Wolf Harvest Management Strategy Evaluation: Annual Report, 2024

FWP- UM Memorandum of Understanding NO. 23-0173
Federal Aid in Wildlife Restoration Grant NO. F23AF01662, Montana Wildlife Management Program

Hannah A. Sipe, Research Scientist, Montana Cooperative Wildlife Research Unit, University of Montana

Sarah N. Sells, U.S. Geological Survey, Montana Cooperative Wildlife Research Unit, Wildlife Biology Program, Ecology & Evolution Program, University of Montana

Justin A. Gude, Research Administrator, Montana Fish, Wildlife and Parks

Kevin M. Podruzny, Wildlife Biometrician, Montana Fish, Wildlife and Parks

Molly Parks, Carnivore Coordinator, Montana Fish, Wildlife and Parks

State: Montana
Agency: Fish, Wildlife & Parks
MOU: Wolf Harvest Management Strategy Evaluation
MOU number: 23-0173
Grant: Montana Wildlife Management Program
Grant number: F23AF01662
Time period: January 1st to December 31st, 2024

Suggested citation: Sipe, H. A., S. N. Sells, J. A. Gude, K. M. Podruzny, and M. Parks. 2024. Wolf Harvest Management Strategy Evaluation. Annual Report to Montana Fish, Wildlife and Parks. Missoula, Montana. https://fwp.mt.gov/conservation/wildlife-management/wolf

INTRODUCTION

Wolf harvest season setting is complicated and controversial. State law requires Montana Fish, Wildlife and Parks (MFWP) to both reduce the wolf population and avoid federal relisting under the Endangered Species Act (Montana Fish, Wildlife and Parks, 2002). Disparate stakeholder groups each have different objectives for wolf management. For instance, big game advocates want to see improved big game populations and hunting opportunities in northwest Montana, while wolf advocates want to see regulations that minimize wolf mortality. Decision making about season setting tries to balance these objectives. Wolf hunting and trapping season decisions are made by the Montana Fish and Wildlife Commission and are informed by annual wolf abundance estimates from an integrated patch occupancy model (iPOM, Sells et al., 2022c) as well as the predictions of wolf abundance into the future under potential constant harvest levels. Parametric uncertainty (uncertainty surrounding the value of a parameter) from the iPOM estimates is propagated through to future projections, providing the Commission with plausible and worst-case outcomes of different levels of public harvest over the short term, i.e., five years into the future, on the wolf population in Montana (Parks et al., 2024).

An alternative approach to inform wolf management and harvest decisions is through adaptive management. Adaptative management is appropriate for decisions that are made iteratively and when monitoring data are collected to learn about the outcomes from decisions, where monitoring data help to reduce critical uncertainties regarding ecosystem function or management outcomes (Walters, 1986; Williams, 2011). Management strategy evaluation (MSE) is one way to develop an adaptive management framework. MSE was developed by fisheries managers and scientists to more accurately and fully incorporate various forms of uncertainty, consider long-term time horizons, and add more transparency in a fisheries context (Punt et al., 2016). It has been used routinely and has become a standard approach for complicated and contentious marine fisheries management situations, yet it has been underutilized in wildlife management (but see Bunnefeld et al., 2013, 2011).

MSE is a forward simulation approach for testing prospective management options or strategies over a wide range of possible states (Punt et al., 2016). A MSE framework captures the 'truth' or what is happening in the system (termed the 'the operating model') and the information available to the decision makers (termed 'the estimation model' or 'management strategy'). More precisely, there are four main processes modeled. First, models are constructed based on current understanding and data to represent 'truth'. Second, the collection of monitoring data is simulated from the 'truth' model. Third, the simulated monitoring data are fit to an estimation model and the next time step's population metrics are predicted from the estimated parameters. Fourth, based on the estimation model results and the predictions, the decision-making process is simulated following a management strategy, whereby a decision is made and the implementation of this decision feeds back into the 'truth' model (Figure 1). This process continues through time. Additionally, each simulation through time is repeated to capture the full range of stochasticity and uncertainty.

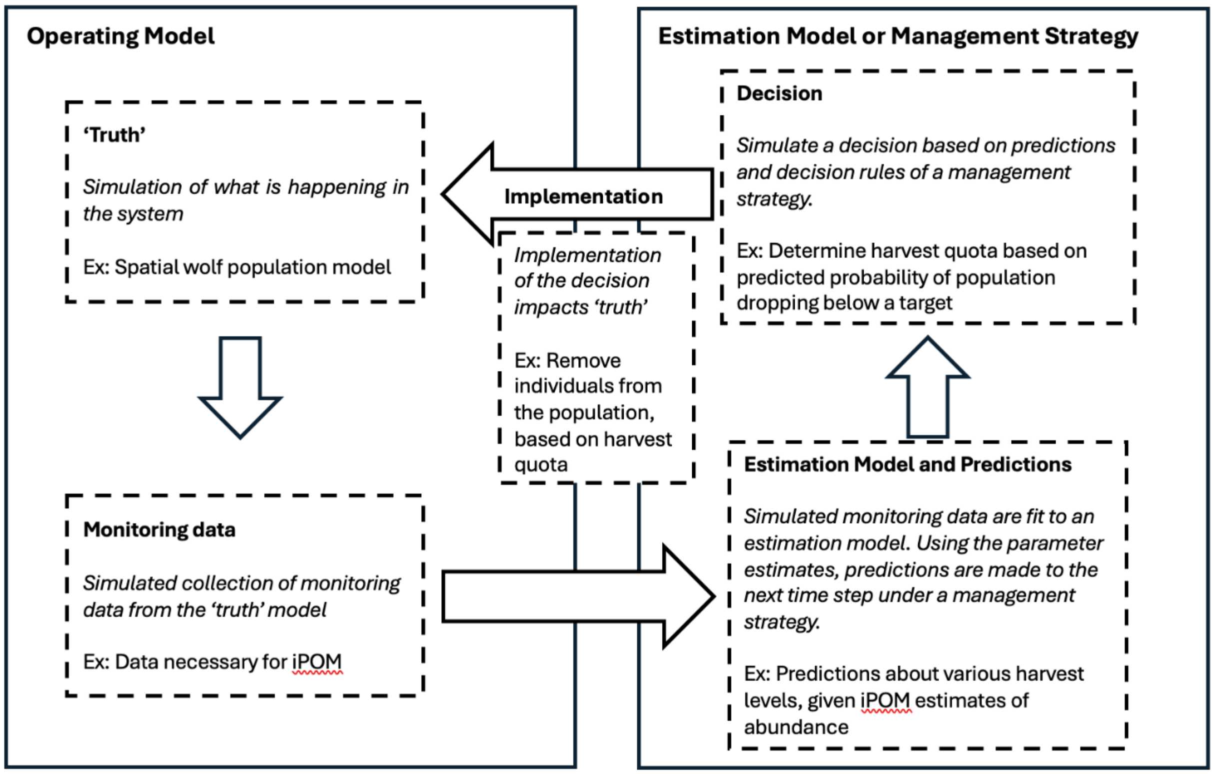

Figure 1. Conceptual diagram of management strategy evaluation, showing the component models that are part of the 'operating model' and 'estimation model'. On the left is the operating model, where the 'truth' is simulated along with the simulated collection of monitoring data from the 'truth'. On the right is the estimation model or management strategy, where monitoring data are fit to a predictive model, a decision is made and implemented, and the implementation impacts the 'truth' model. Note that this process is repeated through time and across repeat simulations. iPOM, integrated patch occupancy model.

Management strategies in the wolf management context are combinations of monitoring programs and harvest control rules (e.g., liberal or restrictive regulations applied when the population is above or below some threshold, respectively). The outcomes from MSE are then assessed in their ability to meet fundamental objectives outlined in the wolf management plan (Runge et al., 2013). Several performance metrics related to fundamental objectives, or other measures of performance, can then be tracked and summarized for each management strategy based on the simulations, essentially allowing for experimental application of different strategies to evaluate relative efficacy over the long term.

Multiple forms of uncertainty can be captured in the MSE. For instance, uncertainty of regulations on realized harvest can be accounted for in simulations to reflect that regulations do not always prescribe exact harvest levels, i.e., partial controllability. The MSE approach is also a useful tool for identifying monitoring and research priorities, and for increasing the efficiency of monitoring programs for a given harvest control rule. The simulation structure allows for 'experimental' implementation of different monitoring programs or levels of monitoring intensity, to examine how or if management decisions (or the population) might be affected with different combinations of monitoring schemes and harvest control rules. Further, incorporating uncertainty in wolf population dynamics (e.g., the relative effect of harvest or density dependence on population trend) along with variable monitoring and harvest control rules in a MSE permits analyses that identify monitoring and research (value of information analyses; Canessa et al., 2015; Runge et al., 2011). Such analyses allow for concrete ties between additional data collection and management.

OBJECTIVES

1. Develop a MSE framework for wolves consisting of operating and estimation models to simulate the real world combined with simulated alternative management strategies (fundamental objectives, monitoring programs, and harvest control rules) and variable monitoring program elements.
2. Based on these results, provide scenario outcomes to MFWP leadership and other decision makers about combinations of harvest control rules and monitoring program elements that are most likely to meet fundamental objectives over long time horizons.

GENERAL PROGRESS

Work on this project began in January 2024. We continue to work closely with MFWP staff to determine various aspects of the MSE. Discussions to date have centered on the general structure of the component models (Figure 1), including how to evaluate management strategies in terms of fundamental objectives that are not direct proxies of wolf abundance (e.g., public attitude objective). Further, we worked with MFWP to identify data and parameter estimates available to parameterize the underlying 'truth' or operating model. We also identified initial management strategies to include in the MSE, but the set will require further refinement as work continues.

An initial prototype of the MSE has been developed, the structure of which will be used to build up each component model. Building up of the 'truth' model is in progress and the structure is expected to be finalized in the coming months. One of the estimation models for the MSE is iPOM (Sells et al., 2022c) since it is the current method for estimating wolf abundance and is used to inform management decisions. Hence, we have taken this opportunity to review iPOM and make updates where necessary.

REVIEW AND UPDATES OF iPOM

iPOM works by integrating three models to estimate wolf abundance in Montana, including a dynamic occupancy model, a mechanistic territory model, and a group size model (Sells et al., 2022c). We took this opportunity to investigate and update each model prior to incorporating iPOM into the MSE. The primary impetus for taking this step was to code iPOM in an entirely Bayesian format and convert the Bayesian estimation code to NIMBLE in R, to enable incorporation into the MSE simulations. Additionally, we are increasing the efficiency of Bayesian estimation using Bayesian updating (Applestein et al., 2022; Ellison, 2004), instead of refitting the entire time series each year, thereby decreasing computation time in the MSE and eliminating slight changes in previous year estimates every time iPOM is fit to the current year's data. Updates to each iPOM model are discussed individually below. The group size model is discussed in greater depth, as this model is close to being finalized.

Dynamic patch occupancy model

The dynamic patch occupancy model (MacKenzie, 2018) in iPOM provides an estimate of whether wolves are present or absent in 600 km^2 grid cells across Montana. Hunter observations of wolf packs greater than 2 and less than 25 individuals, during the 5-week general rifle season (approximately late October-November), provide the observation data for the model. A false positive detection process is used in the occupancy model to account for potential misidentification or misreporting of wolves by hunters (e.g., if a coyote was identified as a wolf; Miller et al., 2011). To allow the model to estimate whether detections are true detections by hunters (wolves were present at the grid cell) or false detections, data from monitoring efforts by MFWP wolf specialists provide information about known packs (Sells et al., 2022c; for previous version of this model, see Miller et al., 2013 and Rich et al., 2013).

We are in the process of testing various aspects of the occupancy model, such as the means to integrate the hunter and MFWP observations. We are also testing different covariate models on each of the occupancy model parameters to identify any improvements to the model's predictive power (e.g., additional environmental covariates or search effort).

Mechanistic territory model

iPOM uses a mechanistic territory size model to predict territory size (km^2; Sells et al., 2022a, 2021; Sells and Mitchell, 2020). The territory model is a spatially explicit, agent-based model that predicts territory size for packs by maximizing food resources and minimizing costs of travel, competition, or mortality risk. The predictions from the mechanistic territory size model for each ecoregion in Montana and across a range of pack densities are employed by iPOM (Sells et al., 2022c). The updated version of iPOM will employ these same data and methods but instead of directly sampling from the predicted territory sizes, we fit the predictions to a statistical distribution and sample from that distribution. Modeling territory size in this way captures a wider range of uncertainty, allowing for more variation, because a statistical distribution is being sampled.

We tested various statistical distributions and found the gamma distribution to fit the predicted territory size data the best. So, for each density and ecoregion, the set of mechanistic model territory sizes are fit to a gamma distribution, then samples from the gamma distribution are used in the overall wolf abundance estimates. Fitting a gamma distribution to the territory sizes results in nearly (within random deviations from sampling) the same estimates as the original iPOM in terms of territory size and overall wolf abundance.

GROUP SIZE MODEL

Introduction

In Sells et al. (2022c), we discussed the future necessity of revisiting the original group size model (Sells et al., 2022b) that was developed for iPOM. The original model was fit to group size data from 2005 – 2018. Selection of covariates was based in part on the data anticipated to be available into the future, and in Sells et al. (2022c), we emphasized the importance of revisiting the model if conditions after 2018 (e.g., harvest regulations) changed beyond that observed from 2005 – 2018.

The original predictive model for group size included five covariates: pack density index (per 1000 km^2; based on pack centroids, measured as the long-term average density over 2005 – 2018), terrain ruggedness, harvest regulations, control removals, and ecoregion. We selected the long-term average for the pack density index because we were uncertain if data to measure annual pack densities would be available. However, we now know that additional data are available and anticipate these data will be available in the future, offering the opportunity to revise the group size model to include an annual density of packs. In addition to group size observation data being available, there have also been updates in covariate information. For instance, the original model's categorical variable for harvest regulations was based on 2005 – 2018 levels (ranging from no harvest to restricted and liberal harvest). Since 2021, harvest regulations have been further liberalized (with limits of ≤20 wolves per person, and snaring, electronic calls, baiting, and private land night hunting now allowed, as well as reimbursement of expenses). Additionally, we have found that data from Wildlife Services for timing and location of control removals have been limited and incomplete since 2018.

We sought to develop and test a new group size model for iPOM given that > 5 years have passed since the first iteration of the group size model was presented. In this time, harvest regulations became more liberalized and new covariate data became available, as have the 5 additional years of response data (pack size counts). We hypothesized that various group characteristics, environmental features, and human-related features would influence group size (Table 1). In general, we hypothesized that groups would be larger in areas with more food resources and areas for security cover (e.g., areas with higher greenness values [Normalized Difference Vegetation Index (NDVI)] and forest cover), and smaller in areas where food resources are more difficult to obtain (e.g., areas with higher ruggedness) and mortality risk is greater (e.g., areas with more roads, high building density, etc.). We included spatiotemporal covariates where available, which we hypothesized influenced group size through direct group additions (presence of pups reported) and removals (mortalities nearby) and effects on dispersal decisions (number of neighboring groups; Sells et al., 2022b).

Table 1. Hypotheses and associated covariate data analyzed for the new group size model for wolves in Montana.

Covariate	Details	Hypothesized effect on group size	Predicted effect on group size	Global Model Result	In final model?	Data type (ST = spatiotemporal; S = spatial; T = temporal)	Data source
Pups	Pups reported at year end (Y/N)	Groups w/ pups are larger due to direct additions to the group	Positive	Positive	Y	ST	MFWP year-end group monitoring records
Ecoregion-II	Categorical based on Level-II ecoregions (mountains vs plains) [A]	Plains groups are often relatively new and easily hunted	Negative	Negative	Y	S	https://www.epa.gov/eco-research/ecoregions (Omernik and Griffith, 2014)
Elevation	Continuous, 1000 meters [B]	Ecological differences across elevations leads to variable group sizes	Variable	None	N	S	Package elevatr (Hollister et al., 2023)
Ruggedness	Vector ruggedness measure [B]	Rugged terrain makes prey more difficult to catch, leading to less food resources in rugged areas and thus smaller groups	Negative	Positive	Y	S	Elevation data & vector ruggedness measure calculations (Sappinton et al., 2007) via package SpatialEco (Evans, 2018)
NDVI	Normalized Difference Vegetation Index [C]	Higher NDVI often means more forage for prey, leading to greater food resources for groups and thus larger groups	Positive	Positive	Y	S	Package MODIStsp (Busetto and Ranghetti, 2016) to access MODIS data 1/16/2025 12:18:00 PM
Density of forest edge	Km of edge per km^2 [D]	Forest edge is adjacent to more human-dominated landscapes and thus associated with smaller groups	Negative	Negative	Y	S	National Land Cover Database (U.S. Geological Survey, n.d.)
Density of riparian	Km of edge per km^2 [E]	Riparian areas provide habitat for prey and security cover for wolves and thus leads to larger groups	Positive	Positive	Y	S	National Hydrography Dataset (U.S. Geological Survey, 2020)
Proportion forest cover	% of grid cell with forest cover	Forest cover is associated with prey resources and security cover and thus leads to larger groups, OR, forest cover is associated with white tailed deer instead of mule deer and elk, supporting fewer wolves per group	Positive or negative	Negative	Y	S	National Land Cover Database (U.S. Geological Survey, n.d.)

Proportion herbaceous cover	% of grid cell with herbaceous cover	Herbaceous cover is associated reduced security cover and thus leads to more mortalities and smaller groups	Negative	Negative	Y	S	National Land Cover Database (U.S. Geological Survey, n.d.)
Neighboring groups	# of groups in the grid cell & 8 surrounding cells	Higher group densities lead to lower dispersal and thus larger groups	Positive	None	N	ST	MFWP year-end monitoring records
Density of buildings	Km of edge per km^2	More human presence leads to more wolf mortalities and group disruptions, and thus smaller groups	Negative	None	N	S	Microsoft Buildings Footprints (Microsoft, 2018)
Density of roads	Km of edge per km^2	“”	Negative	Potentially negative	N	S	TIGER roads data (Bureau USC, 2018)
Human disturbance	Index [B]	“”	Negative		N	S	Montana Natural Heritage Program (Montana Fish, Wildlife and Parks, n.d.)
Harvests nearby	# of wolf harvests in the grid cell & 8 surrounding cells	Harvests leads to direct group size reductions and potential group dissolution/increased dispersals	Negative	Negative	Y	ST	MFWP year-end harvest records
Harvest regulations	Categorical: no harvest, restricted, liberal, aggressive [F]	Liberalized harvest regulations lead to more mortalities and group disruptions, and thus smaller groups	Negative	Negative	Y	T	Hunting regulations

A. Although Sells et al. (2022b) included Level-III ecoregions in their final predictive group size model, the effects of these finer-scale ecoregions overlapped 0; therefore, we opted to include the simplified Level-II ecoregions of mountains (western MT) versus plains (eastern MT) for this new analysis.
B. We transformed several covariates from their default values to bring covariate values into less extreme differences. We transformed elevation by dividing meters by 1000 (such that values represent 1000’s of meters). We multiplied the ruggedness index by 100 units to bring the normally small values to a scale of 0 – 2.4 units. And we divided the human disturbance index by 1000 units to bring the normally large range to a scale of 0 – 1.63.
C. Measured at peak green-up (Jun 15 – Jul 15), from 2005 – 2020, and averaged across years.
D. Identified with the 2016 National Land Cover Database (mrlc.gov) as forest (deciduous, evergreen, mixed forests, and woody wetlands) and non-forest (remaining classes).
E. Identified as riparian edges of waterbody boundaries, rivers, streams, and artificial paths outside waterbody boundaries.
F. Categories: no harvest (≤2008 and 2010), restricted harvest (2009 and 2011, when statewide harvest was limited by a quota, seasons were shorter, bag limits were low, and trapping was prohibited), liberal harvest (2012 – 2021, when statewide harvest quotas were removed, seasons were longer, bag limits were higher, and trapping was allowed), and aggressive harvest (2022 – 2023, when snaring and night hunting became legal, bag limits increased to 20 wolves/person/year, and reimbursement of expenses).

METHODS

Data

To develop the new group size model, we first compiled group size data collected from monitoring efforts from 2005 – 2023. This included 2,110 year-end monitoring reports of group size, which MFWP wolf specialists classified as good, moderate, or poor-quality counts at each reporting year. Good quality counts were based generally on multiple data points collected throughout the year, e.g., using visual surveys, trail cameras, track surveys, etc. Of the 2,110 counts, 943 (approximately 45%) were rated as good quality and served as the basis for our model development. Moderate and poor-quality counts represent data based on fewer observations throughout the year and are expected to be reported with less precision (e.g., may represent undercounts to some unknown degree), so were not used as data in the model. Each group count was also assigned an approximate group centroid associated with the group's territory, based on radio collar data (where available), field surveys, and expert knowledge (Figure 2).

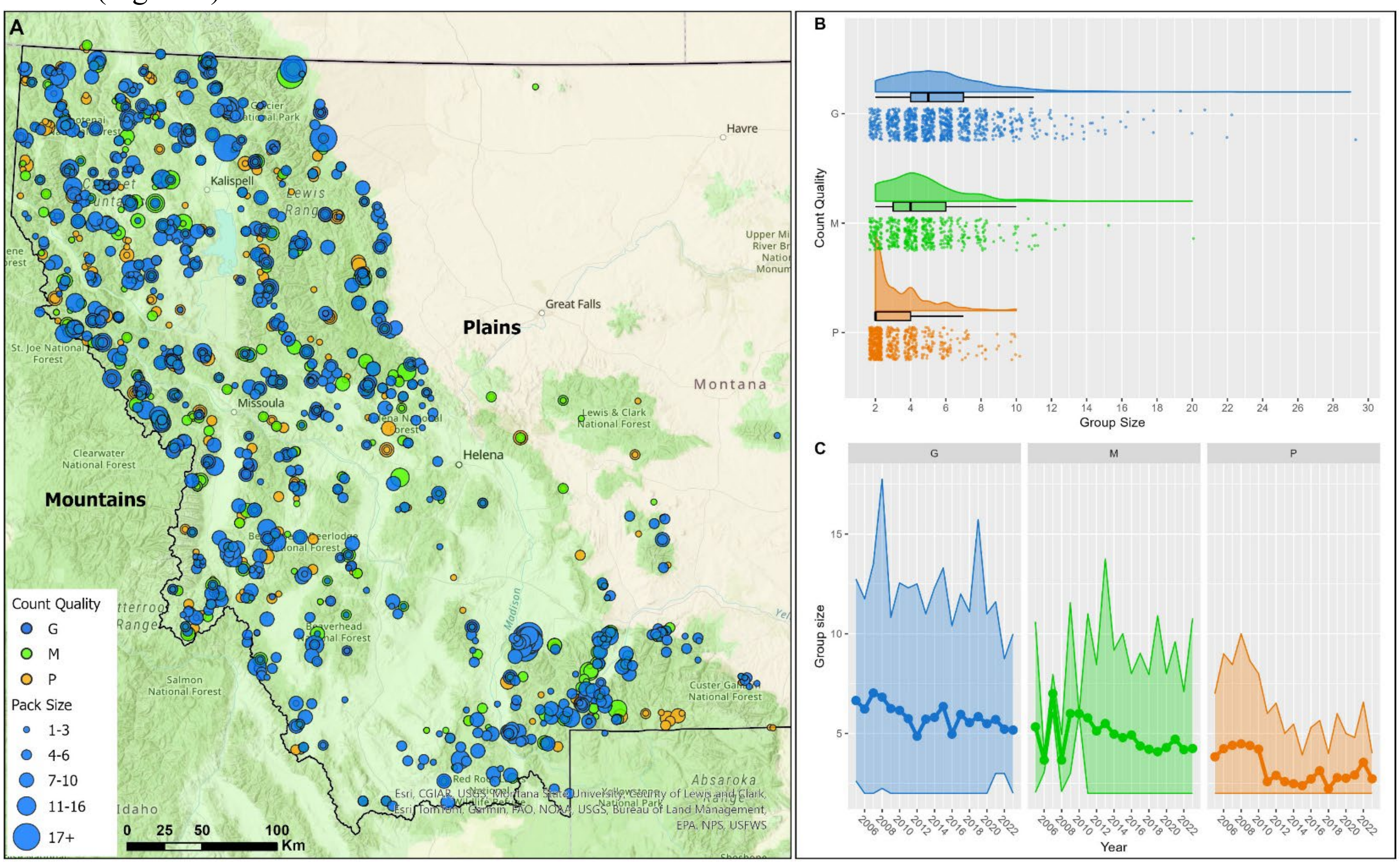

Figure 2. **Panel A:** group sizes and count qualities observed in Montana, 2005 – 2023. **Panel B:** group size in relation to count quality. Points below the density plots represent individual data points. Of the good quality counts, observed group sizes ranged 2 – 29 and averaged 5.8 wolves, with a 50% interquartile range of 4 – 7 wolves per group. **Panel C:** group size varied by year and exhibited a general downward trend. Note that ribbons extend to 2.5 and 97.5% quantiles (where sufficient data were available). In all panels, 'G' = good quality, 'M' = moderate quality, and 'P' = poor quality.

We developed the group size model by focusing on the 600-km^2 iPOM grid as the sample unit, as our end goal was to predict group size in each grid cell for iPOM. Accordingly, we identified which grid cell each group centroid occurred in. If more than one centroid occurred in a cell that year, we calculated the mean value of group sizes therein. We then developed a matrix of 695 grid cells by 19 years (2005 – 2023), where each matrix cell received the associated group size value or an NA if no group size was available.

For each iPOM grid cell, we next compiled data for group characteristics, environmental features, and human-related features that we hypothesized influenced group size (Table 1). For each continuous variable, we measured the mean value within the iPOM grid cell. For categorical variables, we assigned the grid cell to the category with greatest area (e.g., which ecoregion predominated).

Model

We used a Poisson generalized linear mixed effects model to model the size of wolf groups as a function of covariates. The observed good quality counts of wolf packs $y_{i,t}$ at each grid cell i in each year t were modeled as Poisson distributed, namely

$$y_{i,t} \sim Poisson(\lambda_{i,t}) \tag{1}$$

The rate of the Poisson, $\lambda_{i,t}$, was modeled as arising from spatial and temporal covariates, as well as a random effect on each grid cell:

$$log(\lambda_{i,t}) = \alpha + \boldsymbol{X\beta} + \varepsilon_i \tag{2}$$

$$\varepsilon_i \sim normal(0, \sigma)$$

where α is the intercept term for $\lambda_{i,t}$, $\boldsymbol{X}$ is a matrix of covariates, $\boldsymbol{\beta}$ is a vector of coefficient parameters for $\lambda_{i,t}$, and ε_i is the grid cell level effect for grid cell i from a hyperdistribution with $SD = \sigma$.

Model implementation

The group size model was fit in a Bayesian framework. We specified uninformative priors for all parameters. Specifically, priors on the intercept and coefficient terms (i.e., α and $\boldsymbol{\beta}$) were set to $normal(0, SD = 10)$ and the prior for the SD of grid cell level effects (i.e., σ) was $uniform(0, 20)$. Models were fit in NIMBLE (NIMBLE Development Team, 2024; Valpine et al., 2017) and accessed through R (R Core Team, 2024). We used Program R 4.4.0 (R Core Team, 2024) and packages tidyverse (Wickham et al., 2019), terra (Hijmans, 2020), and sf (Pebesma, 2018; Pebesma and Bivand, 2023) to prepare data, package NIMBLE (NIMBLE Development Team, 2024; Valpine et al., 2017) to run models, and package ggplot2 (Wickham, 2016) to plot results. For each model, we ran 3 chains for 250,000 MCMC iterations with the first half as burn-in and a thinning rate of 10 (to decrease file size). We assessed model convergence using the Brooks-Gelman-Rubin convergence statistic (Brooks and Gelman, 1998), ensuring $\hat{R} < 1.1$, and by visually inspected posterior trace plots. We plotted the mean value, median value, and 95% credible intervals (CRIs) of the MCMC chains to examine the density of the posteriors.

Model testing and selection process

Model testing and selection took place in two stages. The first stage reduced the global model based on covariate effects and WAIC (Watanabe-Akaike information criterion; Hooten and Hobbs, 2015; Watanabe, 2013) and the second stage assessed the predictive ability of the models.

We initiated model building by first testing a global model with all covariates included and assessed 85% credible intervals (CRI) of each coefficient to determine which covariates had no effect (85% CRIs centered on 0) and uncertain effect (85% CRIs included but were not centered on 0). We next tested reduced models that each omitted one of the covariates with CRIs overlapping 0. We considered a reduced model to support the omission of that covariate if WAIC decreased compared to the global model WAIC (i.e., if ΔWAIC was negative). However, if ΔWAIC differed by <1, we considered the omission of that covariate to be uncertain and in need of further testing.

We next developed two semi-final candidate models for further testing: one omitting all covariates identified for certain omission in the previous step, and a second model also omitting any covariate whose omission was uncertain (i.e., due to minimal ΔWAIC). To select a final model from the two semi-final models, we predicted group size based on estimated model parameters for each grid cell in each year. For comparison, we did the same using parameter estimates from the global model. For each of these three models (the global model and two semi-final models), we then calculated the mean squared error (MSD) of predicted versus observed group sizes for each set of count quality classifications (i.e., good, moderate, and poor-quality counts, as designated by MFWP wolf specialists each year). We selected the final model as that which minimized MSD across all count quality classifications. We then summarized the percentiles of predictions off by ≤ 1, ≤ 2, and ≤ 3 wolves compared to observed group sizes.

After identifying the final group size model, we used the model to calculate, by grid cell and year, the mean estimate of group size and associated standard deviation. These values will then replace the previous Sells et al. (2022b) group size model estimates in iPOM.

Preliminary results

From 2005 – 2023, MFWP monitored 41 – 147 groups per year within Montana, totaling 2,100 group-years. Of these, 938 were good quality counts, based on 26 – 68 good quality observations per year. Annual mean size was 4.86 – 7.03 wolves per group, averaging 5.78 across all years (Figure 2). Although a handful of groups were large (n = 5 groups with ≥ 20 members), the majority were small; 80% of packs had ≤ 8 wolves and the 50% interquartile range was 4 – 7 wolves.

The global model identified four covariates as having no effect (CRIs centered on 0: elevation, neighboring groups, density of buildings, and density of roads) and three covariates as having uncertain effects (CRIs included but were not centered on 0: proportion herbaceous cover, density of riparian, and human disturbance). Testing reduced models that omitted each of these seven covariates and resulted in lower WAIC for all but proportion herbaceous cover and density of riparian (Models Reduced A – G, Table 2). Most ΔWAICs were approximately −1.5 to −2, except for human disturbance, for which ΔWAIC was −0.4. Accordingly, we tested two new models in our final candidate set, where Model Reduced H omitted elevation, neighboring groups, density of buildings, density of roads, and Model Reduced I omitted these same parameters and human disturbance.

Our final tests of the global model and two semi-final models revealed that Model Reduced H had the lowest MSD for each count quality classification, outperforming both the global model and Model Reduced I (Table 3). Model Reduced H thus became the final model.

Model H's predictions were closely aligned with observations (Figure 3). Compared to the group sizes reported as good quality counts, the model successfully predicted to within +1 wolf of the observed group sizes 55.3% of the time, ≤+2 wolves 78.8% of the time, and ≤+3 wolves 88.9% of the time. Of the moderate quality counts (n = 509), the model successfully predicted to within +1 wolf of the observed group sizes 53.0% of the time, ≤+2 wolves 80.7% of the time, and ≤+3 wolves 91.5% of the time. Of the poor quality counts (n = 658), these values were 40.7%, 75.9%, and 88.2%, respectively. Accuracy did not generally vary by MFWP management region, grid cell, or year (Figure 3). Slight bias was observable for poor quality packs across years (Figures 3C), as expected given that poor quality packs are almost certainly an undercount of true pack sizes (our model predicted these packs were larger than reported from observations alone). Accordingly, the model provided good accuracy in predicted versus observed group sizes, regardless of the count quality classifications, and appears to help correct for undercounts that are probable in poor quality observations. Figures showing relative bias of predictions, predicted group size by year and MFWP management region, and observed vs. predicted group sizes can be found in the Appendix.

Table 2. Models analyzed for the new group size model.

Model	Description	WAIC	Delta WAIC[1]
Global	All	3678.281	
The global model except reduced by one parameter:			
Reduced A	Omit elevation	3675.968	-2.313
Reduced B	Omit neighboring groups	3676.275	-2.006
Reduced C	Omit density of buildings	3676.577	-1.704
Reduced D	Omit density of roads	3676.700	-1.581
Reduced E	Omit proportion herbaceous cover	3680.112	1.831
Reduced F	Omit density of riparian	3679.703	1.422
Reduced G	Omit human disturbance	3677.881	-0.400
Semi-final models (reduced by multiple parameters):			
Reduced H	Omit elevation, neighboring groups, density of buildings, density of roads	3672.081	-6.200
Reduced I	Omit elevation, neighboring groups, density of buildings, density of roads, human disturbance	3672.032	-6.249

1. Compared to the global model

Table 3. The mean squared error (MSD) of observed versus predicted group size, across all good quality counts (G), moderate quality counts (M), and poor quality counts (P), along with the percentile of predicted group sizes within 1, 2, or 3 wolves of the observed group sizes, for the global and top 2 reduced models.

Model	MSD G	MSD M	MSD P	+/- 1 G	+/- 2 G	+/- 3 G	+/- 1 M	+/- 2 M	+/- 3 M	+/- 1 P	+/- 2 P	+/- 3 P
Global	5.66	5.49	5.69	56.10	78.50	88.80	55.00	77.40	88.60	40.80	72.10	88.80
Reduced H	5.61	5.02	5.52	55.30	78.80	88.90	53.00	80.70	91.50	40.70	75.90	88.20
Reduced I	5.79	5.13	5.70	53.70	76.70	89.10	54.40	78.30	89.30	40.70	74.50	87.10

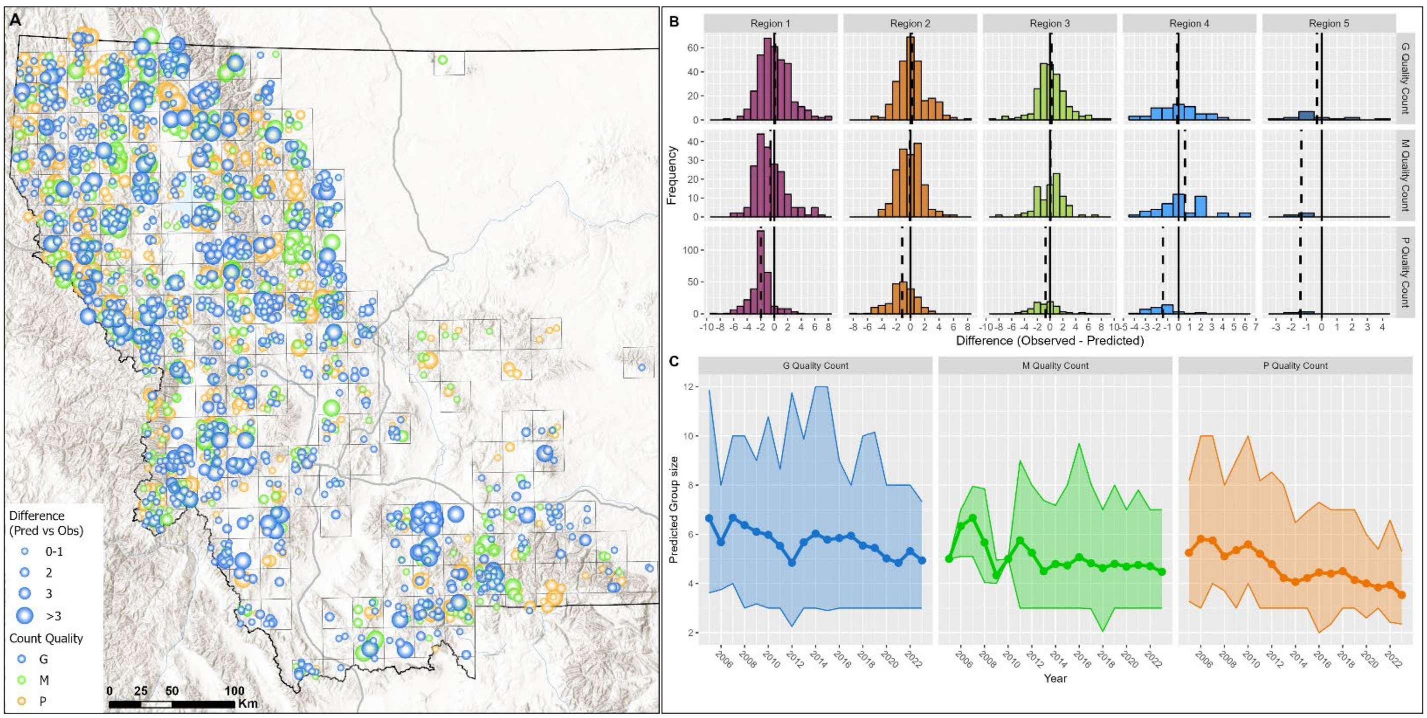

Figure 3. Panel A: absolute difference in predicted (Pred) versus observed (Obs) group sizes based on Model H's predictions. Colors indicate the count quality, and symbols occur for each year in which an observation was recorded, with symbol placement jittered within the grid cell in effort to display all years of data (2005 – 2023). **Panel B:** difference in group size (observed – predicted) by MFWP management region and count quality. Solid vertical lines indicate the 0 intercept (perfect accuracy) and dashed vertical lines indicate the mean difference within that region and count quality. Values < 0 indicate predictions were greater than observations, and values > 0 indicate the opposite. Note that five observations with a difference of >10 were omitted from this plot to constrain the x-axis. These five instances were abnormally large packs reported in Region 1 (packs of 20 in 2008, 20 in 2023, and 29 in 2019), Region 3 (a pack of 19 in 2008) and Region 4 (a pack of 15 in 2013). **Panel C:** mean predicted group size by year and count quality, with 95% confidence intervals. In all panels, 'G' = good quality, 'M' = moderate quality, and 'P' = poor quality.

NEXT STEPS - MSE FUTURE APPROACHES

Currently, the ‘truth’ model simulates an age structured wolf population using survival and fecundity rates from the literature at the pack level. Although remote collar data are available, we were not able to use them to produce reliable estimates of survival through modeling due to a high rate of collar failures and collared wolves being harvested. Hence, we are using survival rates from peer-reviewed literature. There are three age classes in the ‘truth’ model: young of the year (pups born in the current year, <1 year old), first years (juveniles that were born in the previous year that are <2 years old), and adults (2 years old or older). We capture parametric uncertainty, as well as temporal and demographic stochasticity throughout the ‘truth’ model (Regan et al., 2002). Future steps to the ‘truth’ model include adding immigration and emigration throughout the state, dispersal between packs, and simulating the spatial distribution of packs based on the output of updated iPOM (using estimates of occupancy and group size across the state).

From the ‘truth’ model, the collection of monitoring data is simulated and fit to the estimation model. A simple abundance-only integrated population model (IPM; Schaub and Abadi, 2011) was developed as an initial prototype to build the infrastructure of the MSE. Once the updates to iPOM are complete, we plan to simulate requisite monitoring data and fit them to iPOM, generating population size estimates. Other estimation models, e.g., a full IPM, may also be included, if it is determined that the data are feasible to collect. Once the iPOM updates are finalized and included in the MSE, we plan to run trial simulations using initial management strategies that have already articulated. Given the trial simulation output, we plan to review the overall MSE structure and work with MFWP to refine or identify any other management strategies of interest to include.

The simulated management strategies will be compared and evaluated in terms of the fundamental objectives for wolf management. The objectives that have been a part of wolf harvest season setting since 2010 are:

- Maintain a viable and connected wolf population in Montana.
- Gain and maintain authority for state of Montana to manage wolves.
- Maintain positive and effective working relationships with stakeholders.
- Reduce wolf impacts on livestock and big game populations.
- Maintain sustainable hunter opportunities for wolves.
- Maintain sustainable hunter opportunities for ungulates.
- Increase broad public acceptance of sustainable harvest and hunter opportunities as part of wolf conservation.
- Enhance open and effective communication to better inform decisions.
- Learn and improve as we go.

To capture how strategies perform, we plan to develop performance metrics related to each of the above fundamental objectives. Objectives related directly to wolf abundance are straightforward, as they can be measured in terms of the simulated wolf population abundance under each management strategy. Metrics for big game populations and livestock depredations may be associated with wolf abundance, but there is uncertainty about how much of an effect wolf abundance (as compared to other factors) has on either, uncertainty that can be captured by the MSE. The relationship between wolf metrics and public attitude or working relationship objectives need to capture uncertainty in the drivers of public attitudes. As such, we plan to work closely with human dimensions experts to ensure we are capturing these metrics appropriately as they relate to wolf abundance. Given the suite of objectives and performance metrics, the MSE simulation results can be used to assess tradeoffs between conflicting objectives over long time scales.

Acknowledgements

We would like to thank Montana Fish, Wildlife and Parks wolf specialists for consulting on this work, as well as their ongoing data collection efforts, including Subhadeep Bhattacharjee, Wendy Cole, Nathan Lance, Tyler Parks, and Sarah Zielke. We would also like to thank Alexander Metcalf, Elizabeth Metcalf, and Michael Lewis for helping us consider the human dimensions aspects of this work. We would like to thank David Ausband and Brielle Kwarta Thompson for reviewing this report and providing valuable feedback.

Animal Ethics

All wolves were captured, anesthetized, and handled in accordance with MFWP biomedical protocol for free ranging wolves (Montana Fish, Wildlife and Parks, 2005), guidelines from the Institutional Animal Care and Use Committee for the University of Montana (AUP #070–17), and guidelines approved by Sikes et al. (2011).

Data Availability

Data are owned by Montana Fish, Wildlife and Parks. Anyone may request data through the FWP website: https://fwp.mt.gov/aboutfwp/contact-us/records-request. Qualified researchers who wish to collaborate with FWP using these data can contact Justin Gude (jgude@mt.gov).

References

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7, 434. https://doi.org/10.2307/1390675

Bunnefeld, N., Edwards, C.T.T., Atickem, A., Hailu, F., Milner-Gulland, E.J., 2013. Incentivizing monitoring and compliance in trophy hunting. Conserv. Biol. 27, 1344–1354. https://doi.org/10.1111/cobi.12120

Bunnefeld, N., Hoshino, E., Milner-Gulland, E.J., 2011. Management strategy evaluation: a powerful tool for conservation? Trends Ecol. Evol. 26, 441–447. https://doi.org/10.1016/j.tree.2011.05.003

Bureau USC, 2018. TIGER products.

Busetto, L., Ranghetti, L., 2016. MODIStsp : an R package for automatic preprocessing of MODIS Land Products time series. Comput. Geosci. 97, 40–48. https://doi.org/10.1016/j.cageo.2016.08.020

Canessa, S., Guillera-Arroita, G., Lahoz-Monfort, J.J., Southwell, D.M., Armstrong, D.P., Chadès, I., Lacy, R.C., Converse, S.J., 2015. When do we need more data? A primer on calculating the value of information for applied ecologists. Methods Ecol. Evol. 6, 1219–1228. https://doi.org/10.1111/2041-210X.12423

Ellison, A.M., 2004. Bayesian inference in ecology. Ecol. Lett. 7, 509–520. https://doi.org/10.1111/j.1461-0248.2004.00603.x

Evans, J.S., 2018. SpatialEco: spatial analysis and modelling utilities. CRAN Contrib. Packag. https://doi.org/10.32614/cran.package.spatialeco

Hijmans, R.J., 2020. terra: Spatial Data Analysis. CRAN Contrib. Packag. https://doi.org/10.32614/cran.package.terra

Hollister, J., Shah, T, Robitaille, AL, Beck MW, 2023. Elevatr: access elevation data from various apis. CRAN Contrib. Packag. https://doi.org/10.32614/cran.package.elevatr

Hooten, M.B., Hobbs, N.T., 2015. A guide to Bayesian model selection for ecologists. Ecol. Monogr. 85, 3–28. https://doi.org/10.1890/14-0661.1

MacKenzie, D.I., 2018. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence, Second edition. ed. Academic Press, an imprint of Elsevier, London, United Kingdom.

Microsoft, 2018. US building footprints.

Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., Bailey, L.L., Weir, L.A., 2011. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. Ecology 92, 1422–1428. https://doi.org/10.1890/10-1396.1

Miller, D.A.W., Nichols, J.D., Gude, J.A., Rich, L.N., Podruzny, K.M., Hines, J.E., Mitchell, M.S., 2013. Determining occurrence dynamics when false positives occur: estimating the range dynamics of wolves from public survey data. PLoS ONE 8, e65808. https://doi.org/10.1371/journal.pone.0065808

Montana Fish, Wildlife and Parks, 2002. Montana wolf conservation and management planning document (Annual Report).

Montana Fish, Wildlife and Parks, n.d. Montana National Heritage Program. Montana Field Guide.

NIMBLE Development Team, 2024. NIMBLE: MCMC, particle filtering, and programmable hierarchical modeling.

Omernik, J.M., Griffith, G.E., 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. Environ. Manage. 54, 1249–1266. https://doi.org/10.1007/s00267-014-0364-1

Parks, M., Podruzny, K., Cole, W., Parks, T., Lance, N., Sielke, S., Bhattacharjee, S., 2024. Montana gray wolf conservation and management 2023 annual report (Annual). Montana Fish, Wildlife & Parks, Helena, Montana.

Pebesma, E., 2018. Simple features for R: standardized support for spatial vector data. R J. 10, 439. https://doi.org/10.32614/rj-2018-009

Pebesma, E., Bivand, R., 2023. Spatial data science: with applications in R. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

Punt, A.E., Butterworth, D.S., De Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016. Management strategy evaluation: best practices. Fish Fish. 17, 303–334. https://doi.org/10.1111/faf.12104

R Core Team, 2024. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Regan, H.M., Colyvan, M., Burgman, M.A., 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecol. Appl. 12, 618–628. https://doi.org/10.1890/1051-0761(2002)012[0618:ATATOU]2.0.CO;2

Rich, L.N., Russell, R.E., Glenn, E.M., Mitchell, M.S., Gude, J.A., Podruzny, K.M., Sime, C.A., Laudon, K., Ausband, D.E., Nichols, J.D., 2013. Estimating occupancy and predicting numbers of gray wolf packs in Montana using hunter surveys. J. Wildl. Manag. 77, 1280–1289. https://doi.org/10.1002/jwmg.562

Runge, M.C., Converse, S.J., Lyons, J.E., 2011. Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. Biol. Conserv. 144, 1214–1223. https://doi.org/10.1016/j.biocon.2010.12.020

Runge, M.C., Grand, J.B., Mitchell, M.S., 2013. Structured decision making, in: Krausman, P.R., Cain, J.W. (Eds.), Wildlife Management and Conservation: Contemporary Principles and Practices. The Johns Hopkins University Press, published in affiliation with The Wildlife Society, Baltimore, pp. 51–72.

Sappinton, J.M., Longshore, K.M., Thompson, D.B., 2007. Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. J. Wildl. Manag. 71, 1419–1426. https://doi.org/10.2193/2005-723

Schaub, M., Abadi, F., 2011. Integrated population models: a novel analysis framework for deeper insights into population dynamics. J. Ornithol. 11.

Sells, S.N., Mitchell, M.S., 2020. The economics of territory selection. Ecol. Model. 438, 109329. https://doi.org/10.1016/j.ecolmodel.2020.109329

Sells, S.N., Mitchell, M.S., Ausband, D.E., Luis, A.D., Emlen, D.J., Podruzny, K.M., Gude, J.A., 2022a. Economical defense of resources structures territorial space use in a cooperative carnivore. Proc. R. Soc. B Biol. Sci. 289, 20212512. https://doi.org/10.1098/rspb.2021.2512

Sells, S.N., Mitchell, M.S., Podruzny, K.M., Ausband, D.E., Emlen, D.J., Gude, J.A., Smucker, T.D., Boyd, D.K., Loonam, K.E., 2022b. Competition, prey, and mortalities influence gray wolf group size. J. Wildl. Manag. 86, e22193. https://doi.org/10.1002/jwmg.22193

Sells, S.N., Mitchell, M.S., Podruzny, K.M., Gude, J.A., Keever, A.C., Boyd, D.K., Smucker, T.D., Nelson, A.A., Parks, T.W., Lance, N.J., Ross, M.S., Inman, R.M., 2021. Evidence of economical territory selection in a cooperative carnivore. Proc. R. Soc. B Biol. Sci. 288, 20210108. https://doi.org/10.1098/rspb.2021.0108

Sells, S.N., Podruzny, K.M., Nowak, J.J., Smucker, T.D., Parks, T.W., Boyd, D.K., Nelson, A.A., Lance, N.J., Inman, R.M., Gude, J.A., Bassing, S.B., Loonam, K.E., Mitchell, M.S., 2022c. Integrating basic and applied research to estimate carnivore abundance. Ecol. Appl. 32, e2714. https://doi.org/10.1002/eap.2714

U.S. Geological Survey, 2020. National Hydrography Dataset (NHD). https://doi.org/10.3133/70046927

U.S. Geological Survey, 2008. Moderate Resolution Imaging Spectroradiometer (MODIS) Overview. https://doi.org/10.3133/fs20083061

U.S. Geological Survey, n.d. National Land Cover Database.

Valpine, P., Turek, D., Paciorek, C.J., Lang, D.T., Bodik, R., 2017. Programming with models: writing statistical algorithms for general model structures with NIMBLE. J. Comput. Graph. Stat. 26, 403–413.

Walters, C.J., 1986. Adaptive management of renewable resources. Macmillan Publishers Ltd.

Watanabe, S., 2013. A widely applicable Bayesian information criterion. J. Mach. Learn. Res. 14, 867–897.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. Springer-Verlang N. Y. https://doi.org/10.1007/978-0-387-98141-3

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/joss.01686

Williams, B.K., 2011. Adaptive management of natural resources—framework and issues. J. Environ. Manage. 92, 1346–1353. https://doi.org/10.1016/j.jenvman.2010.10.041

Appendix

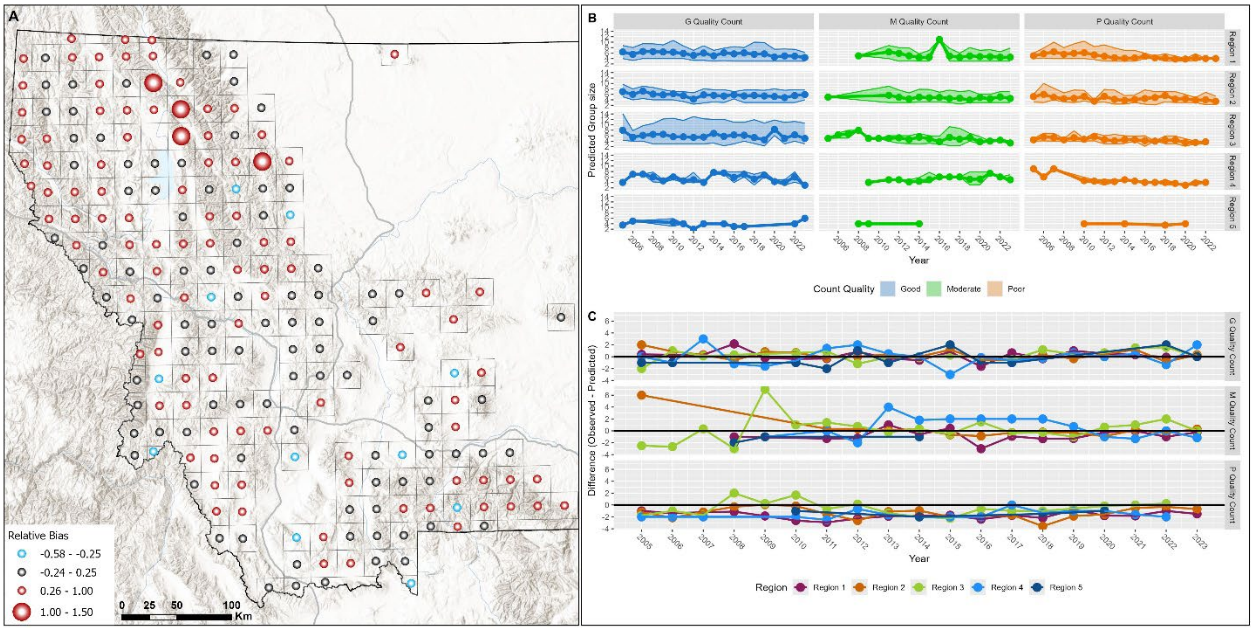

Figure Appendix A1. Panel A: relative bias (calculated as mean((observed – predicted)/observed) by grid cell across years. **Panel B:** predicted group size by year, region, count quality (ribbons depict 2.5 and 97.5% quantiles). **Panel C:** mean difference in group size (observed – predicted) by region and year. Solid horizontal lines indicate the 0 intercept (perfect accuracy); values < 0 indicate predictions were greater than observations, and values > 0 indicate the opposite. In all panels, 'G' = good quality, 'M' = moderate quality, and 'P' = poor quality.

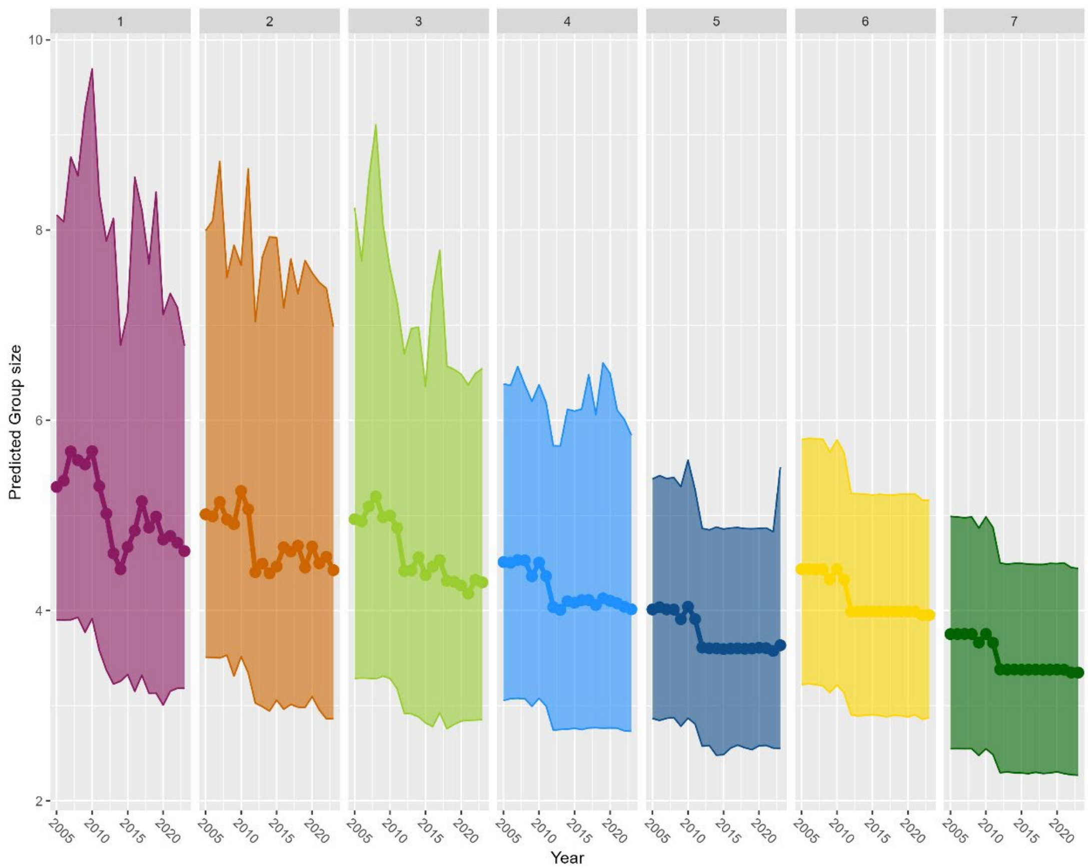

Figure Appendix A2. Mean predicted group size by year and MFWP management region (1 – 7, panels), based on all grid cell predictions (whether a pack was reported there or not). Ribbons depict the 2.5 and 97.5% quantiles.

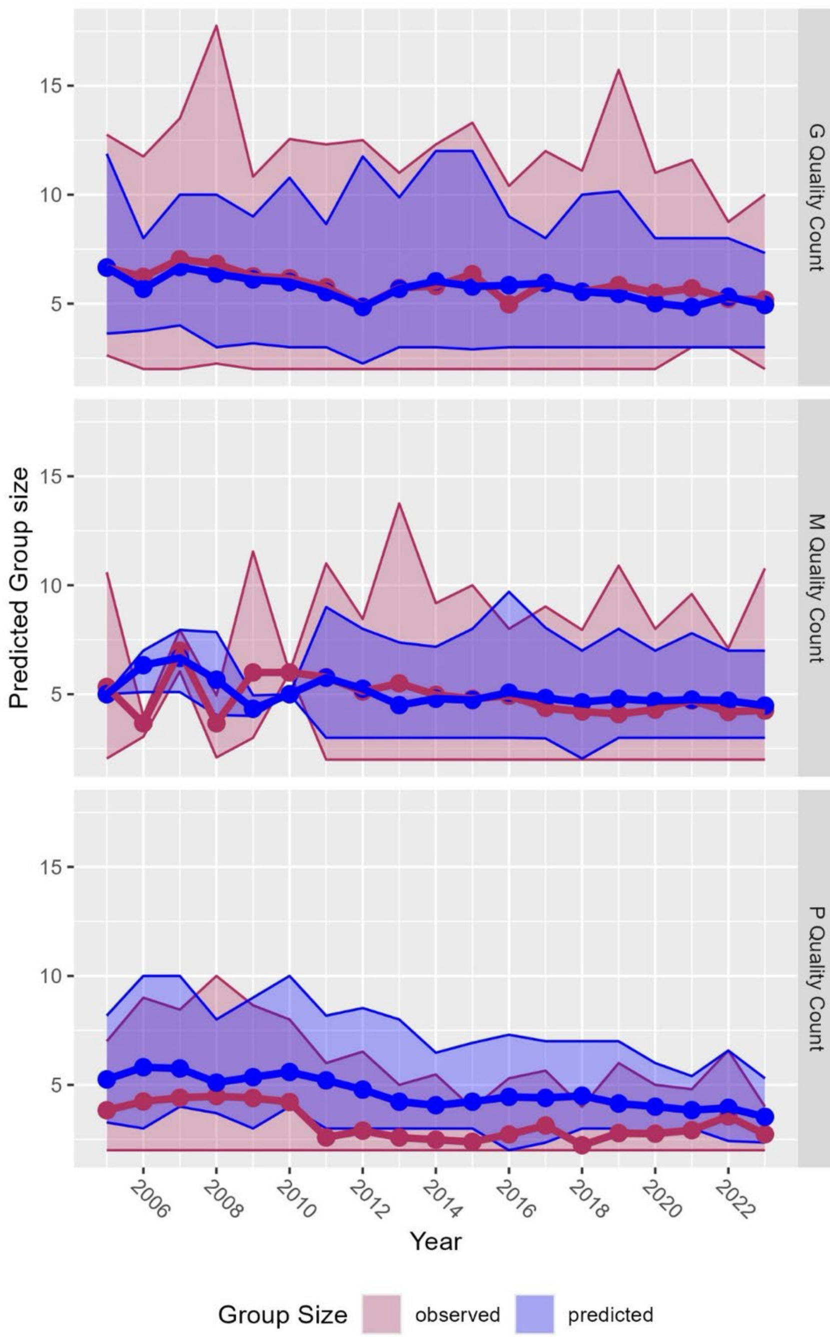

Figure Appendix A3. Observed versus predicted group sizes by year and count quality for reported packs each year. 'G' = good quality, 'M' = moderate quality, and 'P' = poor quality.