Conservation Biology

Contributed Paper

Improving inferences about private land conservation by accounting for incomplete reporting

Matthew A. Williamson,[1] Brett G. Dickson,[2,3] Mevin B. Hooten,[4] Rose A. Graves,[5] Mark N. Lubell,[6] and Mark W. Schwartz[6]

[1]Human Environment Systems, College of Innovation and Design, Boise State University, 1910 University Drive, Boise, ID, 83725, U.S.A., email mattwilliamson@boisestate.edu
[2]Conservation Science Partners, Inc., 11050 Pioneer Trail, Suite 202, Truckee, CA, 96161, U.S.A.
[3]Landscape Conservation Initiative, Northern Arizona University, P.O. Box 5694, Flagstaff, AZ, 86011, U.S.A.
[4]U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Departments of Fish, Wildlife, & Conservation Biology and Statistics, Colorado State University, 1484 Campus Delivery Ft, Collins, CO, 80521, U.S.A.
[5]Global Environmental Change Lab and The Nature Conservancy, Portland State University, 1825 SW Broadway, Portland, OR, 97201, U.S.A.
[6]Department of Environmental Science and Policy, University of California, Davis, One Shields Ave, Davis, CA, 95616, U.S.A.

Abstract: Private lands provide key habitat for imperiled species and are core components of function protected area networks; yet, their incorporation into national and regional conservation planning has been challenging. Identifying locations where private landowners are likely to participate in conservation initiatives can help avoid conflict and clarify trade-offs between ecological benefits and sociopolitical costs. Empirical, spatially explicit assessment of the factors associated with conservation on private land is an emerging tool for identifying future conservation opportunities. However, most data on private land conservation are voluntarily reported and incomplete, which complicates these assessments. We used a novel application of occupancy models to analyze the occurrence of conservation easements on private land. We compared multiple formulations of occupancy models with a logistic regression model to predict the locations of conservation easements based on a spatially explicit social–ecological systems framework. We combined a simulation experiment with a case study of easement data in Idaho and Montana (United States) to illustrate the utility of the occupancy framework for modeling conservation on private land. Occupancy models that explicitly accounted for variation in reporting produced estimates of predictors that were substantially less biased than estimates produced by logistic regression under all simulated conditions. Occupancy models produced estimates for the 6 predictors we evaluated in our case study that were larger in magnitude, but less certain than those produced by logistic regression. These results suggest that occupancy models result in qualitatively different inferences regarding the effects of predictors on conservation easement occurrence than logistic regression and highlight the importance of integrating variable and incomplete reporting of participation in empirical analysis of conservation initiatives. Failure to do so can lead to emphasizing the wrong social, institutional, and environmental factors that enable conservation and underestimating conservation opportunities in landscapes where social norms or institutional constraints inhibit reporting.

Keywords: conservation easements, conservation planning, nonresponse bias, occupancy models, spatial autocorrelation, spatial modeling

Resumen: La incorporación de las tierras privadas a la planeación de la conservación regional y nacional ha sido un reto a pesar de su importancia como hábitat para especies en peligro y como componentes nucleares de las redes funcionales de áreas protegidas. La identificación de las localidades en donde sea probable que los propietarios privados participen en las iniciativas de conservación puede ayudar a evitar conflictos costosos y a aclarar las compensaciones entre los beneficios ecológicos y los costos sociopolíticos. La evaluación empírica

Article impact statement: Occupancy models improve identification of private land conservation opportunities even with incompletely reported data.

Paper submitted November 18, 2019; revised manuscript accepted November 19, 2020.

Conservation Biology, Volume 00, No. 0, 1–12

DOI: 10.1111/cobi.13673

y espacialmente explícita de los factores asociados con la conservación en tierras privadas es una herramienta emergente usada para la identificación de oportunidades de conservación en el futuro. Sin embargo, la mayoría de los datos sobre la conservación en tierras privadas es reportada voluntariamente y está incompleta, lo cual complica realizar estas evaluaciones. Usamos una aplicación novedosa de los modelos de ocupación para analizar la presencia de la mitigación por conservación en tierras privadas. Comparamos diferentes formulaciones de los modelos de ocupación con un modelo de regresión logística para predecir las localidades de la mitigación por conservación con base en un marco de trabajo de un sistema socioecológico espacialmente explícito. Combinamos un experimento de simulación con un estudio de caso sobre datos de mitigación en Idaho y Montana (Estados Unidos) para ilustrar la utilidad del marco de trabajo de ocupación para el modelado de la conservación en tierras privadas. Los modelos de ocupación que consideraron explícitamente la variación en los reportes produjeron estimados de los predictores que estuvieron sustancialmente menos sesgados que los estimados producidos por la regresión logística bajo todas las condiciones simuladas. Los modelos de ocupación produjeron estimaciones para seis predictores que evaluamos en nuestro estudio de caso, los cuales fueron mayores en magnitud pero menos certeros que aquellos producidos por la regresión logística. Estos resultados sugieren que los modelos de ocupación tienen como resultado inferencias cualitativamente diferentes a la regresión logística con respecto a los efectos de los predictores sobre la presencia de mitigación por conservación y resaltan la importancia de la integración de los reportes variables e incompletos sobre la participación dentro del análisis empírico de las iniciativas de conservación. Si se falla en lo anterior se puede terminar enfatizando el factor social, institucional y ambiental equivocado que permite la conservación, además de subestimar las oportunidades de conservación en paisajes en donde las normas sociales o las restricciones institucionales inhiben el reporte de datos.

Palabras Clave: autocorrelación espacial, mitigación por conservación, modelado espacial, modelos de ocupación, planeación de la conservación, sesgo por falta de respuestas

Introduction

Regional planning is an increasingly important component of strategic conservation initiatives (Groves & Game 2016). Many regional initiatives seek to expand existing protected area networks and manage threats to ecosystem processes beyond the boundaries of parks and reserves (Guerrero et al. 2015). The need to develop conservation strategies beyond protected area borders makes privately owned lands a critical component of large-scale conservation initiatives. Private lands provide some of the last remaining habitat for numerous imperiled species (Eichenwald et al. 2020), often occur on locally rare ecosystems (Graves et al. 2019), and are critical for providing connectivity between protected areas (Bargelt et al. 2020). Translating regional plans into local action remains a challenge despite the recognition of the importance of private lands (Pressey et al. 2013). Engaging private landowners in efforts to expand the functional footprint of protected areas can be a time-consuming investment in relationship building (Fischer & Bliss 2009; Yasué et al. 2019). Spatially explicit estimates of conservation opportunity can help identify regions where investments in relationship building are most likely to lead to new conservation (i.e., conservation opportunity) and translate large-scale plans to local action for maintaining ecological structure, function, and services (Guerrero et al. 2015; Guerrero & Wilson 2017). Data sets depicting social, institutional, and environmental conditions coupled with databases of private land conservation can facilitate empirical estimation of conservation opportunity across broad geographies (Williamson et al. 2018).

Analyses of participation in conservation programs on private lands (e.g., locations of conservation easements or covenants, adoption of conservation practices, or regulatory mitigation) based on spatially explicit data sets are increasingly common (e.g., Carter et al. 2015; Baldwin & Leonard 2015; Mishra et al. 2018; Metcalf et al. 2019). Such analyses can provide valuable insight on where future conservation may be possible; however, many such analyses rely on data sets of voluntarily reported information on conservation program participation that are incomplete (Rissman et al. 2017). Nonrandom variation in reporting can induce bias in efforts to quantify the predictors of conservation occurrence (Ferraro & Pattanayak 2006; Kühn 2006; Kormos & Gifford 2014). For example, individual values related to privacy and property rights may affect the likelihood of reporting in ways that are distinct from their effects on participation in the program itself (Olmsted 2011; Arbuckle 2013). Alternatively, some locations may not be as available for conservation due to institutional attributes (e.g., participation requirements, program priorities), leading to difficulty separating unwilling landowners from unavailable landscapes (Merenlender et al. 2004). Such mistakes could be costly if they lead conservation practitioners to devote limited time and resources to working in the wrong locations or attempting to influence factors (e.g., through changes in policy or public opinion) that have little effect on where conservation occurs.

We sought to improve the utility of empirical analyses of conservation on private lands by adapting the widely used occupancy-modeling framework (MacKenzie et al. 2002, 2017; Bailey et al. 2014) for use in situations where information on participation is incomplete. We evaluated the utility of occupancy models for incompletely reported conservation by comparing naive (i.e., not accounting for variation in availability or reporting) logistic regression (NLR) with 4 formulations of occupancy models. We assessed the ability of each model to reduce 3 types of bias: nonreporting, heterogeneity in availability, and spatial autocorrelation. We combined a simulation study with an empirical example of conservation easements from Idaho and Montana (United States) to compare bias in regression coefficients produced by occupancy models with those produced by hierarchical logistic regression; consider the impacts of those biases on inferences regarding the predictors of conservation action in Idaho and Montana; and determine the implications of those biases for identifying locations for future conservation action.

Methods

We used conservation easements to motivate our evaluat the robustness of different formulations of occupancy models to violations of assumptions induced by voluntary reporting of private land conservation. Conservation easements—agreements in which a landowner agrees to limit land use in exchange for direct payments or reduced tax burdens (Cheever & McLaughlin 2014)—are a useful example for 3 reasons. First, they are a common strategy for land protection in the United States, Europe, and Australia (Cheever & McLaughlin 2014; Kamal et al. 2015). Second, easement holders report the locations of many easements in the United States in the National Conservation Easement Database (https://www.conservationeasement.us). Finally, reporting to this database is voluntary and hence incomplete (current estimates suggest the database is approximately 60% complete [https://www.conservationease ment.us]).

Nested Subsamples and Occupancy Estimation

Incomplete geospatial data on conservation easements can be considered analogous to imperfect detection of wildlife in population monitoring and habitat modeling. Occupancy models use repeated surveys to estimate the probability that a site is truly occupied by a species while accounting for imperfect detection (i.e., reporting [Hoeting et al. 2000; MacKenzie et al. 2002, 2017]). Occupancy—the probability that an object of interest (either a species or a conservation action) is present at a sampling location—is modeled as the result of a state process that describes where the object occurs and an observation process that describes how the object is detected at each sample location (MacKenzie et al. 2002; Guillera-Arroita et al. 2014).

We adapted the single-season occupancy model of MacKenzie et al. (2002) to the nested geographies of the U.S. Census (Fig. 1). We relied on subsampling within a spatial unit rather than temporally repeated visits to obtain repeated surveys of the primary sample location (i.e., space-for-time substitution [Kendall & White 2009; Pavlacky et al. 2012; Crosby & Porter 2018]) (Appendix S1). We considered a single U.S. Census tract (defined as a relatively small subdivision of a county containing 1200–8000 people [Fig. 1]) the primary sample unit. In each primary unit (i.e., census tract) there are 1, …, J block groups (a fixed area smaller than a tract). These block groups served as nested subunits analogous to those used in broad-scale wildlife monitoring efforts. The observed absence of an easement in the primary sample unit (tract) may be because there are truly no easements or social–ecological conditions in the tract preclude an easement (i.e., absent or unoccupied); landcover or institutional arrangements preclude an easement regardless of social–ecological conditions (i.e., unavailable); or easements are present, but easement holders have chosen not to report them (i.e., present, but unreported). Repeated spatial subsamples (of the block groups within tracts) helped distinguish between these conditions (Fig. 1 & Appendix S1). Although we used U.S. Census geographic units, nested geographies are used in many population enumeration programs globally, and our approach would be applicable in any situation where a spatial unit can be divided into coherent spatial subunits.

Simulation Study

We simulated the data-generating process depicted in Fig. 1 for easements using U.S. Census tracts as the primary spatial unit and block groups as the spatial subunit. We used these geographies to simulate the true occurrence state for the tract (z_i) and observed data for the block group (y_{ij}) according to

$$y_{ij} \sim \begin{cases} \mathrm{Bern}(p_{ij}) & z_i = 1 \text{ and } v_{ij} = 1 \\ 0, & z_i = 0 \text{ or } v_{ij} = 0 \end{cases}, \tag{1}$$

$$v_{ij} \sim \begin{cases} \mathrm{Bern}(\alpha_{ij}) & z_i = 1 \\ 0, & z_i = 0 \end{cases}, \tag{2}$$

$$z_i \sim \mathrm{Bern}(\psi_i), \tag{3}$$

where y_{ij} is the observed occupancy state (i.e., reported present or absent) of block group j in tract i; z_i is the true (but unobservable) occupancy state of tract i; p_{ij} is the

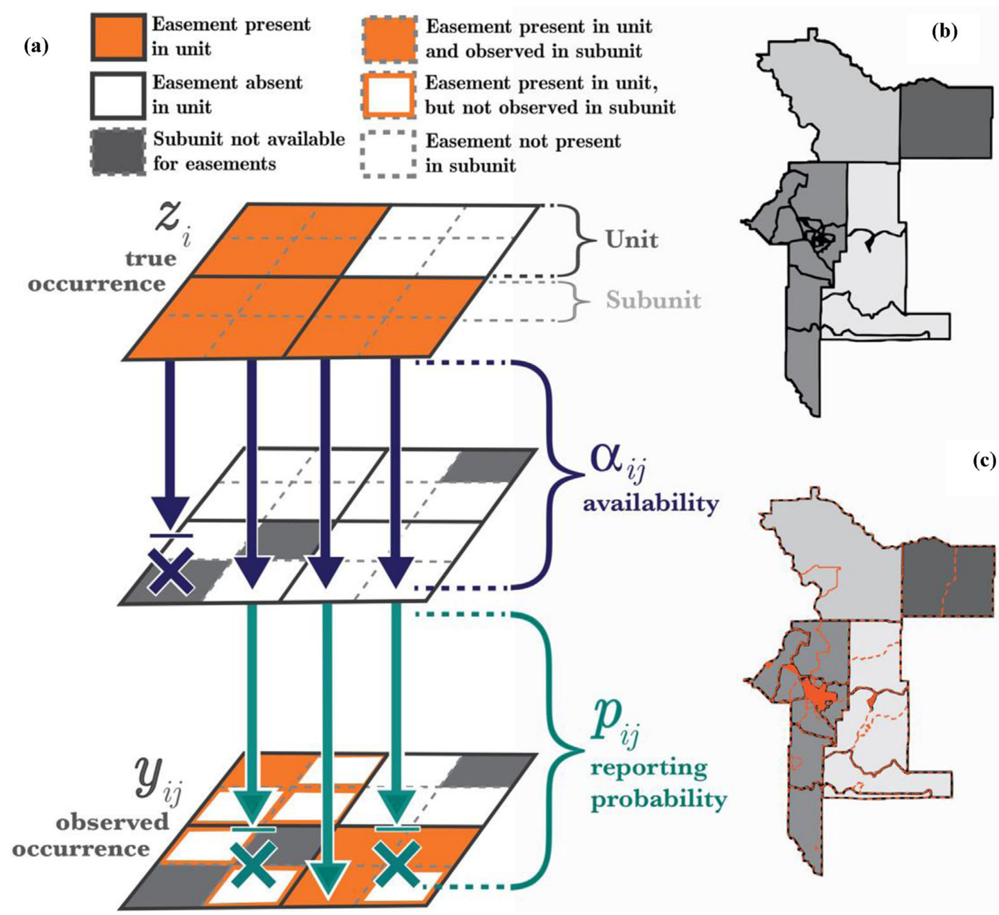

Figure 1. (a) Use of spatial subsampling to determine the probability of easement occurrence (arrows, direction of data-generating process; X, condition where availability [a_{ij}] or reporting probability [p_{ij}] prevents observation [y_{ij}] of an easement that is present). True easement occurrence (z_i) in each spatial unit (here, U.S. Census tracts, [c]) is only partially observable. Spatial subunits (here, U.S. Census block groups, orange-dashed lines in [c]) are treated as repeated visits in the occupancy framework, wherein data observed (y_{ij}) are a function of the true occupancy state for the tract (z_i), the probability that some of the block groups are unavailable for easements (a_{ij}), and the probability that an easement is reported by the easement holder (p_{ij}).

Table 1. Parameters used to model conservation easement occurrence under naive logistic regression (NLR), single-season occupancy model without a conditional autoregressive term (OCC), single-season occupancy model with a conditional autoregressive (CAR) term on both occurrence and detection (OCC-CAR1), single-season occupancy model with a CAR term for occurrence only (OCC-CAR2), and single-season occupancy model with a CAR term on reporting only (OCC-CAR3).

Symbol	*Description*
y_{ij}	Observed presence or absence of an easement for $tract_i$ and $block\ group_j$
v_{ij}	Binary indicator of whether $block\ group_j$ in $tract_i$ is available for an easement
z_i	True (but unobserved) occupancy state for easements in $tract_i$
p_{ij}	Probability of observing an easement in $block\ group_j$ within $tract_i$
α_{ij}	Probability that $block\ group_j$ in $tract_i$ is available for an easement
ψ_i	Probability of easement occurrence for $tract_i$
γ_0, β_0	Average probability of occupancy or reporting, respectively, in the absence of predictor effects (i.e., the intercept)
$\boldsymbol{\gamma}, \boldsymbol{\beta}$	Vector of regression coefficients for predictors of occupancy and reporting, respectively
$\mathbf{x}_i^{'}, \mathbf{w}_{ij}^{'}$	Row vector of predictor values describing occurrence within $tract_i$ and reporting in $block\ group_j$, respectively
η_i, ϕ_{ij}	Conditional autoregressive terms accounting for residual spatial autocorrelation in occurrence and reporting, respectively

probability that an easement is reported; $v_{ij} = 1$ if block group j is available for conservation (with probability α_{ij}) and 0 if not (because not all portions of a census tract may be available for conservation); and ψ_i is the probability of conservation occupancy in tract i (Fig. 1; see Table 1 for a summary of notation). We simulated values for ψ_i and p_{ij} according to

$$\text{logit}(\psi_i) = \boldsymbol{\gamma}_0 + \mathbf{x}'_i\boldsymbol{\gamma} + \eta_i \quad (4)$$

$$\text{logit}(p_{ij}) = \boldsymbol{\beta}_0 + \mathbf{w}'_{ij}\boldsymbol{\beta} + \phi_{ij} \quad (5)$$

where γ_0 is the mean occupancy probability (across all tracts); $\mathbf{x}'_i$ is a vector of spatially varying predictors that describes the occurrence process selected to reflect the hypotheses of interest; and $\boldsymbol{\gamma}$ is a vector of regression coefficients relating the predictors to occurrence. Similarly, β_0 is the mean reporting probability (across all block groups); $\mathbf{w}'_{ij}$ is a vector of spatially varying predictors describing the reporting process; and $\boldsymbol{\beta}$ is a vector of regression coefficients relating the predictors to the probability of reporting (Table 1). We simulated spatial autocorrelation in the occurrence and reporting process by including conditional autoregressive (CAR), latent spatial error terms, η_i and ϕ_{ij}:

$$\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{tracts}}), \quad (6)$$

$$\boldsymbol{\Sigma}_{\text{tracts}} = \left[\tau(\mathbf{D}_{\text{tracts}} - \rho_\psi \mathbf{W}_{\text{tracts}})\right]^{-1}, \quad (7)$$

and

$$\boldsymbol{\phi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{block groups}}), \quad (8)$$

$$\boldsymbol{\Sigma}_{\text{block groups}} = \left[\tau(\mathbf{D}_{\text{block groups}} - \rho_p \mathbf{W}_{\text{block groups}})\right]^{-1}, \quad (9)$$

where τ is a spatially varying precision parameter; $\mathbf{D}$ is a diagonal matrix depicting the number of neighbors for a given location based on either tracts or block groups; $\mathbf{W}$ is the neighbor matrix (based on tracts or block groups); and ρ is the strength of spatial dependence in occurrence (ψ) or reporting (p) (Ver Hoef et al. 2018).

We used a Latin hypercube design (Gentle 2006) to simulate values for each parameter. Ten replicates of 300 uniformly distributed samples (3000 total samples) were generated across a gradient defined by average occupancy probability (γ_0, range: 0.2–0.8), reporting probability (β_0, range: 0.3–0.98), spatial dependence of occupancy (ρ_ψ, range: 0.5–0.999), spatial dependence of reporting (ρ_p, range: 0.5–0.999), precision (τ, range: 0.1–1.0), and the probability that a block group is available for an action (α, range: 0.2–0.8). We generated random values for 3 predictors of occurrence ($\boldsymbol{\gamma}$) and 2 predictors of reporting ($\boldsymbol{\beta}$) because we assumed many studies of this type investigate the strength and relative importance of different predictors on conservation occurrence. We simulated predictor data from $x \sim N(0, [\tau(\mathbf{D} - \rho\mathbf{W})]^{-1})$, where τ, the spatially varying precision parameter, is 1; ρ, the strength of spatial dependence, is 0.3, $\mathbf{D}$ is a diagonal matrix containing the number of neighbors for a given location, and $\mathbf{W}$ is an adjacency matrix. We defined *adjacency* by determining the minimum distance (in meters) necessary to ensure that all locations had at least 1 neighbor and considered locations adjacent if they were within that distance from each other. We generated values for the predictors of detection, w, according to a similar structure, with adjacency based on block groups rather than tracts. Values for ψ_i and p_{ij} were then calculated based on Eqs. 4 and 5, respectively, and used to generate a series of observed data values following Eqs. 1–3 (where α is determined by a sample from the Latin hypercube).

Fitting Models

Characterizing the strength and importance of different predictors is likely to be the focus of many studies of partially reported conservation behaviors. We evaluated the ability of an NLR, standard occupancy model (OCC), occupancy model with spatial autocorrelation in both occurrence and reporting (OCC-CAR1), occupancy model with spatial autocorrelation in occurrence (OCC-CAR2), and an occupancy model with spatial autocorrelation in reporting (OCC-CAR3) to estimate the true regression coefficient values for each simulated data set (Table 2). All models were fit with Stan, a Bayesian modeling platform that implements the No U-turn Hamiltonian Monte Carlo Sampler in R (R Core Team 2020) via Rstan (Carpenter et al. 2017; Stan Development Team 2020; adaptation parameter = 0.98, maximum tree depth = 16, chains = 4, warm-up = 3200, iterations = 3700; all code available at Williamson 2021). We calculated the relative bias for all parameters as

$$\text{RelBias} = \frac{\hat{\beta} - \beta}{|\beta|} \quad (10)$$

Idaho and Montana Case Study

We evaluated the potential for incomplete reporting to produce substantially different inference by fitting the same models (NLR, OCC, OCC-CAR1, OCC-CAR2, and OCC-CAR3) (Table 2) to easements documented in the NCED for Idaho and Montana by comparing the regression coefficients (γ) (Table 1). Both states are home to several flagship protected areas (e.g., Yellowstone and Glacier National Parks, the Bob Marshall and Frank Church/River of No Return Wilderness Areas) and provide habitat for some of the last remaining iconic American wildlife species (e.g., wolves [*Canis lupus*] and grizzly bear [*Ursus arctos*]). Conservation easements are an increasing component of the regional conservation portfolio aimed at reducing the threat of development driven by ready access to amenities (Graves et al. 2019). Idaho and Montana also vary with respect to estimates of completeness for both public (Idaho 63%, Montana 90%) and nonprofit (Idaho 54%, Montana 100%)

Table 2. Model structures used in the comparison of naive (i.e., treating all absences as true absences) logistic regression and occupancy models with and without conditional autoregressive (CAR) terms to account for spatial autocorrelation in the occurrence and reporting process.*

Model Type	*Model Structure*	*Text Reference*
Naive hierarchical logistic regression with a CAR component on occurrence	$y_i \sim \text{Bern}(p_i),$ $\text{logit}(p_i) = \gamma_0 + \mathbf{x}_i'\boldsymbol{\gamma} + \eta_i$	NLR
Occupancy model with a CAR component on occurrence and reporting	$z_i \sim \text{Bern}(\psi_i),$ $\text{logit}(\psi_i) = \gamma_0 + \mathbf{x}_i'\boldsymbol{\gamma} + \eta_i,$ $y_{ij} \sim \begin{cases} \text{Bern}(p_{ij}) & z_i = 1 \\ 0, & z_i = 0 \end{cases},$ $\text{logit}(p_{ij}) = \beta_0 + \mathbf{w}_{ij}'\boldsymbol{\beta} + \phi ij$	OCC-CAR1
Occupancy model with a CAR component on occurrence only	$z_i \sim \text{Bern}(\psi_i),$ $\text{logit}(\psi_i) = \gamma_0 + \mathbf{x}_i'\boldsymbol{\gamma} + \eta_i,$ $y_{ij} \sim \begin{cases} \text{Bern}(p_{ij}) & z_i = 1 \\ 0, & z_i = 0 \end{cases},$ $\text{logit}(p_{ij}) = \beta_0 + \mathbf{w}_{ij}'\boldsymbol{\beta}$	OCC-CAR2
Occupancy model with a CAR component on reporting only	$z_i \sim \text{Bern}(\psi_i),$ $\text{logit}(\psi_i) = \gamma_0 + \mathbf{x}_i'\boldsymbol{\gamma},$ $y_{ij} \sim \begin{cases} \text{Bern}(p_{ij}) & z_i = 1 \\ 0, & z_i = 0 \end{cases},$ $\text{logit}(p_{ij}) = \beta_0 + \mathbf{w}_{ij}'\boldsymbol{\beta} + \phi_{ij}$	OCC-CAR3
Occupancy model without CAR	$z_i \sim \text{Bern}(\psi_i),$ $\text{logit}(\psi_i) = \gamma_0 + \mathbf{x}_i'\boldsymbol{\gamma},$ $y_{ij} \sim \begin{cases} \text{Bern}(p_{ij}) & z_i = 1 \\ 0, & z_i = 0 \end{cases},$ $\text{logit}(p_{ij}) = \beta_0 + \mathbf{w}_{ij}'\boldsymbol{\beta}$	OCC

**See Methods for a complete description of notation and Appendix S1 for priors used in the models.*

easement holders (https://www.conservationeasement.us). The data set we analyzed contained spatial boundaries for 3101 conservation easements (725 in Idaho and 2376 in Montana) established from 1970 to 2017 with easements ranging in size from <1 to >46,000 ha.

We followed the spatially explicit social ecological systems (SpaSES) framework to model easement occurrence as a function of variables that describe the social, institutional, and environmental context of a location (Williamson et al. 2018). We included tract-level estimates of median income and education level as indicators of social support. These variables are often associated with a wide variety of environmental behaviors (e.g., Diamantopoulos et al. 2003; Kroetz et al. 2014) and serve as proxies for the proclivity of individuals to engage in environmental behaviors (Stern 2000). We characterized institutional complexity by including a tract-level estimate of the diversity of land use as an indicator of the number of institutional arrangements (i.e., property rights) in a tract. Institutional diversity is also a measure of walkability (i.e., the spatial proximity of multiple types of land use) (Ewing & Cervero 2010). Walkability contributes to sense of place, a factor often associated with conservation in the western United States (Halpenny 2010; Gosnell & Abrams 2011). We estimated land-use diversity by extracting the values from Theobald's (2014) land-use data set within each tract and calculating the entropy index (E_i) of land-use diversity based on the live, work, play, and shop categories therein. We characterized the environmental conditions based on tract-level estimates of the maximum rarity-weighted species richness (NatureServe 2013) to identify biodiversity hotspots. We also included a tract-level estimate of the variance of the wildness index of Aplet et al. (2000) to characterize tracts that included a mix of developed land juxtaposed with wildland. We also included the log-transformed area of the tract to reflect the hypothesis that larger tracts would coincide with larger parcels that would be more desirable for conservation easements. We did not directly assess parcel size because data for Idaho were not available. Finally, we included a county-level varying intercept because the probability of easement adoption may vary at higher organizational levels than those considered here (e.g., counties may have different incentives or priorities for securing easements). These predictors comprise $\mathbf{x}_i^{'}$ in the occurrence portion of the NLR, OCC, and OCC-CAR1-3 models (Table 2).

Easements generally involve an adjustment to the assessed property value of the parcel and must be processed by the county tax assessor before being reported to national databases. We accounted for county-level differences in reporting requirements, assessor's office staffing, or institutional capacity by including a varying intercept in the reporting component (p_{ij}) of the occupancy models (OCC-CAR1-3 and OCC) (Table 2). We included the log-transformed area of the block group to account for variation in survey effort because we assumed that size of the block group affected the likelihood that an easement occurred and was reported. We also included an estimate of the median percent impervious surface, based on 2011 estimates of the percent impervious surface in the National Land Cover Database (Homer et al. 2015), to account for differences in block group availability (v_i). Highly developed block groups are, by definition, less likely to have land available for conservation easement.

Our primary objective was to compare regression coefficients produced by the different model structures and evaluate the impact of those differences on subsequent inference. We did not attempt an exhaustive evaluation of potential predictors of conservation easements. Rather, we selected a limited number and combination of variables that are commonly used to understand people's environmental behavior and conservation actions.

Results

Simulation

Estimates of the average occupancy probability (i.e., the intercept) for the NLR (model descriptions in Table 2) model were consistently and substantially lower than the true value (median estimate of relative bias $= -4.04$ on the log-odds scale). The magnitude of this underestimate improved slightly at higher levels of reporting but was substantial at low reporting probabilities (Fig. 2). In contrast, estimates of the average occurrence probability produced by all of the occupancy models were accurate (i.e., low relative bias) across the range of reporting probabilities evaluated. The occupancy model with a CAR term on both occurrence and reporting (OCC-CAR1) produced estimates of the intercept that were more precise than the other occupancy models evaluated here (Fig. 2). We observed similar patterns in estimates for regression coefficients. The NLR-based estimates were more biased (median estimate of relative bias $= 0.47$) than those produced by the occupancy models across the range of reporting probabilities we considered (median estimate of relative bias $= 0.003$, 0.003, 0.01, and 0.01 for OCC-CAR1-3 and OCC, respectively) (Fig. 2). The NLR also overestimated the magnitude of the regression coefficients across all reporting probabilities evaluated (Fig. 2); however, the direction of the bias varied with the sign of the true value (Appendix S1). Occupancy models generally produced unbiased estimates across all reporting probabilities regardless of model type (Fig. 2).

Estimates of both the intercept and regression coefficients for NLR models were biased across all values of availability. Results generally followed the same pattern as those for reporting probability (Figs. 2 & 3). All occupancy models produced unbiased estimates of the intercept and regression coefficients across all availability probabilities. Occupancy models with the CAR component on the occurrence and reporting process (OCC-CAR1) and with a CAR component on the occurrence process only (OCC-CAR2) produced more precise estimates of the intercept (Fig. 2). There was no discernible difference among the different occupancy models' ability to estimate regression coefficients. Variation in the probability that a location was available did lead to underestimates of the average reporting probability and regression coefficient estimates for the predictors of the probability of reporting (Appendix S1).

Inclusion of the CAR term in the NLR model did not reduce bias for estimates of the intercept or regression coefficients even at high values of $\rho_{\text{occurrence}}$ (Figs. 2 and 3). Bias for occupancy models (OCC, OCC-CAR1-3) was generally low across the range of spatial autocorrelation we simulated. Explicitly accounting for spatial autocorrelation with a CAR component increased the variance of estimates for the intercept and regression coefficients relative to the occupancy model without the CAR component, suggesting that the OCC model may have overestimated precision when spatial autocorrelation was present (Appendix S1).

Idaho and Montana Case Study

Regression coefficients differed substantially between NLR and the occupancy models for the predictors in the Idaho and Montana case study (Figs. 3 & 4). Posterior distributions of the coefficient estimates were narrower and closer to 0 for NLR compared with those produced by the various occupancy models. In addition, NLR produced negative coefficient estimates for the maximum value of rarity-weighted richness and variation in wilderness character, whereas occupancy models indicated a positive (though uncertain) association with these predictors. The inclusion of CAR terms increased the uncertainty of parameter estimates produced by the

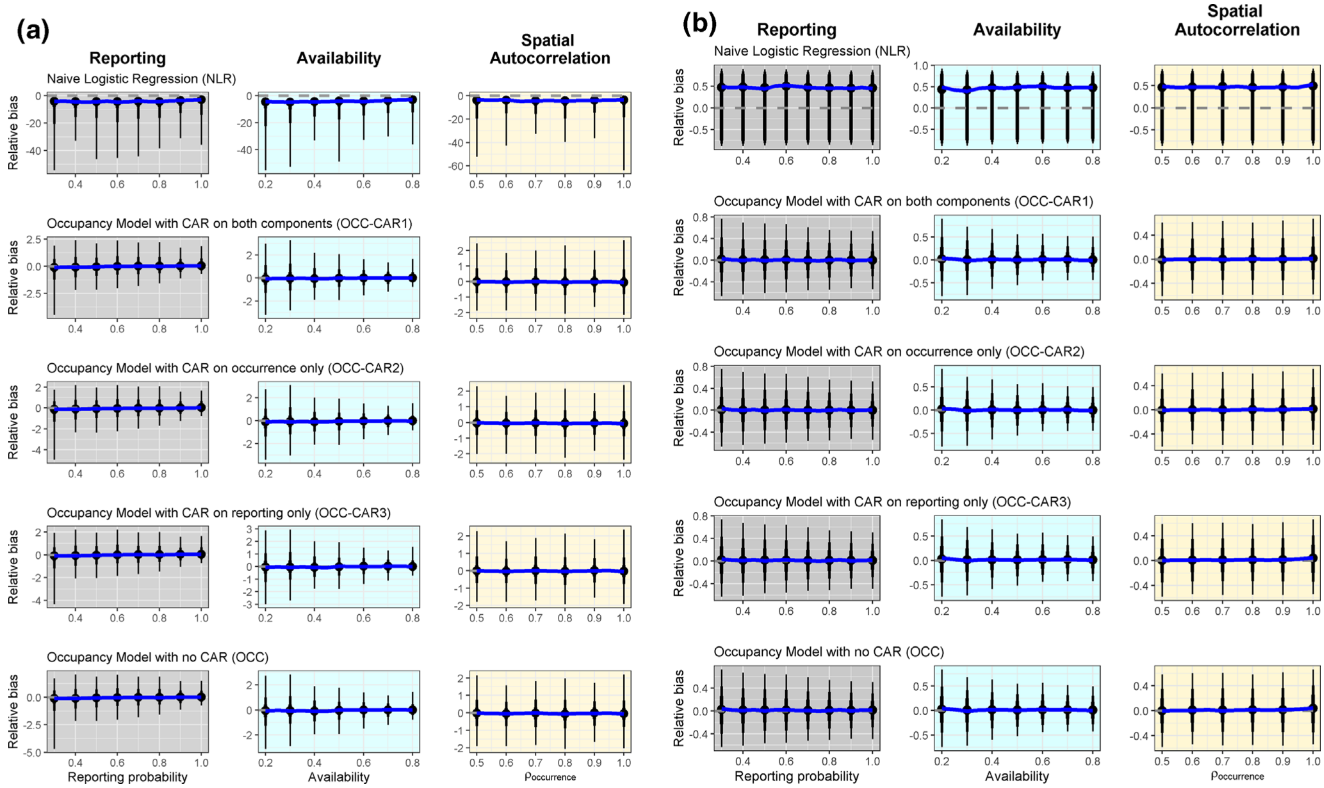

Figure 2. Relative bias of the posterior estimates of the (a) average occurrence probability (i.e., the intercept, γ_0) and (b) regression coefficients for predictors (γ) related to occurrence in the naive logistic regression model and the occupancy component of the occupancy models for each simulated data set (blue lines, relationship between median values of relative bias and reporting probability, probability of availability, and strength of spatial autocorrelation [$\rho_{\text{occurrence}}$]; vertical line ranges, 50th [thickest portion], 75th, and 90th [thinnest portion] percentile estimates of relative bias estimated by grouping all simulations into the nearest 0.1 of the simulated value).

occupancy models (OCC-CAR1-3); the largest increases occurred in models with the CAR component on reporting probability (OCC-CAR1 and OCC-CAR3). Further, inclusion of the CAR component for detection probability led to point estimates that differed substantially from the other occupancy models for the effect of species richness and income.

Although our investigation of predictors was not exhaustive, the models we evaluated indicated that easements were positively associated with the size of the tract and education level (Fig. 3). Occupancy models also indicated a positive, but uncertain relation with maximum value of rarity-weighted richness and land-use diversity. Finally, occupancy models indicated a more negative association with median income than those produced by NLR (Fig. 3).

The marginal effects of 2 of the predictors (maximum value of rarity-weighted richness and education level) further illustrated the differences in inference that arose from the different model structures (Fig. 4). Occupancy models indicated that small increases in rarity-weighted richness led to increases in the probability of easement occurrence, whereas the NLR model indicated that rarity-weighted richness had little effect. In contrast, NLR estimates for the effect of education (i.e., percent with college degrees) indicated a slight increase in occurrence probability, whereas occupancy estimates of occurrence probability indicated a potential threshold where slight changes in the percentage of college-educated individuals produced steep increases in the probability of easement occurrence. More importantly, NLR produced lower and less variable estimates of the average occupancy probability (i.e., intercepts) than OCC-CAR1 for all counties, resulting in different interpretations of the importance of each predictor and different spatial outcomes for the overall probability of easement occurrence (Fig. 4).

Discussion

Private lands contain some of the last remaining habitat for imperiled species, provide linkages between existing protected areas, and provide important contributions to conservation at multiple scales (Graves et al. 2019; Eichenwald et al. 2020; Bargelt et al. 2020).

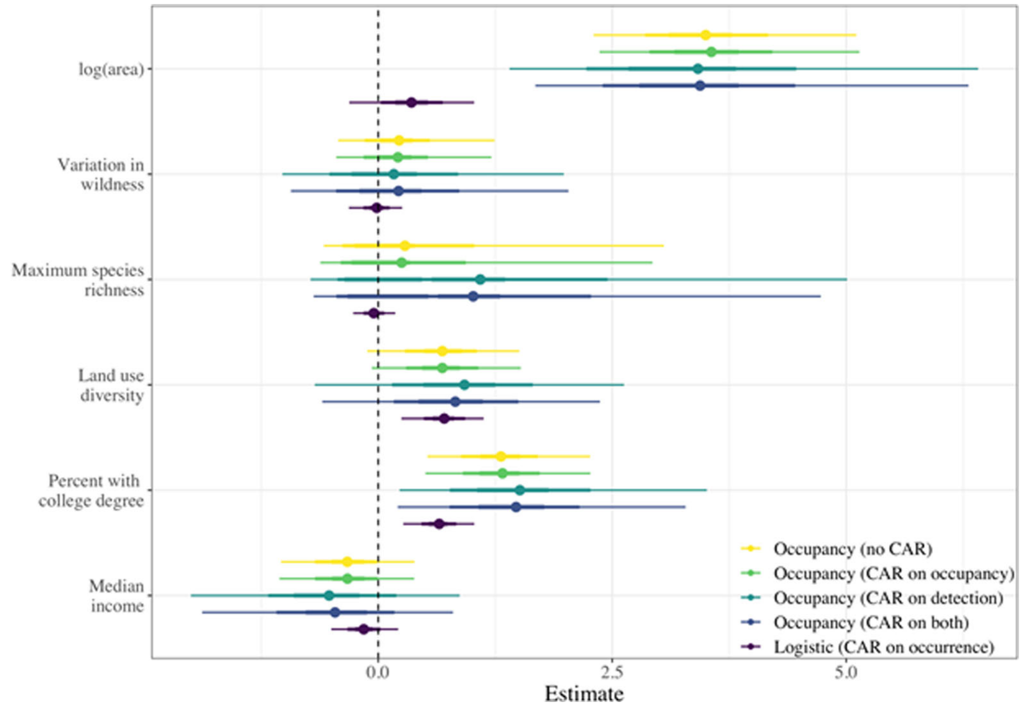

Figure 3. Regression coefficient estimates for models of easement occurrence in Idaho and Montana (United States) based on easement boundaries reported in the National Conservation Easement Database. Predictors include median income (US$), percentage of the population with a college degree, diversity of land use (based on Theobald 2014), maximum rarity-weighted richness (from NatureServe [2013]), variation (SD) of the wildness index (described in Aplet et al. [2000]), and the log-transformed area within a tract (darkest dots, median; horizontal line ranges, 50th [thickest portion], 75th, and 90th [thinnest portion] percentile estimates of the posterior distribution). Estimates are presented for the logistic regression (NLR), single-season occupancy model with a conditional autoregressive (CAR) term on both occurrence and detection, single-season occupancy model with a CAR term on detection only, single-season occupancy model with a CAR term on occurrence only, and single-season occupancy model without CAR terms.

Increased emphasis on strategic selection of private lands to achieve conservation goals requires spatially explicit characterizations of conservation opportunity (i.e., the probability that conservation will occur) on private lands. Although there has been an increasing interest in exploring how social and ecological interactions affect the spatial arrangement of conservation actions (e.g., Carter et al. 2015; Baldwin & Leonard 2015; Williamson et al. 2018), a number of conceptual and methodological hurdles exist to ensure that conservation practitioners do not waste limited resources on strategies based on spurious models (Carter et al. 2020). We have highlighted the potential for incomplete spatial data to bias empirical evaluations of previous conservation actions that may lead conservation planners to incorrectly target locations and interventions. We have also demonstrated a method that may help improve estimates of conservation opportunity on private land based on the widely applied occupancy framework.

Our simulation results suggest 3 key points. First, model structures that explicitly account for incomplete reporting (i.e., occupancy models) resulted in substantially less bias in both estimates for average occurrence probability and regression coefficients for occurrence than those produced by NLR. The consistent underestimation of the average occurrence probability by NLR seems particularly problematic because conservation practitioners may be underestimating their chances of success simply based on incomplete reporting. Further, regression coefficients for occurrence and estimates of occupancy probability were robust despite unmodeled heterogeneity in the reporting process (i.e., availability). This suggests that occupancy models can be used to evaluate the predictors of easement occurrence even when the characteristics that determine availability (e.g., variation in priorities among different land trusts) are not entirely known. Finally, social norms may affect participation in conservation behaviors and incentives to report those behaviors. For example, the finding that the presence of a conservation easement increases the likelihood of easements on neighboring parcels suggests that social processes can induce spatial autocorrelation (Lawley & Yang 2015). The presence of norms violates occupancy model assumptions by inducing autocorrelation in

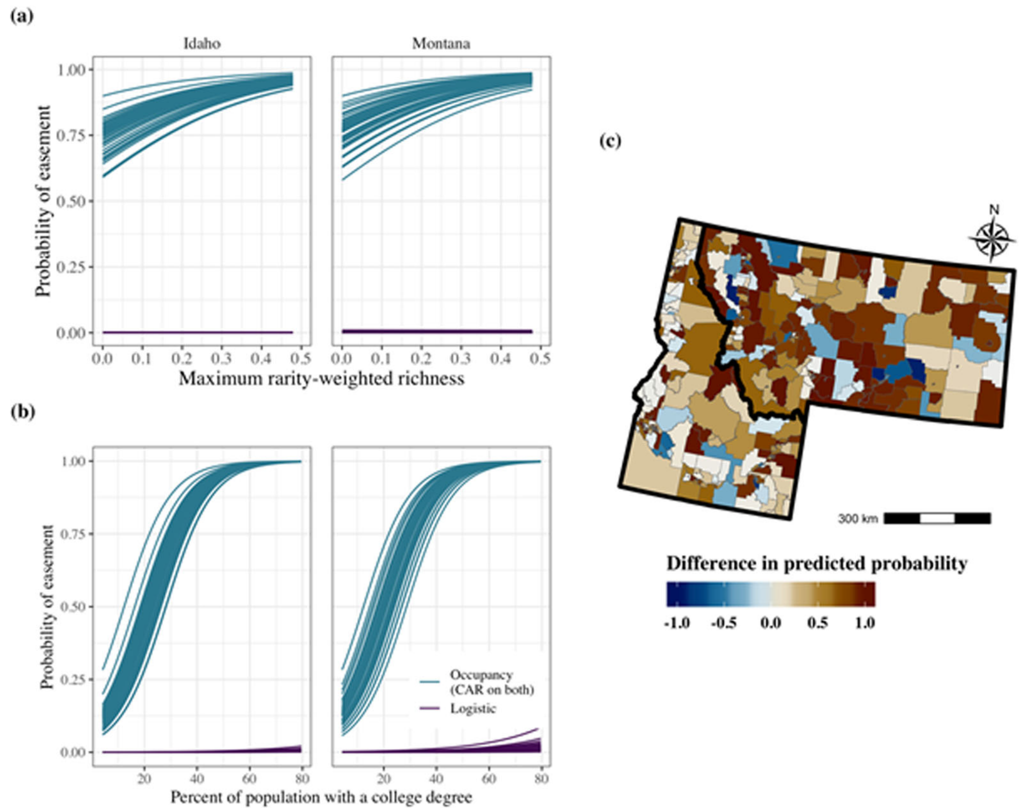

Figure 4. (a) Marginal effects of the maximum value of rarity-weighted richness and (b) percentage of residents with a college degree on easement occurrence in Idaho and Montana (United States) (lines depict the effect of the predictor on the county-level probability of adoption based on the median posterior estimate for each county-level intercept and the median of the regression coefficient for the predictor while holding all other values at their mean). Results are shown for the naive logistic regression (purple) and the occupancy model with conditional autoregressive (CAR) components for both occurrence and reporting (blue). (c) Spatial implications of the different model formulations (the warmer the color, the higher the predicted probability of easement occurrence for the occupancy model with CAR components relative to the naive logistic regression).

both the occurrence and reporting process. Explicitly incorporating autoregressive terms into occupancy models of conservation reduces the impact of violated assumptions. Given that social norms may be prevalent and difficult to model explicitly, we suggest that models of conservation occurrence should include characterization of potentially unmodeled spatial dependence in both the occurrence and detection process a priori.

The occupancy models we fitted did not distinguish between the reporting process and the availability process. Although this misspecification did not affect the regression coefficients for the occupancy component of the model (our primary interest), it did produce coefficients for the reporting process that were biased low. This suggests that these models may need further refinement if the goal is to test hypotheses about the motivations for reporting. In addition, unmodeled differences in availability represent violations of the closure assumption of occupancy models that can bias estimates of occupancy probability (Kendall et al. 2013; Otto et al. 2013). Although prudent choice of covariates may reduce this bias, additional research is necessary to identify strategies for accommodating closure violations in occupancy models that rely on the space-for-time substitution.

Our case study revealed 3 important implications of model choice in studies of private land conservation. First, the choice of model substantially affects interpretation of effects. The explicit inclusion of variation in reporting probability in the occurrence estimates resulted in stronger effects of particular variables (e.g., area, education level, and land-use diversity) than the estimates produced by logistic regression. Second, differences in point estimates produced by models including CAR components for detection suggest that autocorrelation in the reporting process may mask the effect of variables such as income or species richness. This suggests that correlations between income and easement adoption reported elsewhere (e.g., Cho et al. 2005; Baldwin & Leonard 2015) may partially reflect the existence of social norms for reporting that are tied to income. If so, this type of nonresponse bias may result in conservation strategies that overemphasize locations where income levels are high at the expense of areas high in species richness.

Our finding that easements tended to occur in locations of higher biodiversity contrasts with results for counties on the west coast of the United States (Williamson et al. 2018) and parcels in Appalachia (Fouch et al. 2019). These differences may be a function of the different resolutions of the analyses (i.e., the Modifiable Areal Unit Problem [Jelinski & Wu 1996]) or may reflect the importance of accounting for variation in the probability that a location is available for an easement prior to comparing easement locations with and without easements. Finally, and most importantly for conservation practitioners, the fact that the models generated substantially different spatial predictions suggests that conservation planners may be underestimating their ability to secure conservation gains across large portions of Idaho and Montana if their strategies are based on analyses that fail to account for variation in reporting probability.

Identifying locations where individuals may be more likely to participate in conservation can help avoid costly conflict and clarify trade-offs between ecological benefits and sociopolitical costs. Additional work is necessary to determine whether new conservation actions are better predicted by models developed on past behaviors. For models based on past actions, our results suggest that occupancy models produce regression coefficients that are substantially less biased than NLR (i.e., treating all unreported locations as absences). Moreover, our approach appears robust to variation in the probability that a location is available for an action despite the fact that we did not explicitly model the availability process. Our approach should facilitate a broader understanding of the conditions that enable conservation to occur, produce predictions that are statistically valid, and improve the alignment of conservation priorities with conservation action.

Acknowledgments

We thank the editors and 3 anonymous reviewers for their helpful contributions to this manuscript. M. Joseph provided important feedback on multiple elements of the project. We thank A. Simler-Williamson for the illustration in Fig. 1. This work was supported by the National Science Foundation Graduate Research Fellowship Program under award number 1650042, NSF Idaho EPSCoR Program under award number OIA-1757324, The Wilderness Society's Gloria Barron Scholarship, and the Center for Large Landscape Conservation's Conservation Science Fellowship to M.A.W. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government or the funders.

Supporting Information

Additional information is available online in the Supporting Information section at the end of the online article. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

Aplet G, Thomson J, Wilbert M. 2000. Indicators of wildness: using attributes of the land to assess the context of wilderness. Pages 89–98 in McCool SF, Cole DN, Borrie WT, O'Loughlin J, comps. Wilderness science in a time of change conference—volume 2: wilderness within the context of larger systems. Proceedings RMRS-P-15-VOL-2. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. Ogden, Utah.

Arbuckle JG. 2013. Farmer attitudes toward proactive targeting of agricultural conservation programs. Society & Natural Resources **26:**625–641.

Bailey LL, MacKenzie DI, Nichols JD. 2014. Advances and applications of occupancy models. Methods in Ecology and Evolution **5:**1269–1279.

Baldwin RF, Leonard PB. 2015. Interacting social and environmental predictors for the spatial distribution of conservation lands. PLoS One **10:**e0140540.

Bargelt L, Fortin M-J, Murray DL. 2020. Assessing connectivity and the contribution of private lands to protected area networks in the United States. PLoS One **15:**e0228946.

Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. 2017. Stan: a probabilistic programming language. Journal of Statistical Software **76:**1–32.

Carter N, et al. 2020. Integrated spatial analysis for human–wildlife coexistence in the American West. Environmental Research Letters **15:**021001.

Carter SK, Januchowski-Hartley SR, Pohlman JD, Bergeson TL, Pidgeon AM, Radeloff VC. 2015. An evaluation of environmental, institutional and socio-economic factors explaining successful conservation plan implementation in the north-central United States. Biological Conservation **192:**135–144.

Cheever F, McLaughlin NA. 2014. An introduction to conservation easements: a simple concept and a complicated mosaic of law. Journal of Law, Property, and Society **1:**107–186.

Cho S-H, Newman DH, Bowker JM. 2005. Measuring rural homeowners' willingness to pay for land conservation easements. Forest Policy and Economics **7:**757–770.

Crosby AD, Porter WF. 2018. A spatially explicit, multi-scale occupancy model for large-scale population monitoring. The Journal of Wildlife Management **82:**1300–1310.

Diamantopoulos A, Schlegelmilch BB, Sinkovics RR, Bohlen GM. 2003. Can socio- demographics still play a role in profiling green consumers? A review of the evidence and an empirical investigation. Journal of Business Research **56:**465–480.

Eichenwald AJ, Evans MJ, Malcom JW. 2020. US imperiled species are most vulnerable to habitat loss on private lands. Frontiers in Ecology and the Environment **18:**439–446.

Ewing R, Cervero R. 2010. Travel and the built environment. Journal of the American Planning Association **76:**265–294.

Ferraro PJ, Pattanayak SK. 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. PLoS Biology **4:**e105.

Fischer AP, Bliss JC. 2009. Framing conservation on private lands: conserving oak in Oregon's Willamette Valley. Society & Natural Resources **22:**884–900.

Fouch N, Baldwin RF, Gerard P, Dyckman C, Theobald DM. 2019. Landscape-level naturalness of conservation easements in a mixed-use matrix. Landscape Ecology **34:**1967–1987.

Gentle JE. 2006. Random number generation and Monte Carlo methods. 2nd edition. Springer, New York.

Gosnell H, Abrams J. 2011. Amenity migration: diverse conceptualizations of drivers, socioeconomic dimensions, and emerging challenges. GeoJournal **76:**303–322.

Graves RA, Williamson MA, Belote RT, Brandt JS. 2019. Quantifying the contribution of conservation easements to large-landscape conservation. Biological Conservation **232:**83–96.

Groves C, Game ET. 2016. Conservation planning: informed decisions for a healthier planet. Roberts and Company Publishers, Greenwood Village, Colorado.

Guerrero AM, Mcallister RRJ, Wilson KA. 2015. Achieving cross-scale collaboration for large scale conservation initiatives. Conservation Letters **8:**107–117.

Guerrero AM, Wilson KA. 2017. Using a social-ecological framework to inform the implementation of conservation plans. Conservation Biology **31:**290–301.

Guillera-Arroita G, Lahoz-Monfort JJ, MacKenzie DI, Wintle BA, McCarthy MA. 2014. Ignoring imperfect detection in biological surveys is dangerous: a response to 'Fitting and interpreting occupancy models'. PLoS One **9:**e99571.

Halpenny EA. 2010. Pro-environmental behaviours and park visitors: the effect of place attachment. Journal of Environmental Psychology **30:**409–421.

Hoeting JA, Leecaster M, Bowden D. 2000. An improved model for spatially correlated binary responses. Journal of Agricultural, Biological, and Environmental Statistics **5:**102–114.

Homer C, Dewitz J, Yang L, Jin S, Danielson P, Xian G, Coulston J, Herold N, Wickham J, Megown K. 2015. Completion of the 2011 National Land Cover Database for the conterminous United States representing a decade of land cover change information. Photogrammetric Engineering & Remote Sensing **81:**345–354.

Jelinski DE, Wu J. 1996. The modifiable areal unit problem and implications for landscape ecology. Landscape Ecology **11:**129–140.

Kamal S, Grodzińska-Jurczak M, Brown G. 2015. Conservation on private land: a review of global strategies with a proposed classification system. Journal of Environmental Planning and Management **58:**576–597.

Kendall WL, Hines JE, Nichols JD, Grant EHC. 2013. Relaxing the closure assumption in occupancy models: staggered arrival and departure times. Ecology **94:**610–617.

Kendall WL, White GC. 2009. A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. Journal of Applied Ecology **46:**1182–1188.

Knight AT, Cowling RM, Difford M, Campbell BM. 2010. Mapping human and social dimensions of conservation opportunity for the scheduling of conservation action on private land. Conservation Biology **24:**1348–1358.

Kormos C, Gifford R. 2014. The validity of self-report measures of proenvironmental behavior: a meta-analytic review. Journal of Environmental Psychology **40:**359–371.

Kroetz K, Sanchirico JN, Armsworth PR, Banzhaf HS. 2014. Benefits of the ballot box for species conservation. Ecology Letters **17:**294–302.

Kühn I. 2006. Incorporating spatial autocorrelation may invert observed patterns. Diversity & Distributions **13:**66–69.

Lawley C, Yang W. 2015. Spatial interactions in habitat conservation: evidence from prairie pothole easements. Journal of Environmental Economics and Management **71:**71–89.

MacKenzie DI, Nichols JD, Andrew Royle J, Pollock KH, Bailey L, Hines JE. 2017. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. 2nd edition. Academic Press, London.

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Andrew Royle J, Langtimm CA. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology **83:**2248–2255.

Merenlender AM, Huntsinger L, Guthey G, Fairfax SK. 2004. Land trusts and conservation easements: who is conserving what for whom? Conservation Biology **18:**65–76.

Metcalf AL, Phelan CN, Pallai C, Norton M, Yuhas B, Finley JC, Muth A. 2019. Microtargeting for conservation. Conservation Biology **33:**1141–1150.

Mishra B, Gyawali BR, Paudel KP, Poudyal NC, Simon MF, Dasgupta S, Antonious G. 2018. Adoption of sustainable agriculture practices among farmers in Kentucky, USA. Environmental Management **62:**1060–1072.

NatureServe. 2013. Rarity-weighted richness model for critically imperiled and imperiled (G1 or G2) species in the United States. NatureServe, Arlington, Virginia.

Olmsted JL. 2011. The invisible forest: conservation easement databases and the end of the clandestine conservation of natural lands. Law and Contemporary Problems **74:**51–82.

Otto CRV, Bailey LL, Roloff GJ. 2013. Improving species occupancy estimation when sampling violates the closure assumption. Ecography **36:**1299–1309.

Pavlacky DC, Blakesley JA, White GC, Hanni DJ, Lukacs PM. 2012. Hierarchical multi-scale occupancy estimation for monitoring wildlife populations. The Journal of Wildlife Management **76:**154–162.

Pressey RL, Mills M, Weeks R, Day JC. 2013. The plan of the day: managing the dynamic transition from regional conservation designs to local conservation actions. Biological Conservation **166:**155–169.

R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/. Accessed 1 Nov 2020.

Rissman AR, Owley J, L'Roe AW, Morris AW, Wardropper CB. 2017. Public access to spatial data on private-land conservation. Ecology and Society **22:**24.

Stan Development Team. 2020. RStan: the R interface to Stan. Available from http://mc-stan.org/. Accessed 1 Nov 2020.

Stern PC. 2000. New environmental theories: toward a coherent theory of environmentally significant behavior. Journal of Social Issues **56:**407–424.

Theobald DM. 2014. Development and applications of a comprehensive land use classification and map for the US. PLoS One **9:**e94628.

Ver Hoef JM, Peterson EE, Hooten MB, Hanks EM, Fortin M-J. 2018. Spatial autoregressive models for statistical inference from ecological data. Ecological Monographs **88:**36–59.

Williamson MA. 2020. Code for "Improving inferences about private land conservation by accounting for incomplete reporting", v1.0, Zenodo, https://doi.org/10.5281/zenodo.4570929.

Williamson MA, Schwartz MW, Lubell MN. 2018. Spatially explicit analytical models for social–ecological systems. Bioscience **68:**885–895.

Yasué M, Kirkpatrick JB, Davison A, Gilfedder L. 2019. Landowner perceptions of payments for nature conservation on private land. Environmental Management **64:**287–302.