Methods to estimate population sizes of wolves in Idaho and Montana

Dr. Scott Creel, Distinguished Professor of Letters and Science, Department of Ecology, Montana State University, Bozeman MT 59717.

Decisions about the listing or delisting of species under the ESA hinge on information about the size of populations and their rate of change. Animal populations can almost never be counted completely or without error (*i.e.*, one cannot conduct a true 'census'), so population size usually must be estimated. Because inferences about population size are central to conservation and management, methods to estimate population size are well-developed, and there is broad consensus that accurate estimates of population size have three properties (*1-4*):

> First, they rely on methods that provide independent information about number of individuals detected (C) and the probability that an individual will be detected ($\hat{p}$). Estimating the detection probability $\hat{p}$ is critical because changes in the number of animals counted may not reflect a real change in population size (N), but rather a change in survey methods or effort, changes in animal behavior or location, or changes in counting conditions (for example, many animals are more detectable in winter than in summer).
>
> Second, they are based on enough data to provide precise information. *Precision*, in the sense that population ecologists or statisticians use the word, means that the upper and lower bounds on an estimate of population size are not too far apart: the margin of error ('confidence interval' or 'credible interval') is small. Population estimates with low precision cannot detect changes in population size until they become severe.
>
> Third, they are based on data that were collected in a manner that matches the assumptions of the statistical model that is used to estimate population size, and provides data that are representative of the broader population about which inferences will be made. These issues are the domain of sampling design. Many studies have shown that a method of population estimation works well if its assumptions are met, but that problems with sampling design cause large errors, particularly bias, when the assumptions are not met. *Bias*, in the sense that population ecologists or statisticians use the word, means that the estimate is systematically above or below the true population size.

In this report, I use the words bias, precision and accuracy in the sense that they are used in statistics. Precise estimates can be biased, and can create false patterns in population estimates. Unbiased estimates can be too imprecise to allow proper monitoring of population size. Accurate estimates of population size are both unbiased and precise. Also following the conventions of statistics, quantities that can be directly observed (like C) do not have a hat, and unobserved quantities that must be estimated by a statistical model have a hat (like $\hat{p}$).

There are two widely-accepted approaches that estimate population size ($\widehat{N}$) using a count (C) and independent information about the probability that an individual will be counted ($\hat{p}$). With *distance sampling*, one records the distance from a randomly sited transect (or point) at which animals were detected (*1, 5*). If transects are randomly distributed relative to true animal locations, then a decline in the number of animals seen at increasing distances from the transect can be used to determine how many animals were missed, and thus to estimate $\hat{p}$. Population size is then estimated as $\widehat{N} = {}^{C}/_{\hat{p}}$, and population density $\widehat{D}$ is equal to $\widehat{N}$ divided by the total area sampled by the transects. Distance sampling is widely used because it can applied with species that are not individually identifiable. However, it is rarely useful for large carnivores, simply because one cannot obtain enough sightings to produce a precise estimate of population size. Consequently, other methods have been developed to estimate $\hat{p}$ only with data from unmarked individuals and these will be considered below. In contrast, *Capture-mark-recapture* methods rely on individual identification to estimate the detection probability by examining patterns in the re-sighting of individuals over time (*1, 2*). For example, imagine a study that recorded whether individually recognized tigers were sighted in 6 time windows of 2 months each. Using a 1 to denote that a tiger was seen and a 0 to denote that it was not, the pattern 100101 clearly shows that an individual can go undetected in a single 2 month survey. The fact that the tiger was seen in periods 1, 4 and 6 reveals that it was alive but undetected in periods 2, 3 and 5. Extensions of this logic allow for accurate estimates of population size for large carnivores by studies that monitor marked or naturally identifiable individuals (*6-12*), though even with this (gold-standard) method, problems with sampling design can lead to substantial uncertainty or error in population estimates (*13*).

Capture-mark-recapture methods are difficult to apply to sparse, elusive and wide-ranging species. For any large carnivore, it is typical that several years of intensive field study are needed to obtain data that estimate $\hat{p}$ accurately, and thus convert the number of animals counted (C) into a good estimate of population size, $\widehat{N}$ (*9*). For the wolf, this problem is further complicated by the difficulty of identifying individuals, so that data are affected by 'false positives' (where two detections of the same wolf are accidentally recorded as different wolves) and 'false negatives' (where two different wolves are recorded as the same individual). For these reasons, state wildlife management agencies have considered capture-mark-recapture methods to be too difficult, expensive and time consuming for state-wide wolf monitoring programs in the Northern Rockies (*14, 15*), and a range of alternative methods have been assessed. These include: (**A**) Acoustic surveys using 'howl boxes' that play a highly amplified series of wolf howls and then record responses (*16, 17*). (**B**) Genetic capture-recapture methods that use genotypes from fecal samples to apply capture-recapture methods (*18-22*). The logic of this approach is that wolves are difficult to identify by sight, but not by their genotypes. (**C**) Uncorrected total counts of wolves that do not account for variation in sampling effort or detection probability (*23*). (**D**) Integrated population models that use data on survival and pack counts to estimate population size (*24*). (**E**) A method that combines predictions about the area occupied, territory size, and pack size into a predicted population size (*15*). (**F**) Space-to-event models using camera traps (*14, 25*).

Methods **E** and **F** have been employed by the states of Montana and Idaho to estimate wolf numbers in recent years that have seen dramatic liberalization of wolf hunting regulations. Analysis based on these models has concluded that wolf populations in the Northern Rocky Mountains have been little affected by major changes in hunting regulations (*14, 15, 26*). This conclusion runs contrary to prior studies that have detected strong effects of human killing on wolf survival (*27-30*), reproduction (*23, 31, 32*), territory size (*33*), and pack size (*20, 31, 32, 34*)). Direct analyses have typically shown that wolf population growth rates decrease as human killing increases, and that populations decline when the killing rate exceeds 19% to 29% of the population (*23, 27, 28, 35, 36*). One anomalous analysis suggesting that a much higher killing rates is sustainable (*37*) was affected by a statistical error (by including several direct, additive components of population growth in the set of variables used to predict population growth) (*23*). The conclusion that wolf populations have been stable in the face of liberalized hunting regulations also fails to explain why wolf harvests have decreased despite longer seasons, increased limits and a wider range of legal methods of hunting and trapping (e.g., in Montana 272 wolves were killed in 2020-2021, down 17% from the 329 killed the year before, and less than any of the three prior years). Decreasing offtake despite increasing the allowable methods of killing and extending the hunting season is a signal of population decline, and careful evaluation is warranted when newly developed methods to assess population size do not align with other available data.

Below, I evaluate space-to-event modeling for Idaho's wolf population (*14*), and integrated population occupancy modeling for Montana's wolf population (*15*).

Idaho's Space-to-Event Model fit to Camera Trap Data

Idaho's Statewide Wildlife Research Report for 2021 states that the Idaho wolf population was highly stable, with 1556 wolves in 2020, virtually unchanged from an estimate of 1566 in 2019, despite a dramatic liberalization of hunting regulations.

These population estimates were derived from a 'space-to-event' model fit to data from motion-sensitive camera traps (*14, 25*). The space-to-event model is one variant of models that estimate both population size and detection probability from unmarked individuals, with no additional information. While these models can estimate population size when the available data meet their assumptions, data from most real populations do not meet these assumptions, and it is widely recognized that such mismatches lead to inaccurate population estimates (*3, 4, 38-40*). Several of the developers of models that allow simultaneous estimation of detection probability ($\hat{p}$) and population size $\widehat{N}$ from unmarked animals have clearly stated the errors that arise from uncritical application of this class of methods (*3, 4*). Link et al. (*3*) warn that:

> "[these models] allow estimation of population size and detection probability from count data without requiring individuals to be identified… there is, however, a cost to using [these] models: inference is very sensitive to the model's assumptions. … Small violations of assumptions can lead to large biases in estimation. ***In cases where reliable estimates of population size are needed, we encourage investigators to allocate resources to acquiring additional data such as recaptures of marked individuals, for estimation of detection probabilities.***"

Similarly, Barker et al. (*4*) conclude that:

> "Using a capture-recapture perspective, we show that the loss of information that results from not marking individuals is critical … making reliable statistical modeling of *N* and *p* problematic using just count data. This is particularly the case for sparse data … ***reliable inference is only possible for relative abundance in the absence of questionable and/or unrealistic assumptions or with better data than seen in typical applications.***"

The Space-to-Event Model and its Assumptions

Among the set of methods that estimate population seize with data from unrecognized individuals, the space-to-event (STE) model is based on the premise that the area that must be examined before the first individual is seen should go down as population size goes up (*25*).

For the simple logic of the STE model to yield accurate estimates of population size, a number of critical assumptions must hold. In their original development of the space-to-event model, Moeller et al. (*25*) clearly identifed its fundamental assumptions and limitations.

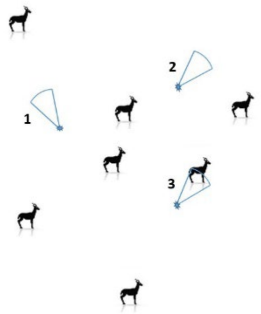


Figure 1. The 'space-to-event' model. A set of cameras (dots) is placed randomly, and each records a photograph of a known area (arcs) at fixed times. At any given time, summing the area for each camera in a random sequence until the first animal is detected (here by camera 3) allows an estimate of population density. As the number of animals increases, the area that must be sampled to detect one animal should decrease.

STE Assumption 1: Moeller et al. emphasize that the times and places at which cameras sample an area must be independent of the locations of animals, if the STE model is to produce accurate estimates of population size. Thus, STE is designed for use with time-lapse camera traps that are set to take photographs at fixed times, and not for data from cameras that are triggered by animal motion:

> "Motion-sensor sensitivity can change with the animal's size, distance, and angle of approach, as well as environmental variables such as ambient temperature and obstructive vegetation. If variable detection is ignored, the area in front of the camera can be hard to calculate, which can bias abundance estimates from models that depend on it… methods that take advantage of time-lapse photographs can be advantageous because they do not need explicitly to model motion-sensor variability." …" we randomly deploy time-lapse cameras that take photographs at predefined times"… "using time-lapse photographs, detection probability is defined as the probability that an animal is captured and correctly identified given it is in the camera's viewshed [when the camera takes a picture at a fixed time]"… "Since we are using time-lapse cameras instead of motion-sensor cameras, the maximum distance r is defined by field landmarks rather than the trigger distance".

STE Assumption 2: Moeller et al. emphasize that random placement of cameras is required for valid population estimates from the STE model:

> "the methods described here rely on random camera placement" … "we take observations of random areas of the landscape at an instant in time"… "we randomly deploy time-lapse cameras that take photographs at predefined times". "We recommend that future studies place cameras completely randomly…" "[future studies] should select sites randomly or systematically to avoid biasing estimates."

STE Limitation 3: Moeller et al. explicitly state that the need to use time-lapse cameras at random locations makes it unlikely to provide accurate estimates for species that are sparse (like the wolf or any other apex carnivore):

> [These methods] "can be quite suitable for common species that produce lots of pictures." … "As with using time-lapse photographs, randomly placed cameras may be more useful for common species than for very rare or elusive animals."

STE Assumption 4: Moeller et al. state that detections of a species by cameras must be independent in space (pairs of adjacent cameras do not produce correlated data) and time (detections at one time are uncorrelated with the next):

> "detections of animals are assumed to be independent in space and time. As long as cameras are randomly deployed, the properties of random sampling mean that animals captured at one camera are not any more or less likely to step in front of the next camera. However, it is slightly more difficult to address independence of animal detections at a single camera. We should consider animal behavior when defining sampling occasions and leave enough time for animals to redistribute across the landscape. We can help address independence of detections by selecting sampling occasions randomly or systematically, but we still may see autocorrelation across observations."

STE Assumption 5: Moeller et al state that the area that is sampled by a photograph must be accurately measured for each camera, because its population estimates are derived by summing these areas.

STE Assumption 6: Moeller et al. state that the STE model assumes that the probability of detection is one, for any animal within the viewshed of a camera, and again emphasize that this is not likely to be true for data from motion sensitive cameras.

> "All models are currently formulated under the assumption that detection probability is 1. When using time-lapse photographs, as in the STE and IS methods, this may be fairly reasonable. … On the other hand, motion-sensor cameras pose a larger issue for detection."

STE Assumption 7: Finally, Moeller et al. note that the STE method assumes that individuals have a Poisson distribution in space. In essence, this statistical assumption translates into an assumption that features of the landscape do not affect the animals' distribution, and that individuals move independently.

> "STE models assume that animals follow a Poisson distribution at the spatial scale of the camera. If animals are clumped due to landscape features, we could [modify the model in ways that the analysis of Idaho wolves did not]." … "we [made] the assumption that animals move independently."
>
> "animal abundance should follow a Poisson distribution at the level of the camera. In other words, animals should move independently at the spatial scale of the camera's viewshed." (*41*)

As the original developers of the STE model emphasized, it will be difficult to obtain data from sparse and elusive species that meet these assumptions. Even in example with data from an abundant and more easily observed species (elk) in the original description of the STE, 95% confidence limits on the elk population estimate from STE (1718 ± 226) did not overlap with the population estimate for the same elk population from an aerial survey (*25*). It should be noted that this example was chosen by the authors because they felt the data matched the model's assumptions adequately, and because elk were abundant enough to provide adequate data.

Application of the STE model to estimate the Idaho wolf population

In the analysis of wolf data from Idaho, many of the model's assumptions were violated (*14*), as described below.

Violation of STE Assumption 1: The STE model is designed for use with time-lapse camera traps that are set to take photographs at fixed times, and not for data from cameras that are triggered by animal motion. The Idaho wolf assessment is based entirely on data from motion sensitive cameras (*14*). It is difficult to assess how the authors used data from motion-sensitive cameras to approximate time-lapse data, as they simply stated that “we developed capture histories using spaceNtime”, which is an R package that can be used to fit the STE model (*41*). This statement does not allow a complete assessment of the assumptions that were made, but it is clear that most cameras in the wolf study would have taken no images of wolves at any fixed time, as 69% of the sampled locations recorded no images of wolves in the entire two year study. The guidelines that accompany the spaceNtime R package again emphasize that, when using the STE method, “time-lapse photography has huge advantages for determining sampling effort and detection probability… sampling effort is difficult to quantify from motion-sensor photography”(*41*).

Violation of STE Assumption 2: Random placement of cameras is required for valid population estimates from the STE model. For the STE method, “practices to increase detections, such as targeting high-use trails, should be avoided as they can bias the abundance estimate”(*41*). In the Idaho wolf assessment, “Cameras were deployed along roads or trails within 500 m of either historical wolf pack rendezvous sites (i.e., pup-rearing sites used in summer) or highly suitable predicted rendezvous site habitat”(*14*).

Assumptions 1 & 2 are fundamental to the logic of the STE model. The entire logic of this method depends on using data from areas that were sampled randomly in space and time, relative to the locations of the animals. The data for the Idaho wolf assessment are basically as far removed from meeting these assumptions as would be possible.

Heavily Censored Data from STE Limatation 3: Because assumptions 1 & 2 are critical, the developers of the STE model explicitly stated that it is “more useful for common species than for very rare or elusive animals” (*25*). The Idaho wolf assessment “violated the STE model assumption of random camera placement in part to ensure an adequate number of detections of a low-density animal” (*14*). Despite violating assumptions 1 & 2, the Idaho wolf assessment detected no wolves at all at 142 of 206 locations for the entire two-year study. This implies that on many sampling occasions, the area sampled probably contained no wolves, and thus the area that would have had to be sampled to detect a wolf remains unknown. In such a situation, the STE model simply assigns a value using a statistical distribution that rests heavily on assumption 7, which is not likely to be valid in the case of the Idaho wolf data (see below).

The magnitude of this hole in the data can be understood by some simple calculations. The area sampled by each camera should be directly measured for a careful implementation of the STE method (*25*), but was not measured in the Idaho wolf study (*14*). However, it was estimated that each camera sampled 106 m^2. Thus, 9,708 cameras would be needed to sample 1 km^2 at a fixed time (47 times more than the 206 occupancy cameras that were used in the Idaho study). The density of wolf populations typically ranges between 5 and 40 individuals/1000 km^2, so at any fixed time one can expect only 0.005 to 0.04 wolves in an area of 1 km^2. Thus, even with a sampling effort two orders of magnitude greater than was used, the vast majority of samples from fixed times will clearly detect no wolves, and therefore provide no direct information on the actual area required to detect a wolf (other than ‘more than was sampled’).

Violation of STE Assumption 4: Detections must be independent in space and time. Moeller & Lukacs unambiguously state that "practices to increase detections, such as targeting high-use trails, should be avoided as they can bias the abundance estimate" (*41*). The Idaho wolf assessment placed cameras on high use trails within areas that were known or strongly suspected to be rendezvous sites (where pups remain) (*14*). At these locations, it is not credible to assume that wolf detections will be independent.

Violation of STE Assumption 5: The area that is sampled should be directly measured for each camera, because STE population estimates are derived by summing these areas. The Idaho wolf assessment "did not estimate viewshed area during camera deployment "(*14*). Instead, they estimated that each camera sampled 106 m^2 at all times, although cameras were set at a downward angle in some years but not others, and individual camera view sheds depend on local topography and vegetation.

Violation of STE Assumption 6: The STE model assumes that an animal in the viewshed of a camera will be detected in a photograph taken at a designated time with 100% success. Using motion sensitive cameras this assumption is almost certain to be incorrect, a point that was emphasized in the development of the model (*25*). Here again, by using motion sensitive cameras, the design of the Idaho wolf study does not meet the assumptions of the STE model.

Violation of Assumption 7: The STE model assumes that animal abundance "follow[s] a Poisson distribution at the level of the camera. In other words, animals should move independently at the spatial scale of the camera's viewshed" (*25*). This assumption is closely related to assumption 4, but is narrowly limited to the statistical distribution that is assumed by the spaceNtime R package when fitting the STE model to a data set. For wolves, coordinated movements by pack members make it very unlikely that the Poisson distribution would provide a good fit to the data (because, as noted by Moeller et al., the use of this distribution assumes that individuals move independently). This problem also intersects with limitation 3. The Poisson distribution often fails to fit a data set because it includes a larger number of zeros than the Poisson distribution allows – this problem is well recognized in population modeling, and is often addressed by fitting a 'zero-inflated Poisson' (ZIP) or negative binomial distribution (*42-44*). As discussed above for limitation 3, it seems clear that Idaho wolf data are substantially affected by zero-inflation. The STE model could be modified to address this problem, but no such modification was incorporated in the Idaho wolf analysis. The consequence of assuming a distribution that does not fit the data is bias in the resulting population estimate.

Extrapolation Assuming that the Area Occupied by Wolf Rendezvous Sites is Unchanged. The Idaho analysis used the STE model to estimate wolf density from cameras deployed in a grid within areas that totaled 10,438 km^2. These areas were predicted to be used by wolves as rendezvous sites. To obtain a total population size for the entire state in recent years, these density estimates were extrapolated to a much larger (~10X) area that Ausband et al.'s 2010 model predicted to be used "using a resource selection function (RSF) that incorporated habitat and landscape characteristics of historical [rendezvous] sites". This RSF model identified the landscape and habitat features associated with rendezvous sites between 1996 and 2006 (*45*). By making an extrapolation based entirely on attributes of the landscape (and not the wolf population using that landscape), this approach is not capable of detecting a population decline that is associated with a change in decisions about rendezvous site locations between 2006 and now. Many species are known to alter habitat use in response to increased predation risk, including risk from humans (*46-49*), and it is plausible that wolves have made such shifts in response to changes in hunting, trapping and snaring regulations and extended seasons. Selectivity in the use of habitats is expected as population size changes (*50*), with greater selectivity as a population declines. The distribution of cameras in areas thought to hold rendezvous sites that detected no wolves in 2020 (Figure 30 in Idaho 2021 Statewide Wildlife Research Performance Report) suggests that a substantial number of predicted rendezvous areas were unused (71% of cameras detected no wolves during the 2020 study).

Detailed methods for this extrapolation have not been described in publicly accessible documents, but it appears that uncertainty in the estimated area occupied by wolves was not propagated into population estimates (Idaho 2021 Statewide Wildlife Research Performance Report). That is, it appears that density estimates for the sampled area were simply multiplied by a constant to obtain an estimate for the entire state (causing the precision of the statewide population estimate to be overstated). Nonetheless, the 95% confidence limits on this estimate range from 1,193 to 2031 wolves, sufficiently wide that a large population decline could go undetected.

Comparing genetic population estimates to STE estimates, the genetic point estimate goes from 184 to 125 between 2017 and 2018, a 32% decrease. The STE point estimate goes from 153 to 188, a 23% increase. These are substantial differences, and monitoring methods that consider a one-year increase of 23% to be comparable to a one-year decrease of 32% are perhaps not sufficiently accurate to guide management.

Taken together, these problems do not allow confidence in the population estimates for Idaho wolves, or the associated inference that greatly liberalized hunting has essentially no effect on population size.

Montana's Integrated Population Occupancy Model fit to Wolf Sightings from Hunters and FWP Biologists

In 2020 Montana began estimating the statewide wolf population using an integrated population occupancy model (iPOM) that purports to "eliminate the need for intensive field-based monitoring" but still "provide a means to reliably estimate wolf abundance" (*15*).

An integrated population model (IPM) is a set of component models that are simultaneously fit to data of multiple types (*44, 51-53*). IPMs normally include component models that are fit to intensive data from field monitoring, including mark-recapture estimates of survival rates, direct data on reproduction, and population counts from structured surveys (for many such examples, see Kery and Schaub's recent book on IPMs: (*51*)). I am aware of no examples other than the Montana iPOM suggesting that population size can be estimated reliably in the absence of direct demographic data and/or population counts.

Integrating a set of component models can provide more accurate estimates of population size because, in a true IPM, component models share common variables. Overlap between component models defines an IPM, and it is this overlap that allows an IPM to provide better estimates of population size than the component models by themselves (*51*). For example, the estimate of population size from a typical IPM must be consistent with data from population counts for that year, and with data on rates of survival and reproduction: the estimate of population size in year 2 must be consistent with the number counted that year, but must also be consistent with the estimated population size for year 1 and patterns of survival and reproduction between the two years. In a true IPM, information sharing between component models improves the estimates from the integrated model.

In contrast, the Montana iPOM is a sequence of unrelated models, so estimation errors simply compound from one model to the next. In the Montana iPOM, “an occupancy model first estimates annual wolf distribution, based on environmental covariates and wolf observations reported by hunters. A mechanistic territory model predicts territory sizes using simple behavioral rules and limited data for prey resources, terrain ruggedness, and human density. Together, these models predict the number of packs in a given area. Finally, a pack size model demonstrates that pack sizes are generally negatively related to terrain ruggedness, local mortalities, and intensity of harvest management. Total abundance estimates are derived by combining the predicted number of packs and pack sizes (*15*).”

The basic logic of this sequence is valid, but the sequential modeling approach inevitably leads to compounding estimation errors. Together with the oblique data on which each of these models rests, there is considerable doubt about the accuracy of population estimates from the iPOM, as described below.

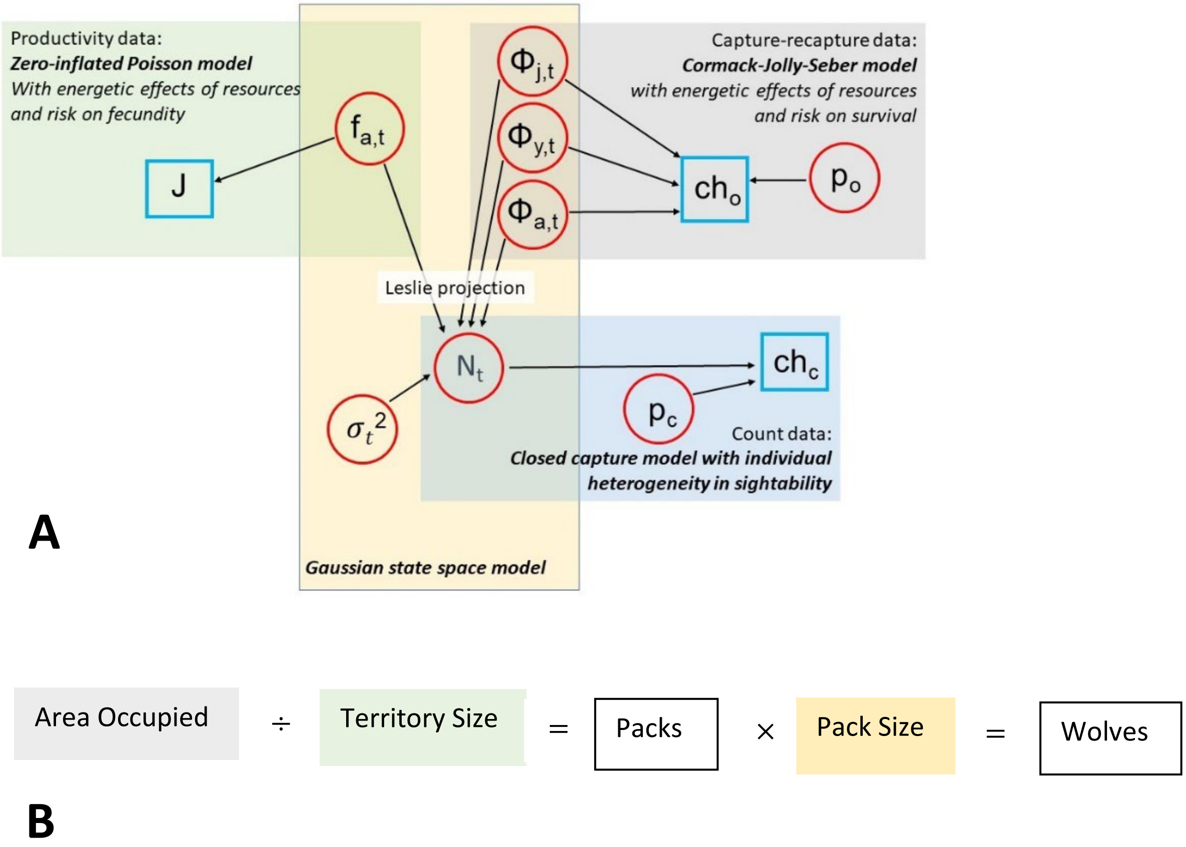


*Figure 2. (**A**) The overlapping relationship between component models in a typical integrated population model (IPM), with dependencies between variables estimated from direct monitoring data shown as a directed acyclic graph (DAG). (**B**) The sequential relationship of the Montana iPOM, with models in shaded boxes and values predicted from those models in open boxes. In **A**, overlap between component models reduces estimation error for population size (N). In **B**, estimation errors compound across a sequence of component models.*

1. Estimation of Area Occupied by Wolves. The iPOM begins by estimating the area occupied by wolves, gridding the state into 600 km^2 cells and using sightings by agency biologists and reports by deer and elk hunters to estimate the total area occupied. The basic patch occupancy model has been demonstrated to provide reasonable estimates of the area that wolves occupy in Montana, by comparison to direct field monitoring data from 2007 to 2010 (*54*). The occupancy model accounts both for failure to detect wolves in an occupied cell and for reports of wolves in cell that does not truly harbor any wolves. These false positives are very common: cells with no wolves had a greater than 95% chance that wolves would be reported, in areas with a large number of hunters (*54*). Although the model attempts to correct for this problem, reliance on raw data with this level of inaccuracy remains a concern for the accuracy of population estimates. As the original developers of this method note, "detection errors in general, and false positives in particular, can have large effects on estimates of range dynamics and other presence-absence processes" (*54*).

In addition, the data used to estimate the area occupied by wolves does not distinguish whether reports from hunters were based on pack-living wolves or floaters: all sightings were included, which would include areas with floaters but no resident packs. In subsequent stages of the iPOM approach, the conversion of area to wolves appears to overestimate population size in two ways. First, the area-to-wolves conversion multiplies by the expected size of a breeding pack, which is greater than the size of floating groups. Second, the resulting number is then multiplied by 1.125 to account for floaters, thus double accounting (because floaters will already have affected the estimate of area occupied).

2. Estimating Territory Size. As Sells et al. (2020) note, prior to the iPOM, Montana estimated wolf population size with an area-to-wolves conversion that used constants for pack size, territory size and territory overlap (*15, 55*), a method that cannot detect a change in population size for any reason other than the elimination of wolves from some areas. To address this issue, the iPOM replaces the constant values for pack size and territory size with predicted values from models, suggesting that this approach "eliminate[s] the need for intensive field-based monitoring". It is true that replacing constant values for territory and pack size with model estimates is likely to improve the accuracy of population estimates, but it is also true that assessing the structure, parameterization and predictive power of these models remains essential, when evaluating the extraordinary claim that they eliminate the need for direct population monitoring.

A. Rich et al. (33) Model of Territory Size: In the original analysis of Montana wolf territory size, territory size could not be estimated for most GPS collared wolves because they were killed by humans, dispersed, or the collar failed (*33*). Thus, estimates of territory size were based on high quality data from GPS collars in only 9 cases, and from VHF collars in 36 cases. These 36 cases included packs for which as few as 10 locations, and with a maximum of only 27 locations (mean = 17). Estimates of territory size depend on the number of locations, and sample sizes this small are likely to underestimate true territory size (*56*). The mean territory size estimated from these small VHF samples (445.3 km^2) was only 80% of the mean from GPS data (556 km^2), even though GPS estimates were for one year but VHF estimates pooled locations over two years. Longer periods yield larger territory sizes, and the mean of 8.5 locations/year (with as few as 5 locations/*year*) provides only a weak description of space use by these packs. Rich et al. (*33*) suggested that differences between sparse VHF data and GPS data (8 fixes/*day*) could be ignored because the 95% confidence limits of a regression relating VHF and GPS territory sizes had a slope that overlapped 1. This inference is flawed, because the regression was based on only 6 points (i.e., six territories with data of both types). Six points contain very little information, and the 95% confidence limits for the slope ranged from 0.43 (i.e., the VHF mean is 43% of the GPS mean) to 2.27. These values bracket a 1:1 relationship, but they are compatible with relationships far from 1:1 – failure to detect a clear difference does not establish equivalence.

The analysis failed to detect an expected effect of prey density on territory size (*33*), at least in part because prey density was not directly measured. Rather, data on the number of elk and deer killed by hunters in districts that overlapped with a territory (or in nearby districts in locations with no data on human offtake) was used as a proxy. It is possible that the size of Montana wolf territories is not affected by the availability of prey. It is also possible that the proxy does not provide an adequate description of the variable of interest.

Without direct data on prey density, the predictor variables in the final model of territory size were human density, terrain ruggedness, the number of wolves killed by humans, and the number of adjacent packs relative to territory size. This last variable is problematic because it includes territory size in a predictor of territory size, creating circularity or 'induced correlation'. Moreover, two of the four predictors (terrain ruggedness and human density) are constants at scale of measurement, and thus will predict a constant territory size for a given location, even if the true territory size changes as other, missing predictors (such as pack size or true prey density) changes.

The 95% confidence limits on the effects of some predictors on territory size are very wide. For example, the estimated effect of terrain ruggedness on territory size ranges from a lower bound of 67.34 to an upper bound more than 9 times greater, 635.03. When substituting the predicted territory size from this model in place of field monitoring, such imprecision reduces the accuracy of downstream estimates of population size.

Moreover, when the output of one model is used as input for another model (as with predicted territory sizes in the iPOM), there is an important distinction between confidence limits and prediction limits. Confidence limits (like those just mentioned) pertain to means. Prediction limits must be used to describe the estimation error for a specific territory size predicted by the model, and prediction limits are invariably much wider than confidence limits. The confidence limits shown in Figure 2 of Rich et al. (which relates observed territory size to the territory size predicted by the model) do not directly reveal prediction limits, but they do reveal that those prediction limits would be broad, reducing the true precision of downstream models that use these predictions to convert an estimate of the total area occupied into a predicted number of packs.

B: Sells and Mitchell (57, 58) Model of Territory Size. Sells & Mitchell first built a generic (non-wolf-specific) simulation model (*57*) that aligned with well-established results (*59-61*) that increases in food supply and competitors should decrease territory size, less-competitive individuals should occupy lower-quality (and thus larger) territories, and that relationships of territory size to environmental attributes are often curvilinear.

Sells et al. (*58*) then used data from GPS collared wolves to: (a) estimate territory size with 95% kernel utilization distributions, and (b) fit a generalized linear mixed model (GLMM) to test how territory size was affected by measures of prey availability, competition from neighboring packs, and human-caused mortality (with a random effect of pack identity appropriately used to account for data from two years being included for some packs). An impressive sample of 93 wolves were GPS collared, but of these, 36 collars failed, 21 were killed by hunters and trappers, 22 were killed by humans in other ways, and 3 dispersed out of the state. This allowed estimation of territory size for 28 packs Most of these fell in two clusters, one adjacent to Yellowstone National Park, and one between Glacier National Park and the Flathead and Blackfeet Reservations.

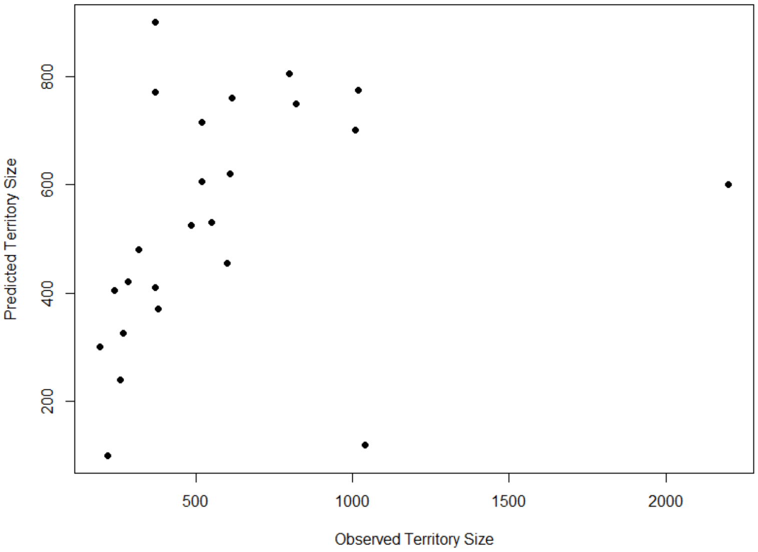
The GLMM assumed a normal error distribution, but the frequency distribution of territory sizes shows strong departures from normality (Figure 3 of ref (*58*)). The authors do not state whether alternative error distributions or transformations were assessed to see if they improved the model's fit.

As described just above for Rich et al.'s analysis, the data used as measures of prey density and human density were both spatially and temporally oblique. Prey density was assessed for a given location by taking the 10-year average for deer, elk and moose counts for the FWP administrative region in which a territory fell. Montana is divided into only 7 such regions, so each region more than 100 times larger than a wolf territory. With a time scale of 10 years and a spatial scale that is several orders of magnitude coarser that of the dependent variable (territory size), this approach is questionable. Sells et al (*58*) made an ad-hoc spatial adjustment to these prey densities using a factor that related the success rate of elk, deer and moose hunters for each district within a region to hunting success for the region as a whole, but the iPOM description (on p. 17) states that this adjustment had only a 'slight' effect (*15*). In some cases, missing data lead to substitution of spatially or temporally extrapolated values for predictor variables.

For the purpose of using the territory size model to estimate wolf abundance using the sequential iPOM approach shown above in Figure 2, it is problematic that group size was found to affect territory size, and that the number of adjacent packs also affects territory size. In the absence of direct data on these predictors, it is not clear how the iPOM could sequentially divide the total area occupied into a number of packs, when the estimate of territory size needed to do this depends on information about the number of packs (and their sizes). Sells et al. (*15*) describe this effort as follows: "We aimed to use only readily available data to demonstrate the model's ability to make predictions absent expensive, difficult-to-collect datasets, including data for wolves." Again, the claim that a population can be monitored without data from that population is extraordinary (and does not align with the approaches that Montana FWP uses for other species).

When comparing the predictions of the territory size model to the size of 26 territories estimated with direct GPS data, Sells et al. circumvented this problem by providing the model with the location of each territory center. Thus, they evaluated only whether the model can predict the size of a territory if its location, and the location of neighboring territories, is already known. This evaluation is not equivalent to testing whether the model could predict the number of territories and their sizes without such information (as the iPOM would require to be used in the proposed manner). With this 'tuning', the model broadly predicted the broad areas of Montana that were used and unused by wolves, but Figure 1.8 of Sells et al. (*15*) clearly shows that many, large areas that were not known to be used by wolves were included in simulated territories. Moreover, the broad alignment of the model's predictions with known wolf territories is affected by the point just mentioned: the model was provided direct data on territory centers for this assessment, which would not be available under the proposed use of the iPOM.

To evaluate the model's performance with this (unrealistic) advantage, Sells et al. (*15*) graphically compared observed and predicted territory sizes for 26 packs (their Figure 1.8), suggesting that "the model successfully predicted the territory sizes observed in real packs". However, the mean predicted territory size differed from observed mean by 202 km^2, relative to a mean of 582 km^2. Predictions with a bias this large would normally be considered inaccurate. Critically, Sells et al. did not report any statistical comparison of observed and predicted territory sizes to accompany their Figure 1.8. To obtain the (unreported) relationship), I ungraphed the data and confirming that the data I obtained replicated the reported mean difference between predicted and observed territory sizes. I then <u>regressed predicted territory size on observed territory size and found that slope is imprecisely estimated, but nonetheless has confidence limits that did not overlap one</u>. As noted above, this estimation error compounds with others in a sequence, but the magnitude of estimation error at this stage alone is so large that it refutes the statement that the model can "make predictions absent expensive, difficult-to-collect datasets, including data for wolves". <u>This model demonstrably did not predict wolf territory size accurately. Because the model underestimates territory size, its use in the iPOM will systematically overestimate the number of packs that occupy a fixed area, and thus overestimate population size.</u>



<u>3. Estimating Pack Size</u>

As for the prior stages in the iPOM sequence, Sells et al adopt a model of pack size that seeks to avoid the use of direct data: "absent annual monitoring data, total wolf abundance could be estimated with a model that predicts pack size. A useful model for MFWP would predict pack sizes with as few data as possible". Thus, all of the issues discussed in sections 1 and 2 above also pertain to this analysis.

Sells et al again argue that "the model reliably estimated annual mean group size" because a "regression of annual mean observed versus predicted group sizes included 1.0". Here again, failure to conclusively show that the slope differs from one does not meaningfully establish that the model makes useful predictions about the size of a given packs. The 95% confidence limits on this regression slope range from 0.32 to 1.31, a more than 4-fold difference. Here again, particularly when compounded with estimation errors at other stages in the iPOM, the prediction limits from a model with this precision are too wide to provide estimates of total population size that are of use for conservation and management of wolves.

Sells et al also plotted estimates of pack size through time and examined them informally to conclude that “our model successfully predicted annual mean pack size for wolves in Montana (Fig. 1.17)”. However, this figure also shows that (due to reliance on predictors that changed little) the model predicted that pack sizes remained completely constant over a period of 7 years from 2012 to 2017, a period in which the observed mean pack size changed by ~25%. This failure to detect changes in one of the critical components of the iPOM (over an interval of many years) again suggests that the iPOM is not likely to provide estimates of population size that can accurately guide conservation and management.

4. Combining component models sequentially in iPOM. The iPOM makes an unprecedented claim that a population can be accurately estimated to assess its conservation and management with methods that do not require direct monitoring. As described above, demonstrable problems with bias and precision affect two of the iPOM’s three component models. Because the models of territory size and pack size are explicitly designed to rely on variables that do not involves wolves, and many of these variables are measured obliquely or at coarse spatial and temporal scales, it is not surprising that when aggregated into the iPOM, no changes in wolf population size are predicted for the decade from 2010 to 2019. In essence this is just a restatement of the fact that many of the predictors used by the component models could not change (e.g., terrain ruggedness, forest cover and the human census data from 2010 are all constant) and other predictors in the component models changed relatively little (e.g., regional assessments of ungulate numbers, hunter success rates). For the reasons outlined above, there is little reason to accept that these demonstrate that the wolf population has been constant, or that greatly liberalized and extended hunting has no essentially no effect on wolf dynamics.

Conclusion

For the reasons laid out above, important changes in the wolf populations of Idaho and Montana are likely to go undetected by the monitoring methods now in use. Apparent changes in estimated population size might also be unrelated to true population size. The population estimates from these methods do not align with considerable prior data on the effects of human killing on wolf populations. As suggested by Barker, with models and data of the type that are now used to monitor wolves in Idaho and Montana, “reliable inference is only possible for relative abundance in the absence of questionable and/or unrealistic assumptions or with better data than seen in typical applications” (*4*). As suggested by Link, “where reliable estimates of population size are needed, we encourage investigators to allocate resources to acquiring additional data such as recaptures of marked individuals, for estimation of detection probabilities” that would allow application of methods whose accuracy is well established. Given the substantial changes in hunting regulations for a recently endangered species, reliable estimates are needed to assess the consequences for wolves in the Northern Rocky Mountains. As almost all population ecologists have concluded for almost all species, this will require direct monitoring of wolves.

References Cited

1. B. Williams, J. D. Nichols, C. J. Conroy, *Analysis and Management of Animal Populations.*, (Academic Press, London, 2002).
2. J. D. Nichols, Capture-Recapture Models. *Bioscience* **42**, 94-102 (1992).
3. W. A. Link, M. R. Schofield, R. J. Barker, J. R. Sauer, On the robustness of N-mixture models. *Ecology* **99**, 1547-1551 (2018).
4. R. J. Barker, M. R. Schofield, W. A. Link, J. R. Sauer, On the reliability of N-mixture models for count data. *Biometrics* **74**, 369-377 (2018).
5. S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, *Distance sampling*. (Wiley Online Library, 2005).
6. K. U. Karanth, J. D. Nichols, N. S. Kumar, J. E. Hines, Assessing tiger population dynamics using photographic capture-recapture sampling. *Ecology* **87**, 2925-2937 (2006).
7. P. Strampelli *et al.*, Camera trapping and spatially explicit capture–recapture for the monitoring and conservation management of lions: Insights from a globally important population in Tanzania. *Ecological Solutions and Evidence* **3**, e12129 (2022).
8. R. Davis *et al.*, Spatial partial identity model reveals low densities of leopard and spotted hyaena in a miombo woodland. *Journal of Zoology* **313**, 43-53 (2021).
9. M. A. Vinks *et al.*, Response of lion demography and dynamics to the loss of preferred larger prey. *Ecological Applications* **31**, e02298 (2021).
10. B. Goodheart *et al.*, Low apex carnivore density does not release a subordinate competitor when driven by prey depletion. *Biological Conservation* **261**, 109273 (2021).
11. A. J. Loveridge, A. W. Searle, F. Murindagomo, D. W. Macdonald, The impact of sport-hunting on the population dynamics of an African lion population in a protected area. *Biological Conservation* **134**, 548-558 (2007).
12. T. Mweetwa *et al.*, Quantifying demographic recovery in lions during a three-year moratorium on trophy hunting. *PloS one* **13(5): e0197030**, (2018).
13. M. W. Tobler, G. V. Powell, Estimating jaguar densities with camera traps: problems with current designs and recommendations for future studies. *Biological conservation* **159**, 109-118 (2013).
14. D. E. Ausband *et al.*, Estimating wolf abundance from cameras. *Ecosphere* **13**, e3933 (2022).
15. S. N. Sells *et al.*, "Improving estimation of wolf recruitment and abundance, and development of an adaptive harvest management program for wolves in Montana. Final Report for Federal Aid in Wildlife Restoration Grant W-161-R-1," (Montana Fish, Wildlife and Parks, Helena, Montana, 2020).
16. D. E. Ausband, J. Skrivseth, M. S. Mitchell, An automated device for provoking and capturing wildlife calls. *Wildlife Society Bulletin* **35**, 498-503 (2011).
17. A. Brennan, P. C. Cross, D. E. Ausband, A. Barbknecht, S. Creel, Testing automated howling devices in a wintertime wolf survey. *Wildlife Society Bulletin* **37**, 389-393 (2013).
18. S. Creel *et al.*, Population size estimation in Yellowstone wolves with error–prone noninvasive microsatellite genotypes. *Molecular ecology* **12**, (2003).
19. D. E. Ausband *et al.*, Monitoring gray wolf populations using multiple survey methods. *The Journal of Wildlife Management* **78**, 335-346 (2014).
20. M. S. Mitchell *et al.*, Estimation of successful breeding pairs for wolves in the Northern Rocky Mountains, USA. *The Journal of wildlife management* **72**, 881-891 (2008).
21. J. L. Stenglein, L. P. Waits, D. E. Ausband, P. Zager, C. M. Mack, Efficient, noninvasive genetic sampling for monitoring reintroduced wolves. *The Journal of Wildlife Management* **74**, 1050-1058 (2010).

22. J. V. López-Bao *et al.*, Toward reliable population estimates of wolves by combining spatial capture-recapture models and non-invasive DNA monitoring. *Scientific Reports* **8**, 2177 (2018).
23. S. Creel *et al.*, Questionable policy for large carnivore hunting. *Science* **350**, 1473-1475 (2015).
24. J. S. Horne *et al.*, Integrated population model to improve knowledge and management of Idaho wolves. *The Journal of Wildlife Management* **83**, 32-42 (2019).
25. A. K. Moeller, P. M. Lukacs, J. S. Horne, Three novel methods to estimate abundance of unmarked animals using remote cameras. *Ecosphere* **9**, e02331 (2018).
26. L. N. Joseph, C. Elkin, T. G. Martin, H. P. Possingham, Modeling abundance using N-mixture models: the importance of considering ecological mechanisms. *Ecological Applications* **19**, 631-642 (2009).
27. L. G. Adams, R. O. Stephenson, B. W. Dale, R. T. Ahgook, D. J. Demma, Population dynamics and harvest characteristics of wolves in the central Brooks Range, Alaska. *Wildlife Monographs* **170**, 1-25 (2008).
28. S. Creel, J. J. Rotella, Meta-analysis of relationships between human offtake, total mortality and population dynamics of gray wolves (Canis lupus). *PloS one* **5**, e12918 (2010).
29. A. M. Sparkman, L. P. Waits, D. L. Murray, Social and demographic effects of anthropogenic mortality: a test of the compensatory mortality hypothesis in the red wolf. *PloS one* **6**, e20868 (2011).
30. D. L. Murray *et al.*, Death from anthropogenic causes is partially compensatory in recovering wolf populations. *Biological Conservation* **143**, 2514-2524 (2010).
31. D. E. Ausband, C. R. Stansbury, J. L. Stenglein, J. L. Struthers, L. P. Waits, Recruitment in a social carnivore before and after harvest. *Animal Conservation*, (2015).
32. D. E. Ausband, M. S. Mitchell, C. R. Stansbury, J. L. Stenglein, L. P. Waits, Harvest and group effects on pup survival in a cooperative breeder. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20170580 (2017).
33. L. N. Rich, M. S. Mitchell, J. A. Gude, C. A. Sime, Anthropogenic mortality, intraspecific competition, and prey availability influence territory sizes of wolves in Montana. *Journal of Mammalogy* **93**, 722-731 (2012).
34. L. Y. Rutledge *et al.*, Protection from harvesting restores the natural social structure of eastern wolf packs. *Biological Conservation* **143**, 332-339 (2010).
35. J. A. Vucetich, "The influence of anthropogenic mortality on wolf population dynamics. Final peer review of documents amending and clarifying the Wyoming gray wolf management plan.," (US Fish and Wildlife Servive, 2012).
36. T. K. Fuller, L. D. Mech, J. F. Cochrane, in *Wolves: Behavior, Ecology and Conservation,* L. D. Mech, L. Boitani, Eds. (University of Chicago Press, Chicago, 2003), pp. 161-191.
37. J. A. Gude *et al.*, Wolf population dynamics in the US Northern Rocky Mountains are affected by recruitment and human–caused mortality. *The Journal of wildlife management* **76**, 108-118 (2012).
38. L. N. Joseph, C. Elkin, T. G. Martin, H. P. Possingham, Modeling abundance using N-mixture models: the importance of considering ecological mechanisms. *Ecological Applications* **19**, 631-642 (2009).
39. A. Duarte, M. J. Adams, J. T. Peterson, Fitting N-mixture models to count data with unmodeled heterogeneity: Bias, diagnostics, and alternative approaches. *Ecological Modelling* **374**, 51-59 (2018).
40. Y. Nakashima, Potentiality and limitations of N-mixture and Royle-Nichols models to estimate animal abundance based on noninstantaneous point surveys. *Population Ecology* **62**, 151-157 (2020).

41. A. K. Moeller, P. M. Lukacs, spaceNtime: an R package for estimating abundance of unmarked animals using camera-trap photographs. *Mammalian Biology*, (2021).
42. A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Savaliev, G. M. Smith, *Mixed effects models and extensions in ecology with R*. (Springer, New York, Berlin, 2009).
43. M. Kery, J. A. Royle, *Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and Static Models*. (Academic Press, 2015).
44. M. Kery, M. Schaub, *Bayesian population analysis using WinBUGS: a hierarchical perspective*. (Academic Press, 2011).
45. D. E. Ausband *et al.*, Surveying predicted rendezvous sites to monitor gray wolf populations. *The Journal of Wildlife Management* **74**, 1043-1049 (2010).
46. L. E. Morgantini, R. J. Hudson, Changes in diets of wapiti during a hunting season. *Journal of Range Management* **38**, 77-79 (1985).
47. J. P. Suraci, M. Clinchy, L. Y. Zanette, C. C. Wilmers, Fear of humans as apex predators has landscape-scale impacts from mountain lions to mice. *Ecology letters* **22**, 1578-1586 (2019).
48. P. Schuette, S. Creel, D. Christianson, Coexistence of African lions, livestock, and people in a landscape with variable human land use and seasonal movements. *Biological Conservation* **157**, 148-154 (2013).
49. S. Creel, J. Winnie, B. Maxwell, K. L. Hamlin, M. Creel, Elk alter habitat selection as an antipredator response to wolves. *Ecology* **86**, 3387-3397 (2005).
50. M. L. Rosenzweig, A theory of habitat selection. *Ecology* **62**, 327-335 (1981).
51. M. Schaub, M. Kery, *Integrated Population Models: Theory and ecological applications with R and JAGS.*, (Academic Press, Elsevier, London, 2022), pp. 622.
52. R. B. Chandler, J. D. Clark, Spatially explicit integrated population models. *Methods in Ecology and Evolution* **5**, 1351-1360 (2014).
53. J. R. Rhodes *et al.*, Using integrated population modelling to quantify the implications of multiple threatening processes for a rapidly declining population. *Biological conservation* **144**, 1081-1088 (2011).
54. D. A. Miller *et al.*, Determining occurrence dynamics when false positives occur: estimating the range dynamics of wolves from public survey data. *PLoS one* **8**, e65808 (2013).
55. L. N. Rich *et al.*, Estimating occupancy and predicting numbers of gray wolf packs in Montana using hunter surveys. *The Journal of Wildlife Management* **77**, 1280-1289 (2013).
56. P. N. Laver, M. J. Kelly, A Critical Review of Home Range Studies. *The Journal of Wildlife Management* **72**, 290-298 (2008).
57. S. N. Sells, M. S. Mitchell, The economics of territory selection. *Ecological Modelling* **438**, 109329 (2020).
58. S. N. Sells *et al.*, Evidence of economical territory selection in a cooperative carnivore. *Proceedings of the Royal Society B* **288**, 20210108 (2021).
59. F. L. Carpenter, D. C. Paton, M. A. Hixon, Weight gain and adjustment of feeding territory size in migrant hummingbirds. *Proceedings of the National Academy of Sciences* **80**, 7259-7263 (1983).
60. M. A. Hixon, F. L. Carpenter, D. C. Paton, Territory area, flower density, and time budgeting in hummingbirds: an experimental and theoretical analysis. *The American Naturalist* **122**, 366-391 (1983).
61. J. L. Brown, Territorial behavior and population regulation in birds: a review and re-evaluation. *The Wilson Bulletin*, 293-329 (1969).