Title: Safari Science: Assessing the reliability of citizen science data for wildlife surveys

Authors: Cara Steger¹, Bilal Butt², Mevin B. Hooten³

¹ Natural Resource Ecology Lab, Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO 80523-1499.

² School for Environment and Sustainability, University of Michigan, Ann Arbor MI 48103.
bilalb@umich.edu

³ U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit; Departments of Fish, Wildlife & Conservation Biology and Statistics, Colorado State University, Fort Collins, CO 80523. mevin.hooten@colostate.edu

Corresponding author: Cara Steger, cara.steger@colostate.edu, 224-261-4184. Natural Resource Ecology Lab, Campus Delivery 1499, Colorado State University, Fort Collins, CO 80523-1499.

Running Title: Safari Science

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1365-2664.12921

This article is protected by copyright. All rights reserved
Summary

1. Protected areas are the cornerstone of global conservation, yet financial support for basic monitoring infrastructure is lacking in 60% of them. Citizen science holds potential to address these shortcomings in wildlife monitoring, particularly for resource-limited conservation initiatives in developing countries – if we can account for the reliability of data produced by volunteer citizen scientists (VCS).

2. This study tests the reliability of VCS data vs. data produced by trained ecologists, presenting a hierarchical framework for integrating diverse datasets to assess extra variability from VCS data.

3. Our results show that, while VCS data are likely to be overdispersed for our system, the overdispersion varies widely by species. We contend that citizen science methods, within the context of East African drylands, may be more appropriate for species with large body sizes, which are relatively rare, or those that form small herds. VCS perceptions of the charisma of a species may also influence their enthusiasm for recording it.

4. Tailored program design (such as incentives for VCS) may mitigate the biases in citizen science data and improve overall participation. However, the cost of designing and implementing high quality citizen science programs may be prohibitive for the small protected areas that would most benefit from these approaches.

5. Synthesis and applications. As citizen science methods continue to gain momentum, it is critical that managers remain cautious in their implementation of these programs while working to ensure methods match data purpose. Context-specific tests of citizen science data quality can improve program implementation, and separate data models should be used when volunteer citizen scientists’ variability differs from trained ecologists’ data. Partnerships across protected areas and between protected areas and other conservation institutions could help to cover the costs of citizen science program design and implementation.

Key-words: Bayesian, East Africa, hierarchical modelling, uncertainty, data reconciliation, validation, citizen science, wildlife surveys, protected areas, monitoring

Introduction
In an attempt to stem biodiversity loss, over 200,000 protected areas have been created worldwide - resulting in a rate of establishment that far exceeds our capacity to monitor (Ervin 2003; Chape et al. 2005; IUCN & UNEP-WCMC 2016). Protected areas have become the cornerstone of global conservation (Caro 2011), yet financial support for basic monitoring infrastructure is lacking in 60% of them (Leverington et al. 2010). Citizen science, or public participation in scientific research, has the potential to address these shortcomings in protected area monitoring by decreasing monitoring costs and increasing public engagement in conservation issues (Bonney et al. 2009). Citizen science programs can cover larger geographic areas and time scales than individual research teams (Silvertown 2009; Dickinson, Zuckerberg & Bonter 2010). A growing number of programs are now being deployed, potentially mitigating the high costs of species monitoring by resource-limited governments and conservation organizations (Danielsen, Burgess & Balmford 2005; Chandler et al. 2012).

Despite the potential of citizen science for species monitoring, questions remain over the reliability and quality of data produced by volunteer citizen scientists (VCS), and how these data differ from those produced by professional scientists and resource managers (Cohn 2008; Bonney et al. 2014). A recent review estimated only 3% of articles related to citizen science were investigations of data validation techniques (Follett & Strezov 2015) and there is varied evidence on the quality, reliability, and utility of citizen science data for wildlife monitoring. For example, Bernard et al. (2013) found that data collected by VCS were of lower quality than scientists’ when looking at the community-level fish abundance in a marine ecosystem, while others found no difference (Darwall & Dulvy 1996). Bernard et al. (2013) report that VCS and scientists produced similar data when counting only common species with high detection probabilities. Dickinson et al. (2010) observe the opposite, where VCS tend to over-report rare species and under-report common species. Delaney et al. (2007) reported age as a significant predictor of VCS observer skill, while Newman et al. (2003) did not find this to be the case. Ericsson and Wallin (1999) suggest that VCS observations of moose (*Alces alces*) reflect the overall population size and fluctuations well, but these data cannot directly replace conventional (and expensive) sampling methods unless they first undergo a calibration process. These findings demonstrate that citizen science data quality and reliability varies widely depending on the species and ecosystem (Dickinson, Zuckerberg & Bonter 2010; Vianna et al. 2014).
Citizen science has great relevance for African countries where funding for basic monitoring is extremely limited. In East Africa, this need is particularly acute given that wildlife densities and distributions are some of the highest in the world (Waithaka 2004), and that there are rapid population declines occurring in and around protected areas (Homewood et al. 2001; Craigie et al. 2010; Ogutu et al. 2016). Scientists have studied these declining populations for decades, yet there remain differing claims as to the drivers of the declines (Homewood et al. 2001; Odadi et al. 2011; Butt & Turner 2012). Part of this confusion is because existing abundance estimates of East African mammals are conducted through aerial surveys with coarse spatial resolution (5km²) (Norton-Griffiths 1978), thus making inference on the causal mechanisms of declines difficult to support (Ogutu et al. 2011; Bhola et al. 2012a; Bhola et al. 2012b; Ogutu et al. 2016).

In cases where new and smaller protected areas are being established, information on species distributions and landscape preferences at local scales are critical to develop adequate conservation plans (Rushton, Ormerod & Kerby 2004; Ogutu et al. 2006). Mismatch between suitable habitat and realized species distributions is predicted by niche theory, as biotic interactions, dispersal, and spatiotemporal habitat variability are thought to restrict the fundamental niche of a species (Hutchinson 1957; Pulliam 2000). Furthermore, ecological inference is impacted by the scale at which habitat selection is monitored; studies conducted at different spatial scales can yield markedly different results (Garshelis 2000). The high degree of spatial and temporal heterogeneity in primary productivity within East African dryland ecosystems requires finer-scale understanding of species density and distributions (Gillson 2004). However, methods that accurately capture wildlife locations at fine scales over large landscapes have proven to be prohibitively difficult, expensive and time-consuming given the financial and logistical constraints of East African protected areas (Witmer 2005; Lung & Schaab 2010). Citizen science therefore appeals greatly to protected area managers, but tests of citizen science data reliability in East African dryland landscapes remain scarce.

Given these contexts, our analysis tests differences in the quality and quantity of data produced from citizen science methods of wildlife monitoring vs. traditional ecological sampling, with results that are directly relevant to protected area managers throughout East Africa.
Materials and methods

Study Site

The study was conducted in the Olare Motorogi Conservancy (1°21′00″S, 35°12′00″E), bordering the Maasai Mara National Reserve in Kenya. Established in 2006, the 137 km² conservancy is comprised of land leased from Maasai landowners (Butt 2014). Elevation ranges between 1525m (SW) and 1762m (NE). Vegetation is open grass plains, with patches of Acacia gerrardii and Terminalia trees, shrublands, and riverine woodlands (Bhola et al. 2012a). The region forms the northern extension of the wildlife rich Serengeti-Mara Ecosystem (Waithaka 2004). Rainfall is bimodal, with the short rains in late November-December, and long rains in March-June (Pennycuick & Norton-Griffiths 1976). Maasai pastoralists sometimes graze their livestock within the protected area and have been historically sympatric with wildlife. Tourists from five tourist lodges within the conservancy regularly go on wildlife viewing game drives (Figure 1). This makes the conservancy one of the least densely visited protected areas in the region.

Fieldwork

Fieldwork was conducted during a 13 week period (May to August 2013) to compare two independent methods of distance sampling of wildlife. In method 1, “Safari Science,” tourists as citizen scientists voluntarily collected data using inexpensive handheld tablet computers equipped with an open source application while on safari game drives. In method 2, “Traditional Sampling,” a research team collected data using ground-vehicle line transect surveys. This method is widely used for scientifically estimating animal populations (Ogutu et al. 2006; Singh & Milner-Gulland 2011).

Safari Science Method: Tourists were recruited from four lodges in the conservancy. Participants were given a 15-minute training program before departing on game drives, explaining the process of using the tablet and software, rangefinder, compass, and estimating animal counts. A card containing the procedure and a set of frequently asked questions was inserted into the casing of the tablet for future reference by participants.
Traditional Sampling Method: Four road transects (each ~ 11.5 km) were selected in collaboration with conservancy management that were representative of the extent and density of tourist visitation within the conservancy (Figure 1).

To account for diurnal fluctuations, both methods involved morning (06:00-10:00) and afternoon (14:00-18:00) sampling sessions. The citizen science method was conducted on an opportunistic basis (given the voluntarily exercise) while each transect was sampled once a week (twice in a single day – morning and evening). Observations were obtained as follows: (i) upon sighting an animal or group of animals, the vehicle stopped and a GPS location was recorded; (ii) distance to the animal (or to the centre of the group) was measured with a Simmons laser rangefinder (model 801405, 548m range) up to a maximum distance of 200m; (iii) direction to the animal (or group) from the observer location in the vehicle was taken with a Suunto compass; (iv) species and numbers of animals were counted (Table 1). We focus on eight species, including elephant (*Loxodonta africana*), giraffe (*Giraffa camelopardalis*), Grant’s gazelle (*Nanger granti*), impala (*Aepyceros melampus*), Thomson’s gazelle (*Eudorcas thomsonii*), topi (*Damaliscus korrigum*), wildebeest (*Connochaetes taurinus*), and zebra (*Equus quagga burchellii*) because they are the most common within the ecosystem and represent a range of guilds and body sizes.

Data Processing

In addition to the wildlife sighting data, we include four environmental covariates in our hierarchical model. These covariates represent the primary environmental determinants of large mammal distributions in East African drylands. A 30-meter ASTER DEM was used to generate slope and elevation layers and calculate the *Topographic Wetness Index* (TWI) which is related to the potential for higher plant productivity (Wilson & Gallant 2000). TWI was calculated as,

\[ TWI = \log \left( \frac{AS}{s} \right) \]

where AS is the catchment area (determined by elevation) and s is the slope. Topographic wetness has been correlated with species abundance in this region (Bhola *et al.* 2012b), and ecologists have noted certain species’ tendency to prefer soils with good drainage and thus firm footing (Estes 1991). Additionally, areas with high TWI are expected to have increased predation risk and low food quality (Ogutu *et al.* 2006; Anderson *et al.* 2010). The DEM was
then used to create streams and calculate distance to river. Distance to river is a commonly used covariate that can represent increased predation risk due to high vegetative cover (Hopcraft, Sinclair & Packer 2005; Anderson et al. 2010). It can also reflect a species inability to go long periods without drinking (Estes 1991). TWI and distance to river values were extracted to the wildlife location points.

The third covariate is the Normalized Difference Vegetation Index (NDVI) – a reliable proxy for above ground biomass. NDVI is used to represent photosynthetically active vegetation (Pettorelli et al. 2005), and has been shown to correlate with species abundances in this area (Bhola et al. 2012b). However, the scale at which NDVI is measured has been shown to affect the strength and sign of the relationship for topi abundance in particular (Bro-Jorgensen, Brown & Pettorelli 2008). NDVI was derived from five Landsat 8 scenes at 30m resolution, which were obtained from the USGS/Earth Explorer interface. We calculated NDVI for each scene, removed the second scene (15 June) due to excessive cloud cover, and calculated a surrogate using the average NDVI of the 30 May and 1 July scenes. The wildlife location data were separated into groups spanning 16 days, which were centred on each of the Landsat scenes, and the NDVI values were extracted to those locations.

Finally, we include a distance to livestock covariate given the sympatricity of wildlife and livestock within the ecosystem. Other studies in this area have used livestock density to explain wildlife abundance, often inferring competition from results that show wildlife occurring in different areas than livestock (Bhola et al. 2012b). Alternatively, many scholars contend that livestock grazing facilitates wildlife grazing and movement (Reid, Galvin & Kruska 2008; Butt, Shortridge & WinklerPrins 2009). Data on the location and abundance of livestock over the study period was separated into the same 16-day groups as the NDVI scenes and Euclidean distance calculations to wildlife were used to generate the variable.

Analysis

We constructed a hierarchical modeling framework for integrating citizen science and traditional sampling wildlife count data that accounts for variability introduced by the citizen science data. The effective combination of datasets of varying strengths and weaknesses is a common problem in ecological modeling, and our approach follows what Hanks et al. (2011) refer to as “Bayesian data reconciliation.” Our approach allows the modeler to assess the quality
of citizen science data in relation to a reliable baseline of traditionally collected data. We model
the citizen science data using the mean of the traditional sampling data in order to understand
whether the citizen science dataset show signs of overdispersion (i.e., whether the variance of the
citizen science data exceeds the mean of the traditional sampling data). In this context, we
interpret overdispersion as extra variability introduced by the observer and reporting biases of
citizen science data. Because our model relates traditional sampling counts to a set of
environmental variables, we are able to predict wildlife abundance at citizen science locations.
Comparing predicted abundance to VCS observations allows us to derive estimates of over- and
under-reporting by VCS across the landscape in relation to wildlife species and landscape
characteristics.

In our Bayesian hierarchical framework, the data model represents the relationship
between the Safari Science data set ($w_i$) and the traditional sampling data set ($y_i$), and can be
written as,

$$w_i \sim \text{negative binomial}(y_i, N_c),$$

where $y_i$ is the mean of the negative binomial distribution, estimated from the traditional
sampling data, and $N_c$ is the overdispersion parameter for the citizen science data. The
overdispersion (or “size”) parameter accounts for extra variability in the data, with smaller
values indicating higher variability. In practice, when $N_c > 10$ the variance approaches the mean
and the negative binomial distribution is virtually identical to a Poisson distribution; however, in
ecological data, $N_c$ is often less than 1, indicating strong overdispersion (Bolker 2008).

The process model represents the distribution of the traditional sampling data,

$$y_i \sim \text{Poisson}(\lambda_{TS}),$$

and relates the log of animal intensity ($\lambda_{TS}$) to a set of four environmental covariates

$$\log(\lambda_{TS}) = X\beta,$$

where $X$ represents a matrix of values for the four covariates. These covariates were standardized
to have mean 0 and standard deviation 1, and the associated coefficients were assigned a
multivariate normal prior,

$$\beta \sim N(0, \sigma^2 I),$$

This article is protected by copyright. All rights reserved
where \( \sigma^2 = 100 \). The regression coefficient for the intercept was assigned a similar prior, but with mean 1. Finally, the overdispersion parameter \( (N_c) \) was assigned a gamma prior with mean 0.5 and variance 0.125,

\[
N_c \sim \text{gamma}(2, 0.25).
\]

We fit the model using a Monte Carlo Markov Chain (MCMC) algorithm with Metropolis-Hastings updates. We generated 100,000 samples using our MCMC algorithm, retaining 90,000 after determining that a burn in of 10,000 samples was adequate. The trace plots showed excellent mixing. A simulation study showed that no bias was introduced through the modeling framework because simulation parameters were recovered by the model within a 95% credible interval (CI). Posterior means and CI were calculated after model convergence (Table 2) and the simulated marginal posterior distributions for \( \mathbf{\beta} \) and \( N_c \) are plotted with means and prior distributions for reference (Figure 2, for topi only; see Supporting Information Figures for other species). The bias of the Safari Science data set was derived as the difference between the predicted \( y_{\text{unobserved}} \) values and the actual Safari Science counts at those locations (Figure 3, for topi only). These errors were plotted spatially to aid our understanding of whether landscape characteristics might influence the over- or under-counting of species by VCS (Figure 4, for topi and elephant only). All analyses were performed in R (R Core Development Team 2008), and the code for our model is provided in the Supporting Information (Appendix S1). This model was fit to data on each of the eight species in our analysis.

**Results**

**Safari Science Effort and Error**

VCS recorded very few data points in this study compared to the Traditional Sampling team (Table 1). Safari Science data are overdispersed for most species, with \( N_c \) values under 1 (Table 2; Bolker 2008). Of our eight species, only Grant’s gazelle and giraffe had \( N_c \) values whose CI exceeded 10, (i.e., their distribution is virtually indistinguishable from a Poisson distribution). Wildebeest data were highly overdispersed, with \( N_c = 0.01 \). Thomson’s gazelle and Grant’s gazelle were both under-counted by VCS, as they reported an average of 1 and 4 individuals fewer (respectively) than the predicted \( y_{\text{unobserved}} \) values at those locations (Table 2). All other species were over-counted, sometimes by an average of over 270 individuals in the
case of wildebeest. The wildebeest model also had the largest RMSE of all species by a large margin, eight times higher than the RMSE in our simulation study (Table 2). All other species displayed RMSE values that were lower than our simulation study, and over-estimated counts ranging from 3 – 10 individuals.

**NDVI**

Elephant, Grant’s gazelle, Thomson’s gazelle, and zebra models reflected positive relationships between species abundance and NDVI, indicating greater abundance in areas of high vegetative productivity. Giraffe and wildebeest displayed negative relationships. Topi and impala showed no significant relationship.

**Distance to River**

Thomson’s gazelle, topi, and zebra models reflected positive relationships between species abundance and distance to river, indicating that those species avoid riparian areas. Grant’s gazelle, impala, and wildebeest displayed negative relationships. Elephant and giraffe showed no significant relationship.

**Topographic Wetness Index**

Grant’s gazelle, impala, topi, and zebra models reflected negative relationships between species abundance and TWI, indicating those species avoid areas with poor drainage. Wildebeest displayed a positive relationship. Elephant, giraffe, and Thomson’s gazelle showed no significant relationship.

**Distance to Livestock**

All models reflected significant relationships between wildlife abundance and distance to livestock. Giraffe, impala, Thomson’s gazelle, wildebeest, and zebra models reflected negative relationships between species abundance and distance to livestock, indicating that larger numbers of those species are often found in close proximity to livestock. Elephant, Grant’s gazelle and topi showed positive relationships.

**Discussion**

This article is protected by copyright. All rights reserved
Citizen science data are increasingly used to model the distribution and abundance of wildlife species (Dickinson et al. 2012; Ruiz-Gutierrez, Hooten & Campbell Grant 2016), yet questions remain as to whether these data can be useful in approximating the quality and quantity of traditionally collected ecological data. Our results demonstrated that, while citizen science data are likely to be overdispersed, the amount of overdispersion varies widely by species. We find that citizen science approaches, within the context of East African dryland systems, may be more appropriate for species with large body sizes, that are relatively rare, or those that form small herds. Additionally, the data model we present is a general guide for future work that seeks to integrate any datasets of varying reliability, such as data produced by multiple laboratory technicians or field assistants. The specific results of our study can also be used as informative priors for other East African hierarchical models using citizen science data, thus improving their convergence and inferential power (Hobbs & Hooten 2015).

Safari Scientists recorded very few data points in this study compared to the traditional sampling team (Table 1), which we attribute to program design (Steger & Butt 2015). The combination of rangefinder, compass, and tablet was overwhelming for several VCS, who were often trying to balance Safari Science use with their own photography. Complex protocols for VCS can result in reduced data accuracy (Dickinson et al. 2010), though simpler protocols do not necessarily produce the type of data needed for system understanding. For example, GPS-enabled cameras may be a viable alternative citizen science method in this area that would be more attractive to VCS. However, this method would result in substantial interpretive work for end users (i.e., counting animals in photos; Maddock & Mills 1994), and the lack of data on direction and distance to animal would make it impossible to determine fine-scale animal locations.

Another barrier in our program design was the lack of incentives, such as personalized maps and wildlife counts for game drives. The addition of these rewards could feasibly increase both recruitment rates and the number of observations recorded by individual VCS (Hochachka et al. 2012), although there are financial and logistical costs associated with designing and implementing such a program (Hamilton et al. 2012). Finally, it is possible that tourists in high-end lodges like the ones in Olare Motorogi Conservancy may be less likely to engage with citizen science if they feel it detracts from their costly and time-constrained holiday experience.
Incorporating research game drives into the packages being sold to tourists may help manage tourists’ expectations before arriving at the lodge by establishing the research as an integral part of the safari experience (Steger & Butt 2015). Designing citizen science programs that collect reliable data while also meeting the interests and motivations of the participants will be a critical challenge for the success of future endeavours.

The ratio of citizen science to traditional sampling observations differs considerably across species, casting doubt as to whether VCS accurately report every species they observe. VCS reported a similar number of elephant sightings as trained ecologists (Table 1), but drastically under-reported Thomson’s gazelle, which were the most common species encountered by the traditional sampling team. Our models reveal overdispersion in six of the eight species-specific models, indicating that data produced by VCS are not equivalent to data collected using trained ecological researchers. However, the level of overdispersion varies widely by species, revealing that VCS in this context display effort at recording only certain species accurately. There are several explanations for overdispersion in the Safari Science data, which we interpret as arising from observer or reporting bias in the count process. While many factors influence overdispersion (e.g., weather, predation risk, and species movement; Linden & Mantyniemi 2011), we suggest that the amount of effort to minimize differences in data collection between our two methods (e.g., restricting observations to the conservancy, synchronizing sampling times) validates our interpretation of overdispersion as a measure of observer error.

Our findings that Grant’s gazelle and giraffe were the only species without highly overdispersed data are attributed to the rarity of these species within the conservancy; game drivers have a tendency to stop and point out rare or uncommon species to tourists, making those species appealing to record. Giraffe and Grant’s gazelle also travel in small herds, making them easier to count than wildebeest or Thomson’s gazelle. Elephants, which are easy to count due to their large body size, showed less overdispersion than the rest of the species. Therefore, the rarity of a species, body size, and small herd size are all characteristics explaining why VCS are better at estimating those numbers than for other species. The charisma of a species also makes it more appealing for VCS to record and observe. VCS in this study showed a strong preference for carnivores, reporting four times as many lion encounters as were observed using traditional
sampling methods (Steger & Butt 2015). These data were often double-counts of the same lion pride, and thus require cleaning by analysts, but the quantity of the data indicates that enthusiasm for a particular species can improve the participation of VCS.

Conversely, the wildebeest model displayed a very high degree of overdispersion in the Safari Science data, which may be attributable to the large numbers that migrate during these months (Serneels & Lambin 2001). When species occurrence is high, it is perhaps unreasonable to expect tourists to accurately estimate numbers within the 200m area closest to the road. It is also likely that game drivers did not give tourists the opportunity to stop and record an individual wildebeest if they encountered it alone, as this would detract from finding and observing more rare species like lion, cheetah, and leopard (Steger & Butt 2015). Participant observation in both methods revealed that the traditional sampling team was likely to record each individual encountered, accounting for the over-estimation seen in the Safari Science dataset. Because 1-2 animals is the most common count in the traditional sampling wildebeest dataset, predictions at Safari Science locations tend to be lower than the large groups tourists stopped to record. Therefore, the Safari Science dataset does not accurately represent the distribution of wildebeest at fine scales, as it is biased toward large aggregations.

Finally, we determined whether certain landscape characteristics influenced VCS count biases in certain locations. A spatial plot of count differences reveals that the largest over-estimations of topi occur on the western side of the conservancy, closest to the tourist lodges along the Ntiakitiak River, while the eastern area have mostly under-estimations (Figure 4). However, this pattern is reversed for elephants, where over-estimates are concentrated on the eastern side of the conservancy, and under-estimates on the western side (Figure 4). Despite the clear spatial correlation of these count differences, we were unable to identify landscape characteristics influencing the ability of VCS to accurately count animals in those areas. One possibility is that vegetation type, rather than vegetation productivity measured as NDVI, could be a physical difference in these locations. However, it is difficult to ascertain land cover in dryland systems due to the dominant spectral signal of the soil over vegetation that experiences only brief growing periods (Bartholomé & Belward 2005).

Management Implications
Our study has several implications for protected area management strategies throughout East Africa. As citizen science methods are gaining momentum in the region, it is important that managers remain cautious in their implementation of these programs while working to ensure methods match the data purpose. The presence of variability does not necessarily preclude citizen science data from being useful to managers, but that variability must be measured and accounted for during decision-making (Milner-Gulland & Shea 2017). In our study, the level of overdispersion varies widely according to species, and managers might be content with the uncertainty in citizen science data for certain species. For example, over-estimating topi counts by six individuals might be an acceptable range of error in a species with herd sizes that range up to 200, but small herds like giraffe have a lower margin of error, and an over-estimation of only 2 individuals might be unacceptable to managers. These findings advance our understanding of citizen science program design and implementation throughout East African drylands, and support programs conducting a similar methodological comparison in their specific contexts before investing in large scale citizen science programs for wildlife monitoring.

It is critical that managers recognize the interests and motivations of VCS when designing site-specific citizen science programs. In the context of this study, narrowing the list of species to charismatic species that are easily counted may have made the task seem more feasible and appealing to VCS, potentially resulting in higher quantities of data and thus mitigating some of the extra variability of those data. Similarly, the implementation of reward systems may result in greater buy-in from VCS. Citizen science monitoring is more cost effective than traditional sampling, but the costs of designing and maintaining a high quality program may still be prohibitive for the small protected areas that would most benefit from citizen science methods (Hamilton et al. 2012). Partnerships that share costs with other protected areas, conservation NGOs, or government agencies hold great potential for the implementation of high quality citizen science programs that transcend logistical barriers in East African drylands.

**Authors’ Contributions**

CS and BB conceived the idea for the study and designed the methodology; CS collected the data; CS and MH specified the model and analysed the data; CS led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for the manuscript.
Acknowledgements

This study was funded by the University of Michigan’s African Studies Center, Department of Afroamerican and African Studies, Rackham Graduate School, and the School of Natural Resources and Environment. We thank our field assistants, the Olare Motorogi Conservancy, and the Safari Scientists who made this all possible. Special thanks to Veronica Burt and Inés Ibáñez for help with the initial model conception. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Data accessibility

We share our data through Dryad Digital Repository http://dx.doi.org/10.5061/dryad.mb7qk (Steger et al. 2017). Although these data have been processed successfully on a computer system at the U.S. Geological Survey (USGS), no warranty expressed or implied is made regarding the display or utility of the data on any other system or for general or scientific purposes, nor shall the act of distribution constitute any such warranty. The USGS or the U.S. Government shall not be held liable for improper or incorrect use of the data described and/or contained herein.

Supporting Information

Appendix S1. R code for model
Figures S1 – S20. Posterior histograms, error plots, and spatial count plots for all species.

References


This article is protected by copyright. All rights reserved


### Tables

**Table 1.** Number and range of observations for both methods. Traditional Sampling followed standardized transects averaging 11.5km, while Safari Science observations were made opportunistically on game drives of varying lengths.

<table>
<thead>
<tr>
<th></th>
<th>Traditional Sampling</th>
<th>Safari Science</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td><em>Mean (SD)</em></td>
<td><em>Mean (SD)</em></td>
</tr>
<tr>
<td>Elephant</td>
<td>33</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>4.55 (3.76)</td>
<td>6.15 (5.66)</td>
</tr>
<tr>
<td>Giraffe</td>
<td>54</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>4.69 (4.58)</td>
<td>5.82 (3.84)</td>
</tr>
<tr>
<td>Grant's gazelle</td>
<td>244</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>4.08 (3.54)</td>
<td>6.44 (5.04)</td>
</tr>
<tr>
<td>Impala</td>
<td>607</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>11.45 (22.87)</td>
<td>21.26 (28.59)</td>
</tr>
<tr>
<td>Thomson’s</td>
<td>1313</td>
<td>14</td>
</tr>
<tr>
<td>Animal</td>
<td>Length (cm)</td>
<td>Width (cm)</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>gazelle</td>
<td>10.70 (13.17)</td>
<td>9.57 (12.53)</td>
</tr>
<tr>
<td>Topi</td>
<td>700</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>7.50 (13.78)</td>
<td>13.91 (35.08)</td>
</tr>
<tr>
<td>Wildebeest</td>
<td>896</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>39.11 (92.58)</td>
<td>310.88 (697.53)</td>
</tr>
<tr>
<td>Zebra</td>
<td>374</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>16.87 (25.71)</td>
<td>25.38 (43.40)</td>
</tr>
</tbody>
</table>
Table 2. Posterior means and credible intervals for all species. Significant species-environment relationships (where the credible interval does not include zero) are highlighted in bold.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Elephant</th>
<th>Giraffe</th>
<th>Grant’s gazelle</th>
<th>Impala</th>
<th>Thomson’s gazelle</th>
<th>Topi</th>
<th>Wildebeest</th>
<th>Zebra</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>(2.5%, 97.5%)</td>
<td>(2.5%, 97.5%)</td>
<td>(2.5%, 97.5%)</td>
<td>(2.5%, 97.5%)</td>
<td>(2.5%, 97.5%)</td>
<td>(2.5%, 97.5%)</td>
<td>(2.5%, 97.5%)</td>
<td>(2.5%, 97.5%)</td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td>1.03</td>
<td>1.48</td>
<td>1.41</td>
<td>1.38</td>
<td>2.43</td>
<td>2.36</td>
<td>1.91</td>
<td>3.51</td>
</tr>
<tr>
<td></td>
<td>(0.98, 1.09)</td>
<td>(1.38, 1.50)</td>
<td>(1.33, 1.38)</td>
<td>(1.35, 1.42)</td>
<td>(2.41, 2.44)</td>
<td>(2.35, 2.37)</td>
<td>(1.91, 1.95)</td>
<td>(3.50, 3.51)</td>
</tr>
<tr>
<td><strong>NDVI (B2)</strong></td>
<td>0.48</td>
<td>0.20</td>
<td>-0.28</td>
<td>0.05</td>
<td>-0.01</td>
<td>0.09</td>
<td>-0.004</td>
<td>-0.59</td>
</tr>
<tr>
<td></td>
<td>(0.45, 0.52)</td>
<td>(0.09, 0.30)</td>
<td>(-0.38, -0.20)</td>
<td>(0.01, 0.09)</td>
<td>(-0.02, 0.006)</td>
<td>(0.08, 0.09)</td>
<td>(-0.02, 0.01)</td>
<td>(-0.60, -0.58)</td>
</tr>
<tr>
<td><strong>Distance to River (B3)</strong></td>
<td>0.48</td>
<td>-0.08</td>
<td>0.03</td>
<td>-0.06</td>
<td>-0.1</td>
<td>0.07</td>
<td>0.39</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>(0.45, 0.52)</td>
<td>(-0.18, 0.01)</td>
<td>(-0.06, 0.10)</td>
<td>(-0.10, -0.02)</td>
<td>(-0.12, -0.09)</td>
<td>(0.06, 0.08)</td>
<td>(0.37, 0.40)</td>
<td>(-0.12, -0.10)</td>
</tr>
<tr>
<td><strong>TWI (B4)</strong></td>
<td>0.5</td>
<td>0.03</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.08</td>
<td>-0.01</td>
<td>-0.04</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.46, 0.53)</td>
<td>(-0.08, 0.14)</td>
<td>(-0.11, 0.05)</td>
<td>(-0.08, -0.003)</td>
<td>(-0.09, -0.06)</td>
<td>(-0.02, 0.002)</td>
<td>(-0.06, -0.02)</td>
<td>(0.08, 0.09)</td>
</tr>
<tr>
<td><strong>Distance to Livestock (B5)</strong></td>
<td>0.48</td>
<td>0.08</td>
<td>-0.17</td>
<td>0.06</td>
<td>-0.05</td>
<td>-0.03</td>
<td>0.04</td>
<td>-0.21</td>
</tr>
<tr>
<td></td>
<td>(0.44, 0.52)</td>
<td>(2e-3, 0.17)</td>
<td>(-0.26, -0.08)</td>
<td>(0.03, 0.10)</td>
<td>(-0.06, -0.03)</td>
<td>(-0.04, -0.02)</td>
<td>(0.03, 0.06)</td>
<td>(-0.21, -0.20)</td>
</tr>
<tr>
<td><strong>Safari Science Error Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>0.1</td>
<td>1.51</td>
<td>5.54</td>
<td>5.65</td>
<td>0.31</td>
<td>0.8</td>
<td>0.27</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.03, 0.19)</td>
<td>(0.58, 3.48)</td>
<td>(1.91, 11.62)</td>
<td>(2.37, 11.05)</td>
<td>(0.18, 0.50)</td>
<td>(0.39, 1.39)</td>
<td>(0.15, 0.43)</td>
<td>(0.005, 0.02)</td>
</tr>
<tr>
<td><strong>Count Difference</strong></td>
<td>-20.1</td>
<td>-1.22</td>
<td>-1.72</td>
<td>4.34</td>
<td>-10.04</td>
<td>1.22</td>
<td>-6.13</td>
<td>-271.15</td>
</tr>
<tr>
<td></td>
<td>(-20.81, -19.34)</td>
<td>(-1.83, -0.58)</td>
<td>(-2.29, -1.10)</td>
<td>(3.77, 4.91)</td>
<td>(-10.79, -9.30)</td>
<td>(0.24, 2.24)</td>
<td>(-6.68, -5.56)</td>
<td>(-272.38, -269.91)</td>
</tr>
<tr>
<td><strong>RMSE</strong></td>
<td>84</td>
<td>5.46</td>
<td>3.65</td>
<td>4.34</td>
<td>29.64</td>
<td>11.82</td>
<td>34.56</td>
<td>738.65</td>
</tr>
<tr>
<td></td>
<td>(82.62, 85.21)</td>
<td>(4.81, 6.18)</td>
<td>(3.09, 4.23)</td>
<td>(3.35, 5.33)</td>
<td>(28.90, 30.34)</td>
<td>(10.78, 12.83)</td>
<td>(33.96, 35.12)</td>
<td>(737.46, 739.81)</td>
</tr>
</tbody>
</table>
Figure Captions

Figure 1. Map of Olare Motorogi Conservancy with active tourism lodges, roads and transects.

Figure 2. Posterior histograms for the topi model. Posterior means are shown using vertical dashed lines and prior distributions are shown in red.

Figure 3. Histograms showing the mean count difference (predicted minus observed) for topi and the root mean squared error (RMSE).

Figure 4. Count differences for topi (left panel) and elephant (right panel) plotted spatially. Purple dots indicate areas where tourists are over-estimating abundance, and yellow dots indicate areas of under-estimation.
This article is protected by copyright. All rights reserved