (e.g. Simons et al., 2007), and aerial surveys that can be manned (e.g. Marsh and Sinclair, 1989; Pollock et al., 2006) or unmanned (e.g. Hodgson et al., 2016; Hodgson et al., 2017; Brack et al., 2018). In many cases, these and other studies have involved analysing double-observer data to quantify the impact of survey conditions on variation in the resultant ecological data. Separately, some studies have considered IOV based on detection of artificial targets, for example of "birds" viewed from above and/or from the ground (Frederick et al., 2003; Hodgson et al., 2018) and diving "marine mammals" viewed from above (Pollock et al., 2006). However, relatively few studies have provided explicit quantitative estimates of IOV in industry-standard mammal surveys under field conditions, nor considered how this co-occurs with survey factors such as detection distance and sighting visibility. This is despite IOV in surveys of other taxa having important applied consequences, including the accurate determination of biodiversity value at proposed development sites (Cherrill, 2015) and informing appropriate habitat management recommendations (Goodenough et al., 2020).

In poor lighting conditions in the African bush, the lead authors of this study once confidently misidentified an ostrich (*Struthio camelus*) as an elephant (*Loxondonta africana*) and anecdotally most field ecologists have made similar identification blunders. Inspired by this, here we consider data from five observers simultaneously undertaking identical walked and driven mammal transects in a South African wildlife reserve with a known mammal assemblage. We quantify the effects of detection distance, visibility, animal group size, and whether a transect was walked or driven, either in tandem with IOV (detection) or directly affecting IOV (enumeration). We also determine consensus versus confusion in identification of animals from the same sighting by the same observers at the same time. Finally, we consider implications of variation for optimising surveying and monitoring.

Methods

The study was undertaken at a 4,750 hectare wildlife reserve in Northwest Province, South Africa. The reserve supported a known assemblage of antelope, including eland (*Taurotragus oryx*), impala (*Aepyceros melampus*), gemsbok (*Oryx gazelle*) and tsessebe (*Damaliscus lunatus*), as well as other herbivores including giraffe (*Giraffa camelopardalis*), zebra (*Equus quagga*), and white rhinoceros (*Ceratotherium simum*). The habitat was a mosaic of savanna grassland (Figure 1). Sward height was generally 5-60 cm, with some taller patches, interspersed with mature trees such as umbrella thorn (*Vachellia tortilis* Forssk.) and marula (*Sclerocarya birrea* Hochst.) There were patches of thicket dominated by sicklebush (*Dichrostachys cinerea* L.), some of which were dense. Uninterrupted horizontal visibility varied considerably based on the combination of local conditions (<5 m to >2 km).

Five observers with similar skills and experience conducted 20 mammal transect surveys on foot (walked; n=11) and from a vehicle (driven; n=9). The observers (co-authors DLB, NF, RJL, NO'R, LHW) were postgraduate researchers who were all experienced ecologists. Before the research, they received classroom-based identification training specific to the mammals that were the subjects of study, followed by an intensive three-day identification masterclass in the field. The observers were accompanied by at least one ranger (co-authors MD, LM) and two members of University staff experienced in ecological fieldwork in Africa (co-authors AEG, AGH, WSC, HT). For walked transects, observers walked 3-6 km in single file ~3 m apart, with observer position randomised for each transect. Driven transects were undertaken in a safari vehicle covering 10-25 km driving at ~20 kph, with observers seated in two rows arranged such that eye-height above ground was approximately equal.

For both transect types, observers carried an inaudible numbered "buzzer" (Retekess T128 wireless pager system, developed for restaurant customers to draw the attention of waiting staff). Each observer covertly pressed their buzzer when they sighted one or more large mammals; ostrich were also recorded as per the industry-standard survey protocol (Bothma and du Toit, 2016). Pressing the buzzer caused an inaudible vibration on a wrist-worn "watch" worn by a sixth person, who was not an active observer and who walked behind the group (walked transects) or sat in the vehicle (driven transects). The buzzer number was displayed on the watch, which also showed the order in which observers detected the sighting if multiple observers pressed their buzzer in quick succession. To ensure all observers had the opportunity to detect the sighting, on walked transects the watch-wearer waited until all observers had passed the point where the sighting was first detected before calling "sighting". After this, no further buzzes were permitted and the observers clustered at the point of first detection. For driven transects, the watch-wearer asked the driver to stop when the first buzzer was activated. During the subsequent slowing down and stopping period, any additional observers who saw a sighting could press their buzzer to register a detection. The driver reversed until the point at which the watch-wearer judged the first buzzer had been activated. For both methods, the first observer to press their buzzer was named by the watch-wearer, and this observer

described the sighting location to all other observers without giving details of species or number. This protocol ensured data on species identification and enumeration could be collected from all observers, even if not all observers had initially detected the sighting. It is recognised that in designing a data collection framework that allowed observers to independently detect live animals in real field conditions, followed by an opportunity for all observers to provide data on identification and enumeration, the amount of time observers had to detect a sighting (the detection period) ended artificially early. This might have supressed frequency of detection. However, all observers had a genuine opportunity to detect each sighting before the detection period ended and many of sightings involved animals at right angles to the transect and/or moving away from observers, such that in many cases the detection period was, in reality, self-limiting.

Once sighting location had been described, all observers, working silently and independently, had 30 seconds to identify the species (singular or plural) and count the number of individuals of each species and thus obtain an estimate of abundance we termed “group size”. To ensure realistic but standardised conditions, all observers utilised identical optimal equipment (8x42 Barr & Stroud Savannah binoculars, Suffolk, UK). During this observation period, the watch-wearer recorded the distance between observers and animal[s] (hereafter “detection distance”) with a laser rangefinder (Leica Rangemaster CRF2400). When the sighting involved a group of animals, the distance to the nearest individual was taken. Each observer who could see the sighting also recorded a subjective visibility score for their view on an ordinal scale (1 = very poor; 5 = excellent); those who were unsighted recorded the sighting as unobservable.

There were 304 unique sightings (n = 140 walked; 164 driven) (Goodenough et al., 2022). Of these, 294 sightings were ultimately observable by ≥ two observers, such that potential variation in species identification and enumeration could be calculated: 258 were single-species sightings while 36 were considered to be multiple-species sightings by at least one observer. For where detection information was recorded (n=278 sightings * 5 observers), we calculated “frequency of detection” as the number of observers who independently registered their detection by pressing their buzzer (minimum = 1 observer; maximum = 5 observers). Variation in enumeration of group size was calculated using Coefficient of Variation: CoV = (standard deviation of the number of animals recorded by each observer / mean number of animals recorded) * 100. This ensured that variation in estimates of group size were independent of the mean group size and thus avoid higher variation being an artefact of larger means (Fowler and Cohen, 1996) and mirrored previous studies of IOV in ecological data (e.g. Pankakoski et al. 1987; Goodenough et al., 2010). CoV calculations used data recorded by all observers who were able to record a sighting, not just those who initially detected it (total number of observations = 1,331 from a theoretical maximum of 1,520 if all observers had been able to record all 304 sightings). The mean visibility score and mean animal group size were calculated from individual observer data.

Before formal analysis, we undertook an internal verification step to assess whether there was a temporal change in frequency of detection or variation in enumeration in relation to day of study, as would be expected if a lack of experience was an issue initially; these were non-significant (Spearman Rank correlation: r_s=0.029, n = 278, p=0.630 and r_s=0.095, n = 294, p=0.103, respectively). Formal analysis took the form of two Generalised Linear Models (GLMs). The first model, to assess variability in detection, used raw binary detection data (1 = detected, 0 = not detected) from the five observers as the dependent variable (n = 1,390; 278 unique sightings where detection was recorded * 5 observers) with a binomial distribution and logit link. The second model, to test variation in enumeration, used data summarised at sighting level using CoV of group size as the dependent variable (n = 294 sightings recorded by ≥ two observers) with a Gamma distribution and log link. In both models, survey type (walked/driven) was entered as a categorical fixed factor; detection distance (min=27 m, max=1,645 m, mean=402 m), mean visibility, and animal group size (min=1, max=25, mean=3.3) were entered as continuous covariates. In GLM Model 1, observer ID (1-5) was also added as a fixed factor so IOV in detection could be quantified explicitly (although variation between observers is inevitable as detection probability decreases for harder-to-observe sightings, this variability would be random rather than systematically linked to specific observers and would not return a significant result for observer ID). In GLM Model 2, IOV was built into the analysis framework directly by using CoV as the dependent variable.

Results and Discussion

Detection data were available for 278 sightings on walked and driven transects. Of these, 53.9% were detected by one observer, 25.3% by two observers, 13.6% by three observers, and 5.0% by four observers: just 2.2% were detected by all five observers. Thus, 97.8% of sightings were missed by ≥ one observer. The average frequency of detection at an individual sighting was 60% (equivalent to three out of five observers) for driven transects, rising to 69% for walked transects.

GLM Model 1 showed that there were significant differences in detection between observers (binary data indicating whether sighting was detected or not detected). Quantifying this using the estimated marginal means function to predict detection probability (95% CI) at mean values for the other predictors – detection distance, visibility, and animal group size – showed substantial differences that were indicative of IOV in detection ability in identical conditions: Observer 1 = 0.36 ± 0.06; Observer 2 = 0.17 ± 0.04; Observer 3 = 0.48 ± 0.07; Observer 4 = 0.63 ± 0.06; Observer 5 = 0.13 ± 0.04. Detection was also significantly affected by detection distance (detection decreased as detection distance increased), mean visibility (detection decreased as visibility decreased) and average animal group size (detection higher for larger groups); there was no difference between survey types (Table 1). That frequency of detection decreased with detection distance and visibility – and increased with group size as animals became more conspicuous – makes intuitive sense, however, survey type being non-significant is interesting. This suggests that variation was the consequence of the properties of the

sighting (detection distance; group size) and the observers (IOV), rather than whether the survey was walked or driven. Given that walked transects are physically demanding, cover substantially less ground than driven surveys per unit time, and record fewer sightings per survey, we suggest driven transects be used for surveys of African land mammals if tracks and vehicle resources are available. The effectiveness of driven transects for establishing species community in an ecologically-meaningful way has been shown previously for detection of nocturnal mammals (Hart et al., 2022).

Of the 294 sightings ultimately *recorded* by multiple observers, there was consensus on the number of species present for 267 (90.8%) and a lack of consensus for 27 (9.2%). Estimates of the number of species only ever differed ± 1 species (1 vs 2 species = 24 sightings; 2 vs 3 species = 3 sightings). However, variability in the total number of individuals recorded per sighting (i.e. enumeration of group size regardless of species) was much more variable. There was consensus on enumeration for 116 sightings out of 294 (39.5%); of the 178 sightings where there were differences in enumeration, between observers, mean CoV of was 30.3% (min=7.2%; max=90.0%). Enumeration has previously been shown to be highly variable in surveys of birds, with observers generally being more likely to underestimate than overestimate numbers (Frederick et al., 2003).

Across all sightings, 18 species were recorded by at least one observer. Of these, four species were rarely encountered (recorded at 7 sightings in total). For the 14 commonly-recorded species, there were 318 observations across the 258 unique sightings: (1) that were recorded by at least two observers; and (2) where all observers who recorded the sighting agreed that only one species was present. Each species was recorded on ≥ 10 occasions. For these species, we tallied the number of consensus and confusion sightings for each species to show species-specific differences (Figure 2). Zebra, giraffe, and warthog (*Phacochoerus africanus*) were never confused with other species and there was little confusion of wildebeest (*Connochaetes taurinus*), ostrich and rhino (Figure 2). However, most sightings involving hartebeest (*Alcelaphus buselaphus*), waterbuck (*Kobus ellipsiprymnus*), eland, and tsessebe involved ≥ one observer recording a different species (Figure 2). We had thought that common confusion species would be likely to be animals with similar size or morphology (e.g. impala confused with blesbok, or tsessebe with hartebeest). In fact, confusion was far more diverse with, for example, confusion between eland and rhino, and between impala and hartebeest. In one sighting, observers agreed on there being one individual of a single species, but each reported a different species identification: gemsbok, nyala (*Tragelaphus angasii*), kudu (*Tragelaphus strepsiceros*), waterbuck, or eland. Such a high degree of confusion between experienced observers working with a known mammal assemblage was much higher than would have been predicted given the relatively straightforward identification of study species in idealised conditions, and clearly shows the potential for IOV under actual field conditions.

Anecdotal observation suggested much of the variation in identification stemmed from very minor differences in the extent of animal visibility in the field. For example, some observers were noted misidentifying hartebeest as impala: these two species differ substantially in size, body morphology and horn profile but colouration can be similar. Very minor differences in line of sight between observers, driven by differences in observer height or intervening vegetation, led to confusion when animals were being identified primarily on coat colour when horns were not seen. Similar problems in species identification were also identified by Hoekman (2021) for seabirds and Hobbs and Waite (2010) for cetaceans. Such species misidentification is important. When there is substantial difference in density of different co-occurring species, misidentification can have profound consequences for rarer species and will typically result in such species being considerably overestimated (Conn et al., 2013).

Modelling variation in estimates of group size (GLM Model 2) showed IOV was significantly affected by mean visibility (variation increased as visibility decreased) and average animal group size (variation higher for larger groups). As for detection, there was no difference between survey types when modelling enumeration; detection distance was also non-significant (Table 1). Thus, while distant animals were less likely to be detected initially by all observers, once they were detected then estimates of group size were unaffected by detection distance. It is interesting that enumeration variation increased with group size as previous work on birds found error rates were independent of group size (Frederick et al., 2003). There were important species-differences, with lowest variability in enumeration of waterbuck and ostrich and much higher levels for wildebeest and eland (Figure 3). Unravelling the factors affecting these species-specific differences was beyond the scope of the current study, but previous research on seabird enumeration found that body size, relative abundance, and other measures of "conspicuousness" may be important (Ryan and Cooper, 1989).

Recommendations

Surveying is resource-intensive, often difficult, and sometimes risky. However, the resultant data are incredibly important for informing effective monitoring, management and conservation (Spellerberg, 2005; Goodenough and Hart, 2017). IOV clearly has the potential to confound data, yet this is rarely explicitly acknowledged. Here, we show that even with a known assemblage of readily-identifiable large mammals, variation in detection, identification and enumeration is considerable. In some cases, errors could have negligible effects but in other cases errors could be cumulative and have considerable consequences leading to ill-informed management decisions as shown in previous studies of other taxa (e.g. Goodenough et al., 2020). The errors that undoubtedly occur in survey data often reflect the difficulty of the task rather than poor performance of surveyors, but, based on this study and drawing on previous research, we make the following recommendations to ensure that wildlife survey data are as robust as possible:

(1) General: Quantifying variability in detection, identification and enumeration in specific field contexts is vital to understand magnitude and determine potential effects. IOV should be minimised, for example, via training or use of automated (Conn et al., 2013) or semi-automated (Hodgson et al., 2018) algorithms for detection, identification, and enumeration.

(2) Detection: In addition to using industry-standard methods to reduce variation linked to sightings themselves (maximum distances of use of species-specific detection correction factors; reviewed by Thomas et al. (2010)), applying modified capture-mark-recapture protocols designed for remote surveying as per Marsh and Sinclair (1989) might be helpful when extrapolating population sizes from detection data. Moreover, distance sampling can account for IOV if observer is included as categorical covariate, random effect, or a predictor.

(3) Identification: Pairing or grouping observers to give double-observer data that allow for lack of consensus in species identification means that the resultant record can be marked as being potentially unreliable or deprioritised in analysis (Conn et al., 2014; Hoekman, 2021); "unknown" identification should be permitted rather than forcing observers to make an identification. Where possible, use spatially explicit models to account for differences in species distributions to account for likely species misidentification (Conn et al., 2014).

(4) Enumeration: Undertaking preliminary analysis specific to the type of survey and surrounding environment where observers enumerate artificial targets to quantify whether they are likely to underestimate (e.g. Frederick et al., 2003) or overestimate (e.g. Simons et al., 2007) group sizes. Consideration should also be given as to whether enumeration variation co-varies with group size, as here, or is group size independent (Frederick et al., 2003). Averaging group sizes from double-observer data might be helpful.

Acknowledgements

We thank Umer Bin Zia for helpful comments during project development and Wesi Kumwenda, Richie Fourie, Quinton Swart, and Willow Dawson for support in the field.

Data Availability Statement

Data available via University of Gloucestershire Research Repository https://eprints.glos.ac.uk/id/eprint/11299

Conflict of interest

The authors declare no conflict of interest.

Funding

No funding was received for this study.

References

Bothma, J. du P. & du Toit, J.G. (2016). Game Ranch Management 6th Edition. Pretoria, South Africa: Van Schaik Publishers.

Brack, I.V., Kindel, A. & Oliveira, L.F.B. (2018). Detection errors in wildlife abundance estimates from unmanned aerial systems (UAS) surveys: synthesis, solutions, and challenges. Methods in Ecology and Evolution, 9, 1864–1873.

Cherrill, A., (2015). Inter-observer variation in habitat survey data: investigating the consequences for professional practice. Journal of Environmental Planning and Management, 59, 1813-1832.

Conn, P.B., McClintock, B.T., Cameron, M.F., Johnson, D.S., Moreland, E.E. & Boveng, P.L. (2013). Accommodating species identification errors in transect surveys. Ecology, 94, 2607-2618.

Conn, P.B., Ver Hoef, J.M., McClintock, B.T., Moreland, E.E., London, J.M., Cameron, M.F., Dahle, S.P. & Boveng, P.L. (2014). Estimating multispecies abundance using automated detection systems: ice-associated seals in the Bering Sea. Methods in Ecology and Evolution, 5, 1280-1293.

Diefenbach, D. R., Brauning, D. W., & Mattice, J. A. (2003). Variability in grassland bird counts related to observer differences and species detection rates. The Auk, 120, 1168-1179.

Fowler, J & Cohen, L. (1996). Statistics for Ornithologists. Thetford, UK: British Trust for Ornithology.

Frederick, P. C., Hylton, B., Heath, J. A., & Ruane, M. (2003). Accuracy and variation in estimates of large numbers of birds by individual observers using an aerial survey simulator. Journal of Field Ornithology, 74, 281-287.

Goodenough A.E., Stafford R., Catlin-Groves C.L., Smith, A.L. and Hart A.G. (2010). Within-and among-observer variation in measurements of animal biometrics and their influence on accurate quantification of common biometric-based condition indices. Annales Zoologici Fennici, 47, 323-335.

Goodenough, A. E., & Hart, A. G. (2017). Applied ecology: monitoring, managing, and conserving. Oxford, UK: Oxford University Press.

Goodenough, A. E., Berry, D. L., Carpenter, W. S., Dawson, M., Furlong, N., Lamb, R. J., MacTavish, L., O'Reilly, N., Toms, H., Whitehead, L. H., & Hart, A.G. (2022). Data underpinning "Do you see what I see? Variation in detection, identification and enumeration of mammals during transect surveys". University of Gloucestershire Research Repository https://eprints.glos.ac.uk/id/eprint/11299.

Goodenough, A.E., Carpenter, W.S., McTavish, L., Blades, B., Clarke, E., Griffiths, S., Harding, N., Scott, R., Walsh, E., Wilson, L. and Hart, A.G. (2020). The impact of inter-observer variability on the accuracy, precision and utility of a commonly-used grassland condition index. Ecological Indicators, 117, 106664.

Hart, A.G., Dawson, M., Fourie R., MacTavish, L., & Goodenough, A.E. (2022). Comparing the effectiveness of camera trapping, driven transects and ad hoc records for surveying nocturnal mammals against a known species assemblage. Community Ecology, 23, 27-39.

Hobbs, R. C., & Waite, J. M. (2010). Abundance of harbor porpoise (*Phocoena phocoena*) in three Alaskan regions, corrected for observer errors due to perception bias and species misidentification, and corrected for animals submerged from view. Fisheries Bulletin, 108, 251-267.

Hodgson, A., Peel, D. & Kelly, N. (2017(. Unmanned aerial vehicles for surveying marine fauna: assessing detection probability. Ecological Applications, 27, 1253–1267.

Hodgson, J.C., Baylis S.M., Mott, R., Herrod, A. & Clarke, R.H. 2016. Precision wildlife monitoring using unmanned aerial vehicles. Scientific Reports, 6, 22574.

Hodgson, J.C., Mott, R., Baylis, S.M., Pham, T.T., Wotherspoon, S., Kilpatrick, A.D., Raja Segaran, R., Reid, I., Terauds, A. and Koh, L.P (2018). Drones count wildlife more accurately and precisely than humans. Methods in Ecology and Evolution, 9, 1160-1167.

Hoekman, S.T., (2021). Multi-observer methods for estimating uncertain species identification. Ecosphere, 12, e03648.

Kissling, M.L. & Garton, E.O. (2006). Estimating detection probability and density from point-count surveys: a combination of distance and double-observer sampling. The Auk, 123, 735-752.

Marsh, H. & Sinclair, D.F. (1989). Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. The Journal of Wildlife Management, 53, 1017-1024.

Pankakoski E., Väisänen R.A. & Nurmi K. (1987). Variability of muskrat skulls: measurement error, environmental modification and size allometry. Systematic Zoology, 36, 35-51.

Pollock, K.H., Marsh, H.D., Lawler, I.R., & Alldredge, M.W. (2006). Estimating animal abundance in heterogeneous environments: an application to aerial surveys for dugongs. The Journal of Wildlife Management, 70, 255-262.

Ronconi, R.A. & Burger, A.E. (2009). Estimating seabird densities from vessel transects: distance sampling and implications for strip transects. Aquatic Biology, 4, 297-309.

Ryan, P. G., & Cooper, J. (1989). Observer precision and bird conspicuousness during counts of birds at sea. South African Journal of Marine Science, 8, 271-276.

Schwarz, L.K., Gerrodette, T. & Archer, F.I. (2010). Comparison of closing and passing mode from a line-transect survey of delphinids in the eastern tropical Pacific Ocean. Journal of Cetacean Research and Management, 11, 253–265.

Shirley, M.H., Dorazio, R.M., Abassery, E., Elhady, A.A., Mekki, M.S. & Asran, H.H. (2012). A sampling design and model for estimating abundance of Nile crocodiles while accounting for heterogeneity of detectability of multiple observers. The Journal of Wildlife Management, 76, 966-975.

Simons, T.R., Alldredge, M.W., Pollock, K.H. & Wettroth, J.M. (2007). Experimental analysis of the auditory detection process on avian point counts. The Auk, 124, 986-999.

Spellerberg, I.F. (2005) Monitoring ecological change. Cambridge, UK: Cambridge University Press.

Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley, S.L., Bishop, J.R., Marques, T.A. & Burnham, K.P. (2010). Distance software: design and analysis of distance sampling surveys for estimating population size. Journal of Applied Ecology, 47, 5-14.

Table 1: GLM models of the factors underpinning variation in sighting detection and enumeration reported by five observers.

	Model 1: Variation in detection of sightings. Dependent variable = binary detected / not detected; n = 1,390 (278 sightings * 5 observers)			Model 2: Variation in enumeration of group size. Dependent variable = coefficient of variation; n = 294 sightings observed by ≥ two observers		
	χ^2	Direction	P	χ^2	Direction	P
Overall model	430.198	N/A	<0.001	45.707	N/A	<0.001
Survey type (walked/driven)	2.520	N/A	0.112	0.005	N/A	0.942
Detection distance (m)	19.425	-	<0.001	1.374	N/A	0.241
Mean visibility (rank 1-5)	83.129	+	<0.001	40.598	-	<0.001
Mean group size	12.890	+	<0.001	10.744	+	0.001
Observer ID	202.407	IOV	<0.001			

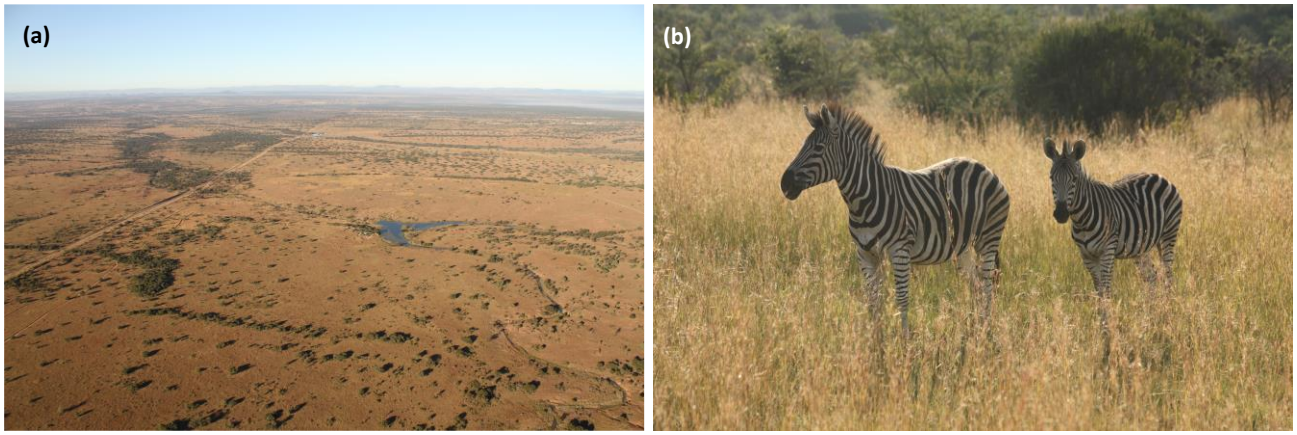

Figure 1: Study site from (a) air and (b) land showing typical habitat and visibility.

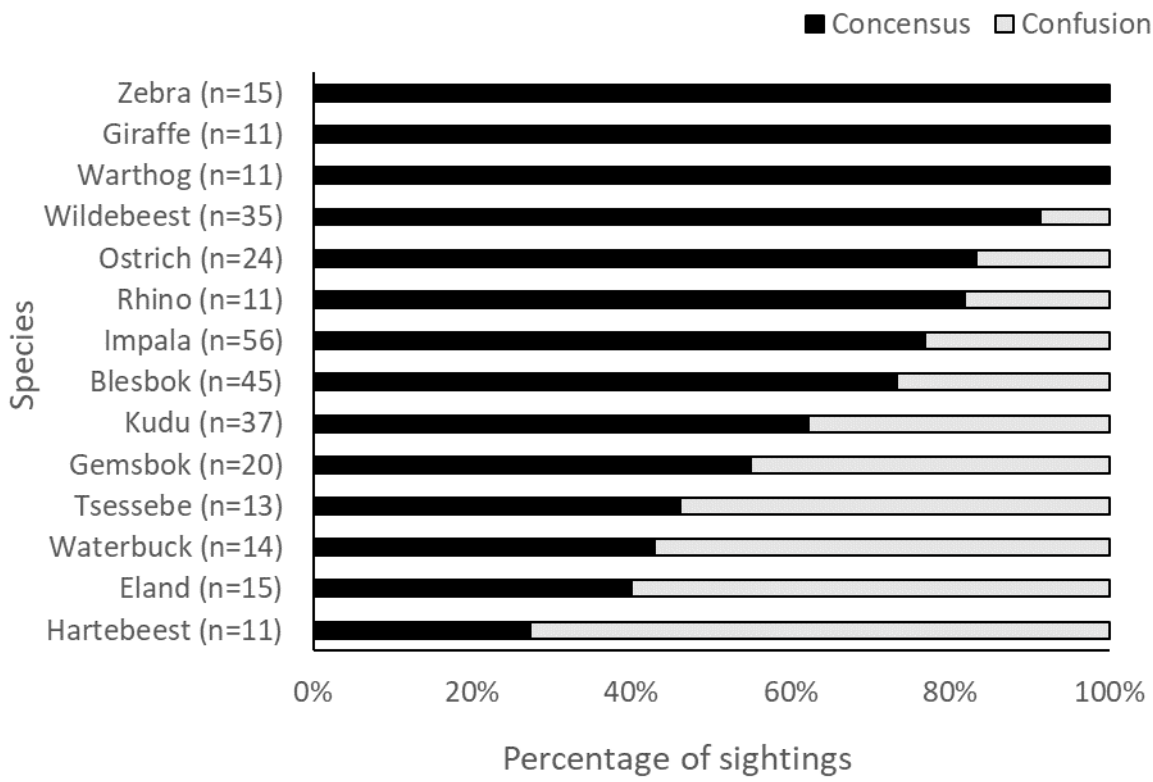

Figure 2: Species-specific differences in whether all observers had recorded the same species identification ("consensus") or whether there were differences between observers ("confusion"). Overall sample size = 318 species identifications across the 258 unique sightings that were recorded by ≥ two observers *and* where all observers agreed only one species was present, such that it was clear what species had been confused. Species-specific sample sizes are displayed on graph; only species where n ≥ 10 sightings were included.

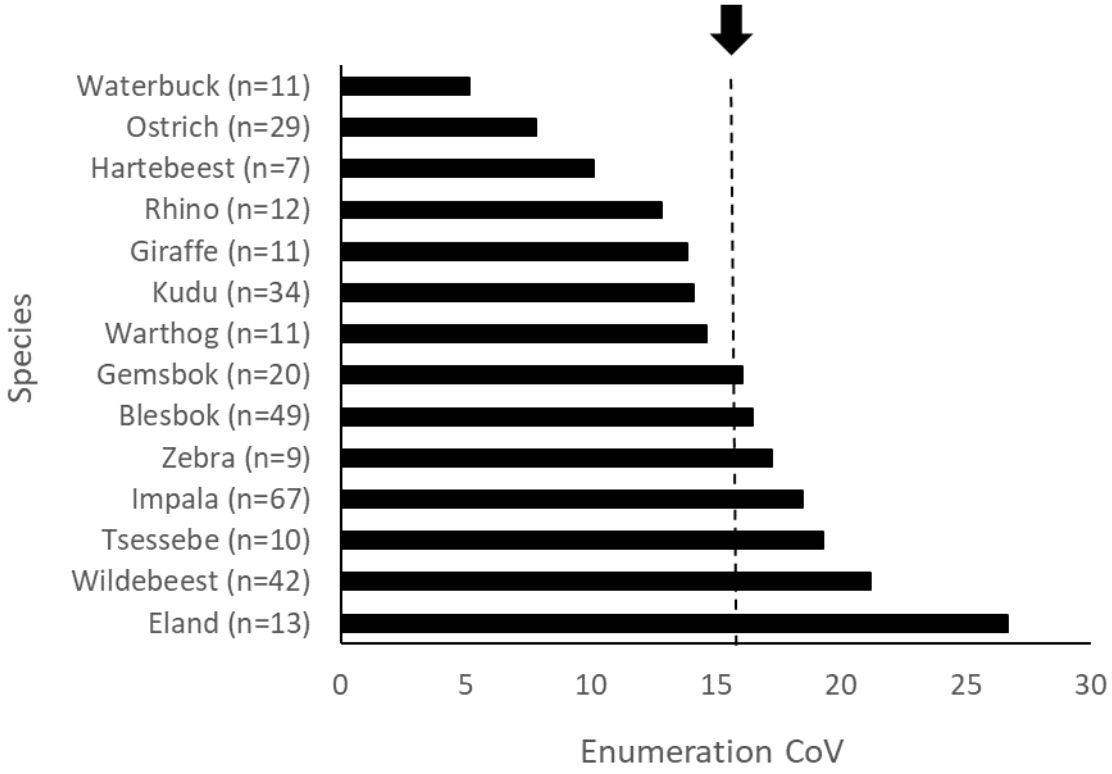

Figure 3: Species-specific differences in inter-observer variation in enumerating group size based on Coefficient of Variation (CoV); the arrow and dashed line identifies overall mean CoV across all species. Overall sample size = 325 species enumeration records across the 294 unique sightings that were recorded by ≥ two observers. Species-specific sample sizes are displayed on graph.