

A hierarchical framework for estimating abundance and population growth from imperfectly observed removals

BRYAN S. STEVENS ^{1,3,†} JAMES R. BENICE,¹ DAVID R. LUUKKONEN,^{1,2} AND WILLIAM F. PORTER¹

¹Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan 48824 USA

²Michigan Department of Natural Resources, Rose Lake Research Center, 563 E. Stoll Road, East Lansing, Michigan 48823 USA

Citation: Stevens, B. S., J. R. Benice, D. R. Luukkonen, and W. F. Porter. 2020. A hierarchical framework for estimating abundance and population growth from imperfectly observed removals. *Ecosphere* 11(5):e03131. 10.1002/ecs2.3131

Abstract. Estimating abundance and growth of animal populations are central tasks in ecology and natural resource management. Removal models for estimating abundance have a long history in applied ecology, and recent developments provided hierarchical extensions that account for spatially replicated sampling and heterogeneous capture probabilities. Measurement error is common to removal data collected from many broad-scale monitoring programs, however, and a general framework for population assessment using removal data in the presence of measurement error is lacking. We developed a hierarchical framework for estimating abundance and population trends from removal experiments that are replicated in space and time that accommodates measurement error, as well as heterogeneity in capture probability and animal density. We describe the model for variable-effort removal sampling and use it to estimate region-specific abundance and population trends for wild turkeys (*Meleagris gallopavo*) in Michigan, USA. We used a Bayesian approach for estimation and inference and fit models using daily hunter harvest and effort estimates collected over 5 management regions for 14 annual hunting seasons. Our analyses provide evidence for spatially heterogeneous capture probabilities among regions and turkey densities that were heterogeneous in both space and time, and show that populations increased slightly over the study. Our framework provides a general approach for population assessment using removal data that are collected over broad scales in resource management contexts (e.g., animal harvesting), facilitating formal abundance estimation instead of reliance on unverified indices for tracking populations of managed species. Thus, we provide a useful tool for monitoring programs to assess populations over broad scales, and therefore inform decision makers about population status at spatial scales similar to those for which regulatory decisions are made.

Key words: catch-effort; depletion sample; harvest management; hierarchical model; *Meleagris gallopavo*; removal sample; wildlife monitoring; wild turkey.

Received 29 January 2020; accepted 28 February 2020; final version received 26 March 2020. Corresponding Editor: Lucas Neal Joppa.

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

³Present address: Idaho Cooperative Fish and Wildlife Research Unit, Department of Fish and Wildlife Sciences, University of Idaho, 875 Perimeter Drive, MS 1141, Moscow, Idaho, 83844 USA.

† **E-mail:** bstevens@uidaho.edu

INTRODUCTION

Estimating abundance and population growth are central tasks in ecology, conservation, and natural resource management. Ecology has been referred to as study of the distribution and

abundance of organisms (Kéry and Schaub 2012), and reliable information about population size is central to understanding a broad array of theoretical and applied concepts, including density dependence (Dennis 2002, Dennis et al. 2006, Lebreton 2009), harvest theory (Ricker 1954,

Schaefer 1954), extinction risks (Pimm et al. 1988), and biogeographic patterns and processes (Brown 1984, Brown et al. 1996). Accurate abundance estimates are also vital in the context of conservation and natural resource management, where they play a critical role in making and evaluating the effectiveness of decisions (Williams et al. 2002, Nichols and Williams 2006). Indeed, accurate information about population size is generally treated as a prerequisite for decision-analytic procedures commonly advocated for state-dependent decision making in conservation and resource management (Johnson et al. 1997, Kendall 2001, Martin et al. 2009). Yet estimates of abundance, rather than just indices that are assumed proportional to abundance, are often challenging for management agencies to generate at broad scales where decisions are made because they are logistically difficult and costly to obtain.

Removal models are a well-established framework for estimating abundance of animal populations (Moran 1951, Zippin 1956, Borchers et al. 2002). Removal models are developed from sampling designs (hereafter referred to as removal experiments) under which individual animals are captured and removed from a population over successive occasions (hereafter referred to as trials), where the population is assumed to be closed to additions or other removals for the duration of the experiment (over all trials). A typical example records the number of animals removed from a population over a series of days; however, removal trials can be defined in other discrete units (e.g., multi-day periods, single electrofishing passes). Sampling intensity is often assumed constant during each successive trial (fixed-effort sampling; Moran 1951, Zippin 1956, Dorazio et al. 2005), but models can be developed with variable sampling effort over trials when data on that effort are available (DeLury 1947, Gould and Pollock 1997, St. Clair et al. 2013). Removal experiments thus measure change in the absolute number of removals or removal rates over time, and as the remaining population is depleted, the expected number of removals for each trial gets smaller. Moreover, changes in both the observed and expected number of removals across trials provide the information needed to estimate abundance at the beginning of the experiment (DeLury 1947,

Moran 1951, Zippin 1956). Given the nature of removal sampling, collection of data under such designs is natural for harvested species. Removal sampling is not limited to harvested animals, however, as a removal itself can be broadly defined. Consequently, removal designs are commonly employed to estimate local abundance in applied ecological studies (e.g., electrofishing surveys of stream fishes; Bohlin and Sundström 1977, Wyatt 2002).

Basic removal models have typically used a multinomial or conditional binomial likelihood to model the data generating process, and in their original form assumed a constant capture probability (Moran 1951, Zippin 1956, Borchers et al. 2002). In reality, capture probabilities can be heterogeneous among individual animals, over trials within the same experiment, or among experiments replicated in space and time (Lewis and Farrar 1968, Bohlin and Sundström 1977, Gould and Pollock 1997, Borchers et al. 2002, Dorazio et al. 2005, Mäntyniemi et al. 2005, St. Clair et al. 2013). Hierarchical statistical models have been used to accommodate many of these heterogeneities, and to explicitly model the distribution of abundance and capture probabilities. For example, Mäntyniemi et al. (2005) developed a model that accommodated heterogeneous capture probabilities among animals by assuming the distribution of capture probabilities among individuals followed a beta distribution. Others explicitly modeled the distribution of capture probabilities and abundance among sites, years, or both for designs where entire experiments were replicated (Dorazio et al. 2005, Royle and Dorazio 2006, Rivot et al. 2008).

Despite recent advances, there often remain practical limitations to the implementation of removal models for estimating abundance at broad spatial extents. With some exceptions (e.g., Rivot et al. 2008), there has been little consideration of experiments that are replicated in time, and data from individual years are often modeled separately (St. Clair et al. 2013) despite the possibility of information sharing among temporal replicates. Existing models also assume the number of animals removed and sampling efforts executed on each trial are perfectly observed (Borchers et al. 2002). Perfect observation is not realistic for many broad-scale monitoring programs collecting removal data,

particularly for managed species where removals occur through recreational or commercial harvest and monitoring produces estimates of harvest and effort from samples. Measurement error in removal data results in overestimation of abundance (Gould et al. 1997) and limits application of removal models with many existing data sets. Thus, a pragmatic framework for the analysis of spatially and temporally replicated removal experiments in the presence of measurement error is needed. To address this need, we developed a hierarchical modeling framework for estimating abundance and population growth when removal data are not perfectly observed. In addition, we specifically sought to accommodate attributes common to harvest monitoring programs implemented by resource management agencies: variable-effort sampling, heterogeneous abundance and capture probabilities, and

sampling error in the measurement of both the number of removals and removal effort. Here, we describe a framework wherein models for fixed-effort or perfectly observed data, as well as models for removal experiments replicated in either space or time, can be considered as special cases. We demonstrate the methods to estimate region-specific abundance and growth of wild turkey (*Meleagris gallopavo*) populations in southern Michigan, USA, and use simulation to evaluate estimator performance under a variety of scenarios.

METHODS

Model description

We developed models for replicated removal experiments based on a model for variable-effort sampling replicated over S sites and T years that

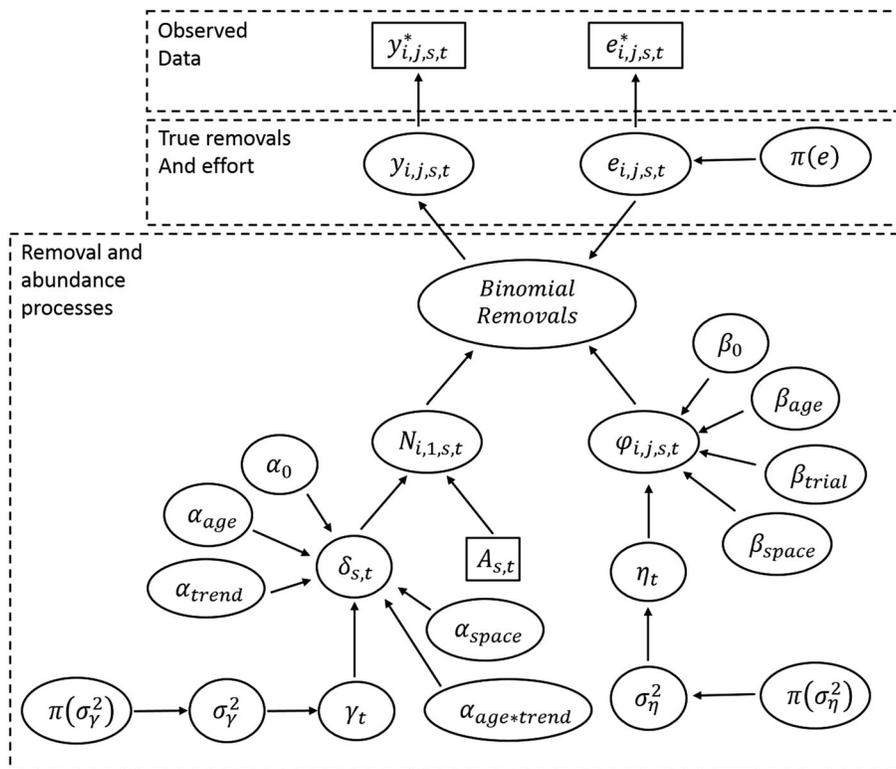


Fig. 1. Conceptual diagram of hierarchical removal model fit to wild turkey harvest data from Michigan, USA. Boxes indicate observed quantities, whereas circles indicate parameters and latent variables or processes. Subscripts denote animal group (i), removal trial (j), sites (s), and year (t), and π denotes a prior distribution for a random quantity (priors for regression coefficients excluded for visual clarity). Symbol definitions are provided in text, and further details of model structure are provided in Box 1.

is extended to accommodate heterogeneity in removal probability and animal density, and also measurement error in removal data (Fig. 1; Box 1). Several individual components of the model we present were described elsewhere (Borchers et al. 2002, Dorazio et al. 2005, Rivot et al. 2008, St. Clair et al. 2013). However, our goal was to synthesize these components and extend the model to a flexible approach for analysis that accommodated imperfectly observed data from spatially and temporally replicated samples (under so-called metapopulation designs; Kéry and Royle 2010).

We started with a conditional binomial likelihood model for removals (Borchers et al. 2002):

$$L(\underline{N}, \underline{\theta}) = \prod_{t=1}^T \prod_{s=1}^S \prod_{j=1}^J \binom{N_{j,s,t}}{y_{j,s,t}} p_{j,s,t}^{y_{j,s,t}} (1 - p_{j,s,t})^{N_{j,s,t} - y_{j,s,t}} \quad (1)$$

where \underline{N} is the vector of abundances at the start of each trial (j), $p_{j,s,t}$ is the probability an individual animal present at the start of trial j in region s and year t is removed during that trial, and $\underline{\theta}$

represents a vector of parameters determining the removal probabilities. Thus, N_j represents the abundance of animals at the start of trial j and y_j is the number of animals removed on that trial, where $N_j = N_{j-1} - y_{j-1}$ for $j > 1$. Under variable-effort sampling, we can model $p_{j,s,t} = 1 - (1 - \varphi)^{e_{j,s,t}}$ as a function of the magnitude of sampling effort (St. Clair et al. 2013), where φ is the per-unit-effort removal probability and $e_{j,s,t}$ is a forcing variable representing the sampling intensity. As such, $1 - \varphi$ represents the probability of not removing an animal for one unit of effort, and $(1 - \varphi)^{e_{j,s,t}}$ is the probability of not removing an animal given all $e_{j,s,t}$ units of effort. With this model for $p_{j,s,t}$, $\underline{\theta}$ in equation 1 only has one element, the per-unit-effort removal probability (φ).

Eq. 1 is easily extended when heterogeneity in removal probability or animal density exists among animal groups (e.g., age or sex classes; see example below):

$$L(\underline{N}, \underline{\theta}) = \prod_{t=1}^T \prod_{s=1}^S \prod_{j=1}^J \prod_{i=1}^I \binom{N_{i,j,s,t}}{y_{i,j,s,t}} p_{i,j,s,t}^{y_{i,j,s,t}} (1 - p_{i,j,s,t})^{N_{i,j,s,t} - y_{i,j,s,t}} \quad (2)$$

for I animal groups, where $p_{i,j,s,t} = 1 - (1 - \varphi_i)^{e_{j,s,t}}$. In this context the primary goal of modeling is to estimate the initial abundances of each animal group for each site (s) and year (t), $N_{i,1,s,t}$, and the per-unit-effort removal probability for each animal group φ_i . Total abundance at the start of each removal sequence is the sum of abundance across all groups ($N_{s,t} = \sum_i N_{i,1,s,t}$). Lastly, un-modeled heterogeneity in removal probability is widely known to affect performance of closed-population abundance estimators (Moran 1951, Bohlin and Sundström 1977, Gould and Pollock 1997, Borchers et al. 2002, Mäntyniemi et al. 2005, St. Clair et al. 2013). To address this, additional structure in φ_i can be modeled explicitly using the logistic function (e.g., for space, time, or trial-level covariates), as is well established in capture-recapture (Pollock et al. 1984, Huggins 1989, Alho 1990, Royle and Link 2002, Royle and Dorazio 2006).

In their basic form, Eqs. 1 and 2 assume no specific structure to the number of animals in the population at the start of removal sampling (i.e., $N_{i,1,s,t}$ are unconstrained). Rather than estimating

Box 1. Summary of model components for hierarchical removal model fit to wild turkey harvest data from Michigan, USA. Subscripts refer to animal groups (i), trials (j), sites (s), and years (t).

Model structure and distributions

Observation models

$$y_{i,j,s,t}^* \sim \text{Poisson}(y_{i,j,s,t})$$

$$e_{j,s,t}^* \sim \text{Poisson}(e_{j,s,t})$$

Removal and abundance process models

$$y_{i,j,s,t} \sim \text{Binomial}(N_{i,j,s,t}, p_{i,j,s,t})$$

$$p_{i,j,s,t} = 1 - (1 - \varphi_{i,j,s,t})^{e_{j,s,t}}$$

$$N_{i,1,s,t} \sim \text{Poisson}(\delta_{i,s,t} A_{s,t})$$

Capture probability and density models

$$\text{logit}(\varphi_{i,j,s,t}) = \beta_0 + \beta_1 \text{Age}_i + \beta_2 j + \beta_s + \eta_i$$

$$\text{log}(\delta_{i,s,t}) = \alpha_0 + \alpha_1 \text{Age}_i + \alpha_2 t + \alpha_3 \text{Age}_i \times t + \alpha_s + \gamma_t$$

Priors and hyper-priors

$$e_{j,s,t} \sim \text{Uniform}(0, 8000)$$

$$\eta_i \sim \text{Normal}(0, \sigma_\eta^2)$$

$$\sigma_\eta^2 \sim \text{Uniform}(0, 0.4)$$

$$\gamma_t \sim \text{Normal}(0, \sigma_\gamma^2)$$

$$\sigma_\gamma^2 \sim \text{Uniform}(0, 1)$$

$$\beta_0 \sim \text{Cauchy}(10)$$

$$\beta_k \sim \text{Cauchy}(2.5)$$

$$\alpha_0 \sim \text{Normal}(0, 1)$$

$$\alpha_k \sim \text{Normal}(0, 1)$$

initial abundances for each removal experiment as unconstrained, we modeled the distribution of initial abundance as an inhomogeneous Poisson process (Dorazio et al. 2005, Royle and Dorazio 2006, Rivot et al. 2008):

$$N_{i,1,s,t} \sim \text{Poisson}(\mu_{i,s,t}) \quad (3)$$

$$\mu_{i,s,t} = \delta_{i,s,t} \times A_{s,t} \quad (4)$$

where $\delta_{i,s,t}$ is the density of animals from group i at the start of sampling at site s , and $A_{s,t}$ represents the area sampled for a given site and year (Rivot et al. 2008, St. Clair et al. 2013). The hierarchical structure of Eqs. 3 and 4 provides a constraint on the values of $N_{i,1,s,t}$ and also provides a flexible framework for modeling spatial-temporal heterogeneity in animal density using log-linear regression (Royle and Dorazio 2008, St. Clair et al. 2013).

We modeled a stochastic observation process to accommodate imperfect observation of variable-effort removal data. Eqs. 1–4 treat removals and effort as known without error and are applicable when perfect observations are feasible (e.g., stream electrofishing surveys, harvest with mandatory reporting). Yet in many practical applications, especially for managed populations where removals are coming from recreational or commercial harvest, observations of $y_{i,j,s,t}$ and $e_{j,s,t}$ are estimated from samples and therefore depart from their true values due to measurement error. Thus, we augment the basic model with an observation model that treats true values of removal and effort on each trial ($y_{i,j,s,t}$ and $e_{j,s,t}$) as imperfectly observed latent variables whose values determine the expectations of the observed data. Under this model, the true value of effort generates the true removal on each trial through a latent first-order Markov process (Eq. 1 or 2), which in turn determines the expectation of the observed removal for trial j . Specifically, we use a Poisson observation model to represent measurement error in the observed values of removal ($y_{i,j,s,t}^*$) and effort ($e_{j,s,t}^*$):

$$y_{i,j,s,t}^* \sim \text{Poisson}(y_{i,j,s,t}) \quad (5)$$

$$e_{j,s,t}^* \sim \text{Poisson}(e_{j,s,t}). \quad (6)$$

Eqs. 5 and 6 can approximate many applications where over- and underestimation of removal and effort are both plausible, and the observation

models could also be adapted for specific applications if additional information on the measurement process were available.

Lastly, we employ a Bayesian paradigm for all model fitting and inference in order to facilitate a straightforward and flexible framework for application. The equations and distributional assumptions about the data and parameters ultimately determine the posterior distribution of parameters of interest, upon which inferences will be made (Box 1). Conceptually, the joint posterior distribution, given the observed data, is proportional to the joint likelihood of the data and the latent states multiplied by the relevant prior distributions:

$$f(\underline{N}, \underline{\varphi}, \underline{\delta} | \underline{y}^*, \underline{e}^*) \propto f(\underline{y}^*, \underline{e}^* | \underline{y}, \underline{e}) f(\underline{y} | \underline{N}, \underline{\varphi}, \underline{e}) f(\underline{N} | \underline{\delta}) \pi(\underline{e}) \pi(\underline{\varphi}) \pi(\underline{\delta}) \quad (7)$$

where $\pi(\underline{e})$, $\pi(\underline{\varphi})$, and $\pi(\underline{\delta})$ represent the prior distributions for true sampling efforts, per-unit-effort capture probability parameters, and animal density parameters, respectively. More specifically, Eqs. 5 and 6 specify probability distributions for the observed data given the latent removal and sampling effort ($f(\underline{y}^*, \underline{e}^* | \underline{y}, \underline{e})$). Eqs. 3 and 4 give probability distributions for initial abundance given density ($f(\underline{N} | \underline{\delta})$), and Eqs. 1 and 2 describe how likely a sequence of latent true removals are, conditioned on the initial abundances, per-unit-effort capture probabilities, and latent values of true sampling effort ($f(\underline{y} | \underline{N}, \underline{\varphi}, \underline{e})$). In practice, the joint distributions and priors of Eq. 7 become functions of additional parameters when structure in φ or δ is modeled using hierarchical regression (Box 1).

Example using wild turkey harvest data

We fit hierarchical removal models to a wild turkey (hereafter turkey) dataset from southern Michigan, USA (Appendix S1). Turkeys are a popular game species in North America and are managed to provide recreational hunting opportunities (Healy and Powell 2000, Harris 2010). The regulatory framework in Michigan consists of 2 discrete annual hunting seasons, with male-only harvests during spring and either-sex harvests in the fall of each year (Kurzejeski and Vangilder 1992, Healy and Powell 2000). Male-only spring harvest is considerably larger than fall

either-sex harvest in Michigan. Estimates of daily harvest and hunter effort across the spring hunting season contain measurement error but enable development of variable-effort removal models (Appendix S1). We developed models using estimates of male-only harvest and hunter effort from recreational spring hunting, and consequently estimated abundance and growth for the male segment of the population. Male-focused monitoring is the current status quo of turkey management (Kurzejeski and Vangilder 1992, Lint et al. 1995, Healy and Powell 2000), as it remains challenging to collect reliable data from females.

We fit removal models using data replicated over 5 geographic sites for a period of 14 yr (2002–2015; Appendix S1: Fig. S1). The 5 sites (hereafter regions) are regions used to manage spring harvest, where individual regions range in size from 6739 to 16,148 km² (Appendix S1: Table S1). Hunting occurs each spring from mid-April through the end of May, and hunting seasons varied annually in length (range 39–45 d; Appendix S1: Table S2) but were consistent among regions each year. Each hunter is legally allowed to harvest one male turkey each spring, but hunter efforts varied among regions and also by day within each hunting season (Appendix S1: Fig. S2, Table S2). Peaks in effort occurred at the beginning of the season and on weekends thereafter but at a reduced magnitude (Appendix S1: Fig. S2, Table S3). Estimates of daily harvest and effort constitute a spatially and temporally replicated but imperfectly observed removal sample from each region and year, where individual days and the number of hunters per day represented removal trials and sampling efforts (Appendix S1).

We fit multiple models to reflect hypothesized heterogeneity in removal probability and turkey density (Appendix S2: Table S1). The most complicated capture probability model we considered contained age-group effects (juvenile and adult male turkeys) and within-experiment trends over removal trials (φ changes linearly over a hunting season), a categorical spatial effect representing differences among regions (removal trial sites), and an annual random intercept to reflect yearly changes in removal probability:

$$\text{logit}(\varphi_{i,j,s,t}) = \beta_0 + \beta_1 \text{Age}_i + \beta_2 j + \beta_s + \eta_t \quad (8)$$

where $\eta_t \sim \text{Normal}(0, \sigma_\eta^2)$ and Age_i is a binary variable indicating juvenile or adult turkeys. In addition, we considered models with reduced versions of Eq. 8, without the annual effect (η_t), and without the site effect (both with and without η_t), resulting in 4 total capture probability models. Both heterogeneity in removal probability among individual animals and behavioral responses to initial removal create a pattern of declining removal probability over trials as more easily captured animals are removed, resulting in negatively biased abundance estimates if not accounted for (Moran 1951, Bohlin and Sundström 1977, Gould and Pollock 1997, Borchers et al. 2002, Mäntyniemi et al. 2005, St. Clair et al. 2013). Our use of the within-experiment time trend effect ($\beta_2 j$) was intended to capture such a pattern if it existed, as has been successfully demonstrated elsewhere (Schnute 1983, Hayes et al. 2007, Kéry and Royle 2010). We expected a decline in removal probability over trials if individual heterogeneity or behavioral responses to removal were present. We also considered 4 hypothesized models of turkey density, where all density models included effects for age-group and linear time trends describing changes in density over the duration of the study, as well as an interaction term that allowed for age-specific trends in abundance over time. The most complicated density model also included a categorical spatial effect and annual random intercepts to reflect hypothesized spatial-temporal heterogeneity in average turkey density:

$$\log(\delta_{i,s,t}) = \alpha_0 + \alpha_1 \text{Age}_i + \alpha_2 t + \alpha_3 \text{Age}_i \times t + \alpha_s + \gamma_t \quad (9)$$

where $\gamma_t \sim \text{Normal}(0, \sigma_\gamma^2)$. We also considered models with reduced versions of Eq. 9, without the annual effect (γ_t), and without the site effect (both with and without γ_t). We fit 16 total models (4 removal \times 4 density; Appendix S2: Table S1) and ranked their relative support using the deviance information criteria (DIC; Spiegelhalter et al. 2002).

We used weakly informative Cauchy priors for all logit-scale regression coefficients (Gelman et al. 2008), normal priors for all log-scale coefficients, and uniform priors for all random effect variance parameters (Gelman et al. 2014; Box 1). Because per-unit effort capture probabilities are

close to zero for the levels of effort observed in this study, we used a Cauchy prior with scale = 10 for the logit-scale intercept parameter, which produces a u-shaped distribution on the real probability scale that is appropriate when mean probabilities are close to zero or one (Appendix S2: Fig. S1; Gelman et al. 2008). Similarly, we used Cauchy priors with scale = 2.5 for additional logit-scale regression coefficients, which produces a distribution on the real probability scale with more support for intermediate values (Gelman et al. 2008). We used standard normal priors for log-scale fixed effects coefficients because when exponentiated these produce realistic priors for the density of male turkeys (Appendix S2: Fig. S2), which rarely exceeds 5 birds per km² in excellent habitats (Vangilder 1992). We used uniform priors for all logit- (Uniform(0, 0.4)) and log-scaled (Uniform(0, 1)) random-intercept variances to restrict their distributions to plausible values, as variances approaching the upper bounds of these distributions result in unrealistically large changes to removal probabilities and densities, respectively (Appendix S2). Lastly, because true daily efforts are unobserved latent states when measurement error is included, prior distributions are also needed for daily efforts for all sites, trials, and years (note that priors for latent removals are specified indirectly through model structure). We assumed a uniform prior for daily hunter effort that was sufficiently broad so-as to encompass all plausible values (Uniform(0, 8000)), given the daily efforts that were observed (e.g., Appendix S1: Fig. S2; Table S3).

We generated Markov Chain Monte Carlo (MCMC) samples to make inferences from posterior distributions of turkey abundance and population trends. We generated MCMC samples using JAGS (version 4.0.1; Plummer 2003) called from within R (version 3.2.0; R Core Team 2015) using the R2jags package (Su and Yajima 2015). For each model, we retained 200,000 MCMC samples from each of 3 chains (600,000 total posterior samples) after an initial burn-in period of 2 million samples and assessed convergence using multivariate Gelman-Rubin statistics (Gelman and Rubin 1992) and trace plots of structural parameters (see Data S1 for an example JAGS model statement). In addition to monitoring structural model parameters and initial

abundances for each region and year, we also generated MCMC samples from posterior distributions of finite rates of population change (i.e., $\lambda_t = N_t/N_{t-1}$) and annual harvest rates ($h_t = \text{Harvest}/N_t$).

Lastly, to understand the implications of model structure for inferences, we conducted a sensitivity analysis using alternative parameterizations of the top model. Specifically, we sought to determine whether accommodating measurement error in the form of an explicit observation model was necessary for inferences from these data, and also the effect of assuming that regional (site-level) abundances shared no structured stochastic dependencies (i.e., abundances not constrained by a hierarchical structure). Thus, we re-fit the top model without measurement error, assuming daily harvest and effort were perfectly observed. We also fit models where there was no explicit structure used to constrain the spatial distribution of turkey abundance, but instead only vague a priori information about population size (but still retaining the removal probability model identical to the original model). The unstructured abundance model assumed discrete uniform priors for initial abundance at the start of each removal experiment and thus placed less constraint on estimates of local abundance (Appendix S2: Table S2). The model with unstructured abundance was fit both including and excluding measurement error, and sensitivity of abundance and population growth estimates was assessed among these alternative parameterizations.

Simulating model performance

We simulated data generation, model fitting, and parameter estimation to determine the ability of our hierarchical models to accurately estimate population sizes and trends. We simulated harvest data using observed estimates of region-specific hunter efforts as true effort, the top model from analyses described above as the data generating model, and posterior mean estimates of parameters from the top model as true values for generating harvest data. We simulated imperfectly observed removal data for each of the 5 management regions under 2 scenarios of temporal replication: (1) one year of sampling, and therefore no temporal replication, and (2) 10 consecutive years of sampling. Performance of

removal models is known to be sensitive to fraction of the population removed, where larger removals result in improved accuracy (Zippin 1956, Gould and Pollock 1997, Dorazio et al. 2005, St. Clair et al. 2013). Thus, we also simulated 6 scenarios where cumulative removal rates were altered through changes to the number of removal trials and daily hunter effort. We considered scenarios where the number of removal trials (i.e., days of hunting) was approximately equal to (40 d), one half of (20 d), or one quarter of (10 d) the length of turkey hunting seasons in southern Michigan. We replicated each of these scenarios over 2 scenarios of hunter effort: (1) high effort, where true region-specific daily hunter efforts were those observed in Michigan, or (2) low effort, where true region-specific daily efforts were half of those observed.

To understand performance of simplified and less constrained versions of the model relative to the model structure described above, we replicated model fitting and parameter estimation for all 12 simulation scenarios described above using 4 different model parameterizations: (1) using the true data generating model with a hierarchical Poisson abundance structure and measurement error in harvest and effort data (i.e., model correctly specified), (2) the correct density model with a hierarchical Poisson abundance structure but excluding measurement error (i.e., making an assumption that data were perfectly observed but the model was otherwise correctly specified), (3) the unstructured abundance model including measurement error, and (4) the unstructured abundance model excluding measurement error (Appendix S2: Table S2). Thus, we assessed performance of removal models under 48 combinations of the temporal replication of experiments, number of removal trials, magnitude of hunter efforts on each trial, and estimation model structure (Appendix S3: Table S1). For each fitted model, we assessed convergence and relative error $((\text{estimate} - \text{truth})/\text{truth})$ of region-specific and total estimates of abundance and the finite rate of population growth over the entire study (only for scenarios with temporal replication; $\lambda = N_{10}/N_1$). We used posterior means as point estimates of abundance and population growth, used the average relative error for each scenario to indicate bias, and also determined the fraction of 95% credible intervals that contained the true

values of abundance and finite rates of change for each scenario. We replicated data generation and parameter estimation 100 times for each scenario, and additional simulation details are provided in Appendix S3.

RESULTS

The top model estimated that turkey abundance in southern Michigan was stable-to-increasing over the study, despite short-term annual fluctuations (Fig. 2). Total abundance peaked at 52,911 males in 2008, which was followed by a short-term decline and then growth to 49,410 males in 2015 (Fig. 2). The best model included a time effect that shifted average turkey density annually, spatial heterogeneity in density among regions, and age-specific density trends over the study (Appendix S2: Table S1, Figs. S4, S6). This model also estimated a larger removal probability for adult males than juveniles and spatial heterogeneity in removal among management regions, but little change in removal probabilities across trials within a hunting season (Appendix S2: Figs. S5, S7–S8). Estimates of population growth from the top model were

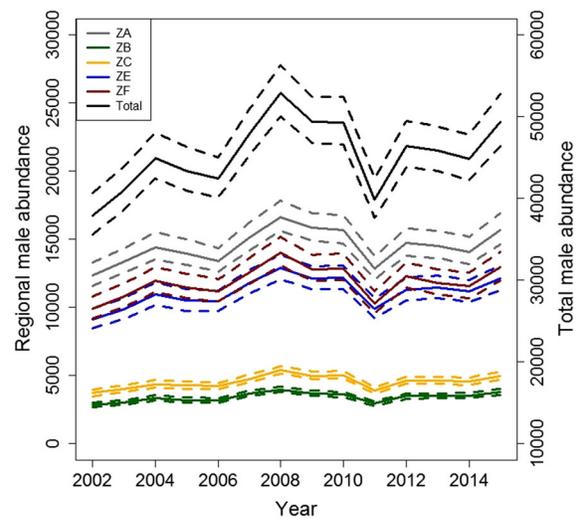


Fig. 2. Region-specific and total population trajectories for male turkey populations (2002–2015) in southern Michigan, USA. Posterior mean estimates are provided by solid lines, whereas dashed lines indicate percentile-based 95% credible intervals (equal tail probabilities).

synchronous among management regions but varied annually despite the spatial heterogeneity in turkey density (Fig. 3). Estimated annual population growth for the entire study area ranged from a low of 0.81 in 2010–2011 to a high of 1.16 in 2011–2012 (Fig. 3). Moreover, estimated annual harvest rates ranged from a low of 0.27 for juvenile males in region ZA to a high of 0.79 for adult males in region ZB (Appendix S2: Fig. S9). Estimated harvest rates were generally larger for adult males than juveniles, but the magnitude and temporal trends of harvest rates

were region specific and affected by patterns of hunter effort (Appendix S1: Table S2).

Sensitivity analyses demonstrated that abundance estimates were generally robust when a model was used to describe spatial–temporal structure in turkey density, whereas ignoring measurement error changed the scale of abundance estimates when an unconstrained abundance model was used (Fig. 4). Estimates of abundance and population growth were similar when the hierarchical density model was used to constrain abundance, irrespective of the inclusion of measurement error in the fitted model (Fig. 4). However, differences in estimates were more pronounced when abundance lacked an explicit structure (i.e., when only constrained through vague uniform priors), and the model lacking both measurement error and an abundance model estimated smaller turkey abundance relative to the other three parameterizations (Fig. 4). Despite differences in the absolute scale of abundance estimates, both models with unconstrained abundance shared similar estimates of population growth over time, irrespective of inclusion of measurement error (Fig. 4b), and these population growth estimates were more variable than the models that included a Poisson abundance model. Region-specific estimates of abundance and population growth usually showed similar sensitivity to model structure, but the magnitude of differences in estimates among models varied by region (Appendix S2: Figs. S10–S11).

Simulation results demonstrated that when specified correctly our model produced estimates of abundance and population growth that were approximately unbiased at the sample sizes and levels of hunter effort observed in this study (Table 1). Simulations assuming 10 yrs of replicated removal experiments also suggested that the model with an abundance hierarchy but ignoring measurement error produced similar abundance estimates as the true data generating model that included measurement error, with a larger proportion of model fits resulting in convergence (Table 2). However, the model including measurement error had better coverage of credible intervals, which was 95% for all management regions (Table 1). The corresponding models without measurement error resulted in credible interval coverage less than 95%

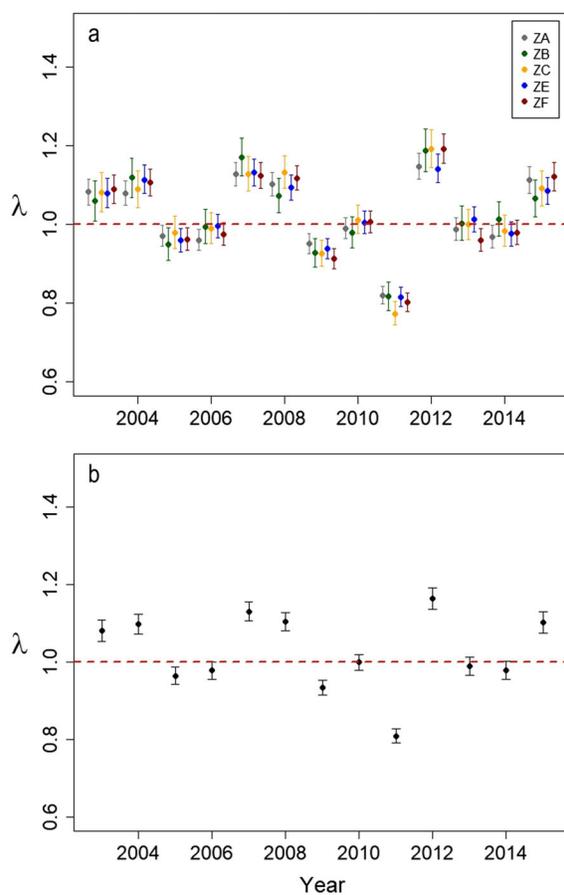


Fig. 3. Annual estimates of the finite rate of change (λ) for male turkey populations in individual regions and the entire southern Michigan, USA, study area for 2002–2015. (a) Regionally, (b) all regions combined. Dots indicate posterior mean point estimates, whereas whiskers represent percentile-based 95% credible intervals (equal tail probabilities).

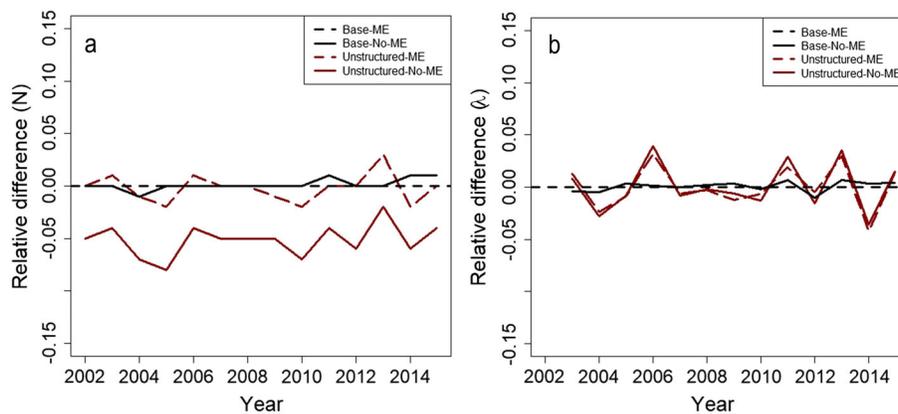


Fig. 4. Differences in estimates of abundance (N) and of finite rate of population change (λ) for turkeys in southern Michigan arising from corresponding models accounting for (ME) and ignoring measurement error (No-ME) in daily hunter effort and harvest data, and using different hierarchical structures for the distribution of abundance (Poisson: Base; no hierarchy: Unstructured). This figure demonstrates sensitivity of estimates to changes in model structure.

(range = 0.84–0.89). Reducing hunter effort by half resulted in worse performance for models including an abundance hierarchy, with fewer model fits converging and larger magnitude of bias (Tables 1–2). The model including measurement error still had reasonable credible interval coverage under the reduced effort scenario (0.90–0.95 for models that converged); however, the model ignoring measurement error had degraded coverage (0.63–0.74). Despite the bias of estimated abundance, estimates of population trends from the base hierarchical abundance models under low effort scenarios were approximately unbiased. Moreover, reducing the number of removal trials degraded performance and reliability of the models with an abundance structure, with decreased rates of convergence and increased bias of abundance estimates (Appendix S3: Tables S2–S3).

Simulations also demonstrated that when the unstructured abundance model was assumed in the presence of measurement error, abundance estimates were biased for all scenarios (Table 1). Convergence was achieved for all models fit assuming an unstructured abundance model; however, these models severely overestimated abundance, and abundance estimates more than twice truth were common. Unstructured models that included measurement error resulted in more severe positive bias for abundance

estimates than unstructured models that simply ignored measurement error, and these biases were typically made worse by reducing hunter effort or the number of removal trials (Table 1, Appendix S3: Tables S2–S3). Despite their inability to accurately estimate absolute abundance, unstructured models (both including and excluding measurement error) were able to estimate population trends over time reasonably well when hunter effort was high (i.e., at levels observed in this study). Nonetheless, negative bias emerged for population growth estimates with unstructured models when hunter effort or the number of removal trials was reduced (Appendix S3: Table S3). Lastly, performance was poor for all model structures fit to data without temporal replication (i.e., only one year of data), where failed convergence, biased abundance estimates, or both were observed under all scenarios (Appendix S3: Table S4).

DISCUSSION

We developed a hierarchical framework for estimating animal abundance and population growth from replicated but imperfectly observed removal experiments. This framework synthesizes existing removal methods and extends the models to accommodate measurement error in sampling intensity and the number of removals

Table 1. Results of simulation study used to assess performance of posterior mean estimates of abundance (\hat{N}) and finite rate of change ($\hat{\lambda} = \hat{N}_{t=10}/\hat{N}_{t=1}$) for removal studies replicated over a ten-year period.

Metric and simulation scenario†	Convergence	Region					Total
		ZA	ZB	ZC	ZE	ZF	
\hat{N}							
High effort							
Measurement error							
Poisson	0.79	0.03	0.05	0.04	0.04	0.04	0.04
Unstructured	1.00	2.03	4.89	4.59	2.03	2.08	2.73
No measurement error							
Poisson	0.89	0.02	0.03	0.02	0.02	0.03	0.02
Unstructured	1.00	1.85	3.97	3.77	1.84	1.89	2.36
Low effort							
Measurement error							
Poisson	0.38	-0.09	-0.08	-0.08	-0.08	-0.08	-0.08
Unstructured	1.00	2.25	5.21	5.03	2.40	2.18	3.01
No measurement error							
Poisson	0.68	-0.12	-0.10	-0.11	-0.10	-0.11	-0.11
Unstructured	1.00	2.16	5.10	4.97	2.33	2.09	2.93
$\hat{\lambda}$							
High effort							
Measurement error							
Poisson	0.79	-0.01	0.01	-0.02	0.00	0.02	0.00
Unstructured	1.00	0.02	0.01	-0.02	-0.01	0.01	0.00
No measurement error							
Poisson	0.89	-0.01	0.01	-0.01	0.00	0.01	0.00
Unstructured	1.00	0.01	0.00	-0.03	-0.02	-0.01	-0.01
Low effort							
Measurement error							
Poisson	0.38	-0.03	0.00	-0.04	-0.01	0.01	-0.01
Unstructured	1.00	-0.02	-0.03	-0.12	0.00	-0.06	-0.05
No measurement error							
Poisson	0.68	-0.02	0.00	-0.04	-0.01	0.01	-0.01
Unstructured	1.00	-0.02	-0.03	-0.12	0.00	-0.06	-0.05

Notes: The fraction of simulated model fits that resulted in convergence (Convergence) as well as average relative error of abundance estimates from year 10 are presented by turkey management region (Region), and for all regions combined (Total). Results are presented for simulation scenarios defined by the magnitude of hunter effort (High, Low), the hierarchical structure of the fitted abundance model (Poisson, Unstructured), and the presence of measurement error in fitted model (Measurement error, No measurement error), for the simulation scenario where the number of trials in each removal experiment was equal to 40 d of hunting.

† Mathematical details of each model structure and simulation scenario are described in Appendix S2: Tables S1–S2, and Appendix S3.

per trial via an explicit observation process. Estimates for our example with wild turkeys in Michigan, USA, suggested stable-to-increasing populations despite spatial-temporal fluctuations in turkey density. Sensitivity and simulation analyses demonstrated these estimates were robust and approximately unbiased at the levels of hunter effort observed. Simulations also demonstrated that accuracy of abundance estimates was sensitive to inclusion of a model to explicitly account for heterogeneity in density and abundance, implying weak ability to

determine the absolute scale of abundance in the presence of measurement error if heterogeneous density was not modeled directly but instead abundances were unconstrained (e.g., the common approach where removal data from individual years are analyzed separately; St. Clair et al. 2013). Broad-scale estimates of population growth from our model were also robust and relatively insensitive to model parameterization. Moreover, the framework we describe is flexible and additional data or prior information could be used by tailoring the models to attributes of

Table 2. Results of simulation study used to assess performance of posterior mean estimates of abundance (\hat{N}) and finite rate of change ($\hat{\lambda} = \hat{N}_{t=10}/\hat{N}_{t=1}$) for removal studies replicated over a 10-yr period.

Metric and simulation scenario†	Region					Total
	ZA	ZB	ZC	ZE	ZF	
\hat{N}						
High effort						
Measurement error						
Poisson	0.95	0.95	0.95	0.95	0.95	0.95
Unstructured	0.00	0.00	0.00	0.00	0.00	0.00
No measurement error						
Poisson	0.88	0.84	0.89	0.87	0.87	0.89
Unstructured	0.00	0.00	0.00	0.00	0.00	0.00
Low effort						
Measurement error						
Poisson	0.95	0.95	0.95	0.95	0.95	0.90
Unstructured	0.00	0.00	0.00	0.00	0.00	0.00
No measurement error						
Poisson	0.63	0.74	0.66	0.69	0.68	0.66
Unstructured	0.00	0.00	0.00	0.00	0.00	0.00
$\hat{\lambda}$						
High effort						
Measurement error						
Poisson	0.94	0.99	0.97	1.00	0.91	1.00
Unstructured	0.97	1.00	0.99	1.00	1.00	0.99
No measurement error						
Poisson	0.80	0.93	0.94	0.91	0.78	0.85
Unstructured	0.92	0.95	0.87	0.83	0.97	0.88
Low effort						
Measurement error						
Poisson	0.89	1.00	0.74	1.00	1.00	0.97
Unstructured	0.98	0.96	0.55	1.00	0.81	0.59
No measurement error						
Poisson	0.76	0.97	0.34	0.96	0.97	0.78
Unstructured	0.86	0.87	0.29	0.97	0.53	0.26

Notes: The fraction of simulated model fits that resulted in 95% credible intervals that contained the true abundance or finite rate of change (for model fits that converged) are presented by turkey management region (Region), and for all regions combined (Total). Results are presented for simulation scenarios defined by the magnitude of hunter effort (High, Low), the hierarchical structure of the fitted abundance model (Poisson, Unstructured), and the presence of measurement error in fitted model (Measurement error, No measurement error), for the simulation scenario where the number of trials in each removal experiment was equal to 40 d of hunting.

† Mathematical details of each model structure and simulation scenario are described in Appendix S2: Tables S1–S2, and Appendix S3.

specific study systems and monitoring programs.

The hierarchical model we describe provided estimates that were robust and approximately unbiased in the presence of measurement error under sample sizes and removal rates observed in our study. The model used an inhomogeneous Poisson process to explicitly model changes in density and induce stochastic dependence on the distribution of abundances, and provided estimates of abundance and population growth that were approximately unbiased. Surprisingly, this

model also produced abundance estimates that were robust to measurement error in the removal data, even when it was ignored by excluding the observation model. The Poisson distribution has been used by numerous authors to provide structure and constraint to local abundance estimates arising from studies under a metapopulation design (Dorazio et al. 2005, Royle and Dorazio 2006, 2008, Rivot et al. 2008, Kéry and Royle 2010, St. Clair et al. 2013). Our results imply such models can be used to constrain abundance estimates to realistic values and ameliorate the bias

that typically results from measurement error in removal data. This appears to be a novel finding with respect to removal models, as we are not aware of other studies evaluating the implications of measurement error when a hierarchical model is used to represent deterministic (i.e., covariate effects) and stochastic structure in regional abundance for spatially replicated removal experiments. However, excluding the observation model did result in overestimation of precision in the presence of measurement error. Thus, there remain tangible benefits of using the more complicated model that includes an explicit observation process, as it resulted in a more reliable assessment of uncertainty.

Absent a model inducing structure on the distribution of density and abundance, the implications of measurement error for estimating abundance via removal methods were substantial. Eliminating the model for abundance resulted in greater sensitivity of estimates to observation model structure (i.e., including and excluding an observation process), and also greater variability of estimates over time for turkey populations in Michigan (i.e., no shrinkage; Royle and Link 2002). The effect of measurement error on estimator performance when abundance was effectively unconstrained, except by a vague prior, was to inflate abundance estimates substantially. Similarly, Gould et al. (1997) demonstrated positive bias in abundance estimates when removal data from single experiments were subject to measurement error, and found this directional bias was consistent over an order-of-magnitude change in the true abundance. Collectively, these results imply that estimating the absolute scale of abundance via removal models is more difficult when removals and effort are not perfectly observed, that inappropriate scaling caused by measurement error results in large positive bias in abundance estimates if ignored, and that without the structure induced by appropriately modeling spatial-temporal changes in density, the estimators are suspect.

Our analyses demonstrate that estimation problems induced by measurement error can be ameliorated when information is shared among replicated experiments and the systematic changes to animal density are explicitly modeled. Estimation challenges for closed-population

abundance estimators with latent states are well described in the literature (e.g., mixture models in capture–recapture; Coull and Agresti 1999, Dorazio and Royle 2003). Estimation challenges are commonly used to justify modeling abundance hierarchically under metapopulation designs as we did in this study (Royle and Dorazio 2006, 2008, Royle et al. 2007, Kéry and Royle 2010), which acts along with prior distributions to regulate parameter estimates derived from posterior distributions under a Bayesian paradigm (Hooten and Hobbs 2015). This effectively allows researchers to mathematically constrain abundance estimates to biologically plausible ranges. This is done via use of hierarchical structures and prior distributions on animal abundance and density, respectively, as we did in this study to prevent unrealistically large densities of turkeys. This approach is defensible for application of complex hierarchical models in ecology, as one is using what is already known to develop constraints on parameters in order that they remain biologically feasible (Hooten and Hobbs 2015). Moreover, comparison of posterior and prior distributions from our example analyses (Appendix S2) clearly shows the data were sufficient for estimating abundance under the model that included both an abundance hierarchy and an explicit observation process. Thus, while the hierarchical model acted to constrain abundance to plausible values, the data were also informative about spatial-temporal changes in turkey abundance within the constraints of the model.

Despite the sensitivity of abundance estimator performance to inclusion of an explicit density model, estimates of population growth were more robust to model structure. Our simulations demonstrated that all of the model parameterizations considered were able to reasonably estimate decadal-scale population trends in the presence of measurement error at the observed sample sizes and removal rates. Even models with unstructured regional abundances that overestimated the scale of abundance for a given year could reliably depict the finite rate of change over a 10-yr period (i.e., positive bias in abundance was consistent over time). That is, simulations showed that the models were capable of reliably estimating long-term lambda even when estimates of absolute abundance were biased. Thus, assessing long-term population

trends may be easier than accurately scaling abundance when removal data are imperfectly observed. Few studies have considered models for temporally replicated removal experiments (but see Rivot et al. 2008), and their usefulness for assessing population trends for managed species appears to be underappreciated. Although trends provide less information to decision makers than absolute abundance (Nichols and Williams 2006), we believe the reliable trend estimates produced by our model are more informative and useful for decision making than traditional indices (e.g., catch-per-unit effort) used to monitor species like the wild turkey, as the models can explicitly account for changes to both animal density and per-unit-effort removal probability. Moreover, estimates of population trends are often used by practitioners to inform natural resource management decisions (e.g., when adjusting harvest regulations), as estimates of absolute abundance can lack context to stakeholders and decision makers without comparable historical information on population size.

We demonstrated reasonable performance for hierarchical removal models that accommodate measurement error under conditions similar to those observed for our example, but performance was sensitive to the fraction of the population removed and the availability of replicated data. Reducing the fraction of the population removed, manifested through reduction to either the magnitude of sampling effort or the total number of removal trials, degraded performance considerably. Under these conditions, the models often failed to converge and bias of abundance estimates increased. These issues are widely recognized in the literature on capture–recapture and removal models, where performance is tightly linked to the overall fraction of the population sampled (Zippin 1956, Gould and Pollock 1997, Coull and Agresti 1999, Dorazio and Royle 2003, Dorazio et al. 2005, Mäntyniemi et al. 2005, St. Clair et al. 2013). Our simulations indicate that problems associated with removing only a small fraction of the population will likely manifest themselves through failed model convergence. In this light and given the wide variety of contexts for which removal models may be applied, we suggest it wise to use simulation to assess estimator performance for specific applications when removal data are subject to

measurement error. In addition, larger samples of simulated parameter estimates would benefit future studies. We chose to simulate a large number of scenarios (48), which came at the expense of large numbers of iterations of data generation and model fitting per scenario (100) due to computational limitations and demands of model fitting. While we believe this was adequate for assessment of central tendency and bias of estimators, we acknowledge that more simulation replicates would be useful to provide improved assessment of precision and coverage of credible intervals.

While simulations indicated reasonable performance for removal experiments replicated in space and time similar to our example dataset, our models were data hungry and did not perform reliably in the face of measurement error when spatially replicated removal experiments were not also replicated in time. This suggests substantial replication is likely a prerequisite for applying these models, which could limit their application. In many applied settings, spatial replicates are fixed and represent the entire area of interest (e.g., a set of management regions). Additional replication of such experiments can only come in time, and our simulations suggest this replication is necessary. However, it is not entirely clear what amount of temporal replication will be necessary for reliable application of the model under different levels of spatial replication. It seems likely that replicated observations of the removal processes and information sharing among experiments is what is needed, and thus the necessary amount of temporal replication may depend on the number of spatial replicates, and vice versa. In other contexts, the number of spatial replicates could be much larger and require less temporal replication for reliable inferences. Thus, while the reliable performance of models required considerable replication of removal data, we expect the necessary amount of replication to depend on the context of the study. This further supports our suggestion that simulation-based assessment of estimator performance for specific applications is wise when removal data are subject to measurement error.

Although models for removal data affected by measurement error have some limitations, the hierarchical framework described provides a

high degree of flexibility that can enable further tailoring of models to specific species and sampling protocols. In our example, we focused on relatively simple model structures to describe spatial-temporal heterogeneity in removal and animal density (fixed effects and iid normal random intercepts), yet the framework is flexible enough to accommodate a variety of model structures. For example, this could include logit- or log-scale multivariate normal effects that induce a correlation structure on realized values of density and removal probability (Royle and Link 2002, Royle and Dorazio 2006, 2008), or conditional autoregressive models to induce spatial correlation among regions (Lichstein et al. 2002, Webster et al. 2008). In addition, alternative observation models could be tailored to the characteristics of specific monitoring programs. For instance, one might be able to use sampling theory (Cochran 1977) to derive observation-error variances for estimated removal data under specific survey designs, instead of assuming the variance-mean relationship of a Poisson distribution. Lastly, given the variety of contexts for which removal models are applied, extensions for sampling specific taxonomic groups may also be possible, where examples include the analysis of removal data from avian point counts and stream electrofishing surveys with measurement error due to species misidentification. Consequently, the framework described here provides a flexible starting point for fine-tuning population assessment for a variety of species using imperfectly observed data arising from removal sampling protocols.

Lastly, our analyses indicated male turkey populations in Michigan were stable-to-increasing, with heterogeneous density among management regions that also varied annually, and heterogeneous per-unit-effort removal probabilities by animal group and among regions. We estimated fluctuations in abundance consistent with known dynamics of turkey populations, where previous authors suggested population fluctuations of up to 40% annually are possible (Kurzejeski and Vangilder 1992, Healy and Powell 2000). These large annual fluctuations were synchronous across southern Michigan despite a smaller magnitude of spatial heterogeneity in density, which is again consistent with earlier work suggesting broad-scale synchrony in turkey abundance is common (Fleming and Porter

2007). Higher removal probabilities for adult males than for juvenile males were also anticipated, as multiple studies reported that adult males are more vulnerable to harvest than juveniles (Vangilder 1992, Healy and Powell 2000, Chamberlain et al., 2012). Previous research also suggested harvest rates for males might vary among management regions (Diefenbach et al. 2012), which would be expected even in the absence of spatially heterogeneous capture probabilities given variation of hunter effort in space (Clawson et al. 2015, Stevens et al. 2020).

We suspected a priori that removal probability would decline over the hunting season due to capture heterogeneity among individual animals, behavioral responses to harvest, or some combination of these and other factors, yet we found little evidence of this pattern. It is possible that other parameterizations could better portray systematic shifts to capture heterogeneity over a hunting season. For example, a categorical effect over trials could capture behavioral shifts of animals during the hunting season (i.e., early season vs late season), or linear trends over the season may be modeled as a function of cumulative hunter effort (and therefore cumulative exploitation) up to a given trial, instead of declining at a constant rate over trials. Nonetheless, our estimates appear robust and indicate turkey populations in Michigan are stable. This directly contradicts, at least for our study region, a widespread belief that turkey populations in the eastern USA are in decline (Kobilinsky 2018). Such beliefs are often based on monitoring turkey populations using raw harvest or catch-per-unit effort indices that do not account for spatial-temporal changes in both density and removal probability. Thus, we provide assessment tools that enable a more refined understanding of population trends for harvested species like wild turkeys, which in this case facilitated a shifting perception of the status of populations across the study region. In the context of removal via commercial or recreational harvest, our models also provide the ability to estimate abundance and exploitation rates arising from a given set of management regulations, both of which are useful for simulation-based assessment of sustainable harvest strategies (Quinn and Collie 2005, Deroba and Bence 2008, Punt et al. 2016). Our models therefore provide value beyond the

assessment of status and should further enable rigorous assessment of management decisions for target species.

ACKNOWLEDGMENTS

Funding for this work was provided by the Michigan Department of Natural Resources (MDNR), the Boone and Crockett Quantitative Wildlife Center, and Quantitative Fisheries Center at Michigan State University. Funding for this project was also provided by the U.S. Fish and Wildlife Service through the Pittman-Robertson Wildlife Restoration Act Grant MI W-155-R. We thank the Michigan Wild Turkey Stakeholder Group, and especially Michael Jones and Al Stewart, for insights about population assessment and turkey management that informed this project. This is publication 2020-10 of the Quantitative Fisheries Center at Michigan State University.

LITERATURE CITED

- Alho, J. M. 1990. Logistic regression in capture-recapture models. *Biometrics* 46:623–635.
- Bohlin, T., and B. Sundström. 1977. Influence of unequal catchability on population estimates using the Lincoln index and the removal method applied to electro-fishing. *Oikos* 28:123–129.
- Borchers, D. L., S. T. Buckland, and W. Zucchini. 2002. Estimating animal abundance: closed populations. Springer Science + Business Media, London, UK.
- Brown, J. H. 1984. On the relationship between abundance and distribution of species. *American Naturalist* 124:255–279.
- Brown, J. H., G. C. Stevens, and D. M. Kaufman. 1996. The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics* 27:597–623.
- Chamberlain, M. J., B. A. Grisham, J. L. Norris, N. J. Stafford III, F. G. Kimmel, and M. W. Olinde. 2012. Effects of variable spring harvest regimes on annual survival and recovery rates of male wild turkeys in Southeast Louisiana. *Journal of Wildlife Management* 76:907–910.
- Clawson, M. V., J. R. Skalski, J. L. Isabelle, and J. J. Millspaugh. 2015. Trends in male wild turkey abundance and harvest following restoration efforts in the southeast region of Missouri, 1960–2010. *Wildlife Society Bulletin* 39:116–128.
- Cochran, W. G. 1977. *Sampling techniques*. Third edition. John Wiley & Sons, New York, New York, USA.
- Coull, B. A., and A. Agresti. 1999. The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics* 55:294–301.
- DeLury, D. B. 1947. On the estimation of biological populations. *Biometrics* 3:145–167.
- Dennis, B. 2002. Allee effects in stochastic populations. *Oikos* 96:389–401.
- Dennis, B., J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples. 2006. Estimating density dependence, process noise, and observation error. *Ecological Monographs* 76:323–341.
- Deroba, J. J., and J. R. Bence. 2008. A review of harvest policies: understanding relative performance of control rules. *Fisheries Research* 94:210–233.
- Diefenbach, D. R., M. J. Casalena, M. V. Schiavone, M. Reynolds, R. Eriksen, W. C. Vreeland, B. Swift, and R. C. Boyd. 2012. Variation in spring harvest rates of male wild turkeys in New York, Ohio, and Pennsylvania. *Journal of Wildlife Management* 76:514–522.
- Dorazio, R. M., H. L. Jelks, and F. Jordan. 2005. Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics* 61:1093–1101.
- Dorazio, R. M., and J. A. Royle. 2003. Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* 59:351–364.
- Fleming, K. K., and W. F. Porter. 2007. Synchrony in a wild turkey population and its relationship to spring weather. *Journal of Wildlife Management* 71:1192–1196.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian data analysis*. Third edition. CRC Press, Taylor and Francis, Boca Raton, Florida, USA.
- Gelman, A., A. Jakulin, M. G. Pittau, and Y. Sung Su. 2008. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* 2:1360–1383.
- Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7:457–511.
- Gould, W. R., and K. H. Pollock. 1997. Catch-effort maximum likelihood estimation of important population parameters. *Canadian Journal of Fisheries and Aquatic Sciences* 54:890–897.
- Gould, W. R., L. A. Stefanski, and K. H. Pollock. 1997. Effects of measurement error on catch-effort estimation. *Canadian Journal of Fisheries and Aquatic Sciences* 54:898–906.
- Harris, A. 2010. Turkey hunting in 2006: an analysis of hunter demographics, trends, and economic impacts: addendum to the 2006 national survey of

- fishing, hunting, and wildlife-associated recreation. USFWS Report 2006–7. Arlington, Virginia, USA.
- Hayes, D. B., J. R. Bence, T. Kwak, and B. Thompson. 2007. Biomass, density, and yield estimators. Pages 327–374 in M. Brown and C. Guy, editors. Analysis and interpretation of freshwater fisheries data. American Fisheries Society, Bethesda, Maryland, USA.
- Healy, W. M., and S. M. Powell. 2000. Wild turkey harvest management: biology, strategies, and techniques. USFWS Biological Technical Publication BTP-R5001-1999.
- Hooten, M. B., and N. T. Hobbs. 2015. A guide to Bayesian model selection for ecologists. *Ecological Monographs* 85:3–28.
- Huggins, R. M. 1989. On the statistical analysis of capture experiments. *Biometrika* 76:133–140.
- Johnson, F. A., C. T. Moore, W. L. Kendall, J. A. Dubovsky, D. F. Caithamer, J. R. Kelly Jr., and B. K. Williams. 1997. Uncertainty and the management of mallard harvests. *Journal of Wildlife Management* 61:202–216.
- Kendall, W. L. 2001. Using models to facilitate complex decisions. Pages 147–170 in T. M. Shenk and A. B. Franklin, editors. Modeling in natural resource management: development, interpretation, and application. Island Press, Washington, D.C., USA.
- Kéry, M., and J. A. Royle. 2010. Hierarchical modelling and estimation of abundance and population trends in metapopulations designs. *Journal of Animal Ecology* 79:453–461.
- Kéry, M., and M. Schaub. 2012. Bayesian population analysis using WinBUGS: a hierarchical perspective. First edition. Academic Press, Waltham, Massachusetts, USA.
- Kobilinsky, D. 2018. Are turkeys in trouble? Managers search for answers. *Wildlife Professional* 12:18–26.
- Kurzejeski, E. W., and L. D. Vangilder. 1992. Population management. Pages 165–184 in J. G. Dickson, editor. *The Wild Turkey: Biology and Management*. Stackpole Books, Mechanicsburg, Pennsylvania, USA.
- Lebreton, J. D. 2009. Assessing density-dependence: where are we left? Pages 19–32 in Thompson, D. L., E. G. Cooch, and M. J. Conroy, editors. Modeling demographic processes in marked populations. Springer Series in Environmental and Ecological Statistics, Springer Science + Business Media, New York, New York, USA.
- Lewis, J. C., and J. W. Farrar. 1968. An attempt to use the leslie census method on deer. *Journal of Wildlife Management* 32:760–764.
- Lichstein, J. W., T. R. Simons, S. A. Shriner, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72:445–463.
- Lint, J. R., B. D. Leopold, and G. A. Hurst. 1995. Comparison of abundance indexes and population estimates for wild turkey gobblers. *Wildlife Society Bulletin* 23:164–168.
- Mäntyniemi, S., A. Romakkaniemi, and E. Arjas. 2005. Bayesian removal estimation of a population size under unequal catchability. *Canadian Journal of Fisheries and Aquatic Sciences* 62:291–300.
- Martin, J., M. C. Runge, J. D. Nichols, B. C. Lubow, and W. L. Kendall. 2009. Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications* 19:1079–1090.
- Moran, P. A. P. 1951. A mathematical theory of animal trapping. *Biometrika* 38:307–311.
- Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. *Trends in Ecology and Evolution* 21:668–673.
- Pimm, S. L., H. L. Jones, and J. Diamond. 1988. On the risk of extinction. *American Naturalist* 132:757–785.
- Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vienna, Austria.
- Pollock, K. H., J. E. Hines, and J. D. Nichols. 1984. The use of auxiliary variables in capture-recapture and removal experiments. *Biometrics* 40:329–340.
- Punt, A. E., D. S. Butterworth, C. L. de Moor, J. A. A. De Oliveira, and M. Haddon. 2016. Management strategy evaluation: best practices. *Fish and Fisheries* 17:303–334.
- Quinn II, T. J., and J. S. Collie. 2005. Sustainability in single-species population models. *Philosophical Transactions of the Royal Society B* 360:147–162.
- R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ricker, W. E. 1954. Stock and Recruitment. *Journal of the Fisheries Research Board of Canada* 11:559–623.
- Rivot, E., E. Prévost, A. Cuzol, J. L. Baglinière, and E. Parent. 2008. Hierarchical Bayesian modelling with habitat and time covariates for estimating riverine fish population size by successive removal method. *Canadian Journal of Fisheries and Aquatic Sciences* 65:117–133.
- Royle, J. A., and R. M. Dorazio. 2006. Hierarchical models of animal abundance and occurrence. *Journal of Agricultural, Biological, and Environmental Statistics* 11:249–263.

- Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations, and communities. Academic Press, San Diego, California, USA.
- Royle, J. A., M. Kéry, R. Gautier, and H. Schmid. 2007. Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs* 77:465–481.
- Royle, J. A., and W. A. Link. 2002. Random effects and shrinkage estimation in capture-recapture models. *Journal of Applied Statistics* 29:329–351.
- Schaefer, M. B. 1954. Some considerations of population dynamics and economics in relation to the management of commercial marine fisheries. *Journal of the Fisheries Research Board of Canada* 14:669–681.
- Schnute, J. 1983. A new approach to estimating populations by the removal method. *Canadian Journal of Fisheries and Aquatic Sciences* 40:2153–2169.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B* 64:583–639.
- St. Clair, K., E. Dunton, and J. Giudice. 2013. A comparison of models using removal effort to estimate animal abundance. *Journal of Applied Statistics* 40:527–545.
- Stevens, B. S., D. R. Luukkonen, C. A. Stewart, W. F. Porter, J. R. Bence, and M. L. Jones. 2020. Spatial-temporal dynamics of wild turkey hunter effort in Michigan. *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0230747>.
- Su, Y. S., and M. Yajima. 2015. R2jags: A Package for Running jags from R. R package version 0.5-7. <http://CRANR-project.org/package=R2jags>
- Vangilder, L. D. 1992. Population dynamics. Pages 144–164 in J. G. Dickson, editor. *The wild turkey: biology and management*. Stackpole Books, Mechanicsburg, Pennsylvania, USA.
- Webster, R. A., K. H. Pollock, and T. R. Simons. 2008. Bayesian spatial modeling of data from avian point count surveys. *Journal of Agricultural, Biological, and Environmental Statistics* 13:121–139.
- Williams, B. K., J. D. Nichols, and M. J. Conroy 2002. *Analysis and management of animal populations*. Academic Press, San Diego, California, USA.
- Wyatt, R. J. 2002. Estimating riverine fish population size from single- and multiple-pass removal sampling using a hierarchical model. *Canadian Journal of Fisheries and Aquatic Sciences* 59:695–706.
- Zippin, C. 1956. An evaluation of the removal method of estimating animal populations. *Biometrics* 12:165–189.

SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3131/full>

Appendix S1: Supplemental description of wild turkey harvest management and data.

Appendix S2: Supplemental results tables and figures.

Appendix S3: Description of analyses and results of simulation study.

Data S1: Example JAGS model statement for fitting hierarchical removal model with imperfectly observed removal and effort data.