

Pairing field methods to improve inference in wildlife surveys while accommodating detection covariance

JOHN CLARE,^{1,5} SHAWN T. MCKINNEY,² JOHN E. DEPUE,^{3,4} AND CYNTHIA S. LOFTIN²

¹*Department of Wildlife, Fisheries, and Conservation Biology, University of Maine, Orono, Maine 04469-5755 USA*

²*Maine Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey, Orono, Maine 04469-5755 USA*

³*Maine Department of Inland Fisheries and Wildlife, Bangor, Maine 04401-5654 USA*

⁴*Michigan Department of Natural Resources, Baraga, Michigan 49908 USA*

Abstract. It is common to use multiple field sampling methods when implementing wildlife surveys to compare method efficacy or cost efficiency, integrate distinct pieces of information provided by separate methods, or evaluate method-specific biases and misclassification error. Existing models that combine information from multiple field methods or sampling devices permit rigorous comparison of method-specific detection parameters, enable estimation of additional parameters such as false-positive detection probability, and improve occurrence or abundance estimates, but with the assumption that the separate sampling methods produce detections independently of one another. This assumption is tenuous if methods are paired or deployed in close proximity simultaneously, a common practice that reduces the additional effort required to implement multiple methods and reduces the risk that differences between method-specific detection parameters are confounded by other environmental factors. We develop occupancy and spatial capture–recapture models that permit covariance between the detections produced by different methods, use simulation to compare estimator performance of the new models to models assuming independence, and provide an empirical application based on American marten (*Martes americana*) surveys using paired remote cameras, hair catches, and snow tracking. Simulation results indicate existing models that assume that methods independently detect organisms produce biased parameter estimates and substantially underestimate estimate uncertainty when this assumption is violated, while our reformulated models are robust to either methodological independence or covariance. Empirical results suggested that remote cameras and snow tracking had comparable probability of detecting present martens, but that snow tracking also produced false-positive marten detections that could potentially substantially bias distribution estimates if not corrected for. Remote cameras detected marten individuals more readily than passive hair catches. Inability to photographically distinguish individual sex did not appear to induce negative bias in camera density estimates; instead, hair catches appeared to produce detection competition between individuals that may have been a source of negative bias. Our model reformulations broaden the range of circumstances in which analyses incorporating multiple sources of information can be robustly used, and our empirical results demonstrate that using multiple field-methods can enhance inferences regarding ecological parameters of interest and improve understanding of how reliably survey methods sample these parameters.

Key words: *American marten; density; false positive detection; Martes americana; monitoring; multi-method dependence; noninvasive methods; occupancy; spatial capture–recapture methods; survey design.*

INTRODUCTION

Species distribution and abundance are the canonical parameters for applied ecologists (Kéry and Royle 2016), and estimates of distribution and abundance are often used to justify resource management decisions (e.g., Fuller et al. 2016). Robust decision making is improved by reliable information, and there have been three major methodological branches of recent applied ecological

research aimed at improving the information used to inform decision making. One branch has focused on developing statistical models to reduce the bias of distribution or abundance estimates and properly characterize estimate uncertainty (e.g., MacKenzie et al. 2006, Royle et al. 2014). A second has focused on optimizing study design in order to maximize the statistical power or estimate precision of statistical models, given that limited resources are available to devote to data collection (e.g., MacKenzie and Royle 2005). Finally, a third branch has focused on developing field methods that ensure the requisite data needed to fit statistical models are collected as efficiently as possible without violating estimator assumptions (e.g., Long et al. 2007a, Pauli et al. 2008).

Manuscript received 11 November 2016; revised 8 March 2017; accepted 10 May 2017. Corresponding Editor: David A. W. Miller.

⁵E-mail: jclare2@wisc.edu

The causes of bias and uncertainty in distribution or abundance estimates are increasingly well described. Few survey efforts detect all animals in a population or all species present at a location, and statistical methods to account for false absences associated with observed distribution or abundance are now widely used (Kéry and Royle 2016). The precision of distribution or abundance estimates that account for imperfect detection is improved by either increasing the spatial or temporal sampling effort (greater sample size) or increasing sampling efficiency (increasing the probability of detection; MacKenzie and Royle 2005). Other causes of bias have only more recently received attention: survey efforts subject to even a small number of false-positive detections or that non-representatively sample a subset of the target population can produce estimate bias (e.g., Miller et al. 2011, Graves et al. 2012). Although field method comparisons often seek to find methods that allow strong inference at reduced cost, selecting a single method often requires practitioners to weigh trade-offs between financial or logistical constraints to data collection vs. the risk of biased or uncertain inference (Gompper et al. 2006, Long et al. 2007b, Chambert et al. 2015).

Field method comparisons and state parameter inference can be greatly enhanced by models that explicitly incorporate multiple survey methods (multi-method models). Multi-method models permit direct comparison of method-specific detection probabilities, allow the estimation of false-positive detection probabilities and evaluation of method-specific sampling biases, provide more precise estimates of distribution or abundance, enable estimation of additional parameters such as availability for detection, and permit practitioners to formally combine different pieces of method-specific information into an integrated model (Nichols et al. 2008, Miller et al. 2011, Gopalaswamy et al. 2012, Sollman et al. 2013, Fisher and Bradbury 2014). Because multi-method models allow simultaneous field-method comparison, bias correction, and improved estimation of state parameters, they represent very powerful tools for a variety of applied objectives if correctly implemented. However, a critical assumption of existing multi-method models is that different methods produce independent detections.

Meeting this assumption can present logistical and inferential challenges. Maintaining independence between the detections produced by different methods means that the methods must operate in different places or at different times. Because methodological comparisons or multi-method inferences are often most critical for species such as carnivores that are inherently difficult to detect and expensive to effectively survey because they exist at low density over broad spatial extents, maintaining independence between methods can require substantial additional field effort or sampling duration. Indeed, many carnivore-sampling efforts specifically pair multiple detection methods at the same location to reduce logistical costs, and because pairing methods provides ancillary benefits

like laboratory cost-savings associated with culling genetic samples deposited by non-target species (Magoun et al. 2011). Furthermore, maintaining independence between different detection methods can induce additional causes of bias. Prolonging a survey duration to maintain temporal independence (*sensu* Gopalaswamy et al. 2012, Sollman et al. 2013) increases the likelihood that population size or distribution changes over the survey duration, while staggering detection methods in space can induce additional detection variance associated with local habitat conditions or the location of individuals with respect to specific methods (Graves et al. 2012). Thus, it is unsurprising that many multi-method survey efforts collocate methods (e.g., Magoun et al. 2011, Fisher et al. 2016).

The assumption that different methods provide independent detections is tenuous when methods are collocated, because two paired methods should jointly detect individuals or species more often than expected under the assumption of independence. Treating positively covarying detections as independent may not effect relative estimates of detection probability or estimates of detection bias, but is expected to overstate the overall probability of detecting a species or individual, and consequently lead to negative bias in state parameters. Thus, developing models that can allow covariance between the detections produced by different field methods is an important research need for preserving the benefits associated with multi-method models without biasing estimates of the parameters of primary managerial interest. Below, we formally describe occupancy and spatial capture–recapture models (MacKenzie et al. 2006, Royle et al. 2014) that can accommodate observation covariance between separate survey methods.

Our motivation for developing these models was based on a need to evaluate survey methods for American marten (*Martes americana*) in order to inform effective survey designs for providing baseline estimates of population density at specific localities of managerial interest and for monitoring changes in distribution over broader extents in Maine, USA. Martens are predicted to have undergone recent state-wide decline due to habitat loss (Simons 2009), and mitigating this decline may require making changes to existing fur-harvest or forest management practices. Because physically capturing marten requires tremendous field effort (Buskirk and Ruggiero 1994), passive detection techniques such as snow tracking, non-invasive genetic sampling, and remote cameras are common survey methods used for occupancy or capture–recapture efforts (Williams et al. 2009, Moriarty et al. 2011, Siren et al. 2016). These methods incur different costs and carry different potential shortcomings. We were concerned that tracks of the closely related fisher (*Pekania pennanti*) could be misclassified as marten due to overlap in track and gait measurements (Halfpenny et al. 1995). Although independent observers appear capable of distinguishing individual martens in remote-camera photograph for

capture–recapture analysis using remote cameras (Siren et al. 2016), the technique is still relatively untested. Hair collection and subsequent genetic analysis produces reliable detections, but can be inefficient due to increased expense and lower probabilities of detection (Fisher and Bradbury 2014). Our goal was to generate robust estimates of marten state parameters and compare alternative methods to inform future designs while working under logistical constraints that made collocating survey methods most feasible.

METHODS

Model descriptions

We use a hierarchical Bayesian framework to describe models to estimate animal occurrence or density when multiple methods produce correlated detections, but the models could also be fit by maximizing the likelihood after marginalizing out latent variables. Throughout, we index individual animals i , spatial replicates (sites or stations) j , and temporal replicates corresponding to sampling occasions k . We consider a sampling situation in which there are two collocated methods that may detect a species or individual organism. We consider the observed data y ($y_{j,k}$ or $y_{i,j,k}$, depending on whether species or individual-level detections are considered) as a categorical random variable to reflect four possible detection outcomes that can occur at a station during a specific occasion: detection by method 1 but not method 2, detection by method 2 but not method 1, detection by both methods, and detection by neither method. The vector π denotes the probability of each outcome.

Following MacKenzie et al. (2006), we assume that the latent occupancy state at site j , z_j , is distributed as Bernoulli (ψ_j), where ψ_j is the expected occupancy at site j and may vary across space in relation to environmental covariates as logit (ψ_j) = $\beta_0 + \beta_1 X_j$. Under a standard single-method design, detections during repeated visits to sites are Bernoulli outcomes with probability ($z_j \times p_{j,k}$), where $p_{j,k}$ is the probability of detecting a species present at site j during interval k and may vary across spatial and/or temporal replicates: for example, logit ($p_{j,k}$) = $\beta_0 + \beta_1 X_j + \beta_2 X_k + \beta_3 X_{j,k}$. Our reformulation considers $y_{j,k} \sim \text{Categorical}(\pi)$, and if methods are assumed to be independent, the vector π can be described using two parameters corresponding to the probabilities of detection with method 1 and method 2 (Table 1). To accommodate covariance, we estimate four real parameters (the vector ϕ_p) associated with the probabilities of all outcomes $y_{j,k}|z_j = 1$ (Table 1): if a species is present, there are distinct probabilities associated with detecting the species using method 1 but not method 2, detection with method 2 but not method 1, detection with both methods, and detection with neither. Thus the elements of π , the probability vector that describes the distribution of $y_{j,k}$, are equal to $\{z_j \times \varphi_1\} \{z_j \times \varphi_2\} \{z_j \times \varphi_3\} \{z_j \times \varphi_4 + (1 - z_j)\}$. We define $\phi_p \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_4)$, where the subscript p

TABLE 1. Categorical observation probabilities associated with specific observation outcomes $y_{j,k}$ for standard and false-positive occupancy models when two separate detection methods are considered independent or covarying.

$y_{j,k}$	Description	Pr($y_{j,k}$)		False positive	Covarying
		Standard	Covarying		
1	detection by method 1 only	$(z_j)(p_1)(1 - p_2)$	$(z_j)(\varphi_1)$	$(z_j)(p_1)(1 - p_2)$	$(z_j)(\varphi_1)$
2	detection by method 2 only	$(z_j)(p_2)(1 - p_1)$	$(z_j)(\varphi_2)$	$(z_j)(p_2)(1 - p_1) + (1 - z_j)(p_{fp})$	$(z_j)(\varphi_2) + (1 - z_j)(p_{fp})$
3	detection by both methods	$(z_j)(p_1)(p_2)$	$(z_j)(\varphi_3)$	$(z_j)(p_1)(p_2)$	$(z_j)(\varphi_3)$
4	not detected	$(z_j)(1 - p_1)(1 - p_2) + (1 - z_j)$	$(z_j)(\varphi_4) + (1 - z_j)$	$(z_j)(1 - p_1)(1 - p_2) + (1 - z_j)(1 - p_{fp})$	$(z_j)(\varphi_4) + (1 - z_j)(1 - p_{fp})$

Notes: False-positive models assume that only method 2 is possible of falsely observing a species when not present (probability p_{fp}); z_j indicates the binary occupancy state of a specific site, p_1 and p_2 are method-specific probabilities of detection conditional on occupancy, and φ is a vector of probabilities associated with each specific observation outcome conditional on occupancy.

indexes a particular detection outcome (out of total possible outcomes P) that occurs with probability φ_p . Following Royle and Dorazio (2008), we induce the Dirichlet prior distribution for φ_p by treating each $\alpha_p \sim \text{Gamma}(1, 1)$ and calculating each probability within $\varphi_p = \alpha_p / \sum_{p=1}^P \alpha_p$.

Method-specific probabilities of detection p_1 and p_2 can be respectively derived as $\hat{p}_1 = \hat{\varphi}_1 + \hat{\varphi}_3$ and $\hat{p}_2 = \hat{\varphi}_2 + \hat{\varphi}_3$. The additional real parameters allow the covariance between derived quantities p_1 and p_2 to take on values other than 0, as according to probability theory, $\text{Cov}(\hat{p}_1, \hat{p}_2) = \hat{\varphi}_3 - (\hat{\varphi}_1 + \hat{\varphi}_3) \times (\hat{\varphi}_2 + \hat{\varphi}_3)$. The vector φ_p can vary in relation to covariates, for example, a categorical covariate with two levels (A, B) can be accommodated by defining two different vectors for φ_p : $\varphi_{p,A} \sim \text{Dirichlet}(\alpha_{1,A}, \dots, \alpha_{4,A})$, and $\varphi_{p,B} \sim \text{Dirichlet}(\alpha_{1,B}, \dots, \alpha_{4,B})$. Estimates of p_1 and p_2 for each categorical level can be derived the same as when there are no covariates (e.g., $\hat{p}_{1,B} = \hat{\varphi}_{1,B} + \hat{\varphi}_{3,B}$), and the effect of the categorical covariate on a specific method can be estimated by calculating the difference between the posterior distributions of $\hat{p}_{1,A}$ and $\hat{p}_{1,B}$. Alternatively, both continuous or categorical variables can be accommodated by using a multinomial logistic link function such that $\alpha_p = \beta_{0,p} + \beta_{1,p}X_j$, where $\varphi_p = e^{\alpha_p} / (1 + \sum_{p=1}^{P-1} e^{\alpha_p})$ and one value of φ_p is treated as a reference and equal to $1 / (1 + \sum_{p=1}^{P-1} e^{\alpha_p})$. Modeling φ_p as varying in relation to covariates implicitly allows the covariance between derived method-specific detection probabilities to vary as well.

The model can be easily extended to estimate false-positive detection probability within a site-confirmation design (Miller et al. 2011, Chambert et al. 2015, Table 1). We assume that one of the two methods produces uncertain detections of the focal species of interest, and that the species is truly absent from a site, false-positive detections occur with probability p_{fp} . The other method is assumed to produce certain detections that provide auxiliary information about species presence. For example, if method 2 is prone to false positives and methods are not assumed to be independent, then $\pi = [\{z_j \times \varphi_1\} \quad \{z_j \times \varphi_2 + (1 - z_j) \times p_{fp}\} \quad \{z_j \times \varphi_3\} \quad \{z_j \times \varphi_4 + (1 - z_j) \times (1 - p_{fp})\}]$. The probability of falsely detecting a species can be allowed to vary in relation to predictors as logit($p_{fp,j}$) = $\beta_{0,fp} + \beta_{1,fp}X_j$.

Alternatively, φ_p can be considered conditional on detection rather than conditional on the state parameter, and defined as a vector of length three corresponding to the probability of detecting a species in a certain manner given that it was detected at all. Under this formulation, φ_1 , φ_2 , and φ_3 are equivalent to the probability that a given detection occurs with method 1 only, method 2 only, or both methods. The vector $\varphi_p \sim \text{Dirichlet}(\alpha_1 \dots \alpha_3)$, and $y_{j,k} \sim \text{Categorical}(\pi)$, with $\pi = [\{\varphi_1 \times p \times z_j\} \quad \{\varphi_2 \times p \times z_j\} \quad \{\varphi_3 \times p \times z_j\} \quad \{(1 - p) \times z_j + (1 - z_j)\}]$. In essence, φ_p is reformulated to represent the expected proportional probability of specific positive detection outcomes conditional on a single pooled detection model, and p now represents the probability of detection in any fashion. Outcome-specific detection probabilities can be

derived as $\varphi_p \times p$, and covariance between derived method-specific detection probabilities can be further derived as $\hat{\text{Cov}}(p_1, p_2) = \hat{\varphi}_3 \times \hat{p} - ((\hat{\varphi}_1 \times \hat{p}) + (\hat{\varphi}_3 \times \hat{p})) \times ((\hat{\varphi}_2 \times \hat{p}) + (\hat{\varphi}_3 \times \hat{p}))$. This formulation precludes estimation of additional parameters such as false-positive detection probability or multi-scale occurrence because there is a single detection process. However, it can drastically reduce the number of parameters that need to be estimated if there are several methods in operation or several detection covariates, as it makes it possible to model variation in the overall probability of detection without needing to also model variation in or covariance between the probability of specific detection outcomes.

We adopt this formulation for implementation of multi-method spatial capture–recapture models (SCR; Borchers and Efford 2008, Royle et al. 2014) when methods produce covarying detections. In describing SCR, we reuse certain variable names used in the occupancy description above to retain consistency with published descriptions. SCR models assume that an individual i has a latent activity center s_i where its detection probability during a specific sampling occasion is equal to parameter p_0 . Individual detection probability at specific locations j during specific intervals k decreases as a function of distance ($d_{i,j}$) between location j and s_i , such that $p_{i,j,k} = f(d_{i,j})$. The decay function is commonly considered bivariate normal, with $p_{i,j,k} = p_0 \times \exp(-d_{i,j}^2 / 2\sigma^2)$, and σ describes the rate of detection decay given distance from the activity center and relates to the spatial area that an individual uses (Royle et al. 2014). Variation in p_0 and σ can be modeled using individual, trap, or temporal covariates: e.g., logit(p_0) = $\beta_0 + \beta_1 X_{1,j}$, or $\log(\sigma) = \beta_0 + \beta_1 X_{1,i}$. Bayesian implementation of SCR models requires supplementing the observed individual detection histories with a number of all-zero detection histories such that there are M observed and augmented individuals (Royle et al. 2014). The crux of the model is to estimate the total number of the M individuals that have an activity center within a user-defined space ($\|S\|$): the existence of individual animals $z_i \sim \text{Bernoulli}(\psi)$, and abundance (\hat{N}) is estimated as $\sum_{i=1}^M \hat{z}_i$. With a single survey method, individual detections at specific stations and specific times $y_{i,j,k} \sim \text{Bernoulli}(p_{i,j,k} \times z_i)$.

Following our occupancy description above, our multi-method reformulation instead considers $y_{i,j,k} \sim \text{Categorical}(\pi)$, with four potential outcomes: $y_{i,j,k} = 1$ indicates an individual has been detected by the first method but not the second, $y_{i,j,k} = 2$ indicates an individual has been detected by the second method but not the first, $y_{i,j,k} = 3$ indicates detection by both methods, and $y_{i,j,k} = 4$ indicates no detection. If methods are assumed to produce independent detections, π can be formulated using method-specific parameters for p_0 , and σ can either be considered fixed or method specific (Gopalaswamy et al. 2012, Table 2). To accommodate covariance between methods, we define $\pi = [\{\varphi_1 \times p_{i,j,k} \times z_i\} \quad \{\varphi_2 \times p_{i,j,k} \times z_i\} \quad \{\varphi_3 \times p_{i,j,k} \times z_i\} \quad \{(1 - p_{i,j,k}) \times z_i + 1 - z_i\}]$, where φ_1 , φ_2 , and φ_3 are equivalent

TABLE 2. Categorical observation probabilities for spatial capture–recapture models when two separate detection methods are considered independent or covarying.

$y_{i,j,k}$	Description	$\Pr(y_{i,j,k})$	
		Independent	Covarying
1	detection method 1 only	$(z_i)(p_{i,j,k,1})(1 - p_{i,j,k,2})$	$(z_i)(\phi_1)(p_{i,j,k})$
2	detection method 2 only	$(z_i)(p_{i,j,k,2})(1 - p_{i,j,k,1})$	$(z_i)(\phi_2)(p_{i,j,k})$
3	detection both methods	$(z_i)(p_{i,j,k,1})(p_{i,j,k,2})$	$(z_i)(\phi_3)(p_{i,j,k})$
4	not detected	$(z_i)(1 - p_{i,j,k,1})(1 - p_{i,j,k,2}) + (1 - z_i)$	$(z_i)(1 - p_{i,j,k}) + (1 - z_i)$

Notes: All values of $p_{i,j,k} = p_0 \times \exp(-d_{i,j}^2/2\sigma^2)$, where p_0 is equivalent to detection probability if a detector is placed directly at an individual's activity center, and σ represents the rate of detection decay given distance from the activity center; z_i indicates the binary occurrence state of an individual animal, and ϕ is a vector of probabilities associated with each specific observation outcome conditional on it being detected.

to the probability that an individual was detected with method 1 only, method 2 only, or both methods given that it was detected at all, the vector $\phi_p \sim \text{Dirichlet}(\alpha_1 \dots \alpha_3)$, and $p_{i,j,k} = p_0 \times \exp(-d_{i,j}^2/2\sigma^2)$. Under this formulation, p_0 is equivalent to individual detection by any possible means given that a set of detection devices are placed at the activity center, and σ describes the decline in detection probability in any fashion as distance from the activity center increases. We adopt the formulation of ϕ_p as conditional on detection and state parameters rather than as a detection probability vector conditional on individual existence and location because we consider it unlikely that the relative probability of a specific detection outcome given detection at all changes across space. However, method-specific p_0 estimates can be derived for comparison or subsequent power analysis, e.g. $\hat{p}_{0,1} = (\hat{\phi}_1 + \hat{\phi}_3) \times \hat{p}_0$, and the covariance between method specific detection probabilities can be estimated as $\hat{\text{Cov}}(p_{i,j,k,1}, p_{i,j,k,2}) = \hat{\phi}_3 \times \hat{p}_{i,j,k} - ((\hat{\phi}_1 \times \hat{p}_{i,j,k}) + (\hat{\phi}_3 \times \hat{p}_{i,j,k})) \times ((\hat{\phi}_2 \times \hat{p}_{i,j,k}) + (\hat{\phi}_3 \times \hat{p}_{i,j,k}))$. Modeling variation in the covariance between methods can be accommodated by incorporating additional dimensions of ϕ_p or using log-linear models for α_p within a multinomial logistic link function.

Simulation study

We used simulation to evaluate the influence of between-method detection covariance on multi-method occupancy and SCR estimators assuming methods operated independently, and we evaluated the ability of different posterior predictive checks (Gelman et al. 1996) to detect departures from model assumptions (Appendix S1 contains details). We considered an occupancy design with 100 sites, three visits, and two methods. We simulated 500 data sets for each of 18 scenarios corresponding to different values for occupancy probability, overall detection probability, and covariance between detection methods. Our simulation for SCR models was more limited because the models are highly dimensional and computationally more intensive. We considered a design with 36 detectors placed in a 6×6 square with one unit spacing, and five sampling occasions. We fixed $\sigma = 0.5$ and

$p_0 = 0.3$, with $N = \text{Binomial}(100, 0.3)$, and ϕ_p varying as either $\{\{0.1\} \{0.1\} \{0.8\}\}$, or $\{\{0.2\} \{0.2\} \{0.6\}\}$, because we could not imagine that individual identification could be reconciled across methods if they did not strongly positively covary. We simulated 150 data sets for each scenario. We fit models assuming between-method independence and accommodating between-method covariance using Markov chain Monte Carlo simulation in JAGS (Plummer 2003) via the R Library jagsUI (R Development Core Team 2014, Kellner 2015). We evaluated relative bias of point estimates (posterior mean or mode) and frequentist coverage of the 95% credible intervals in order to evaluate sensitivity to violation of observation independence. Simulation convergence was assessed by visualizing trace plots and considering a value of 1.05 for the Gelman-Rubin statistic (Gelman and Rubin 1992) as a threshold for convergence.

We assessed the ability of three posterior predictive checks (Gelman et al. 1996) to correctly reject a model assuming independence between method-specific detections: one based on the marginal log-likelihood of the observations, one based on the log-likelihood of the observations conditional on the latent occupancy parameters, and one based on the distribution of joint and method-specific detections (Appendix S1). We report average Bayesian P values across simulations as a description of how powerful the checks were to detect departure from assumptions. For SCR models, we only used the third posterior predictive check because iteratively tracing the log-likelihood required an enormous amount of computer memory.

Empirical study

We surveyed martens in north-central Maine, USA (adjacent townships T5R11, T6R10, T6R11, T7R10 centered at 46.149°N , 69.124°W ; Fig. 1). The region is humid (mean annual precipitation 105 cm; NOAA 2015a) and characterized by cold winters and mild summers (mean temperature, January -11°C , July 17°C ; NOAA 2015a). Forest composition in the study area was typified by red and white spruce (*Picea rubens* and *P. glauca*), balsam fir (*Abies balsamea*), American beech

(*Fagus grandifolia*), and sugar maple (*Acer saccharum*) managed for industrial or mixed-use purposes.

During 7 January–4 April 2015, we deployed 238 detection stations with multiple detection methods (Fig. 1). We used Monte Carlo simulation to evaluate how different station spacing and deployment time influenced spatial capture–recapture estimators in order to optimize survey design with the R library *secr* (R Development Core Team 2014, Efford 2014) and simulated marten detection parameters drawn from the results of Siren (2013). Simulations suggested that 750-m spacing and 15-d deployment time produced satisfactory estimation, and our planned design was to rotate stations across the study area every 15 d. In reality, stations were sampled for 9–31 consecutive 24-h periods ($\bar{x} = 19$) owing to an abrupt loss of snowpack at the end of the season, and because we extended one rotation to evaluate whether we had dramatically underestimated latency to detection. Mean nearest neighbor distance between stations was 602 m, and the minimum convex polygon surrounding the outer camera stations was 306 km².

Each station consisted of a bait platform built out of 2.5 × 10 cm boards screwed or zip tied to a standing tree 50–100 cm above snow level, and baited with an open sardine can (Fig. 2). Following the designs of Magoun et al. (2011) and Siren et al. (2016), we wired an access ramp (downed wood with >7 cm diameter) to the bait station to provide access from the snow level. We set a single camera (Bushnell Trophy Cam, Overland

Park, Kansas, USA) facing the bait station 50–100 cm opposite the bait to capture images of unique ventral markings (Fig. 2). We collected hair samples on two pieces of glue-board (Tomcat Indoor Rodent Trap, Madison, Wisconsin, USA) cut into 2.5 × 10 cm strips and stapled below the sardine tin on the platform. All stations were located between 50 and 100 m from unpaved roads or snowmobile trails, and we deployed a call lure made out of skunk quill mixed into liquefied petroleum jelly (ratio of quill to petroleum jelly 1:12 ounces). Although the proximity to roads creates the risk of a biased sample, because sites were baited, nearly all potential marten home ranges within the study area contain roads, and martens exhibit only extremely fine-scale avoidance of roads, we expect our design did not seriously compromise estimates of marten occurrence or density (Chapin et al. 1998, Robataille and Aubry 2000, Zielinski et al. 2008).

We visited the detection stations every 3–4 d ($\bar{x} = 3.18d$) to replace bait, camera batteries, and glue strips as needed. If hair was deposited, glue strips were replaced. We collected hair samples if we could not exclude marten deposition (e.g., deposition by another species was observed on camera). We stored samples in coin envelopes with silica desiccant. During each visit, we recorded any marten tracks encountered along the foot trail to the station or underneath the station. We marked tracks when observed to avoid subsequent double counting.

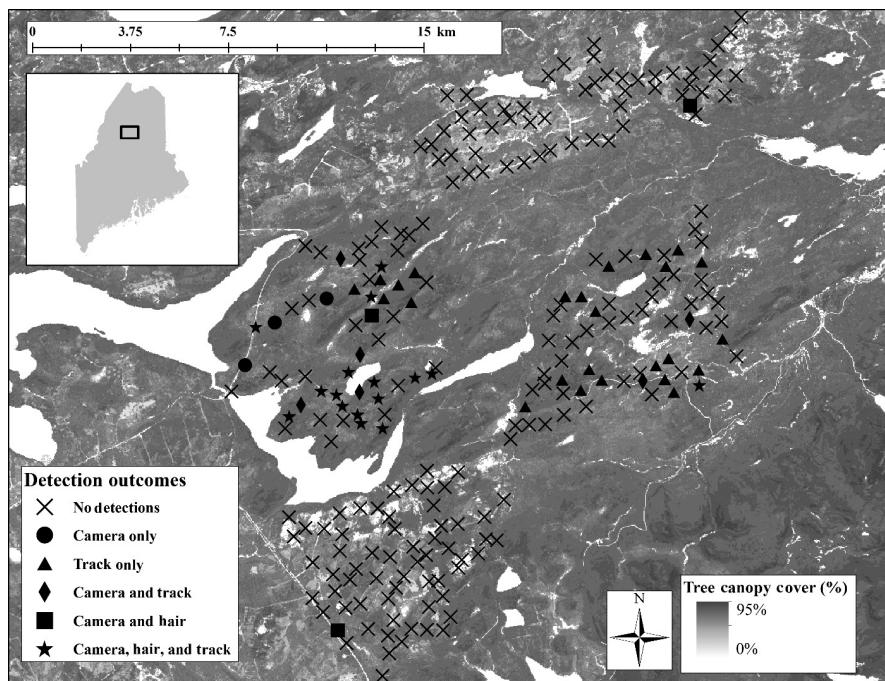


FIG. 1. Location of stations used to sample American marten during winter 2015, and recorded detections by sampling method in northern Maine, USA relative to existing tree cover derived from LANDFIRE. Detections by method are aggregated across visits (e.g., camera and track indicates that both methods produced detections, not necessarily that both methods produced a simultaneous detection).

We identified individual martens detected by cameras and hair catches. Two independent observers identified individual martens from photographs based on the presence of unique ventral markings (Siren et al. 2016). When both observers agreed, we treated this as a specific individual encounter, and when observers were either ambiguous regarding individual assignment or disagreed, we discarded the image sequence from analysis. Hair samples were confirmed as marten using standard mtDNA protocols (Riddle et al. 2003) and individually assigned based on genotypes from nine microsatellite loci and sex-specific markers (Appendix S2). A primary difference between the methods was that although hair samples could only possibly detect one individual visiting a station during a check, a camera could detect multiple individuals visiting the station.

We cross-compared marten images with the detection histories produced by genotypes to evaluate the accuracy of our photo-identification. The evaluation was imperfect, because there were several possible comparative outcomes: (1) a marten was identified on camera and directly observed depositing hair during the encounter, (2) a single marten was identified on camera during a specific check and a genotype was available for that check, (3) multiple martens were identified on camera during a specific check, and a genotype was available, (4) a genotype was available without a photographic identification, and (5) a photographic identification was available without a genotype. We treated outcomes one and two as absolute matches, and first evaluated any inconsistency between detection histories. We cross-checked the third outcome by assessing if an individual with linked photographic and genotypic signatures was observed during the check. Ultimately, we found no unambiguous disagreements between individual assignments (see Appendix S3 for outcomes).

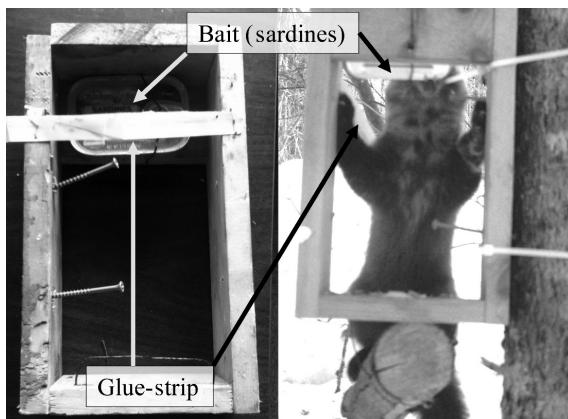


FIG. 2. Detection station design used to sample American marten in northern Maine during winter 2015. On right, a marten visiting a detection station and depositing hair on a glue strip.

Fitting occupancy models

We considered six distinct models for estimating marten occupancy based on the combination of camera images and snow tracks: three models considered method-specific observations to be independent, and three models allowed methods to produce covarying detections. Within each model, we considered variance in occupancy as $\text{logit}(\psi_i) = \beta_0 + \beta_1 \times \text{FOREST}_j$, where FOREST_j represents average vertically projected tree canopy cover (LANDFIRE 2015) within a 1-km circular buffer around a specific station. Martens select against using landscapes with limited forest regionally (e.g., Chapin et al. 1998), and the extent of this buffer was selected based on previous studies of marten habitat associations (e.g., Simons 2009, Wasserman et al. 2012). When methods were assumed to be independent, variance in detection probability conditional on presence ($p_{j,k,m}$, where m indexes method) was modeled as $\text{logit}(p_{j,k,m}) = \beta_{0,m} + \beta_{1,m} \times \text{EARLY}_j + \beta_{2,m} \times C_{j,k,m}$. EARLY_j was a site-specific binary variable distinguishing sites sampled during early winter (January and February) and late winter (March and April) to account for temperature-specific changes in marten activity, seasonal effects with regard to detecting active marten such as changes in snow compaction, temperature effects on the stickiness of glue traps and camera triggering, and real changes in marten behavior associated with females giving birth in March and early April (Powell et al. 2003). We used this binary variable to capture numerous correlated continuous variables (snow depth and compaction, temperature) for which we lacked data: February 2015 was a historically cold month in the study area, after which the temperature increased rapidly (NOAA 2015b), and initial assessment suggested that detection rates were better reflected by two distinct averages than a linear or quadratic trend. $C_{j,k,m}$ was a time- and site-specific binary indicator of whether a specific method had previously detected marten, included to accommodate potential trap-happy responses to the bait. When methods were considered to have covariance, we defined four probability vectors φ to describe the four possible detection outcomes for each possible covariate combination of $C_{j,k}$ and EARLY_j . Because we were explicitly concerned about the potential for track-based false positive detections, four models (two assuming independence, two allowing covariance) included a track-specific false positive probability. In two models, p_{fp} was considered constant. In two other models, $p_{fp,i,j}$ varied as a logit-linear model based on whether a fisher was photographed at the station during the same occasion. For general comparison, we also fit single-method models (which could not incorporate false-positive probability) based on track and photographic detections; we make comparisons based on estimates of finite-sample occupancy (i.e., the proportion of detection stations within the sample estimated as occupied; Royle and Kéry 2007).

We fit all models using JAGS through the R Library jagsUI (R Core Team 2014, Kellner 2015). Prior

distributions for the inverse-logit of the intercept parameters were set as Uniform (0, 1), and prior distributions for effects on the logit scale were defined as Uniform (-5, 5) to reflect the limited range of plausible effect sizes (Gelman et al. 2008). Simulation settings included three chains of 25,000 iterations, with the first 15,000 iterations discarded as burn-in and the last 10,000 iterations thinned by five. Thus, 6,000 posterior samples were used for inference, and model convergence was assessed by visual inspection of the chain trace-plots and with estimates of the potential scale reduction factor (Gelman and Rubin 1992).

We first evaluated the suitability of different model structures (independent vs. covarying methods), using the results of our simulation study and posterior predictive checks to assess whether assuming independence between method-specific detections was a reasonable assumption. Once a model structure was selected, we compared the three candidate models with indicator variable selection methods (Kuo and Mallick 1998). Indicator variable selection typically associates a model coefficient term with a latent indicator variable g that has a prior distribution of Bernoulli (0.5) that determines whether the predictor is included within the model likelihood. We modified the formulation to treat g_2 (the indicator associated with a fisher effect on false-positive probability) as conditional on g_1 (the indicator associated with there being any false-positive probability at all): when the iterative sample for $g_1 = 1$, $g_2 \sim \text{Bernoulli}(0.5)$, but was otherwise fixed as equal to zero. Thus, an intercept-only false-positive probability entered the model likelihood when $g_1 = 1$ and $g_2 = 0$, and when both g_1 and $g_2 = 1$, false-positive probability was modeled as varying in relation to visit-specific fisher photographs. The posterior mean of the indicator variables is used to evaluate model support (Kuo and Mallick 1998, Royle and Dorazio 2008): the proportion of simulations in which $g_1 = 0$ is equivalent to the support for a model lacking a false-positive component, the proportion of iterations in which both $g = 1$ is equivalent to the support for a false-positive model and a term for fisher, and the remainder is the support for an intercept-only false-positive model. Because there were three models under consideration, we used a Bernoulli (0.67) prior for g_1 and a Bernoulli (0.5) prior for g_2 to represent uniform prior support for each specific model. Post-hoc refitting suggested little sensitivity to the prior distributions used for g .

Comparing method-specific density estimates

We attempted to fit two spatial capture–recapture models: one assuming independence between genetic and photographic detections and one allowing methods to produce covarying detections. The model that assumed independence between genetic and photographic detections exhibited poor mixing even after extending chain length, and we consequently make inference with the single model incorporating method-

specific detection covariance. We considered the same sources of detection variation used for occupancy modeling, trap-specific behavioral responses and a detector-specific factor coding for whether it was deployed during early or late winter. We also modeled sex-specific differences in p_0 and σ to account for differences between male and female marten space use (Katnik et al. 1994), treating the sex of individual marten as a random variable following Sollman et al. (2011). Thus, the model for individual detection probability was $p_{i,j,k} = p_{0,i,j,k} \times \exp(-d_{i,j}^2/2\sigma_{\text{sex}[i]}^2)$, where $\text{logit}(p_{0,i,j,k}) = \beta_{0,\text{sex}[i]} + \beta_{1,\text{sex}[i]} \times \text{EARLY}_j + \beta_{2,\text{sex}[i]} \times C_{i,j,k}$. Observed data $y_{i,j,k} \sim \text{Categorical } (\pi)$, with $\pi = [\{\varphi_1 \times p_{i,j,k} \times z_i\} \quad \{\varphi_2 \times p_{i,j,k} \times z_i\} \quad \{\varphi_3 \times p_{i,j,k} \times z_i\} \quad \{(1 - p_{i,j,k}) \times z_i + 1 - z_i\}]$. The prior distribution for σ was defined as Uniform (0, 10). MCMC settings, assessments of convergence, and prior distributions for probability or beta parameters were identical to those used for occupancy modeling except that all 30,000 samples after the burn-in were used for inference. For comparison, we also fit SCR models using exclusively camera or hair detections. The single-method model based on hair detections incorporated the same covariate structure for $p_{i,j,k}$, but the single-method camera model did include sex-specific parameters because we could not distinguish sex photographically and were specifically interested in understanding whether failure to do so might bias estimates. We considered φ_p as a constant vector because earlier model fitting attempts suggested that there were too few detections during earlier winter to effectively identify separate levels of φ_p , and that φ_p exhibited little change after a marten was initially detected. Code for fitting occupancy and SCR models with empirical data is provided in the supporting information.

RESULTS

Simulation results

As we hypothesized, multi-method occupancy and spatial capture–recapture estimators that assume methods produce independent detections performed poorly when this assumption was violated. Occupancy models that assume methodological independence and that accommodate inter-method detection covariance performed similarly when the simulated covariance between inter-method detections was small. When simulated covariance was large, models that assumed methodological independence tended to underestimate occupancy and overestimate detection probability, although the effect of inter-method covariance did not seem to compromise model performance when the overall probability of detection was large enough that few occupied sites lacked detections (Appendix S1: Fig. S3). Models accommodating inter-method detection covariance exhibited approximately nominal coverage and much less bias across the high-covariance simulations, although estimates of method-specific detection

probability were positively biased when the number of observations within the simulated data was very sparse (i.e., $p = 0.3$, $\psi = 0.3$; Fig. 3). Spatial capture–recapture models that assumed methodological independence when the models featured positive observation covariance also exhibited permissive coverage, negatively biased estimates of both population size and σ , and positively biased estimates of p_0 (Table 3). Models accommodating detection covariance methods exhibited approximately nominal coverage and minimal bias.

None of the posterior predictive checks we considered were perfect tools for determining when assumption

violations led to estimator bias (Appendix S1). The check based on a chi-square discrepancy between the sums of observed and simulated detections by method rejected the occupancy models assuming independence when methods produced strongly positively covarying detections regardless of whether inter-method covariance led to estimator bias, but exhibited limited power to reject improperly structured SCR models (Appendix S1). The other posterior predictive checks based on model deviance exhibited limited power to reject models that incorrectly assumed methods produced independent detections, although power to detect poor goodness-of-fit tended to increase as

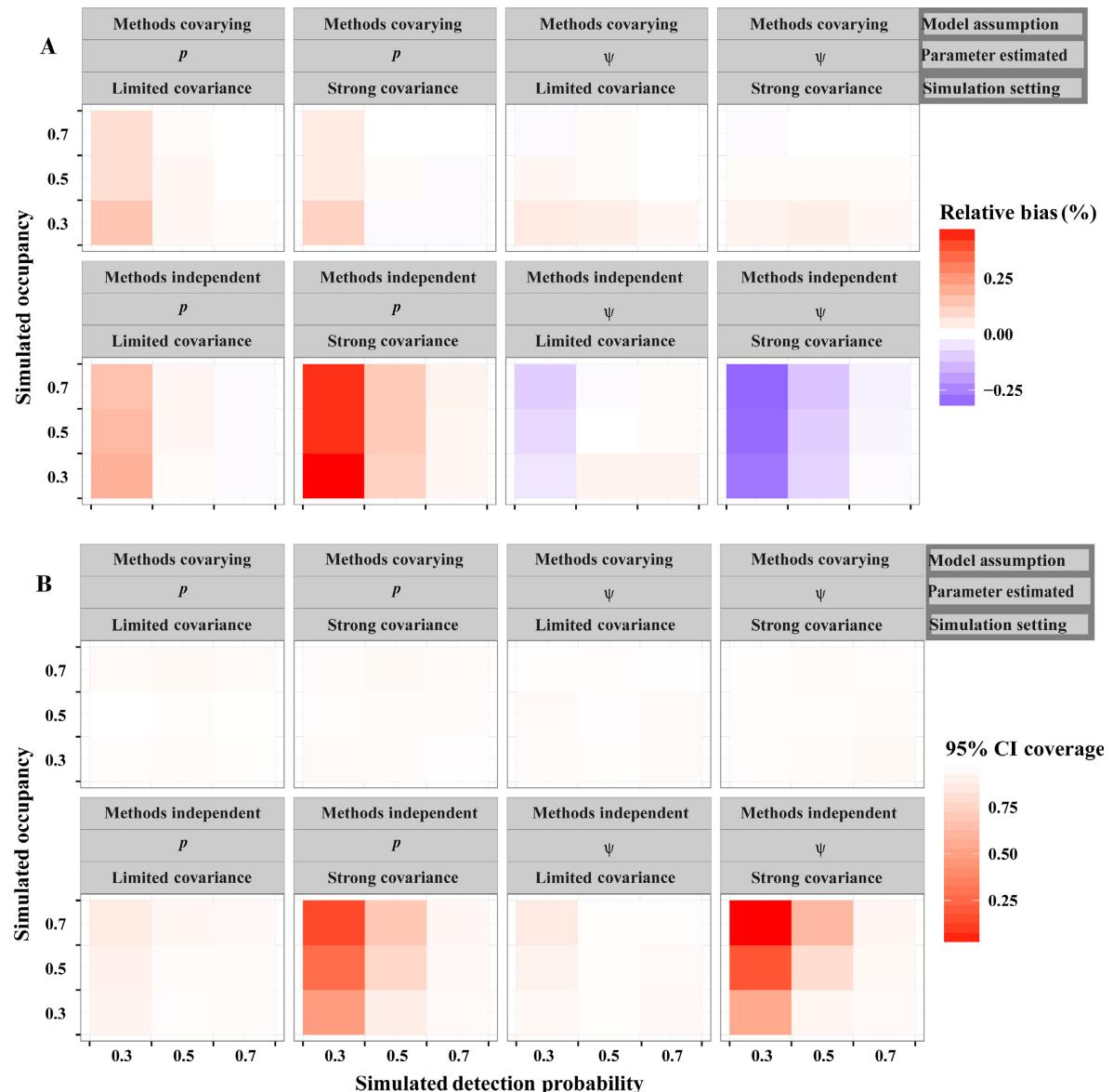


FIG. 3. (A) Relative bias of occupancy and method-specific detection probability point estimates and (B) frequentist coverage of 95% credible intervals when multi-method occupancy models assumed independence between methods or allowed covariance between methods. Axes correspond to simulated occupancy probability and the combined detection probability of the two simulated methods per occasion.

TABLE 3. The relative bias and coverage of spatial capture–recapture models that either assume multiple methods produce independent observations or allow method-specific observations to covary.

Detection vector (ϕ), model type, and, parameter	Relative bias				\overline{BPV}	
	Posterior mean (%)	Posterior mode (%)	95% CRI coverage (%)	—		
[{0.1} {0.1} {0.8}]						
Independent methods						
p_0	75.44	—	30.67	0.204		
σ	-16.44	—	34.00			
N	-14.14	-16.90	56.00			
Covarying methods						
p_0	6.04	—	96.00	—		
σ	0.95	—	94.00			
N	6.17	-0.75	93.30			
[{0.2} {0.2} {0.6}]						
Independent methods						
p_0	57	—	44.00	0.201		
σ	-13.39	—	45.30			
N	-11.59	-14.37	70.00			
Covarying methods						
p_0	5.51	—	93.30	—		
σ	0.97	—	97.30			
N	3.43	-1.50	92.70			

Notes: Simulated models include three potential successful detection outcomes that make up specific proportions of the observed detections (ϕ): the first two values represent the proportion of detections that are exclusively achieved with a specific method, and the third value represents the proportion of detections in which both methods detect an individual. Parameters p_0 , σ , and N correspond to method-specific detection probability of an individual when a method is located directly at its activity center, the half-normal decay of individual detection probability as distance between an individual and detector increases, and the number of individuals. \overline{BPV} corresponds to the average value of a posterior predictive check based on a comparison of the empirical chi-square discrepancy between the number of methods detecting an individual at a specific trap and occasion vs. iteratively simulated data. CRI, credible interval. Dashes represent values that were not estimated.

estimator bias increased (Appendix S1). Although this was initially surprising, such checks would be better suited for detecting a misspecification of the variability in occupancy or detection probability, and our simulations considered all parameters constant across sites and occasions. Thus, the most robust way to ensure assumptions were not being violated in ways that hindered inference appeared to be fitting both models to evaluate estimate similarity.

Empirical study

Occupancy.—Total sampling effort was 4,112 nights, corresponding to 1,296 individual station visits. Cameras recorded martens during 52 (4%) station visits, tracks were recorded on 76 (6%) visits, and verified marten hair was collected on 39 (3%) visits. In total, martens were detected at 49 (21%) out of the 238 stations, although detection outcomes were variable by method (cameras = 24, hair = 20, tracks = 47; Fig. 1). Of the three positive detection outcomes we considered within multi-method camera and track occupancy models, camera-only detections were recorded 13 times, track-only detections were recorded 37 times, and joint camera and track detections were recorded 39 times. Fishers were detected on camera at 18 stations during 21 check intervals.

Estimates of marten occupancy using multiple methods were ubiquitously smaller and more precise than those

based on single-method observations, but they varied considerably depending on assumptions of methodological independence and false-positive detections (Fig. 4). Models assuming independence between methods provided more precise estimates of marten occupancy than models allowing methodological covariance, but also provided much lower estimates of occupancy (Fig. 5), a strong indication that the assumption of methodological independence was inducing bias. Furthermore, the model assuming methodological independence and no false-positive detections appeared to be producing more joint-method detections than iteratively simulated data (Bayesian P value < 0.01), while exhibiting poorer fit for the detection sub-model than iteratively simulated data (Bayesian P value = 0.88, Table 4). Consequently, we further considered only models allowing methodological covariance.

Indicator variable selection suggested there was extremely little support for assuming that snow tracking did not produce false-positive detections (posterior model probability = 0.01), with relatively equitable support for models that assumed snow tracks resulted in false-positive marten detections at a constant rate (posterior model probability = 0.50) or that the likelihood of false-positive track detections increased during check intervals when fishers were observed on camera (posterior model probability = 0.49). We use the averaged model produced by indicator variable selection when presenting

TABLE 4. Model selection and goodness of fit results for the six multi-method occupancy models fit considering method-specific detection to be either independent or covarying.

Model	Methods	BPV (detections)	BPV (marginal log \mathcal{L})	BPV (conditional log \mathcal{L})	Model weight
$\psi(\text{forest}), p(\text{early} + C)$	independent	<0.01	0.54	0.88	—
$\psi(\text{forest}), p(\text{early} + C), \text{fp}(.)$	independent	0.44	0.52	0.65	—
$\psi(\text{forest}), p(\text{early} + C), \text{fp(fisher)}$	independent	0.48	0.53	0.26	—
$\psi(\text{forest}), p(\text{early} + C)$	covarying	—	0.26	0.25	0.01
$\psi(\text{forest}), p(\text{early} + C), \text{fp}(.)$	covarying	—	0.45	0.36	0.50
$\psi(\text{forest}), p(\text{early} + C), \text{fp(fisher)}$	covarying	—	0.46	0.37	0.49

Notes: BPV represents the proportion of Markov chain Monte Carlo (MCMC) iterations in which the data set produced a less extreme chi-square statistic based on cell-specific detections than a simulated data set (Detections), a smaller log likelihood than simulated data based on iterative parameter values (BPV Marginal Log \mathcal{L}), and a smaller log-likelihood given iterative parameter values and the latent occurrence states of specific sites (BPV Conditional Log \mathcal{L}). Model weight represents the proportion of posterior MCMC iterations in which that model was used for estimation. Dashes indicate no values were estimated. C is an indicator of whether a marten had been detected previously at a specific trap location, and fp represents false-positive detection probability.

estimates of detection and state parameters. Realized marten occupancy within the sampling area was low (finite sample occupancy = 0.26, 95% credible interval [CRI] 0.15–0.42), but positively associated with surrounding forest canopy cover ($\hat{\beta} = 1.51$, 95% CRI 0.50–2.77). Cameras and snow tracks had comparable probabilities of detecting present martens: both methods were more likely to detect present martens in late winter and after initial marten detection (Fig. 5). The overall probability of detecting a present marten by any means during a specific visit ranged from 0.08 (95% CRI 0.02–0.17) in early winter for initial detections to 0.57 (95% CRI 0.43–0.70) in late winter after a marten had been previously detected. The constant probability of falsely detecting marten tracks during a check at a site where the species was absent was estimated as 0.02 (95% CRI 0.01–0.03); when the additional term associated with concurrent fisher photographs was included within the model, there was weak indication that this probability increased to nearly 0.08 ($\hat{\beta} = 1.30$, 95% CRI –1.61–3.68; Fig. 5).

Density.—Individual martens were detected at specific stations during specific checks 55 times (24 camera only, 4 hair only, 27 combined). We identified 15 individual martens visiting camera stations based on photographs, with 75% observer agreement on individual assignment across all photographic events. A substantial proportion (66%) of the unassigned photographic events contained no individual markings or were too blurry to be useful, with the remaining events subject to observer disagreement. Based on the patterns of assigned individuals (i.e., several successive visits over a single occasion), we expect that few data were lost, because most of the unassigned photographic events were flanked by identifiable events within the same occasion and the average time between inter-individual visits has been reported as >24 h (Siren et al. 2016). Genetic assignments based on hair distinguished eight individuals with an equal sex ratio (4:4), with one individual we assume was distinct and not identified photographically given that nearby

martens identified photographically were linked to genotypes (Appendix S3).

We estimated marten density using a multi-method model as 7.7 or 8.0 individuals/100 km² (respectively, the mode and mean of the posterior distribution; 95% CRI 4.4–13.1). Camera-only estimates of marten density were comparable (posterior mode = 6.0, posterior mean = 6.8, 95% CRI 3.8–11.1), while hair-only estimates were lower (posterior mode = 2.9, posterior mean = 3.5, 95% CRI 1.50–6.60; 95% CRI for differences between posterior density samples using hair-only and or multi-method models did not overlap 0). Although male martens appeared to use more space than female martens ($\sigma_{\text{male}} > \sigma_{\text{female}}$), they were typically less detectable at any specific point within the area they used ($p_{0,\text{male}} > p_{0,\text{female}}$; Table 5). Both sexes were detected less frequently during early winter and detected more frequently at specific detectors after initial encounter, and we could not detect whether the sex ratio of martens within the study area was different than 1:1 (Table 5). Individual martens were more likely to be detected photographically than genetically, although combined genetic and photographic detections were as likely as photographic detections alone (Table 5).

DISCUSSION

Using multiple survey methods and multi-method analysis can enhance the rigor of wildlife survey inference by strengthening state parameter estimates, incorporating information that may be provided only by specific methods such as individual sex, and correcting for specific methodological shortcomings such as false-positive detections (Nichols et al. 2008, Miller et al. 2011, Sollman et al. 2013, Fisher and Bradbury 2014). Additionally, multi-method models serve as valuable tools for evaluating the strengths and weaknesses of different field methods, allowing practitioners to more effectively design and efficiently implement survey or monitoring efforts. Our reformulations relax the need

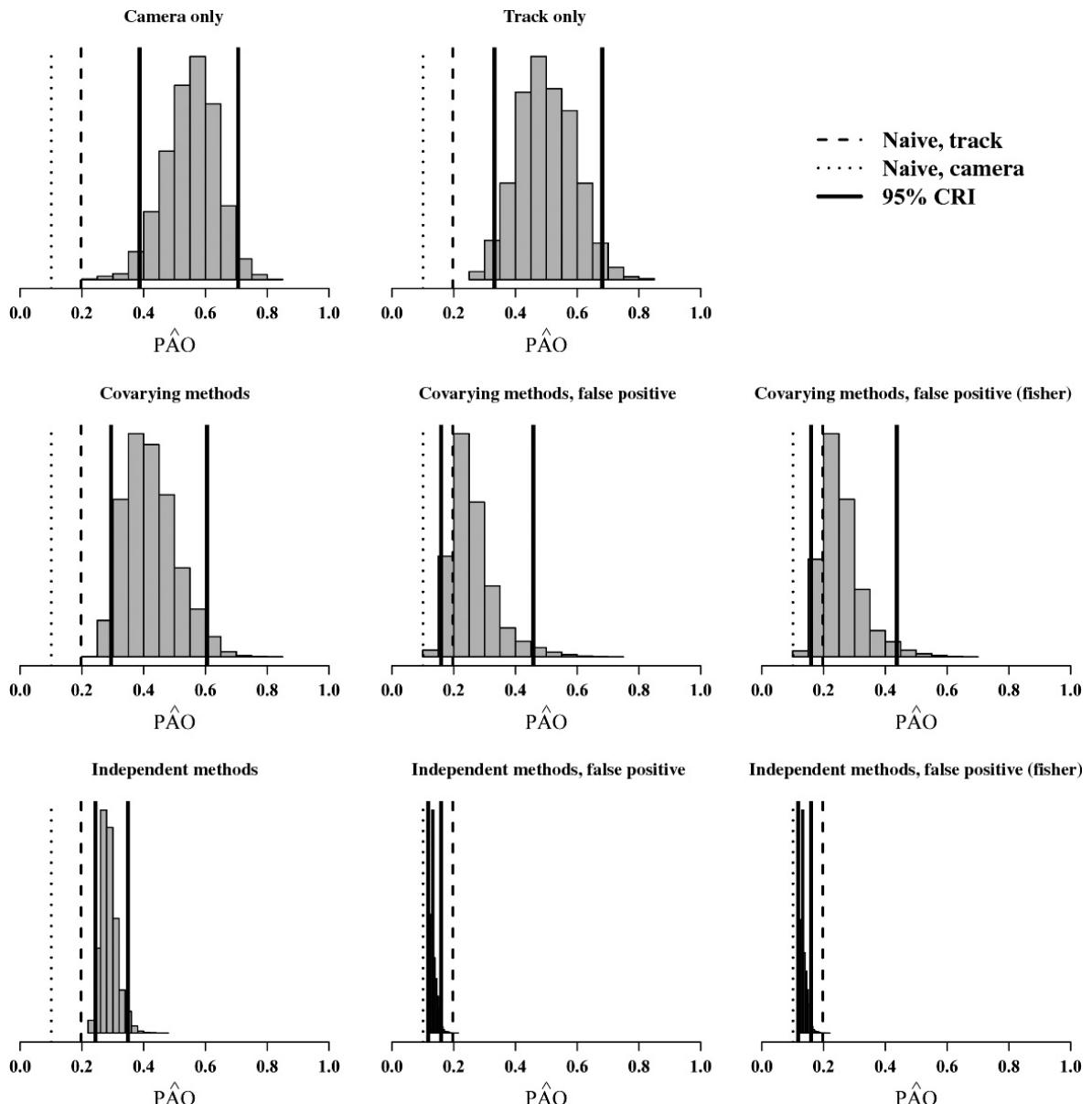


FIG. 4. Posterior distributions of finite-sample occupancy (proportion of sites occupied, or \hat{PAO}) for American martens sampled using multiple methods in Maine during winter 2015. Vertical lines correspond to observed occupancy by method (snow tracks or remote camera detections), or 95% credible intervals (CRI). Top panels depict estimates derived from a single method, middle panels depict estimates derived from multi-method models incorporating methodological covariance, and bottom panels depict estimates derived from multi-method models assuming methodological independence. Multi-method models either assumed no false positive detections (left), constant probability of false positive detections using tracks (middle), or that false positive detections were associated with observed fishers (right).

for methods to operate independently, providing practitioners considerably more flexibility to take advantage of existing models. Furthermore, our application highlights the broader value of using multiple field-methods when implementing wildlife surveys.

Relaxing the assumption of methodological independence when using multi-method analysis provides two primary benefits. First, our simulations suggest that

violation of this assumption can result in precisely biased estimates of state and detection parameters using previously described model structures, meaning that practitioners using these analyses without consideration for the assumptions may gain the pyrrhic ability to estimate new parameters or compare relative method-specific detection parameters while losing the ability to robustly estimate the state or detection parameter values motivating the

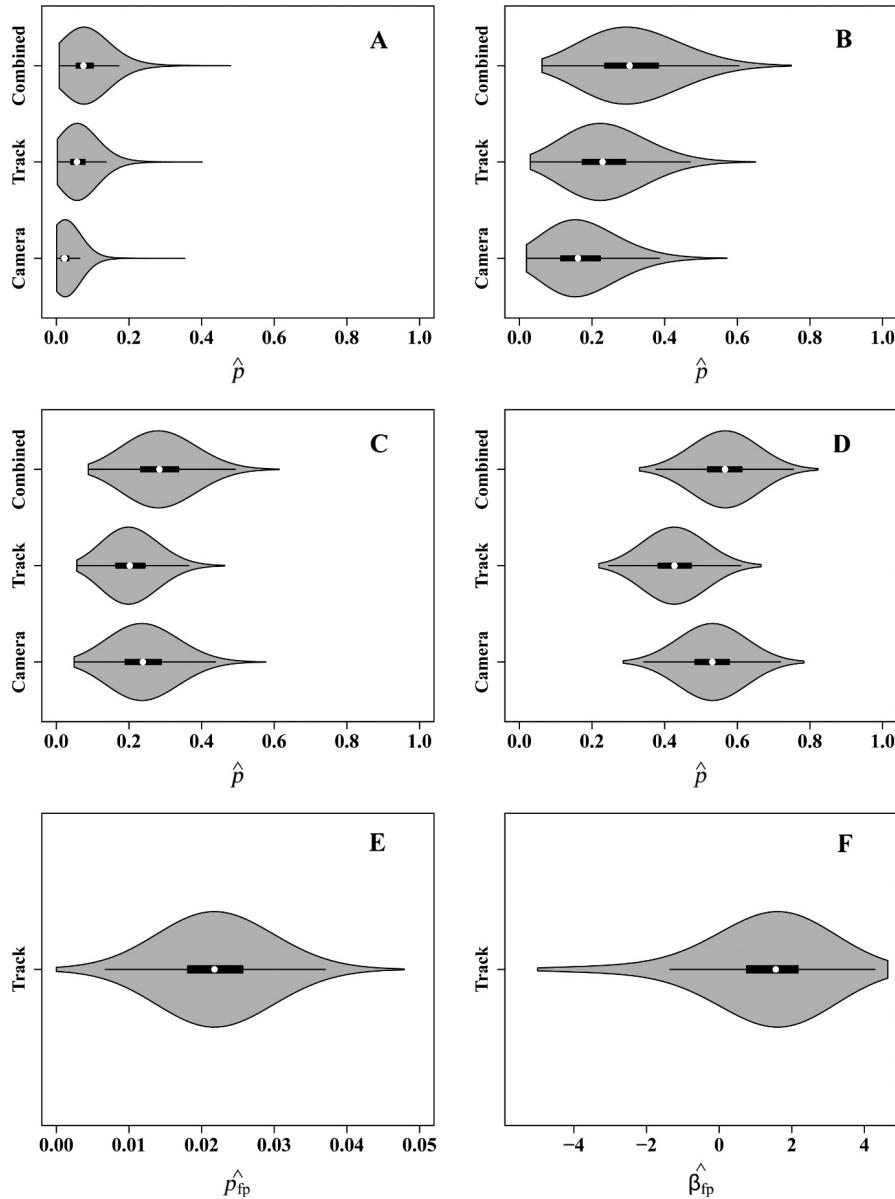


FIG. 5. Posterior distributions of method-specific detection and false-positive detection parameters drawn from a model-selected multi-method occupancy model estimated for American martens in Maine during winter 2015. Panels depict (A) initial detection probability during early winter, (B) detection probability after initial detection during early winter, (C) initial detection probability during late winter, (D) detection probability after initial detection during late winter, (E) probability of falsely detecting a marten when not present at a station, and (F) the effect of a fisher being photographed at a station on false-positive probability.

survey effort. Simulation results suggest that models assuming independence produce robust parameter estimates when there is not strong inter-method covariance or when overall detectability is high. Unfortunately, these parameters are difficult to accurately guess a priori (Guillera-Arroita et al. 2014), and thus the biased estimates produced by models assuming independence cannot necessarily be used to diagnose whether one operates in a problematic part of the parameter space. This may be particularly problematic for efforts like our own that seek

to use initial parameter estimates to inform subsequent survey or monitoring designs. The combination of negative bias and permissive coverage increases the likelihood of overstating population vulnerability and survey ability to detect changes in population parameters across time or space. As many changes in management or policy are triggered when specific population objectives are achieved, this precise bias increases the risk of using survey or monitoring designs that may both fail to detect the achievement of a positive distribution or abundance

TABLE 5. Parameter estimates (presented as effects) and 95% credible intervals (upper, UCRI; lower, LCRI) for male and female (M, F) American marten derived from multi-method spatial capture–recapture models considering hair and camera detections in northern Maine, USA.

Parameter	Posterior mean	LCRI	UCRI
p_0 , intercept (M) [†]	-1.93	-2.62	-1.12
p_0 , intercept (F) [†]	-1.34	-2.31	-0.27
p_0 , early winter (M) [†]	-1.51	-3.20	0.63
p_0 , early Winter (F) [†]	-1.50	-2.97	-0.11
p_0 , previously detected (M) [†]	1.99	0.93	3.01
p_0 , previously detected (F) [†]	1.37	0.18	2.61
σ (M) [‡] [§]	1.31	0.88	1.81
σ (F)	0.66	0.43	0.91
Proportion female	0.68	0.32	0.94
φ_1 , camera only	0.43	0.31	0.56
φ_2 , hair only	0.09	0.02	0.16
φ_3 , both camera and hair	0.48	0.35	0.61
Proportion camera detections	0.91	0.84	0.98
Proportion genetic detections	0.57	0.45	0.69
Density [¶]	8.00	4.40	13.10

Notes: Early winter indicates sampling during January and February rather than March and April, previously detected indexes the effect of being previously detected at a specific detection station, p_0 represents the probability of detecting an individual at its activity center, σ represents detection probability decay with distance, and φ represents the probability of a specific detection type occurring conditional on any detection at all.

[†]Estimates are presented as effects on the logit scale.

[‡]95% credible intervals for differences in sex-specific parameters do not overlap 0.

[§] σ is estimated on the real scale in kilometer units.

[¶]Individuals per 100 km².

objective as well as misconstrue random sampling variation with negative distribution or abundance trends (e.g., Gibbs et al. 1999). These risks are pointed within our own study system, as our estimates corroborate previous predictions of local marten population decline (Payer 1999, Simons 2009), increasing the urgency of implementing monitoring practices and considering alterations to management practices.

Of course, such bias would be a non-issue if there were no justification for pairing different survey methods or practical constraints precluding maintaining inter-method independence, the second advantage provided by our reformulation. We had distinct motivations for pairing cameras and hair catches: a desire to limit costs associated with laboratory processing of non-target hair samples, and a desire to externally evaluate whether robust photographic identification of marten individuals was possible. Cameras allowed us to remove more than 50% of hair samples from lab analysis, and the savings essentially recouped our initial camera investment. In turn, the observed agreement between genetic and photographic identifications provided additional assurance that martens could be reliably photographically distinguished, an external validation that is largely absent from the vast majority of photographic capture–

recapture efforts (Foster and Harmsen 2012). We also had distinct constraints that limited our ability to maintain spatial or temporal independence between the different candidate field methods we deployed: most pertinently, financial constraints required us to reduce sampling extent, intensity, or duration. Even without these considerations, maintaining methodological independence may have increased the risk that differences in sampling efficacy would have been confounded with ecological differences (e.g., Graves et al. 2012). We expect that our use of bait would have limited the risk of local spatial environmental heterogeneity confounding method efficacy, as martens will willingly use many otherwise risky or low quality habitats if there is a food reward (Moriarty et al. 2016). On the other hand, had we operated distinct methods during different time intervals, our interpretation of the relative sampling efficacy of different methods may have been strongly compromised by the strong seasonal changes in detection probability we observed. Lengthening our survey duration to achieve the same spatial coverage may have further exacerbated confounding of method-specific sampling efficacy with marten population or behavioral changes, because our sampling period would then have overlapped with either commercial fur-harvest or parturition (Powell et al. 2003). In essence, had we maintained independence between our field methods, our survey effort may have been both more expensive and less effective at evaluating field methods and estimating population parameters.

The benefits associated with implementing multi-method models in any fashion depend on there being measurable value in using or considering multiple sampling methods. External constraints often motivate practitioners to evaluate trade-offs between the costs and efficacy of different field techniques (e.g., Long et al. 2007a), and this was one objective of our field effort. On the basis of cost-efficiency alone, we may have identified snow tracking as the most efficient means for monitoring changes in marten occupancy, and remote cameras as the most efficient means for estimating marten density. But more broadly, different field methods provide different lenses through which ecological patterns are imperfectly sampled, and method-specific foibles can lead to distinct biases that can be difficult to evaluate without using multiple methods, and can change the calculus associated with survey design. We leveraged our multi-method approach to evaluate *a priori* concerns that snow tracking could produce false-positive marten detections, that inability to distinguish marten sex photographically could negatively bias camera-based density estimates, and, as discussed above, that photographic marten identification could be prone to misclassification. This more nuanced evaluation paid strong dividends. Despite the affordability of snow tracking, evidence that many snow-track detections were false positives reinforces previous warnings that exclusive reliance on inexpensive ephemeral detection methods to

survey organisms sympatric to very similar species can carry substantial risk of overestimating distribution or mischaracterizing trends (Miller et al. 2015). Robust use of snow tracking or other ephemeral sign as exclusive survey techniques may generally be contingent on implementing more conservative protocols to avoid misclassification, or using model frameworks that can account for false-positive error by leveraging additional information such as experimental assessments of observer error or ancillary validation of plastered or photographed tracks (Halfpenny et al. 1995, Chambert et al. 2015, Ruiz-Gutierrez et al. 2016).

In contrast, cameras provided a greater probability of detecting individual martens than hair snares with reduced long-term cost, and similarity between density estimates derived from camera detections only and a combination of genetic and photographic detections suggests photographic inability to distinguish marten sex was not severely biasing density estimates. This may be because, although male martens used more overall space than female martens, they were detected less frequently at specific locations, a compensatory heterogeneity in detection that is not expected to lead to serious bias in density estimates (Efford and Mowat 2014). Consequently, using exclusively cameras for marten capture-recapture studies (Siren et al. 2016) may be a cost-effective alternative to more prevalent genetic capture-recapture techniques.

Other benefits provided by using multiple field methods may be serendipitous. The ability to observe hair deposition on camera revealed potential limitations of our hair-catch implementation. The first marten to deposit hair while visiting a station often appeared to rip the staple holding one part of the glue-strip off of the bait structure, precluding deposition by subsequent martens. This limits individual cross-contamination (Pauli et al. 2008), however, it is problematic because it implies a type of detection competition between individuals recognized as leading to underestimating density using SCR in live-trapping studies (Distiller and Borchers 2015). In fact, this potential bias might be stronger when sampling with single-catch hair detectors rather than single-catch live traps because one individual can prevent the detection of other individuals at multiple locations, and our hair-based estimates of marten density were notably smaller than estimates produced by cameras only or multiple methods. Consequently, we caution that density estimates derived from exclusive use of single-catch hair detectors may not be robust. To a lesser degree, cameras delivered an additional unanticipated benefit by providing direct information about station visitation by fishers, a strongly suspected cause of misclassification error. The effect of fisher detection by camera on track false-positive probability was not precisely estimated, perhaps due to infrequent fisher detection, but modeling variation in false-positive detection probability can improve occupancy estimates (Miller et al. 2015). Observing fishers on camera, and not martens,

when marten tracks were detected also provided strong evidence of false-positive error independent of parameter estimates.

However, the method-specific concerns we identified should not over-shadow the value of the information each technique provided. For example, although some snow-tracks appeared to be false-positive detection, many detections were true positive, providing additional information regarding marten occurrence that improved the precision of occupancy estimates and required negligible additional cost or effort. Many surveys for different taxa may permit the collection of opportunistic but potentially unreliable detection information such as tracks or calls, and the framework we describe explicitly enables practitioners to include this information within models while accounting for the risk that this opportunistic information is unreliable. We can think of few means by which a survey effort can increase the probability of detecting target species and improve the precision of state parameter estimates that requires less expenditure. Similarly, hair catches were less likely to detect martens than cameras and there was no compelling evidence that ignoring sex-specific differences in marten space use biased density estimates. Yet, continued validation of photographic identification of individual martens could prove extremely useful if the analyzing observers changed, and the genetic and isotopic information contained in hair samples provides considerable potential to evaluate more targeted reasons for variability in marten population parameters (e.g., Schwartz et al. 2007). Our paired-method design appears to allow this information to be collected with substantial cost savings.

Instead, a broader epiphany provided by this analysis was that there was no compelling inferential or logistical reason to monitor martens using only cameras or snow tracks, or estimate marten density using only either cameras or hair snares. Admittedly, the synergistic benefits of using different field methods may not always outweigh the financial costs of doing so. However, pairing multiple methods and using models that appropriately accommodate covariance between these methods provides practitioners the potential to improve understanding of ecological parameters and how well these parameters are sampled while simultaneously limiting the investment that could be lost. No information comes for free in wildlife sampling, but collocating methods is a practical approach to providing additional information for limited expenditure.

ACKNOWLEDGMENTS

Funding was provided by the Maine Department of Inland Fisheries and Wildlife (MDIFW) through the Federal Aid in Wildlife Restoration Act (award # 301331); at the time of this publication, the data have not been published by MDIFW. Additional support was provided by the University of Maine and MDIFW through the Cooperative Agreement with the U.S. Geological Survey, Maine Cooperative Fish and Wildlife

Research Unit, and J. Clare was partially supported by NESSF #NNX16AO61H during manuscript preparation. This research was approved by University of Maine Institutional Animal and Care Use Committee (protocol #A2014-01-03). Mention of trade names and commercial products does not constitute endorsement or recommendation for use by the U.S. Government. R. Canarr, J. Eddy, B. Peterson, and L. Skora provided field assistance; W. Jakubas assisted with logistics, improved research focus, and provided helpful review; A. Siren provided insight into detection station design; J. Hoekwater, J. Bissell, R. Morrell, E. Syiptowski, B. Burgason, A. Cowperwaite, and M. McKeague graciously assisted with logistics and land access. M. Schwartz and K. Pilgrim graciously implemented laboratory methods to determine marten genotypes. D. Linden, D. Miller, and three anonymous reviewers provided helpful commentary on previous drafts.

LITERATURE CITED

- Borchers, D. L., and M. G. Efford. 2008. Spatially-explicit maximum likelihood methods for capture–recapture studies. *Biometrics* 64:377–385.
- Buskirk, S. W., L. F. Ruggiero. 1994. American marten. Pages 7–37 in L. F. Ruggiero, et al., editors. The scientific basis for conserving forest carnivores: American marten, fisher, lynx and wolverine in the western United States. General Technical Report RM-254. Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado, USA.
- Chambert, T., D. A. W. Miller, and J. D. Nichols. 2015. Modeling false positive detections in species occurrence data under different study designs. *Ecology* 96:332–339.
- Chapin, T. G., D. J. Harrison, and D. D. Katnik. 1998. Influence of landscape pattern on habitat use by American marten in an industrial forest. *Conservation Biology* 12:1327–1337.
- Distiller, G., and D. L. Borchers. 2015. A spatially explicit capture–recapture estimator for single-catch traps. *Ecology and Evolution* 5:5075–5087.
- Efford, M. G. 2014. secr: Spatially explicit capture–recapture models. <https://CRAN.R-project.org/package=secr>
- Efford, M. G., and G. Mowat. 2014. Compensatory heterogeneity in spatially explicit capture–recapture data. *Ecology* 95: 1341–1348.
- Fisher, J. T., and S. Bradbury. 2014. A multi-method hierarchical modeling approach to quantifying bias in occupancy from noninvasive genetic tagging studies. *Journal of Wildlife Management* 78:1087–1095.
- Fisher, J. T., N. Heim, S. Code, and J. Paczkowski. 2016. Grizzly Bear noninvasive genetic Tagging surveys: estimating the magnitude of missed detections. *PLoS ONE* 11:e0161055.
- Foster, R. J., and B. J. Harmsen. 2012. A critique of density estimation from camera-trap data. *Journal of Wildlife Management* 76:224–236.
- Fuller, A. K., D. W. Linden, and J. A. Royle. 2016. Management decision making for fisher populations informed by occupancy modeling. *Journal of Wildlife Management* 80:794–802.
- Gelman, A., A. Jakulin, M. Pittau, and Y. Su. 2008. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* 2:1360–1383.
- Gelman, A., X. L. Meng, and H. Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 6:733–807.
- Gelman, A., and D. B. Rubin. 1992. Inference form iterative simulation using multiple sequences. *Statistical Science* 7: 457–472.
- Gibbs, J. P., H. L. Snell, and C. A. Causton. 1999. Effective monitoring for adaptive wildlife management: lessons from the Galapagos Islands. *Journal of Wildlife Management* 63:1055–1065.
- Gompper, M. E., R. W. Kays, J. C. Ray, S. D. LaPoint, D. A. Bogan, and J. A. Cryan. 2006. A comparison of non-invasive techniques to survey carnivore communities in northeastern North America. *Wildlife Society Bulletin* 34:1142–1151.
- Gopalaswamy, A. M., J. A. Royle, M. Delampady, J. D. Nichols, K. U. Karanth, and D. W. Macdonald. 2012. Density estimation in tiger populations: combining information for strong inference. *Ecology* 93:1741–1751.
- Graves, T. A., J. A. Royle, K. C. Kendall, P. Beier, J. B. Stetz, and A. C. Macleod. 2012. Balancing precision and risk: Should multiple detection methods be analyzed separately in N-mixture models? *PLoS ONE* 7:e49410.
- Guillera-Arroita, G., J. J. Lahoz-Monfort, D. I. MacKenzie, B. A. Wintle, and M. A. McCarthy. 2014. Ignoring imperfect detection in biological surveys in dangerous: a response to ‘fitting and interpreting occupancy models’. *PLoS ONE* 9: e99571.
- Halfpenny, J. C., R. W. Thompson, S. C. Morse, T. Holden, and P. Rezendes. 1995. Snow tracking. Pages 91–163 in W. J. Zielinski and T. Kucera, editors. American marten, fisher, lynx, and wolverine: survey methods for their detection. General Technical Report PSW-GTR-157. USDA Forest Service, Pacific Southwest Research Station, Albany, California, USA.
- Katnik, D. D., D. J. Harrison, and T. P. Hodgman. 1994. Spatial relations in a harvested population of marten in Maine. *Journal of Wildlife Management* 58:600–607.
- Kellner, K. 2015. jagsUI: a wrapper around rjags to streamline JAGS analyses. github.com/kenkellner/jagsUI
- Kéry, M., and J. A. Royle. 2016. Applied hierarchical modeling in ecology, Volume 1. Academic Press, London.
- Kuo, L., and B. Mallick. 1998. Variable selection for regression models. *Sankha* 60B:65–81.
- LANDFIRE. 2015. LANDFIRE: existing vegetation cover, U.S. Department of Agriculture and U.S. Department of the Interior. www.landfire.gov/NationalProductionDescriptions_23.php
- Long, R. A., T. M. Donovan, P. MacKay, W. J. Zielinski, and J. S. Buzas. 2007a. Effectiveness of scat detection dogs for detecting forest carnivores. *Journal of Wildlife Management* 71:2007–2017.
- Long, R. A., T. M. Donovan, P. MacKay, W. J. Zielinski, and J. S. Buzas. 2007b. Comparing scat detection dogs, cameras, and hair snares for surveying carnivores. *Journal of Wildlife Management* 71:2018–2025.
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. Pollock, L. Bailey, and J. E. Hines. 2006. Occupancy estimation and modeling—inferring patterns and dynamics of species occurrence. Elsevier Publishing, London, UK.
- MacKenzie, D. I., and J. A. Royle. 2005. Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology* 42:1105–1114.
- Magloum, A. J., C. D. Long, M. K. Schwartz, K. L. Pilgrim, R. E. Lowell, and P. Valkenburg. 2011. Integrating motion-detection cameras and hair snares for wolverine identification. *Journal of Wildlife Management* 75:731–739.
- Miller, D. A., L. L. Bailey, E. H. C. Grant, B. T. McClintock, L. A. Weir, and T. R. Simons. 2015. Performance of species occurrence estimators when basic assumption are not met: a test using field data where true occupancy status is known. *Methods in Ecology and Evolution* 6:557–565.
- Miller, D. A., J. D. Nichols, B. T. McClintock, E. H. C. Grant, L. L. Bailey, and L. A. Weir. 2011. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology* 92:1422–1428.
- Moriarty, K. M., C. W. Epps, M. G. Betts, D. J. Hance, J. D. Bailey, and W. J. Zielinski. 2016. Experimental evidence that

- simplified forest structure interacts with snow cover to influence functional connectivity for Pacific martens. *Landscape Ecology* 30:1865–1877.
- Moriarty, K. M., W. J. Zielinski, and E. D. Forsman. 2011. Decline in American marten occupancy rates at Sagehen Experimental Forest, California. *Journal of Wildlife Management* 75:1774–1787.
- Nichols, J. D., L. L. Bailey, N. W. Talancy, E. H. Cambell Grant, A. T. Gilber, E. M. Annand, T. P. Husband, and J. E. Hines. 2008. Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology* 45:1321–1329.
- NOAA. 2015a. 1981–2010 Climate Normals. <https://www.ncdc.noaa.gov/cdo-web/datatools/normals>. Accessed May 30, 2015.
- NOAA. 2015b. State of the climate: national overview for February 2015. <https://www.ncdc.noaa.gov/sotc/national/201502>
- Pauli, J. N., M. B. Hamilton, E. B. Crain, and S. W. Buskirk. 2008. A single-sampling hair trap for mesocarnivores. *Journal of Wildlife Management* 72:1650–1652.
- Payer, D. C. 1999. Influences of timber harvesting and trapping on habitat selection and demographic characteristics of marten. Dissertation. University of Maine, Orono, Maine, USA.
- Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using GIBBS sampling. Pages 20–22 in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. <http://www.ci.tuwien.ac.at/Conferences/sDSC-2003/Drafts/Plummer.pdf>
- Powell, R. A., S. W. Buskirk, W. J. Zielinski. 2003. Fisher and Marten. Pages 635–649 in G. A. Feldhamer, et al., editors. *Wild mammals of North America: biology, management, and conservation*. Johns Hopkins University Press, Baltimore, Maryland, USA.
- Riddle, A. E., K. L. Pilgrim, L. S. Mills, K. S. McKelvey, and L. F. Ruggiero. 2003. Identification of mustelids using mitochondrial DNA and non-invasive sampling. *Conservation Genetics* 4:241–243.
- Robataille, J. F., and K. Aubry. 2000. Occurrence and activity of American martens in relation to roads and other routes. *Acta Theriologica* 45:137–143.
- R Development Core Team. 2014. R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. <http://www.r-project.org>. Accessed February 22, 2014.
- Royle, J. A., R. B. Chandler, R. Sollman, and B. Gardner. 2014. *Spatial capture–recapture*. Elsevier, Waltham, Massachusetts, USA.
- Royle, J. A., and R. M. Dorazio. 2008. *Hierarchical modeling and inference in ecology*. Academic Press, London, UK.
- Royle, J. A., and M. Kéry. 2007. A Bayesian state-space formulation of dynamic occupancy models. *Ecology* 88:1813–1823.
- Ruiz-Gutierrez, V., M. B. Hooten, and E. H. Campbell Grant. 2016. Uncertainty in biological monitoring: a framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution* 7:900–909.
- Schwartz, M. K., G. Luikart, and R. S. Waples. 2007. Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution* 22:25–33.
- Simons, E. M. 2009. Influences of past and future forest management on the spatiotemporal dynamics of habitat supply for Canada lynx and American martens in Northern Maine. Dissertation. University of Maine, Orono, Maine, USA.
- Siren, A. P. K. 2013. Population ecology of American marten in New Hampshire: impact of wind farm development in high elevation habitat. Thesis. University of New Hampshire, Durham, New Hampshire, USA.
- Siren, A. P. K., P. J. Pekins, P. L. Abdu, and M. J. Ducey. 2016. Identification and density estimation of American martens (*Martes americana*) using a novel camera-trap method. *Diversity* 8:art. 3.
- Sollman, R., M. M. Furtado, B. Gardner, H. Hofer, A. T. A. Jacomo, N. M. Torres, and L. Silveira. 2011. Improving density estimates for elusive carnivores: accounting for sex-specific detection and movements using spatial capture-recapture models for jaguars in central Brazil. *Biological Conservation* 144:1017–1024.
- Sollman, R., N. M. Torres, M. M. Furtado, A. T. A. Jacomo, F. Palomares, S. Roques, and L. Silveira. 2013. Combining camera-trapping and noninvasive genetic data in a spatial capture–recapture framework improves density estimates for the jaguar. *Biological Conservation* 167:242–247.
- Wasserman, T. N., S. A. Cushman, and D. O. Wallin. 2012. Multi-scale habitat relationships of *Martes americana* in northern Idaho, USA. *Rocky Mountain Research Station Research Paper RMRS-RP-94*. U.S. Department of Agriculture, Forest Service, Fort Collins, Colorado.
- Williams, B. W., D. R. Etter, D. W. Linden, K. F. Millenbah, S. R. Winterstein, and K. T. Scribner. 2009. Noninvasive hair tagging and genetic sampling of co-distributed fishers and American martens. *Journal of Wildlife Management* 73:26–34.
- Zielinski, W. J., K. M. Slauson, and A. E. Bowles. 2008. Effects of off-highway vehicle use on the American marten. *Journal of Wildlife Management* 72:1558–1571.

SUPPORTING INFORMATION

Additional supporting information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/eap.1587/full>

DATA AVAILABILITY

Data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.4mt00>