

## Accounting for imperfect detection in observational studies: modeling wolf sightability in Yellowstone National Park

MATTHEW C. METZ<sup>1,2,†</sup>, JEREMY SUNDERRAJ,<sup>1</sup> DOUGLAS W. SMITH,<sup>2</sup> DANIEL R. STAHLER<sup>1,2</sup>,  
MICHEL T. KOHL<sup>1,3</sup>, KIRA A. CASSIDY,<sup>2</sup> AND MARK HEBBLEWHITE<sup>1</sup>

<sup>1</sup>Wildlife Biology Program, Department of Ecosystem and Conservation Sciences, W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, Montana 59812 USA

<sup>2</sup>Yellowstone Wolf Project, Yellowstone Center for Resources, Yellowstone National Park, Wyoming 82190 USA

<sup>3</sup>Warnell School of Forestry and Natural Resources, University of Georgia, Athens, Georgia 30602 USA

**Citation:** Metz, M. C., J. SunderRaj, D. W. Smith, D. R. Stahler, M. T. Kohl, K. A. Cassidy, and M. Hebblewhite. 2020. Accounting for imperfect detection in observational studies: modeling wolf sightability in Yellowstone National Park. *Ecosphere* 11(6):e03152. 10.1002/ecs2.3152

**Abstract.** Imperfect detection is ubiquitous among wildlife research and is therefore commonly included in abundance estimation. Yet, the factors that affect observation success are largely unknown for rare and elusive species, such as large carnivores. Here, we took advantage of intensive ground-based monitoring effort and an extensive GPS data set (2000–2018) and developed a winter sightability model for gray wolves (*Canis lupus*) in northern Yellowstone National Park, Wyoming, USA. Our resulting sightability model indicated that observation success was positively affected by the topographic nature of where wolves were in relation to observer locations (viewshed), areas being less forested (openness), and wolf group size, and negatively affected by distance from observer locations. Of these, viewshed had the strongest effect on the probability of observing a wolf, with the odds of observing a wolf being four times more likely when wolves were in the predicted viewshed. Openness was the next most influential covariate, and group size was the least influential. We also tested whether a wolf being harvested from a pack when they were outside of Yellowstone National Park had an effect on wolf sightability. We did not, however, find support for human-induced mortality affecting wolf sightability inside of Yellowstone National Park. Our results indicate that the ability to observe wolves was greatly affected by ecological and landscape-level factors, a finding that is likely to generally extend to other large carnivores. As such, our sightability model highlights the importance of considering landscape structure and variation in large carnivore use of the landscape when conducting observational-based studies.

**Key words:** *Canis lupus*; harvest; resource selection probability function; wildlife observation; wolves.

**Received** 23 October 2019; revised 12 March 2020; accepted 20 March 2020. Corresponding Editor: Joseph D. Holbrook.

**Copyright:** © 2020 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** matthew.metz@umontana.edu

### INTRODUCTION

Direct observations of wildlife provide foundational data for understanding how species are distributed across the globe and how these species use landscapes. This is even true in modern wildlife research where expensive global positioning system (GPS) data and camera traps are increasingly used to study wildlife populations

(Kays et al. 2015, Steenweg et al. 2017) because of the importance of combining observational data with remote methods (Patzelt et al. 2014, Kohl et al. 2018). Observing wildlife is also a key form of nature-based tourism (Newsome et al. 2005), the economic consequences of which are significant for local economies (Naidoo and Adamowicz 2005). Moreover, observations collected by the general public (i.e., citizen or

community science) are important for monitoring and research (Bonney et al. 2009, van Strien et al. 2013, Sun et al. 2019). Regardless of the purpose, wildlife observations are affected by many factors, and a great deal of variation therefore exists in the success of attempts to observe wildlife.

Variation in observation success is driven by factors related to the species' ecology and the observation process. For example, marine mammals such as whales spend much of their time underwater, surfacing only rarely, which affects their availability to be detected during surveys (Hain et al. 1999). Even upon surfacing, boat-based detection is affected by factors such as distance and group size (Williams et al. 2016). Similarly, for terrestrial species, distance from observers, vegetation characteristics (e.g., forest cover), and group size often affect detection (Samuel et al. 1987, Buckland et al. 2001, Peters et al. 2014). It is surprising then that detection is sometimes assumed to be perfect (Hutto 2016), with any consequences of imperfect detection therefore underestimated (Peters et al. 2014). The problem of imperfect detection generally permeates all wildlife studies, but is particularly severe for rare and elusive species.

A bit ironically, however, the observation of rare and elusive species, such as large carnivores, is often highly desired by both the general public and researchers. In the case of large carnivores, such observations, including those collected through citizen science efforts (Farhadinia et al. 2018), contribute information for estimating their distribution, abundance, and ecological effects (Ripple et al. 2014). Direct observations have also provided the backbone for many studies of large carnivores and their prey in African systems (Schaller 1972, Creel and Creel 2002). But conclusions from observational studies are affected by the relationship between large carnivore observations, the factors affecting the observation process, and how large carnivores use the landscape. Much of the initial research on cheetahs (*Acinonyx jubatus*), for example, focused on their use of open plains, and often overlooked their capacity to skillfully use woodland areas (Mills et al. 2004, Bissett and Bernard 2007). To date, we are aware of no studies that directly assessed the detection of large carnivores using only ground-based observations.

Perhaps nowhere else in North America is the desire to see large mammals, including gray wolves (*Canis lupus*), greater than it is in Yellowstone National Park (YNP). Opportunities to observe wolves have been uniquely abundant since their reintroduction in 1995, and wolf watching has, in turn, benefited local economies (Duffield et al. 2008) and wildlife research. Observations of wolves during long-term behavioral studies have revealed heretofore unknown aspects of wolf ecology. For example, how wolf hunting success is affected by individual characteristics of predators (MacNulty et al. 2009) and prey (Metz et al. 2018), and how the social structure of wolf groups affects aggressive interactions between wolf packs (Cassidy et al. 2015). Such studies obviously rely on observations, but the degree to which wolves are successfully observed could vary through time as ecological conditions change. In fact, the number of summer wolf observations have recently declined, with the decline having been associated with the effects of human harvest of wolves outside of YNP (and Denali National Park; Borg et al. 2016). However, a comprehensive assessment of factors that affect the sightability of wolves has not yet occurred, although Borg et al. (2016) found that the number of wolf observations declined concurrently with declining wolf abundance. Smith et al. (2004) also previously showed how increased distance from park roads negatively affected the detection of wolf kills. The frequent observations of wolves in YNP provide a unique opportunity to identify the factors that affect large carnivore sightability.

Here, we employed a logistic regression framework to evaluate factors affecting the likelihood of successful ground-based observations of GPS-collared wolves in northern YNP during winter from 2000 to 2018. We hypothesized that wolf group sightability would be affected by wolves being in more or less forested areas, how far wolves were from observer locations, whether wolves were in areas where observers could physically view (i.e., whether the view was not obstructed by topographic features), and how many wolves were present in the group. We also evaluated the effect of human harvest on sightability by testing whether sightability of wolves in a pack declined following harvest. Finally, we predicted that the opportunity to successfully

observe wolves may have changed over time due to changes in wolf and elk (*Cervus canadensis*) abundance. We therefore explored whether wolf use of sightable locations, as determined through our initial analysis, has changed through time. Understanding sightability of large carnivores such as wolves will generally improve monitoring of these ecologically important species, primarily through providing a tool that can account for spatial–temporal variation in sightability.

## METHODS

### Study area

Our study area was defined by the movements of 13 wolf packs that resided primarily within northern YNP during 17 winters (November 2000–March 2018; Fig. 1; Appendix S1: Fig. S1). The study area included much of Yellowstone's well-known Northern Range, which is generally characterized by elevations between 1500 and 2400 m (Houston 1982). Accordingly, snow generally covered the ground during our monitoring efforts. Lower elevations of northern Yellowstone are generally dominated by large open valleys

and shrub-steppe vegetation, while higher elevations are increasingly characterized by coniferous forests (Houston 1982). The portion of the study area inside YNP is protected from human harvest. Researchers and tourists do, however, observe wildlife and recreate within YNP, which also affects animal behavior (Cassirer et al. 1992). Outside of YNP, recreationists and hunters use the landscape for both non-consumptive and consumptive (hunting) use.

### Winter study

Winter study is a long-term (i.e., beginning in November 1995) research program that investigates wolf–prey relationships through visual observation of wolf packs feeding on carcasses (Smith et al. 2004). Elk, and more recently, bison (*Bison bison*) are the most important food sources for wolves in northern YNP (Metz et al. 2012, Metz et al. 2016). To observe wolves, wolf packs were monitored through aerial and ground-based radio-tracking for 30-d periods during an early (mid-November to mid-December) and late (March) winter period. Observations therefore rely on radio-collaring, accomplished through

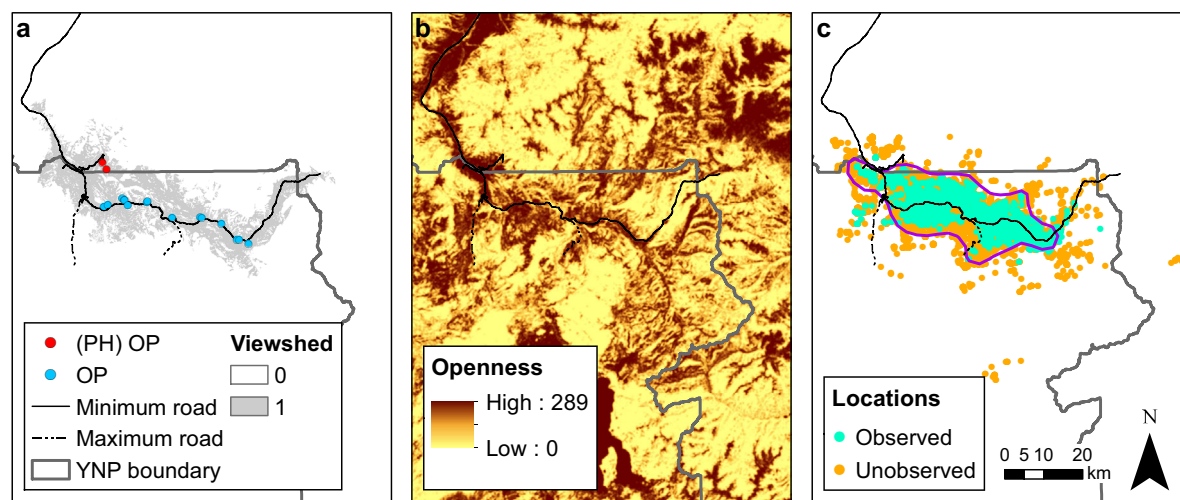


Fig. 1. Spatial distribution of observation points, covariates, and wolf GPS data. Panels display (a) common observation points (OP; red circles indicate observation points used only before Montana wolf harvest season was implemented), the minimum and maximum road, viewshed (from the minimum road and all common observation points; 1 = viewshed-predicted viewable, 0 = viewshed-predicted not viewable), and the Yellowstone National Park boundary (roads and boundary shown in all panels), (b) openness (2001), and (c) wolf GPS locations (i.e., observed-unobserved wolf locations). Note that the purple line in panel (c) displays an 85% population-level utilization distribution that was used in Fig. 4.

helicopter darting or net-gunning. Wolf capture and handling conformed to National Park Service animal capture and handling policies, as well as University of Montana IACUC protocol AUP MH-043-15. Both very high frequency (VHF) and GPS collars were routinely deployed. Here, we used data only for packs that contained GPS-collared wolves during a 30-d period (but see Appendix S2 for a complementary sightability model developed using VHF data).

Our study relied on ground-based observations collected by teams of 2–3 field technicians (hereafter, observation crew). Each observation crew was assigned to a single wolf pack, and together, they observed 1–3 GPS-collared wolf packs (Appendix S1: Fig. S1) during each 30-d period. Daily, observation crews located their assigned wolf pack at daybreak using standard radio telemetry and then used high-powered spotting scopes to observe the wolf pack for as much of daylight as possible (Smith et al. 2004, Martin et al. 2018). We visually observed GPS-collared wolf packs on 21.7 ( $\pm 0.6$  standard error [SE]) days, on average. Observation crews did not have access to real-time GPS locations.

#### *GPS data preparation*

Among included wolf packs, GPS data were available from 17 winters (27 30-d periods), 13 packs, and 45 wolves. We included wolves only if they were usually with other members of their pack, which was known from our aforementioned daily observations. Some packs contained multiple GPS-collared wolves, but we included only one wolf per pack for each 30-d period. We used a multi-step process to thin our data to a single wolf, stopping whenever a single wolf remained. First, we excluded wolves who only tangentially belonged to a pack. If multiple wolves were core members of their pack, we selected wolves whose collars acquired hourly locations. Then, we censored wolves that did not survive the 30-d period. Finally, when multiple wolves still remained, we randomly selected a wolf and removed the remaining individual(s). Such thinning of our data resulted in data from 33 wolves.

We further filtered the GPS data in two ways. First, we included GPS data only for days that observation crews attempted to monitor the wolf

pack (i.e., rarely, severe weather events precluded ground-based monitoring for a day). Next, we included only daylight wolf GPS locations that occurred well after or before sunrise and sunset, respectively, by including GPS locations acquired only from the 09:00–16:00 hours during early winter (mid-month sunrise  $\approx 07:40$  and sunset  $\approx 16:45$ ) and the 08:00–18:00 hours during late winter (mid-month average of sunrise  $\approx 06:30$  and sunset  $\approx 18:30$ ). Data on sunrise and sunset were obtained from the United States Naval Observatory (<http://aa.usno.navy.mil/data>). GPS data were recorded at varying intervals, although most collars recorded a location each hour. On average, we included nine successfully acquired locations per wolf per day. Overall, GPS location fix success was high, with GPS collars successfully recording a location 98.0% of the time (11,999 of 12,243 possible locations). We determined whether the wolf group containing the GPS-collared wolf was observed or not observed when a GPS location was recorded through comparing our daily observations to each acquired wolf GPS location. Of the 11,999 included GPS locations, 3082 were observed, and 8917 were unobserved.

#### *Wolf sightability model*

For each GPS location, we included covariates that we hypothesized may affect the observation of the wolf group in a logistic regression framework (Hosmer and Lemeshow 2000). The first covariate that we included was wolf group size (group size), which represented the number of wolves present, not necessarily the number of wolves seen at the precise time when the GPS collar recorded the spatial location. We did not record group size for 46 of 3082 observed locations, and we were unable to observe group size for the 8917 unobserved locations. For these 8963 locations (i.e., the 46 observed and 8917 unobserved locations), we assigned group size in the following manner. First, if the wolf group was both observed on the day and had only one group size recorded for the day, we assigned this group size. Next, if wolves were observed on the day but group size differed during the day (36 of 696 observed wolf group days), we assigned the mean group size. Finally, when we did not observe the wolf pack on a given day, we randomly sampled with replacement from the daily



values of the single-daily and mean group sizes for that pack during that study period.

We also included covariates that described the distance from the nearest road or observation point (distance), openness (openness), and whether the location was viewable (viewshed). These spatial covariates were extracted from Geographic Information System layers in Program R (version 3.4.1; R Core Team 2017) using the package raster (Hijmans et al. 2015; we converted all continuous covariates to rasters).

A road transecting the study area, and plowed during winter, was our main observation platform (Fig. 1). Observation crews occasionally used an additional section of the road, dependent on the specific movements of a pack and road conditions (Appendix S1: Fig. S1). Additionally, observation crews took advantage of various higher-elevation observation points, typically located near the road (Fig. 1a). Some of these observation points were commonly used, while others were less commonly used because they were physically difficult to reach. For simplicity, we considered only the two roads (i.e., minimum or maximum; Fig. 1a) and the 17 common observation points (Fig. 1a; but see Appendix S2). Two of these 17 observation points were included only through the winter of 2008–2009 because observation crews did not use these after Montana wolf harvest seasons were implemented (see below in Wolf sightability model). These roads and observation points affected the values of distance and viewshed.

For distance, we determined the minimum distance (measured in kilometers) to the road (i.e., minimum or maximum) used by the observation crew for a pack during a particular study period (hereafter, pack-road) or a common observation point. We similarly determined whether the wolf location was in (1) or out (0) of a model-predicted viewshed (Fig. 1a). We created our viewshed layers with the Viewshed 2 tool in ArcGIS version 10.3.1 and digital elevation model layers of our study area (pixel size equal to ~9.7 m). We set the observer offset to 2 m (assuming an observer height of 2 m), the vertical upper and lower angles to  $-90$  to  $+90$ , respectively (assuming an observer could look up and down), and the maximum viewable distance to 6 miles (i.e.,

~9.66 km), nearly equivalent to the maximum distance where wolves were observed.

Openness values ranged from 0 to 289 and characterized the openness of a  $500 \times 500$  m window, with the pixel as the central location (0 = heavily forested, 289 = fully open; see Kohl et al. 2018; Fig. 1b). Vegetation data from the LANDFIRE program (landfire.gov) for all available years (2001, 2008, 2010, 2012, 2014) were used to create the openness layers (2001 for 1995–2005, 2008 for 2006–2009, 2010 for 2010–2011, 2012 for 2012–2013, and 2014 for 2014–2017).

Lastly, we included a covariate that described whether a wolf was harvested from the pack during the immediately preceding or ongoing hunting season because previous work has suggested that wolf harvest affects the number of wolf sightings in Yellowstone (Borg et al. 2016). Wolf harvest became seasonally legal in the autumn of 2009 and has been so since (with the exception of the winter of 2010–2011). Human harvest is permissible only outside of YNP and was often limited by a quota in the relevant geographic area (see Fig. 1c and Smith et al. 2016). Most harvest that affected wolves in our study occurred in Montana, where the harvest season generally began in September and lasted through March, unless the quota was filled. The time and location of wolf harvest events were reported to the governing state wildlife agency, and, together with that agency, we used near-daily observations of pack movements and counts to assign harvests as having occurred within a pack that commonly frequented Yellowstone National Park. We assigned harvest as having occurred in a pack beginning on the day after the initial harvest event for each pack, and allowed the potential effect of harvest to continue throughout winter.

We estimated wolf group sightability with the logistic link function in the package lme4 (Bates et al. 2014) using mixed-effect logistic regression models (binomial; observed = 1, unobserved = 0; random effect for pack and biological year) through comparing GPS locations where wolf groups were observed to GPS locations where wolf groups were not observed (Gillies et al. 2006), akin to a true used–unused resource selection probability function (RSPF) design (Manly et al. 2002). Our specific model was

$$P(\text{wolf group sighting}) = \text{logit}(\beta_{0i} + \beta_{1i}\text{viewshed} + \beta_{2i}\text{distance} + \beta_{3i}\text{openness} + \beta_{4i}\text{group size} + \beta_{5i}\text{harvest} + \varepsilon_i) \quad (1)$$

where  $\beta_x$  is the effect of that covariate's coefficient on the probability of observing the wolf group at wolf location  $i = 1, \dots, n$ ,  $\varepsilon$  is the error, and  $\beta_0$  is the baseline probability of observing the wolf group containing the GPS-collared wolf independent of covariates because of the true used–unobserved (observed–unobserved) design (Manly et al. 2007).

We screened for association and collinearity among covariates, of which we found none (e.g., maximum  $r = 0.191$ ). We scaled continuous covariates to facilitate model convergence, evaluation of covariate effects, and comparison among models. We created a priori candidate model sets based on the hypothesized importance of our covariates. We created all combinations of additive models using viewshed, openness, distance, group size, and harvest. We did not include models with interactions because they increased complexity without improving model-based predictions (M. Metz, *unpublished data*). We evaluated the models in the package AICcmodavg (Mazerolle 2017) using Bayesian information criterion (BIC). We used BIC instead of AIC (Akaike information criterion) due to its superior performance in identifying the best model when data are abundant (Aho et al. 2014), as was the case with our GPS data.

We evaluated the performance of the top models using standard logistic regression diagnostics, and report overall classification success, sensitivity (probability of correctly classifying observed locations), specificity (probability of correctly classifying unobserved locations), and the area under the curve (AUC) that measures overall model performance (values of 0.7–0.8 suggest acceptable discrimination, while values of 0.8–0.9 suggest excellent discrimination; Hosmer and Lemeshow 2000). Finally, we externally validated our top-ranked wolf sightability model. To do so, we predicted the response,  $p(\text{sighting})$ , for an average pack (i.e.,  $\text{re.form} = \text{NA}$ ) for 1258 GPS locations from the winter of 2018–2019 ( $n = 5$  packs; again limited to a single wolf per pack). We then used the optimal cutpoint to evaluate classification success, sensitivity, and specificity.

## RESULTS

### *Covariates affecting sightability*

The 3082 observed locations were, on average, 2.3 km ( $\pm 0.02$  SE) from the closest road or common observation point, while the 8917 unobserved locations were 3.3 km ( $\pm 0.03$  SE) away (Appendix S1: Fig. S2). Mean openness values were 215 ( $\pm 1.1$  SE) and 156 ( $\pm 0.9$  SE) for the observed and unobserved locations, respectively. Of the 3082 observed locations, 2671 (86.7%) were in viewable locations (i.e., viewshed = 1). Conversely, 4897 of 8917 (54.9%) unobserved locations were in viewable locations. Finally, of the 49 wolf pack-study periods in our observed–unobserved data set, nine experienced a harvest event (Appendix S1: Fig. S1).

### *Observed–unobserved wolf sightability model*

The top wolf sightability model included viewshed, distance, openness, and group size, as well as the random effects for pack and biological year (i.e., winter). Harvest was not included in our top model. Among our candidate models, the top model was strongly supported, with 97% of the BIC weight (Appendix S1: Table S1). The beta coefficients for the top model, all of which were significant (Fig. 2), indicated a positive effect of viewshed ( $\beta = 1.37$ ; odds ratio [OR] = 3.95 [95% CI: 3.49, 4.47]), openness ( $\beta = 0.81$ ; OR = 2.25 [2.13, 2.39]), and group size ( $\beta = 0.23$ ; OR = 1.26 [1.19, 1.34]) on the probability of sighting a wolf group. Conversely, distance had the expected negative effect ( $\beta = -0.56$ ; OR = 0.57 [0.52, 0.63]). The combined influence of these covariates resulted in, for example, the probability of seeing a group of 12 wolves that were in a location that was viewable, fairly open (openness = 242), and 500 m from the nearest observer location that was approximately 14 times greater than seeing a group of 6 wolves in a location that was non-viewable, fairly forested (openness = 112), and 3 km from the nearest observer location (Fig. 3).

### *Predictions: the confusion matrix, space, and time*

Our top model showed excellent discriminatory power as the AUC was 0.80 (Table 1). Using the optimal cutpoint (see Table 1), we predicted whether each observation in our data set was observed or unobserved (Fig. 4a). Upon doing

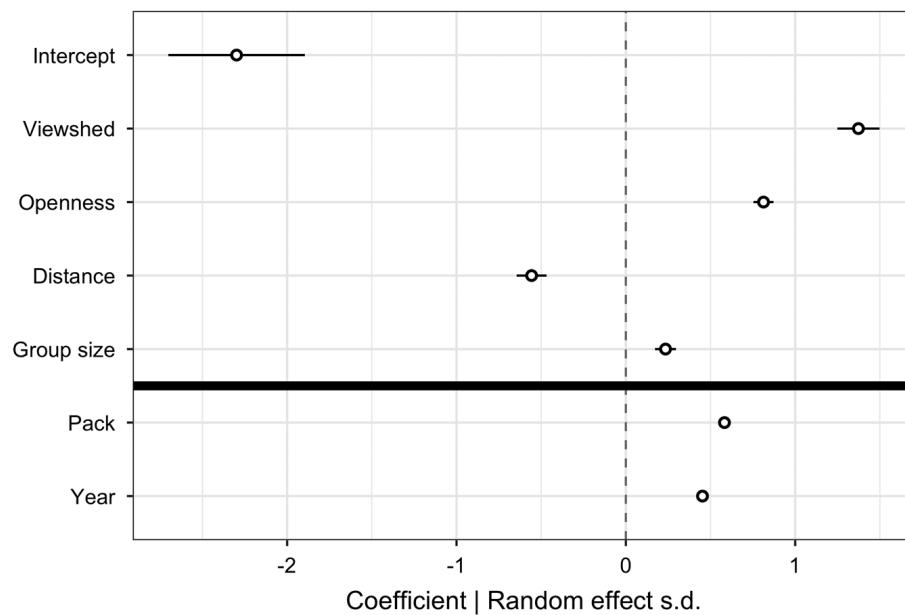


Fig. 2. Beta coefficient estimates for top-ranked wolf sightability model. Error bars represent 95% confidence intervals. Below the solid black line displays the standard deviation for the random effects (pack, year), which is directly comparable to the beta coefficient strength (Harrell Jr 2001).

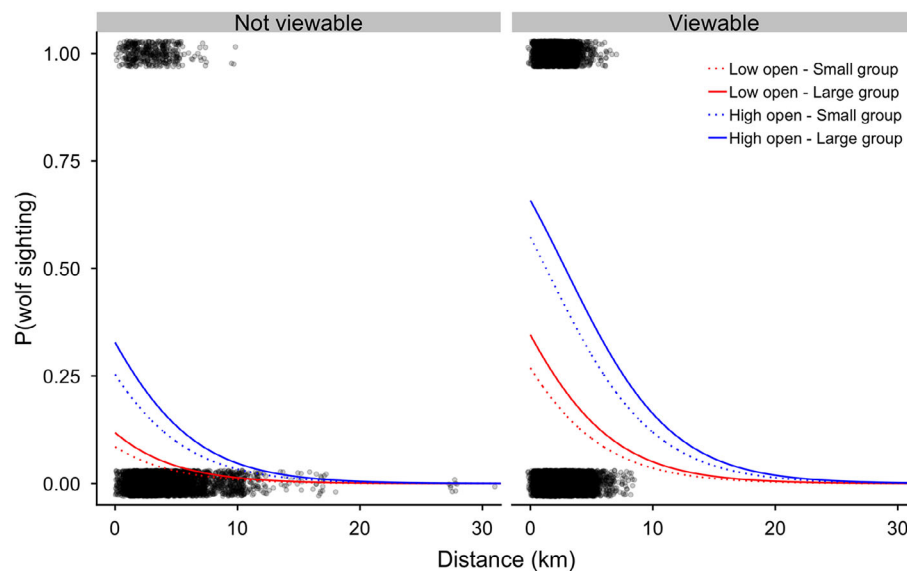


Fig. 3. Predicted probability of sighting a wolf group. Values for low and high represent the lower and upper quartiles (openness: 111.5, 241.8; group size: 6, 12). The displayed low and high openness values describe moderately forested and pretty open areas, respectively.

so, 72% of wolf GPS locations were correctly predicted (Table 1). Sensitivity (i.e., the correct prediction of 1s) was 0.77, while specificity (i.e., the correct prediction of 0s) was 0.70 (Table 1).

Our external validation of the top model similarly suggested that this model did a good job of predicting whether a GPS-collared wolf group would be observed. Specifically, 76% of the 1258

Table 1. Classification success of wolf sightability model.

Data set	Locations	Predict	Prop. correct	AUC	Cutpoint	Specificity	Sensitivity
Model	11999	8620	0.72	0.80	0.26	0.70	0.77
Validation	1258	955	0.76	...	...	0.74	0.79

Notes: Row for Model reports the classification success of the observed–unobserved data that were used to develop the wolf sightability model, while the row for Validation reports the classification success for the external validation. Locations indicate the total number of locations in the data set, Predict indicates the number of locations correctly predicted as a 1 (i.e., above the cutpoint), and Prop. correct is the proportion of total locations correctly predicted. Specificity indicates the proportion of 0 (i.e., unobserved) locations correctly predicted, and, sensitivity, the same for 1 (i.e., observed) locations.

GPS locations were correctly predicted, with a sensitivity of 0.79 and a specificity of 0.74 (Table 1).

Of course, an end-product from our sightability model is a map of northern YNP that visualizes the variation in where wolves were most, and least, likely to be seen (Fig. 4b). But whether wolves were observed is the product of this sightability map (or a representation of it that varied across time; see Appendix S3) and the probability that a wolf group was in a sightable location. We therefore used the predictions from our observed–unobserved model to show how wolves were less likely, over time, to be in spatial locations where they were predicted to be observed (i.e., where the prediction for each wolf GPS location was greater than the optimal cutpoint; Table 1). This pattern was clearer if the data from the earliest years, when sample sizes were smallest, were ignored (Fig. 5).

## DISCUSSION

Our study showed how wolf group sightability was driven by both ecological (i.e., forest cover, group size) and landscape (i.e., topography, distance) factors (Figs. 2–4), but not human harvest (Appendix S1: Table S1). Viewshed, followed by openness and then distance, most affected the success of our ground-based observation attempts (Fig. 2). In our case, the relative importance of distance was likely influenced by our use of high-powered spotting scopes. But our results nonetheless suggest that distance, most commonly employed in sightability models (Buckland et al. 2001), may not always be the most important factor affecting detection. Ultimately, our work highlights how sightability of large carnivores is substantially and intuitively affected by multiple factors, and that accounting for these factors may be important for reducing bias in observational-based studies.

Larger group sizes increase the search image for an observer, and while group size did have a positive effect on sightability, its importance was much less than other factors (Fig. 2). The relatively small effect of wolf group size tends to differ from previous work with other species. For example, even small changes in group size (i.e., from one to two individuals) may double boat-based detection of whales (Williams et al. 2016), and aerial detection has been estimated to be essentially one once group size reaches at least fifteen in elk (Samuel et al. 1987). Group size may more strongly influence large carnivore detection when snow is not commonly present in a study area or for a species such as the spotted hyena (*Crocuta crocuta*) that tends to live in much larger groups (Kruuk 1972). Determining whether group size does more substantially influence detection in other instances could be useful for other applications, such as estimating abundance from camera traps (Burton et al. 2015).

Our sightability model correctly classified 72% of wolf GPS locations in our original data set as being observed or not, indicating strong predictive power (Table 1). This classification success occurred despite our model deriving the viewshed and distance covariates from fixed observer locations (Appendix S2). Further, observation success or failure was also affected by stochastic events such as heavy snowfall, while the collar was recording a fix, which we could not include in our sightability model. Despite these limitations, our model performed similarly well during our external validation, correctly predicting 76% of the 1258 locations. This strong validation of our model indicates its usefulness in predicting wolf sightability in northern YNP, and that models developed for other large carnivore populations could be similarly useful.

Factors that affected wolf sightings changed relatively little over time in our system (Appendix S1: Fig. S3), but we observed a general temporal



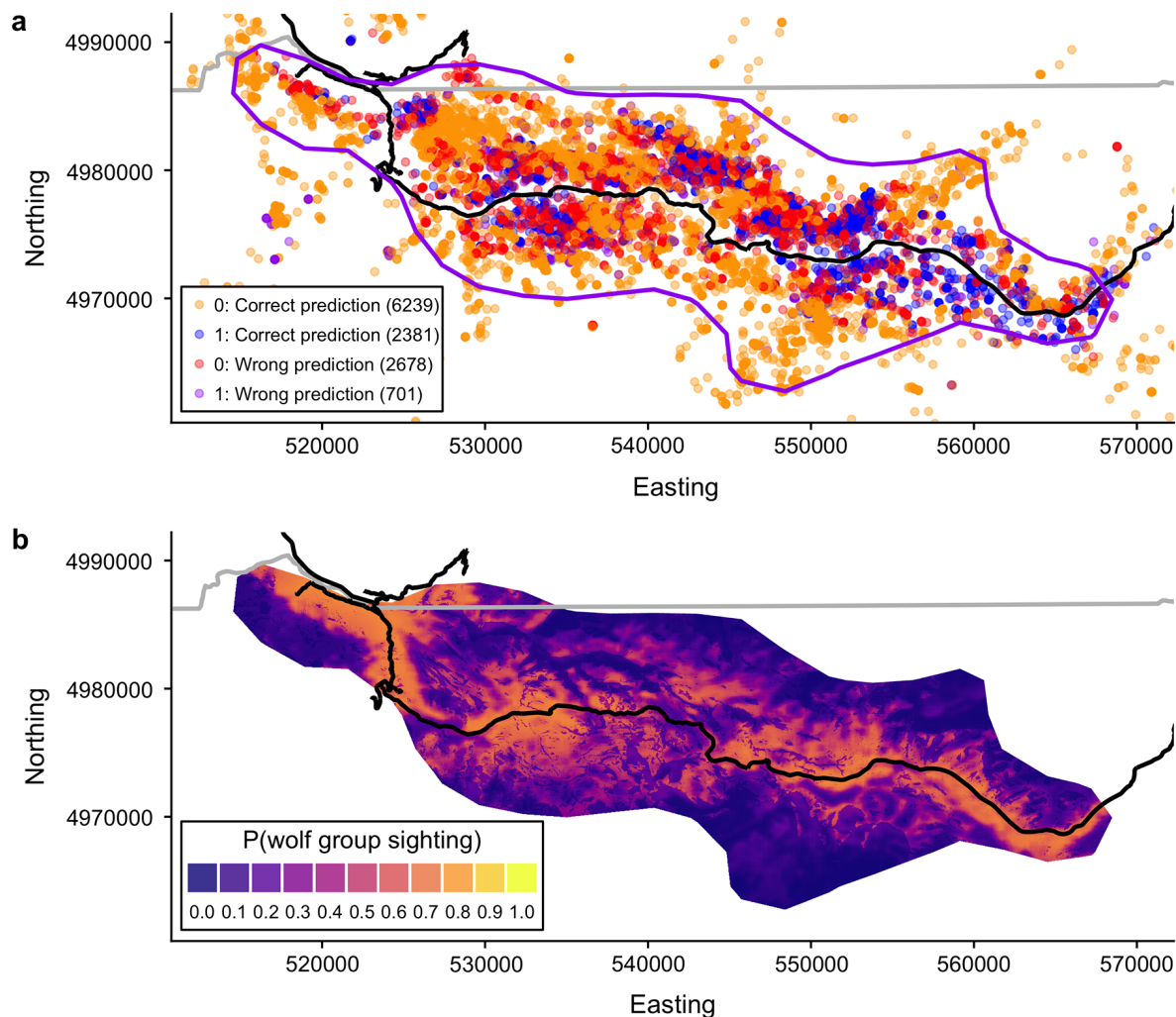


Fig. 4. Spatial predictions for wolf sightability model. Panel (a) displays the classification success for model-based predictions for the actual data set, using the optimal cutpoint (see Table 1). The legend reports the confusion matrix for the 11,999 GPS locations. In panel (b), we used the beta coefficients to create spatial predictions (see further details in Appendix S3) for a pack that was monitored from the minimum road and common observation points (from before wolf harvest was implemented; see Fig. 1a), openness values from 2001 (see Fig. 1b), and the high value for group size (12; see Fig. 3). Note that the purple line in panel (a) displays the 85% population-level utilization distribution that was used for panel (b), and that the black and gray lines in each panel display the minimum road and Yellowstone National Park boundary, respectively.

decline in wolf use of sightable locations (Fig. 5). At least three relevant changes during our study period could explain this pattern. The first two, declining wolf and elk abundance, are linked together through the numeric response of predator-prey dynamics (i.e., wolf abundance changes in response to prey abundance; Messier 1994). In turn, GPS-collared wolf territory sizes have increased over time in northern YNP (Appendix S1: Fig. S4;

Kittle et al. 2015), which results in many territories including more area further from the road and out of the viewshed. Any effect of elk on wolf use of sightable locations also likely resulted from changes in elk (and other prey) distribution across northern Yellowstone (White et al. 2012, 2015). Wolves' increased use of bison in northern YNP (Metz et al. 2016) further complicates wolf sightability in northern YNP.

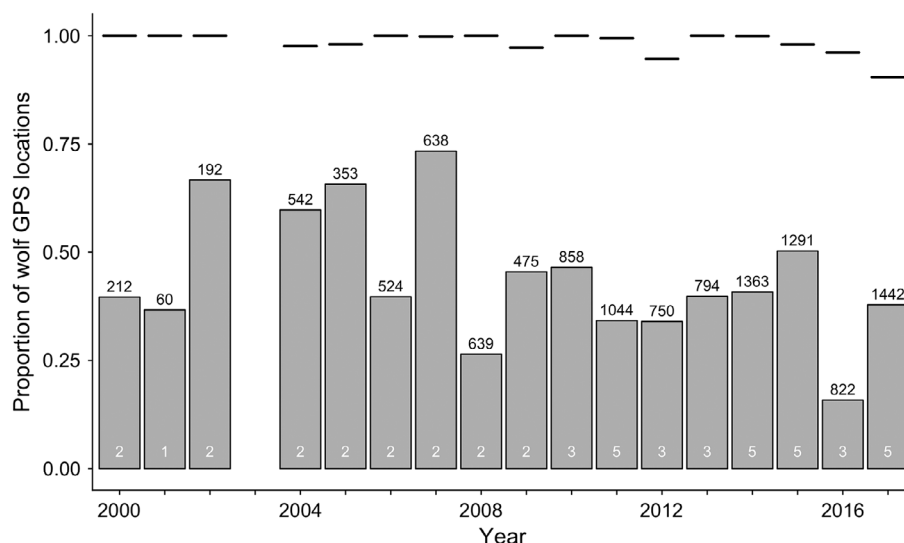


Fig. 5. Proportion of wolf GPS locations predicted to be sighted through time. The white numbers along the  $x$ -axis indicate the number of wolf-study periods, and the black numbers on top of each bar indicate the total number of GPS locations acquired, during that winter (see also Appendix S1: Fig. S1). Note that GPS collars generally attempted to record a fix once per hour beginning during the winter of 2004–2005 (year = 2004 on  $x$ -axis). Black dashes indicate the proportion of GPS locations inside of Yellowstone National Park.

The last relevant change was the implementation of a wolf harvest season outside of YNP (Smith et al. 2016). A harvest effect has been suggested by previous work evaluating wolf sightings during summer in YNP (Borg et al. 2016) and elsewhere where wolves are subject to harvest (Thurber et al. 1994, Hebblewhite and Merrill 2008, Borg et al. 2016). We, however, found no effect of human harvest on wolf sightability during winter, seemingly indicating that wolf packs did not display substantial behavioral changes following a harvest event that had any lasting (i.e., throughout winter) effect on wolf sightability when a wolf group was in a similar covariate space. Moreover, GPS-collared wolves used spatial locations outside of YNP more frequently in recent years (black dashes in Fig. 5), although we did not account for whether harvest seasons were ongoing. This pattern of increased use of locations outside of YNP, along with concurrent shifts in elk abundance (White et al. 2012), suggests elk abundance and distribution could in fact be the driving force behind the general temporal decline of wolf use of sightable spatial locations during winter. It is difficult, however, to tease apart these three effects (elk and wolf abundance, hunting) because each has

occurred concurrently and, in the case of wolf and elk abundance, are clearly linked. We should also note that some of our results could be affected by the distribution of wolf GPS collars, although our inclusion of a random effect for the baseline sighting of a wolf pack minimized this potential effect. Similarly, the random effect for biological year also minimized the effect of other factors that likely affected wolf sightability, such as fine-scale temporal variation in snow cover.

In addition to our observed–unobserved logistic regression model, we developed a similar observed–available logistic regression model (Appendix S2), which is the model structure that researchers without GPS data would need. The top-ranked observed–available model was similar, but did not include group size (Appendix S2: Table S1). The most important difference between the observed–unobserved and observed–available model was how the relative strengths of the remaining beta coefficients differed (Appendix S2: Fig. S3). Likely due to the strength of viewshed being markedly different, the observed–unobserved model was far superior in its overall classification success (Appendix S2: Table S2). Our results nonetheless suggest that other researchers could develop carnivore

sightability models without concurrent GPS data, which could be useful in study areas where cost and/or sample size limit the use of GPS data.

Our observed–unobserved model was developed with data collected over a nearly two-decade period. During that time, hundreds of thousands of park visitors have viewed wolves. Our results show that the potential to successfully observe wolves has differed across time (Fig. 5), but where on the landscape wolves are most sightable is essentially fixed by the covariates that most strongly affect sightability (Fig. 2). Our model could therefore be used to identify areas of future development that enhance visitor experience, while minimizing any negative visitor effect on wolves. Visitation to YNP to attempt to observe wolves is not trivial. In fact, a decade ago when there were ~3 millions visits (2018, ~4.1 million visits), Duffield et al. (2008) estimated that ~300,000 visitors observed wolves annually, and that these ~300,000 visitors (plus those who had hoped to observe wolves) brought ~35 million dollars annually to local economies (Duffield et al. 2008). These observations, however, often result in wildlife jams that present significant management challenges to YNP staff (Haroldson and Gunther 2013); the use of our spatial predictions (Fig. 4b) would allow for easy identification of appropriate places to increase vehicle parking that could reduce these management challenges, for example.

Wolf group sightability was predictably affected by factors (viewshed, openness, distance) related to the study area and, to a lesser extent, the species' ecology (group size). Previous work in our system identified that only distance affected ground-based detection of wolf-killed prey (Smith et al. 2004), but our sightability model suggests that distance would not necessarily be the most important covariate in an updated version of this kill–recovery model. Moreover, given that the potential to view wolves and their interactions with each other and other species varies through time (Fig. 5), our work also suggests that future work in our system may need to explicitly account for the imperfect observation process. Temporal evaluation of wolf–prey (Martin et al. 2018) or wolf–wolf (Cubaynes et al. 2014) interactions may be affected by temporal changes in observation success, for example. The importance of including

sightability models in large carnivore observation-based studies would seemingly extend to other systems as well, such as African large carnivore–prey systems. Given the frequency with which GPS collars are now used, such work could similarly develop detection models such as ours to account for large carnivore sightability.

## ACKNOWLEDGMENTS

We thank E. Stahler for data management, B. Cassidy for GIS assistance, and P. Mahoney for statistical advice. We also thank A. Nelson (Montana Fish Wildlife and Parks) and K. Mills (Wyoming Game and Fish) for their assistance in assigning pack affiliation for harvested wolves. Finally, we thank all of the Yellowstone winter study field technicians who dedicatedly collected the observational data that were the foundation of this work. Our work was supported by Yellowstone Forever (Tapeats Fund, Valerie Gates, Frank and Kay Yeager, many individual donors), National Science Foundation (NSF) Long Term Research in Environmental Biology grants DEB-0613730 (D.W.S.) and DEB-1245373 (D.R.S. and D.W.S.). M.C.M. was also supported by the Wesley M. Dixon Graduate Fellowship and Bertha Morton Scholarship from the University of Montana and a Graduate Enhancement Fellowship from the Montana Institute on Ecosystems funded through NSF Established Program to Stimulate Competitive Research (EPSCOR).

## LITERATURE CITED

- Aho, K., D. Derryberry, and T. Peterson. 2014. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* 95:631–636.
- Bates, D., M. Maechler, B. Bolker, and S. Walker. 2014. lme4: linear mixed-effects models using Eigen and S4. R Package Version 1:1–23. <https://CRAN.R-project.org/package=lme4>
- Bissett, C., and R. T. F. Bernard. 2007. Habitat selection and feeding ecology of the cheetah (*Acinonyx jubatus*) in thicket vegetation: Is the cheetah a savanna specialist? *Journal of Zoology* 271:310–317.
- Bonney, R., C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk. 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience* 59:977–984.
- Borg, B. L., S. M. Arthur, N. A. Broman, K. A. Cassidy, R. McIntyre, D. W. Smith, and L. R. Prugh. 2016. Implications of harvest on the boundaries of protected areas for large carnivore viewing opportunities. *PLOS ONE* 11:e0153808.

- 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, New York, New York, USA.
- Burton, A. C., E. Neilson, D. Moreira, A. Ladle, R. Steenweg, J. T. Fisher, E. Bayne, and S. Boutin. 2015. Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology* 52:675–685.
- Cassidy, K. A., D. R. MacNulty, D. R. Stahler, D. W. Smith, and L. D. Mech. 2015. Group composition effects on aggressive interpack interactions of gray wolves in Yellowstone National Park. *Behavioral Ecology* 26:1352–1360.
- Cassirer, E. F., D. J. Freddy, and E. D. Ables. 1992. Elk responses to disturbance by cross-country skiers in Yellowstone National Park. *Wildlife Society Bulletin* 20:375–381.
- Creel, S., and N. M. Creel. 2002. The African wild dog: behavior, ecology, and conservation. Princeton University Press, Princeton, New Jersey, USA.
- Cubaynes, S., D. R. MacNulty, D. R. Stahler, K. A. Quimby, D. W. Smith, and T. Coulson. 2014. Density-dependent intraspecific aggression regulates survival in northern Yellowstone wolves (*Canis lupus*). *Journal of Animal Ecology* 83:1344–1356.
- Duffield, J. W., C. J. Neher, and D. A. Patterson. 2008. Wolf recovery in Yellowstone: park visitor attitudes, expenditures, and economic impacts. *George Wright Forum* 25:13–19.
- Farhadinia, M. S., R. J. Moll, R. A. Montgomery, S. Ashrafi, P. J. Johnson, L. T. B. Hunter, and D. W. Macdonald. 2018. Citizen science data facilitate monitoring of rare large carnivores in remote montane landscapes. *Ecological Indicators* 94:283–291.
- Gillies, C. S., M. Hebblewhite, S. E. Nielsen, M. A. Krawchuk, C. L. Aldridge, J. L. Frair, D. J. Saher, C. E. Stevens, and C. L. Jerde. 2006. Application of random effects to the study of resource selection by animals. *Journal of Animal Ecology* 75:887–898.
- Hain, J., S. Ellis, R. D. Kenney, and C. K. Slay. 1999. Sightability of right whales in coastal waters of the southeastern United States with implications for the aerial monitoring program. Pages 191–207 in G. Garner, S. Amstrup, J. Laake, B. Manley, L. McDonald, and D. Robertson, editors. Marine mammal survey and assessment methods. A.A. Balkema, Rotterdam, The Netherlands.
- Haroldson, M. A., and K. A. Gunther. 2013. Roadside bear viewing opportunities in Yellowstone National Park: characteristics, trends, and influence of whitebark pine. *Ursus* 24:27–41.
- Harrell Jr, F. E. 2001. Regression modeling strategies: with applications to linear models, logistic regression and survival analysis. Springer, New York, New York, USA.
- Hebblewhite, M., and E. Merrill. 2008. Modelling wildlife–human relationships for social species with mixed-effects resource selection models. *Journal of Applied Ecology* 45:834–844.
- Hijmans, R., J. van Etten, J. Cheng, M. Mattiuzzi, M. Sumner, J. Greenberg, O. Lamigueiro, A. Bevan, E. Racine, and A. Shortridge. 2015. Package ‘raster’. R package. <https://CRAN.R-project.org/package=raster>
- Hosmer, D. W., and S. Lemeshow. 2000. Applied logistic regression. Wiley, New York, New York, USA.
- Houston, D. B. 1982. The Northern Yellowstone Elk: ecology and Management. Macmillan Publishing, New York, New York, USA.
- Hutto, R. L. 2016. Should scientists be required to use a model-based solution to adjust for possible distance-based detectability bias? *Ecological Applications* 26:1287–1294.
- Kays, R., M. C. Crofoot, W. Jetz, and M. Wikelski. 2015. Terrestrial animal tracking as an eye on life and planet. *Science* 348:aaa2478.
- Kittle, A. M., M. Anderson, T. Avgar, J. A. Baker, G. S. Brown, J. Hagens, E. Iwachewski, S. Moffatt, A. Mosser, and B. R. Patterson. 2015. Wolves adapt territory size, not pack size to local habitat quality. *Journal of Animal Ecology* 84:1177–1186.
- Kohl, M. T., D. R. Stahler, M. C. Metz, J. D. Forester, M. J. Kauffman, N. Varley, P. White, D. W. Smith, and D. R. MacNulty. 2018. Diel predator activity drives a dynamic landscape of fear. *Ecological Monographs* 88:638–652.
- Kruuk, H. 1972. The spotted hyena: a study of predation and social behavior. University of Chicago Press, Chicago, Illinois, USA.
- MacNulty, D. R., D. W. Smith, L. D. Mech, and L. E. Eberly. 2009. Body size and predatory performance in wolves: Is bigger better? *Journal of Animal Ecology* 78:532–539.
- Manly, B. F. L., L. McDonald, D. Thomas, T. L. McDonald, and W. P. Erickson. 2007. Resource selection by animals: statistical design and analysis for field studies. Springer Science & Business Media, New York, New York, USA.
- Martin, H. W., L. Mech, J. Fieberg, M. Metz, D. MacNulty, D. Stahler, and D. Smith. 2018. Factors affecting gray wolf (*Canis lupus*) encounter rate with elk (*Cervus elaphus*) in Yellowstone National Park. *Canadian Journal of Zoology* 96:1032–1042.
- Mazerolle, M. J. 2017. AICcmodavg: model selection and multimodel inference based on (Q) AIC (c). R Package Version 2.1.1. <https://CRAN.R-project.org/package=AICcmodavg>



- Messier, F. 1994. Ungulate population models with predation: a case study with the North American moose. *Ecology* 75:478–488.
- Metz, M. C., D. J. Emlen, D. R. Stahler, D. R. MacNulty, D. W. Smith, and M. Hebblewhite. 2018. Predation shapes the evolutionary traits of cervid weapons. *Nature Ecology & Evolution* 2:1619–1625.
- Metz, M. C., D. W. Smith, D. R. Stahler, J. A. Vucetich, and R. O. Peterson. 2016. Temporal variation in wolf predation dynamics in Yellowstone: lessons learned from two decades of research. *Yellowstone Science* 24:55–60.
- Metz, M. C., D. W. Smith, J. A. Vucetich, D. R. Stahler, and R. O. Peterson. 2012. Seasonal patterns of predation for gray wolves in the multi-prey system of Yellowstone National Park. *Journal of Animal Ecology* 81:553–563.
- Mills, M., L. S. Broomhall, and J. T. du Toit. 2004. Cheetah *Acinonyx jubatus* feeding ecology in the Kruger National Park and a comparison across African savanna habitats: Is the cheetah only a successful hunter on open grassland plains? *Wildlife Biology* 10:177–186.
- Naidoo, R., and W. L. Adamowicz. 2005. Economic benefits of biodiversity exceed costs of conservation at an African rainforest reserve. *Proceedings of the National Academy of Sciences USA* 102:16712–16716.
- Newsome, D., R. K. Dowling, and S. A. Moore. 2005. *Wildlife tourism*. Channel View Publications, Clevedon, UK.
- Patzelt, A., G. H. Kopp, I. Ndao, U. Kalbitzer, D. Zinner, and J. Fischer. 2014. Male tolerance and male–male bonds in a multilevel primate society. *Proceedings of the National Academy of Sciences USA* 111:14740–14745.
- Peters, W., M. Hebblewhite, K. G. Smith, S. M. Webb, N. Webb, M. Russell, C. Stambaugh, and R. B. Anderson. 2014. Contrasting aerial moose population estimation methods and evaluating sightability in west-central Alberta, Canada. *Wildlife Society Bulletin* 38:639–649.
- R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ripple, W. J., J. A. Estes, R. L. Beschta, C. C. Wilmers, E. G. Ritchie, M. Hebblewhite, J. Berger, B. Elmhagen, M. Letnic, and M. P. Nelson. 2014. Status and ecological effects of the world's largest carnivores. *Science* 343:1241484.
- 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.
- Schaller, G. B. 1972. *The Serengeti lion: a study of predator-prey relations*. University of Chicago, Chicago, Illinois, USA.
- Smith, D. W., T. D. Drummer, K. M. Murphy, D. S. Guernsey, and S. B. Evans. 2004. Winter prey selection and estimation of wolf kill rates in Yellowstone National Park, 1995–2000. *Journal of Wildlife Management* 68:153–166.
- Smith, D. W., P. White, D. R. Stahler, A. Wydeven, and D. E. Hallac. 2016. Managing wolves in the Yellowstone area: balancing goals across jurisdictional boundaries. *Wildlife Society Bulletin* 40:436–445.
- Steenweg, R., M. Hebblewhite, R. Kays, J. Ahumada, J. T. Fisher, C. Burton, S. E. Townsend, C. Carbone, J. M. Rowcliffe, and J. Whittington. 2017. Scaling-up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment* 15:26–34.
- Sun, C. C., J. A. Royle, and A. K. Fuller. 2019. Incorporating citizen science data in spatially explicit integrated population models. *Ecology* 100:e02777.
- Thurber, J. M., R. O. Peterson, T. D. Drummer, and S. A. Thomasma. 1994. Gray wolf response to refuge boundaries and roads in Alaska. *Wildlife Society Bulletin* 22:61–68.
- van Strien, A. J., C. A. van Swaay, and T. Termaat. 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology* 50:1450–1458.
- White, P., K. M. Proffitt, and T. O. Lemke. 2012. Changes in elk distribution and group sizes after wolf restoration. *American Midland Naturalist* 167:174–187.
- White, P. J., R. L. Wallen, and D. E. Hallac. 2015. *Yellowstone Bison: conserving an American Icon in Modern Society*. Yellowstone National Park, Wyoming, USA.
- Williams, S. H., S. M. Gende, P. M. Lukacs, and K. Webb. 2016. Factors affecting whale detection from large ships in Alaska with implications for whale avoidance. *Endangered Species Research* 30:209–223.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3152/full>