Integrating occupancy models and structural equation models to understand species occurrence

Maxwell B. Joseph,[1,3] Daniel L. Preston,[2] and Pieter T. J. Johnson[1]

[1]*Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA*
[2]*Department of Integrative Biology, Oregon State University, Corvallis, Oregon 97331, USA*

Abstract. Understanding the drivers of species occurrence is a fundamental goal in basic and applied ecology. Occupancy models have emerged as a popular approach for inferring species occurrence because they account for problems associated with imperfect detection in field surveys. Current models, however, are limited because they assume covariates are independent (i.e., indirect effects do not occur). Here, we combined structural equation and occupancy models to investigate complex influences on species occurrence while accounting for imperfect detection. These two methods are inherently compatible because they both provide means to make inference on latent or unobserved quantities based on observed data. Our models evaluated the direct and indirect roles of cattle grazing, water chemistry, vegetation, nonnative fishes, and pond permanence on the occurrence of six pond-breeding amphibians, two of which are threatened: the California tiger salamander (*Ambystoma californiense*) and the California red-legged frog (*Rana draytonii*). While cattle had strong effects on pond vegetation and water chemistry, their overall effects on amphibian occurrence were small compared to the consistently negative effects of nonnative fish. Fish strongly reduced occurrence probabilities for four of five native amphibians, including both species of conservation concern. These results could help to identify drivers of amphibian declines and to prioritize strategies for amphibian conservation. More generally, this approach facilitates a more mechanistic representation of ideas about the causes of species distributions in space and time. As shown here, occupancy modeling and structural equation modeling are readily combined, and bring rich sets of techniques that may provide unique theoretical and applied insights into basic ecological questions.

Key words: cattle grazing; imperfect detection; indirect effects; latent variable; occupancy model; structural equation model.

Introduction

Much of ecology is concerned with explaining and predicting where species occur in space and time. While species distributions are directly and indirectly influenced by a diverse suite of abiotic and biotic factors, most statistical treatments do not formally differentiate between direct and indirect effects (Guisan and Thuiller 2005). Structural equation modeling (SEM) provides one means by which to formally represent causal assumptions in the form of direct and indirect causal pathways, helping to close the gap between biological mechanisms and statistical methodology (Bollen 1989, Grace 2006).

Structural equation models combine causal assumptions with observations to generate causal inferences (Pearl 2000). Often, these assumptions are displayed in a path diagram that represents model components and causal relationships (i.e., how one variable affects

another). These path diagrams encode both direct and indirect effects, such as those involved in complex processes that structure communities (Clough 2012, Alsterberg et al. 2013). As a result, SEM provides a framework for clarifying, representing, and evaluating hypotheses in ecology that can outperform traditional associational statistical methods. For example, the relationship between productivity and biodiversity has historically been evaluated statistically in a bivariate context, despite the multidimensionality of the hypotheses that have been put forward to explain the relationship. By explicitly representing causal relationships between productivity and diversity, along with disturbance and stressors in a SEM, the mechanistic relationship between productivity and biodiversity becomes more clear, with more predictive power than bivariate approaches (Grace et al. 2014). In addition, SEM is flexible, with recent developments facilitating nonlinear effects and a wide range of distributions, broadening the potential applicability of SEM in ecology (Lee 2007). Finally, SEM differentiates between latent and observable variables. Latent variables cannot be observed directly, but observations can provide information on

latent quantities. A classic example is the presence or absence of a species.

In practice, studies of species occurrence are notoriously plagued by imperfect detection because species may be present but unobserved. This can be problematic when trying to explain species distributions because false absences create bias in estimated coefficients for covariates that explain occurrence. For instance, if turbidity increases the probability of occurrence of an aquatic organism, but decreases detectability (i.e., the organism is harder to observe), detection events will not appear to be related to turbidity. However, adjusting for imperfect detection facilitates unbiased estimation of the true relationship between such a factor and species occurrence. Such adjustments can be made via occupancy modeling, a method in which repeat surveys are used to estimate probabilities of detection and draw inferences on true occupancy states (MacKenzie et al. 2002). Occupancy models are classic hierarchical models, where the latent binary occurrence state is partly observed, and the extra zeros in detection histories are modeled as potentially arising from a combination of true and false absences. As the appreciation and application of occupancy models has increased, they have been extended to include multiple species, temporal dynamics including colonization and extinction, estimation of species richness, habitat ephemerality, and even disease dynamics (Dorazio et al. 2010, MacKenzie et al. 2011, Miller 2012). These developments have been instrumental in improving inferences about the processes that govern species distributions. However, many existing models require independent covariates for occurrence and detection. This limitation represents a challenge to understanding even moderately complex processes that drive species distributions that may include a combination of direct and indirect effects on occurrence and detectability.

Here, we show that combining SEM with occupancy models overcomes this limitation and provides a means by which to more directly represent processes that influence species occurrence. We aim to illustrate the practical and conceptual compatibility of occupancy modeling and structural equation modeling by developing a causal model of species occurrence and community composition that explicitly accounts for non-detection. We use pond-breeding amphibians as an applied case example. First, we outline the logic behind the model by summarizing the relevant empirical knowledge for our system in terms of the effects of pond permanence, cattle grazing, nonnative fishes, water chemistry, and vegetation on amphibian occurrence. We then discuss the data collection scheme and formally relate our observations to the processes under consideration by developing observation, process, and parameter (prior) models. As a model verification step, we demonstrate the recovery of parameters from our SEM using simulation (Rykiel 1996). Last, we present results for the case study and discuss the value, limitations, and future directions for this approach.

STUDY SYSTEM

The amphibian community in freshwater ponds of the San Francisco Bay Area of California, USA includes the Pacific chorus frog (*Pseudacris regilla*), western toad (*Anaxyrus boreas*), California newt (*Taricha torosa*), California tiger salamander (*Ambystoma cailforniense*), California red-legged frog (*Rana draytonii*), and nonnative American bullfrog (*Lithobates catesbeianus*) (Stebbins 2003). Of the native species, two are federally protected and are of broad conservation concern (*A. californiense* and *R. draytonii*), and *A. boreas* has declined in specific regions (Lannoo 2005). Some populations of these species are currently threatened by invasive species including nonnative fishes and bullfrogs, which potentially act as predators and competitors for native amphibians (Fisher and Shaffer 1996, Lawler et al. 1999, Preston et al. 2012). Further, widespread cattle grazing in this region has the potential to affect the population dynamics of amphibians at breeding sites directly via trampling, and indirectly by removing vegetation and increasing the concentration of nitrogenous compounds in wetlands (Robins and Vollmar 2002, Roche et al. 2012). The ponds we considered included both temporary and permanent water bodies up to 7 m in depth.

DATA COLLECTION

From 2009 to 2013, field crews used standardized methods to assess amphibian site occupancy in 171 wetlands in the San Francisco Bay Area, USA (Contra Costa, Alameda, Santa Clara, and San Mateo counties) (Johnson et al. 2013). For simplicity, we used only one year of data from each site (with years chosen at random). The sites consisted of both artificially created livestock ponds and natural wetlands. Crews surveyed wetlands twice per summer using a combination of visual encounter, seine, and dipnet surveys (Crump and Bury 1994). Visual encounter surveys involved a single observer walking the perimeter of the wetland and recording all species observed. Dipnet sweeps were conducted every 3 m around the pond perimeter (1.4 mm mesh and 2600 cm2 net area) and three to four seine net hauls were performed in the deepest regions of each pond (4 mm mesh size and 1×2 m net area). For each survey, we recorded all life stages of each observed amphibian species. We considered sites "occupied" if larval amphibians were present, which indicates adult presence, successful fertilization, some degree of larval development, and potential for metamorphosis (i.e., breeding activity).

We quantified cattle grazing intensity, percentage of shoreline vegetated, pond permanence, water chemistry, and the presence of nonnative fish. To measure cattle grazing intensity, we recorded the number of cow paddies within 3 m of shoreline and recorded shoreline perimeter with a handheld GPS unit to quantify cow paddy density at each pond. As a second measure of grazing intensity, we made a qualitative judgment of whether
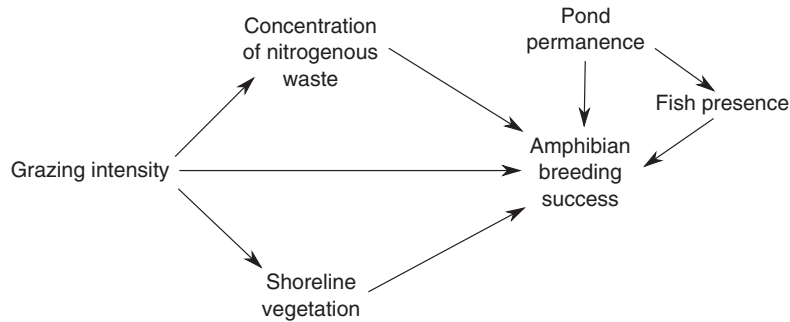
FIG. 1. Conceptual model showing influences on amphibian breeding success. Nodes represent concepts of interest, and directional arrows represent causal effects (e.g., grazing intensity affects shoreline vegetation).

the wetland was disturbed by cattle based on evidence of tracks and trampling. Water samples were collected in acid-washed Nalgene bottles, filtered, frozen, and analyzed for ammonium ($NH4^+$) and total dissolved nitrogen (TDN) concentrations using standard protocols (available online).[4] Imagery from Google Earth was used to help determine whether wetlands were permanent, based on consistent presence of water year round. Last, crews estimated the percentage of pond shoreline that was vegetated and recorded whether fish were present or absent using observations from all sampling methods. The most common fish species were nonnative, including mosquitofish (*Gambusia affinis*), largemouth bass (*Micropterus salmoides*), and bluegill sunfish (*Lepomis macrochirus*).

CONCEPTUAL MODEL

Drawing upon previous literature, we developed a multivariate hypothesis about the drivers of amphibian community composition in this system (Fig. 1). We expected fish to strongly affect community composition, particularly species with poor avoidance strategies and high palatability (Kruse and Stone 1984, Kats et al. 1988, Adams 2000, Welsh et al. 2006). Further, pond permanence should increase bullfrog occurrence because bullfrogs have a multi-year larval development period that is longer than most western native species (Collins 1979). We hypothesized that livestock would physically alter wetland ecosystems via trampling and grazing, and chemically alter wetlands via inputs of nitrogenous waste products in urine and feces (Kauffman et al. 1983, Jansen and Healey 2003, Knutson et al. 2004, Schmutzer et al. 2008, Adams et al. 2009). Previous studies indicate that characteristics of shoreline vegetation influence amphibian breeding and that water chemistry can alter reproductive success and occupancy probability (Freda and Dunson 1986, Rowe and Dunson 1995, Rouse et al. 1999, Brodman et al. 2003, Jansen and Healey 2003, Egan and Paton 2004, Burne and Griffin 2005, Earl and Whiteman 2009). Separating

[4] http://snobear.colorado.edu/Kiowa/Kiowaref/procedure.
html.

out the causal pathways through which grazing alters amphibian occupancy is important from a practical standpoint, as each one might be targeted differently with management interventions (e.g., limiting cattle-induced damage to vegetation versus limiting the total intensity of grazing).

MODEL FORMALIZATION

To formalize the model of Fig. 1, we adopted a hierarchical Bayesian approach, developing model components related to observations, processes in the system, and priors for unknown parameters (Lee 2007, Cressie et al. 2009, Dorazio et al. 2010).

Observation model

While early structural equation models assumed normally distributed observed indicator variables, current methods are extremely flexible in terms of likelihood functions and relationships between variables, broadening the applicability of these methods in ecology (Grace 2006, Lee 2007). We consider the presence/absence state of a species at a location to be a hidden (latent) binary variable which gives rise to imperfect detection/nondetection data. The use of a latent state variable is a key conceptual link between the fields of SEM and occupancy modeling. The flexibility of this approach is demonstrated in the following sections where we apply a variety of observation models from the exponential family, a moderately complex finite-mixture distribution, and a combination of continuous and discrete latent variables.

We considered grazing intensity by cattle to be a continuous latent quantity that cannot be observed directly, whose value is indicated by evidence of disturbance and the density of cow paddies in the vicinity of a wetland. Specifically, cattle disturbance was modeled as a Bernoulli random variable with latent grazing intensity (ξ) as a continuous covariate

$$Y_1[j,k] \sim \text{Bernoulli}\,(p_{Y_1}[j])$$

$$\text{logit}\,(p_{Y_1}[j]) = \beta_{Y_1,0} + \beta_{Y_1,1}\xi[j]$$

for the jth site and the kth survey, with square brackets representing indexing. Observed cow paddy counts are treated as a second indicator (specifically, a multi-method indicator) (Grace 2006) and modeled as a Poisson random variable with an offset for shoreline perimeter of site j ($\mu_{\text{perim}}[j]$) in meters:

$$Y_2[j,k] \sim \text{Poisson}(\lambda[j,k])$$

$$\frac{\lambda_{j,k}}{\mu_{\text{perim}}[j,k]} = e^{\beta_{Y_2,0} + \beta_{Y_2,1}\xi[j]}.$$

Perimeter observations are subject to measurement error and variation within a season. Therefore, we modeled latent mean perimeter values for each site (μ_{perim}), which represents the expected pond perimeter value for site j:

$$\log(\text{perim}[j,k]) \sim N(\mu_{\text{perim}}[j],\sigma_{\text{w}})$$

$$\mu_{\text{perim}}[j] \sim N(\alpha_{\text{perim}},\sigma_{\text{a}})$$

where σ_{w} represents measurement error and variation within a season, α_{perim} is the (log) average perimeter across all wetlands, and σ_{a} is the variability in perimeter among wetlands.

Shoreline vegetation, v, ranged from 0 to 100% with a non-negligble number of 0%, 100%, and intermediate observations. Therefore, we treated these observations as arising from a zero-one inflated beta distribution, which is a finite mixture distribution with a Bernoulli component that produces 0s and 1s, and a beta component that produces values on the interval (0,1) (Ospina and Ferrari 2012):

$$P(v;\alpha,\mu_v,\phi) = \begin{cases} \alpha(1-\mu_v) & v = 0 \\ \alpha\mu_v & v = 1 \\ (1-\alpha)f(v;\mu_v,\phi) & 0 < v < 1 \end{cases}$$

$$\text{logit}(\mu_v[j]) = \beta_{v,0} + \beta_{v,1}\eta_1[j]$$

$$\text{logit}(\alpha) = a_0 + a_2\mu_v^2.$$

Here, α determines the extent to which the beta or binomial mixture components dominate the probability density function. The second-degree polynomial term with coefficient a_2 causes extreme values of the logit-expected shoreline vegetation cover μ_v to increase the probability of an observer recording either 0% or 100% shoreline vegetation cover. This formulation essentially imposes a minimum probability of observers recording a discrete value when the true shoreline vegetation is 50%, and increases the probability of 0% or 100% observations as the true cover approaches those values. Last, $f(v;\mu_v,\phi)$ is the probability density function of the beta distribution, parameterized in terms of its mean (μ_v) and variance (ϕ), which represents the combination of observation error and within-summer variation in true shoreline vegetation cover (Ospina and Ferrari 2012).

Log-transformed concentrations of ammonium and total dissolved N in the water were used as multi-method indicators of latent N concentration. Because both total dissolved N and NH_4^+ concentrations tend to increase over the course of the summer due to exogenous inputs and pond drying, we also included an effect of survey number k, coded as an indicator variable for the 2nd survey, such that $I(k=2)=1$, and $I(k=1)=0$

$$\log(NH_4^+[j,k]) \sim N(\beta_{NH_4^+,0} + \beta_{NH_4^+,1}\eta_2[j] + \beta_{NH_4^+,2}I(k),\sigma_{NH_4^+})$$

$$\log(N[j,k]) \sim N(\beta_{N,0} + \beta_{N,1}\eta_2[j] + \beta_{N,2}I(k),\sigma_N).$$

We adopt an occupancy modeling approach for the data model describing amphibian detection and non-detection. Observations of the ith species at the jth site on the kth repeat survey are represented by $Y[i,j,k]$. We treated these observations as Bernoulli random variables with probability $p[i,j,k]z[i,j]$, where p is the probability of detection and z is the latent binary presence/absence state. The true occurrence state z is only partly observed. If species i was seen at site j on any survey, it was present, but if it was not seen on any survey, it is possible that it was present but unobserved (MacKenzie et al. 2002)

$$Y[i,j,k] \sim \text{Bernoulli}(p[i,j,k]z[i,j]).$$

We treated fish presence and pond permanence as directly observed quantities, which is supported by consistency of observations within and across years in this system.

Process model

Our process model represents the latent processes connecting the latent quantities: grazing intensity ξ, shoreline vegetation η_1, N concentration η_2, the true occupancy states z, and the probability of detection P (Fig. 2). In the SEM lexicon, process models are sometimes referred to as structural models. Note that there is an additional pathway from fish to nitrogen that was not initially hypothesized, but revealed in the process of model evaluation (see *Model Assessment*).

We treat grazing intensity as an exogenous latent variable, unaffected by the other latent quantities, with mean 0 and standard deviation 1 as an identifiability constraint. Due to cattle grazing on and trampling of shoreline vegetation, we modeled a linear effect of cattle grazing intensity ξ on shoreline vegetation η_1:

$$\eta_1 \sim N(\gamma_1\xi,1)$$

where the standard deviation term, similarly set to 1 for identifiability, represents the influence of other, unmodeled factors on shoreline vegetation cover. Nitrogenous inputs from cattle excretion in and around wetlands are treated similarly. Following a graphical check of independence assertions, we included an effect of fish presence (γ_3) on nitrogen:

$$\eta_2 \sim N(\gamma_2\xi + \gamma_3\text{fish},1).$$

We represented true occupancy states as Bernoulli random variables with probability of occupancy $\psi[i,j]$ for the
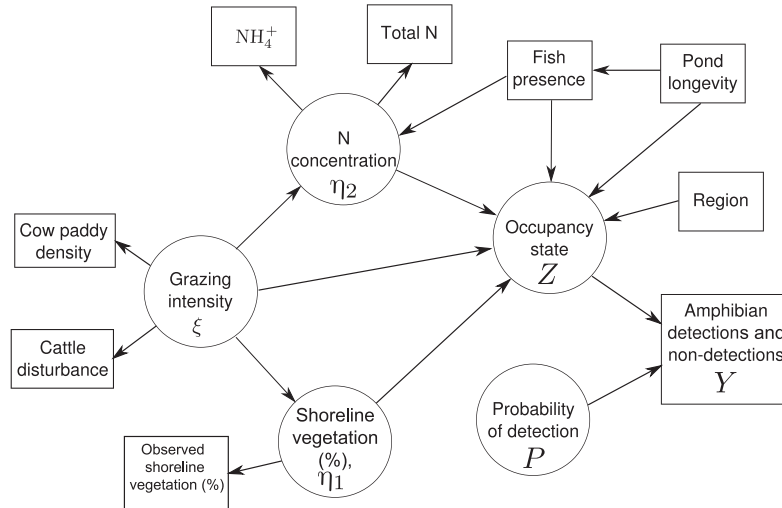
Fig. 2. Directed acyclic graph illustrating relationships between unknown quantities (circles) and observed indicator variables (rectangles).

ith species at the jth site, such that $z[i,j] \sim$ Bernoulli $(\psi[i,j])$. We use a logit-link to model the effects of observed and latent covariates on ψ:

$$\text{logit } (\psi[i,j]) = \alpha_{\text{species}}[i] + \alpha_{\text{region}}[i,r[j]] + \alpha_{\text{site}}[i,j]$$

where α_{species} accounts for species-specific differences in overall occupancy across the entire study area, α_{region} accounts for regional differences in occupancy rates within each species (with region r being indexed by site j). Last, α_{site} represents the local effects of fish, pond permanence, grazing, shoreline vegetation and nitrogen concentrations. These terms can be decomposed as follows:

$$\alpha_{\text{region}}[i,j] \sim N(0, \sigma_{\alpha_{\text{region}}}[i]).$$

This varying intercept term accounts for among-region variation within species in occupancy, with varying species-specific standard deviations, to account for the fact that the degree of regional variation in occupancy rates varies among species. Local covariate effects enter the final term:

$$\alpha_{\text{site}}[i,j] = \beta_{\psi}[i,1] \text{ fish } [j] + \beta_{\psi}[i,2] \text{ perm } [j] + \beta_{\psi}[i,3]\eta_1[j]$$
$$+ \beta_{\psi}[i,4]\eta_2[j] + \beta_{\psi}[i,5]\xi[j].$$

Species vary in their detection probabilities and p may also vary between first and second visits:

$$\text{logit } (p[i,j,k]) = \alpha_p[i] + \beta_p k$$

where α_p is a species-specific mean and the last term represents the effect of early vs. late summer surveys.

Priors

We assumed logit-normal species responses to covariates with covariate-specific community means and variance parameters that represent among-species variability, such that:

$$\beta_{\psi} \sim N(\mu_{\beta_{\psi}}, \sigma_{\beta_{\psi}})$$

$$\beta_p \sim N(\mu_{\beta_p}, \sigma_{\beta_p}).$$

Further, we assumed that for each species, mean detection probabilities would be logit-normally distributed around community level means:

$$\alpha_p \sim N(\mu_{\alpha_p}, \sigma_{\alpha_p}).$$

Hierarchical parameters corresponding to community-level variance terms received semi-informative half-Cauchy priors that were weighted towards small values to reduce bias relative to commonly used uniform priors (Gelman 2006). We adopt vague priors for all other parameters except the loading terms for indicator variables which were constrained to be positive (e.g., increases in latent grazing intensity ξ correspond to increases in its indicators Y_1 and Y_2). Last, prior information based on previous work has entered the model in the form of included effect pathways (e.g., we assume that amphibian community composition does not affect grazing intensity).

Estimation

We used Stan, and the R package rstan to draw samples from the joint posterior distribution of all parameters (R Core Team 2014, Stan Development Team 2014*a,b,c*). Although this model could be implemented with WinBUGS, OpenBUGS, or JAGS, the simultaneous updating of all parameters via the No-U-Turn sampler in Stan results in faster convergence and more efficient sampling (Hoffman and Gelman 2014). Running three chains in parallel with 30 000 iterations took about 80 min on a

quad-core i7 laptop. Convergence was assessed using visual inspections of trace plots and the Gelman-Rubin potential scale reduction factor (Gelman and Rubin 1992, Brooks and Gelman 1998).

Model assessment

As this is a new method, we conducted a simulation analysis to ensure adequate recovery of parameter estimates across a range of known values. We simulated approximately 100 data sets with structure identical to our model, and the same amount of information (observations) present in our data set. We then attempted to recover the known parameters by fitting the model to our simulated data sets (Gimenez et al. 2012). Any simulations that did not reach convergence at the Markov chain Monte Carlo (MCMC) step were discarded.

Independence assumptions were evaluated graphically using scatter plots to detect missing causal pathways (Grace et al. 2012). This revealed a positive correlation between fish presence and N (both total dissolved N and NH_4^+), leading to the inclusion of an additional effect of fish on N that we had not initially hypothesized. Mechanistically, this pathway may represent the joint effects of fish locking up N in their tissues, and suppressing zooplankton through predation, leading to

release of phytoplankton from grazing and greater uptake of N from the water (Andersson et al. 1978, Henrikson et al. 1980).

RESULTS

Parameter recovery

Our simulations demonstrated parameter recovery with 95% highest density posterior intervals (HDIs) including the true population-level parameters over 90% of the time (Fig. A1). Furthermore, we were able to recover effects of local factors (β_ψ), identifying large effects as being non-zero (Appendix S1: Fig. S2). These results increased confidence in parameter estimates for the empirical data and gave an indication of the power that we might have to detect effects of local drivers of occurrence with a data set of comparable size.

Empirical results

Nonnative fish and pond permanence directly affected amphibian community composition, with fish exerting the most consistent and strongest effects. Nonnative fish reduced the probability of occurrence for four of five native amphibians (all species except western toads and bullfrogs; Fig. 3). Species-specific effects were observed for
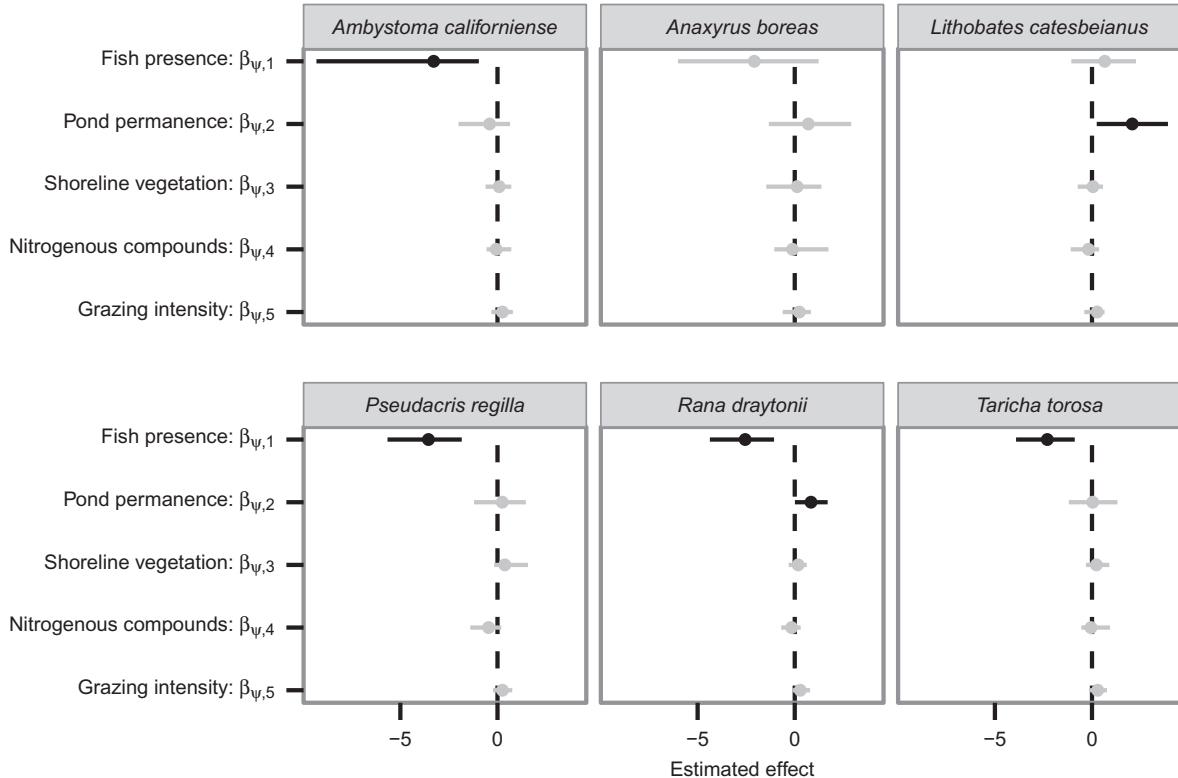

FIG. 3. Estimated direct effects on occurrence probabilities for each species. Black corresponds to parameters for which highest density posterior intervals (HDIs) excluded zero, and gray corresponds to HDIs including zero.

pond permanence, which had a strong positive effect on bullfrog and red-legged frog occurrence. Shoreline vegetation, nitrogenous compounds, and cattle grazing exerted relatively weak direct effects, with all HDIs including zero. Similarly, all HDIs for indirect effects of fish (via nitrogen) and cattle (via nitrogen and shoreline vegetation) on amphibian occurrence included zero.

Consistent with our expectations and previous work, cattle grazing decreased shoreline vegetation (HDI: [−0.556, −0.206]) and increased nitrogenous compounds in the water column (HDI: [0.097, 0.450]). The grazing submodel combined information from cow paddy density counts and disturbance classifications to generate values of latent grazing intensity for each site (Fig. 4A,B). As we expected based on field observations, many sites experience moderate to high levels of grazing, while fewer experience very low levels of grazing.

Regionally, some species were far more variable in their distribution among parks than others. For example, although local factors tended to have minimal effects on the distribution of western toads (Fig. 3), among-region variability was quite high for this species (Fig. A3). In contrast, Pacific chorus frogs had low among-park variability, being nearly ubiquitous.

Most amphibian species were easier to detect during the first survey, probably due to metamorphosis occurring before or during the late-summer visit when evidence of successful breeding is no longer detectable (Fig. A4). The inclusion of the visit number covariate (first vs. second) accounted for this discrepancy in the sense that species-specific covariate effects can still be recovered (Fig. A2).

The shoreline vegetation submodel performed well, capturing the fact that some shorelines were either devoid or completely covered by vegetation, with variability between these two extremes (Fig. 4C). The latent N variable and observed log-transformed total dissolved N and NH_4^+ concentrations showed good fit, with increasing N concentrations in late summer compared to early summer (Fig. 4D,E). Graphical checks of independence assertions indicated no further causal pathways for inclusion (Fig. A5).

Discussion

Our approach demonstrates the integration of SEM with occupancy modeling using a large-scale survey of pond breeding amphibians to gain a better understanding of the drivers of community composition. Importantly, this approach allows for explicit differentiation between observed data and underlying processes, accounting for a variety of measurement error models including the imperfect measurement process that gives rise to species
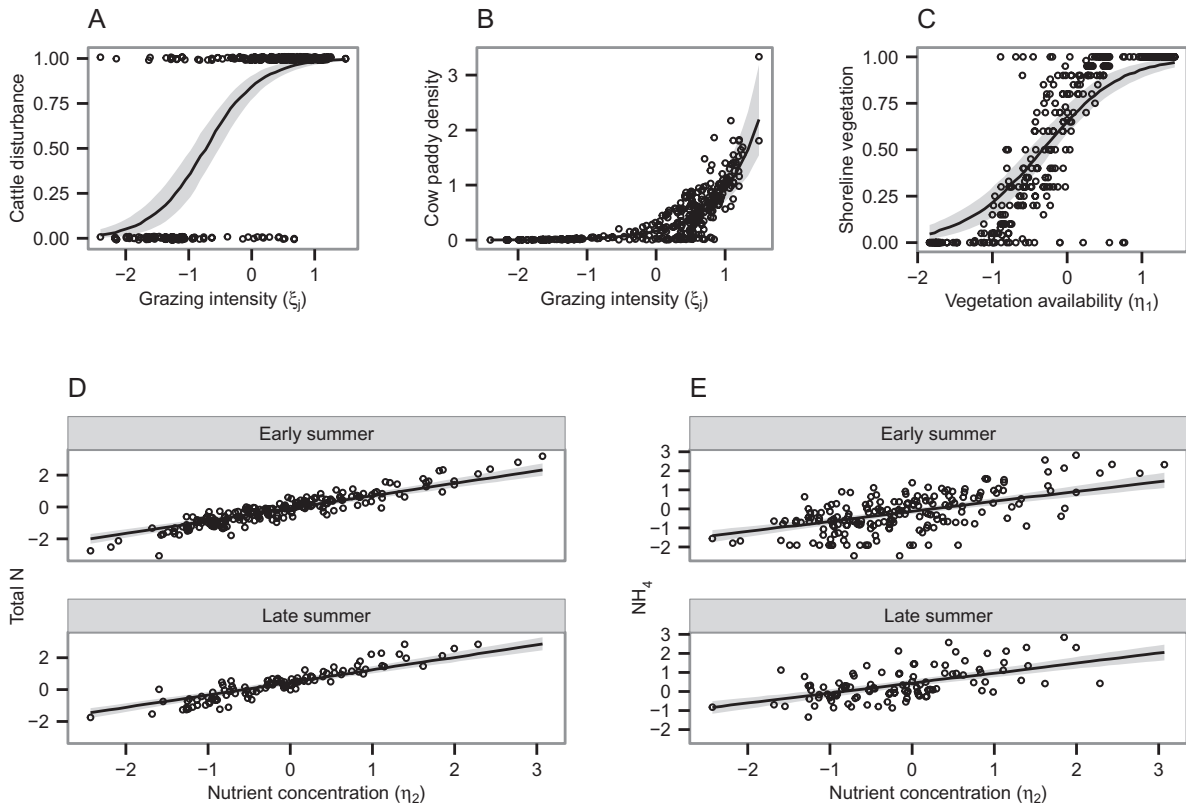

Fig. 4. Fit of the submodels to the observed indicator variables. Shaded regions encompass the 95% HDI, with observed data shown as jittered points (x-axis values represent posterior medians).

detection data. We also embedded other types of observation models to account for more complicated likelihood functions, including the zero-one inflated beta distribution that was used to model observations of shoreline vegetation cover. With this approach, latent ecological processes hypothesized to drive occurrence can be explicitly represented and their direct and indirect consequences formally evaluated.

The potential advantages of integrating occupancy and structural equation frameworks include (1) developing more mechanistic approaches for understanding species distributions influenced by simultaneous related processes; (2) inheriting a formal method for evaluating potential outcomes that would result from management interventions; and (3) clarifying hypotheses by requiring that assumptions be represented in a formal causal model. Unlike purely associational methods, the causal assumptions embedded within structural equation models facilitate unique predictions that can be used to answer applied management questions.

Well-specified SEMs yield unique management insights based on "do" operators, which can estimate the consequences of targeted interventions from effect decompositions (Pearl 1998). "Do" operators could be used to predict changes in occupancy following a management intervention that reduces grazing intensity or removes fish. Indeed, this was one of the original arguments for the development of structural equation models in the mid-20th century (Marschak 1950, Koopmans 1953). Causal inferences drawn from such analyses can be no better than the validity of the assumptions used to construct the model, however. Processes that drive the species distributions can be represented best when there is substantive knowledge to construct such models. This brings up two issues: how mechanistic is "mechanistic enough," and how valid are the assumed causal relationships? First, it would be unreasonable to expect to construct an infinitely accurate and specific causal model for the occurrence of any species (Shipley 2002). However, an adequate model may capture the most important drivers of occurrence and provide "good-enough" predictions for management interventions. This point underlines the role of interventional tests of such models – if the true outcome of management intervention deviates from anticipated effects, then the validity of the underlying causal model must be questioned.

Although not a substitute for interventions, tests of model fit to observational data can also be useful for identifying potentially missing pathways based on independence relationships among variables in the causal network. Here, we employed graphical checks of independence assumptions for observed variables, but better methods could be developed. Information theoretic and d-separation tests exist for path analytic models, and other indices of model fit have long been used for traditional linear structural equation models that assume multivariate normality of errors (Grace 2006, Shipley 2013). However, methods for evaluating occupancy model fit are still relatively new and mostly rely on out-of-sample data (Zipkin

et al. 2012) or bootstrapping (MacKenzie and Bailey 2004). Developing methods to evaluate the fit of hierarchical occupancy-type structural equation models with non-normal indicators and binary, partially observed latent variables is a non-trivial task, but one that would increase the utility of this method. Advances in the study of probabilistic graphical models may provide solutions, but such advances are only just beginning to be applied by ecologists (Koller and Friedman 2009, Grace et al. 2012). Aside from d-separation tests, management experiments provide another way to test the predictive power of these types of models.

Management implications

Our results have direct relevance to the management of threatened amphibian populations within lowland wetlands in California, particularly when land managers are faced with multiple potential challenges simultaneously. Livestock grazing, which is common throughout the western United States (Fleischner 1994), has been a topic of uncertainty with regard to amphibian conservation. While several studies indicate negative overall effects of grazing on populations of specific amphibian species (Knutson et al. 2004, Schmutzer et al. 2008), others indicate the potential for positive effects of grazing on diversity and the perseverance of native communities (Marty 2005).

Based on our analysis, livestock grazing in the Bay Area of California had minimal effects on the occurrence of six amphibian species, two of which are native species of conservation concern. This finding suggests that current grazing levels employed on these parks may be compatible with management aimed at conserving threatened amphibians. However, we did not explore the effects of grazing on amphibian abundance or temporal dynamics such as persistence and colonization. Such approaches may reveal effects of grazing that were not seen in this study.

Most ponds in this data set were initially constructed to serve as watering sites for livestock. Because most original natural wetlands in California have been destroyed for agriculture and development, particularly in the Central Valley (Garone 2011), such livestock ponds may now serve as vital habitat refuges for declining species. This trend is particularly important for the California tiger salamander and California red-legged frog, which are the focus of considerable conservation efforts due largely to habitat destruction (Lannoo 2005).

In sharp contrast to the minimal effects of grazing, we found strong negative effects of nonnative fish on native amphibian occupancy for all species except western toads and nonnative bullfrogs. This finding supports a large body of research showing that native pond-breeding amphibians that lack evolutionary history with fishes are unlikely to persist once predatory fish have been introduced to a breeding site. The most common fish species at our field sites were mosquitofish and centrarchids (bass, bluegill, and other sunfish), which are native to the eastern United States. These species likely prevent amphibian

reproduction through direct predation on multiple life stages of native amphibians. The lack of an effect on toads is consistent with prior work showing that toxicity and schooling behavior of toad larvae provides resistance to predation by fish (Kruse and Stone 1984, Welsh et al. 2006). Bullfrogs, which require permanent water bodies to complete metamorphosis, have coevolved with fish in their native range, are unpalatable, and are not strongly affected by fish presence (Walters 1975, Kruse and Francis 1977, Szuroczki and Richardson 2011).

Taken together, our results indicate that land management strategies should prioritize removal of nonnative fish rather than limitation of livestock grazing. Fish removal via pond draining has been shown to be effective in restoring populations of threatened amphibians within the study region (Alvarez et al. 2003). Considering the needs of interests groups involved and the prevalence of grazing, this strategy is perhaps more feasible than dramatically limiting access of livestock to ponds, which may have unintended side effects including elimination of pond breeding habitat due to overgrowth by vegetation. While more research is needed to evaluate both temporal effects of grazing, and how grazing affects amphibian abundance, our analysis indicates that fish introductions have more severe impacts than cattle grazing.

Conclusion

Structural equation modeling provides a framework to evaluate why species occur in some areas and not others, allowing a more direct confrontation of some of the most fundamental questions in ecology. Occupancy modeling provides a solution to the problem of imperfect detection and can account for many different processes giving rise to detection data. Combining these approaches provides a means to evaluate complex causal processes driving occurrence while accounting for false absences in empirical occurrence data. This framework potentially facilitates deeper insights into biological processes, making a clear separation between imperfectly observed data and underlying states. From a pragmatic standpoint, it is clearly advantageous to be able to represent both direct and indirect determinants of species occurrence rather than being limited to treating covariates as being independent. This approach also provides the practical advantage of inheriting a suite of methods to anticipate the effects of management interventions ("do" operators), and to account for many different sampling schemes in the occupancy modeling literature. Future extensions of this method could deepen connections between SEM and other classes of occupancy models, including dynamic multi-year occupancy models, models of abundance such as N-mixture models, Dail-Madsen models, their multi-species extensions, and spatial models (Royle 2004, Royle and Kéry 2007, Dail and Madsen 2011, Dorazio and Connor 2014, Lamb et al. 2014).

The full value of this approach may be most apparent, and perhaps palatable to individuals unfamiliar with SEM, when coupled with controlled field experiments, where the validity of the underlying causal model may be tested directly. Compared to fields like sociology and economics, ecology as a field is perhaps in a unique position to reap benefits from SEM because it is both complex and relatively amenable to experimental manipulation.

Literature Cited

Adams, M. J. 2000. Pond permanence and the effects of exotic vertebrates on anurans. Ecological Applications 10:559–568.

Adams, M. J., C. A. Pearl, B. Mccreary, S. K. Galvan, S. J. Wessell, W. H. Wente, W. Chauncey, and A. B. Kuehl. 2009. Short-term effect of cattle exclosures on columbia spotted frog (*Rana luteiventris*) populations and habitat in northeastern Oregon. Journal of Herpetology 43:132–138.

Alsterberg, C., J. S. Eklöf, L. Gamfeldt, J. N. Havenhand, and K. Sundbäck. 2013. Consumers mediate the effects of experimental ocean acidification and warming on primary producers. Proceedings of the National Academy of Sciences USA 110:8603–8608.

Alvarez, J. A., C. Dunn, and A. F. Zuur. 2003. Response of California Red-legged Frogs to removal of non-native fish. Transactions of the Western Section of the Wildlife Society 39:9–12.

Andersson, G., H. Berggren, G. Cronberg, and C. Gelin. 1978. Effects of planktivorous and benthivorous fish on organisms and water chemistry in eutrophic lakes. Hydrobiologia 59:9–15.

Bollen, K. A. 1989. Structural equations with latent variables. John Wiley & Sons Inc., Hoboken, New Jersey, USA.

Brodman, R., J. Ogger, T. Bogard, A. J. Long, R. A. Pulver, K. Mancuso, and D. Falk. 2003. Multivariate analyses of the influences of water chemistry and habitat parameters on the abundances of pond-breeding amphibians. Journal of Freshwater Ecology 18:425–436.

Brooks, S., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7:434–455.

Burne, M. R., and C. R. Griffin. 2005. Habitat associations of pool-breeding amphibians in eastern Massachusetts, USA. Wetlands Ecology and Management 13:247–259.

Clough, Y. 2012. A generalized approach to modeling and estimating indirect effects in ecology. Ecology 93:1809–1815.

Collins, J. 1979. Intrapopulation variation in the body size at metamorphosis and timing of metamorphosis in the bullfrog, *Rana catesbeiana*. Ecology 60:738–749.

Cressie, N., C. A. Calder, J. S. Clark, J. M. Ver Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecological Applications 19:553–570.

Crump, M., and R. Bury. 1994. Visual encounter surveys. Pages 84–92 in W. Heyer, *M. Donnely, R. McDiarmid, L. Hayek, and M. Foster, editors.* Measuring and monitoring biological diversity, standard methods for amphibians. Smithsonian Institution Press, Washington, D.C., USA.

Dail, D., and L. Madsen. 2011. Models for estimating abundance from repeated counts of an open metapopulation. Biometrics 67:577–587.

Dorazio, R. M., and E. F. Connor. 2014. Estimating abundances of interacting species using morphological traits, foraging guilds, and habitat. PLoS One 9:e94323.

Dorazio, R. M., M. Kéry, J. A. Royle, and M. Plattner. 2010. Models for inference in dynamic metacommunity systems. Ecology 91:2466–2475.

Earl, J. E., and H. H. Whiteman. 2009. Effects of pulsed nitrate exposure on amphibian development. Environmental Toxicology and Chemistry 28:1331–1337.

Egan, R., and P. Paton. 2004. Within-pond parameters affecting oviposition by wood frogs and spotted salamanders. Wetlands 24:1–13.

Fisher, R. N., and H. B. Shaffer. 1996. The Decline of amphibians in California's Great Central Valley. Conservation Biology 10:1387–1397.

Fleischner, T. L. 1994. Ecological costs of livestock grazing in western North America. Conservation Biology 8:629–644.

Freda, J., and W. Dunson. 1986. Effects of low pH and other chemical variables on the local distribution of amphibians. Copeia 1986:454–466.

Garone, P. 2011. The fall and rise of the wetlands of California's great central valley, First edition. University of California Press, Oakland, California, USA.

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1:515–533.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–511.

Gimenez, O., T. Anker-Nilssen, and V. Grosbois. 2012. Exploring causal pathways in demographic parameter variation: path analysis of mark-recapture data. Methods in Ecology and Evolution 3:427–432.

Grace, J. B. 2006. Structural equation modeling and natural systems. Cambridge University Press, Cambridge, United Kingdom.

Grace, J. B., D. R. Schoolmaster, G. R. Guntenspergen, A. M. Little, B. R. Mitchell, K. M. Miller, and E. W. Schweiger. 2012. Guidelines for a graph-theoretic implementation of structural equation modeling. Ecosphere 3:1–44.

Grace, J. B., P. B. Adler, W. Stanley Harpole, E. T. Borer, and E. W. Seabloom. 2014. Causal networks clarify productivity-richness interrelations, bivariate plots do not. Functional Ecology 28:787–798.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993–1009.

Henrikson, L., H. G. Nyman, H. G. Oscarson, and J. A. E. Stenson. 1980. Trophic changes,without changes in the external nutrient loading. Hydrobiologia 68:257–263.

Hoffman, M. D., and A. Gelman. 2014. The No-U-Turn Sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research 15:1593–1623.

Jansen, A., and M. Healey. 2003. Frog communities and wetland condition: relationships with grazing by domestic livestock along an Australian floodplain river. Biological Conservation 109:207–219.

Johnson, P. T. J., D. L. Preston, J. T. Hoverman, and K. L. D. Richgels. 2013. Biodiversity decreases disease through predictable changes in host community competence. Nature 494:230–300.

Kats, L. B., J. W. Petranka, and A. Sih. 1988. Antipredator defenses and the persistence of amphibian larvae with fishes. Ecology 69:1865–1870.

Kauffman, J., W. Krueger, and M. Vavra. 1983. Effects of late season cattle grazing on riparian plant communities. Journal of Range Management 36:685–691.

Knutson, M. M. G., W. W. B. Richardson, D. M. Reineke, B. R. Gray, J. R. Parmelee, and S. E. Weick. 2004. Agricultural ponds support amphibian populations. Ecological Applications 14:669–684.

Koller, D., and N. Friedman. 2009. Probabilistic graphical models: principles and techniques. MIT Press, Cambridge, Massachusetts.

Koopmans, T. C., and W. E. Hood. 1953. The estimation of simultaneous linear economic relationships. Studies in Econometric Method, Cowles Commission Monograph 14:112–199.

Kruse, K., and M. Francis. 1977. A predation deterrent in larvae of the bullfrog, *Rana catesbeiana*. Transactions of the American Fisheries Society 106:248–252.

Kruse, K. P., and B. M. Stone. 1984. Largemouth bass (*Micropterus salmoides*) learn to avoid feeding on toad (*Bufo*) tadpoles. Animal Behaviour 32:1035–1039.

Lamb, E. G., K. L. Mengersen, and K. J. Stewart. 2014. Spatially Explicit Structural Equation Modeling. Ecology 95:2434–2442.

Lannoo, M. J. 2005. Amphibian declines: the conservation status of United States species. University of California Press, Oakland, California, USA.

Lawler, S. P., D. Dritz, T. Strange, and M. Holyoak. 1999. Effects of introduced mosquitofish and bullfrogs on the threatened California red-legged frog. Conservation Biology 13:613–622.

Lee, S. 2007. Structural equation modeling: a bayesian approach. Wiley, Hoboken, New Jersey, USA.

MacKenzie, D. I., and L. L. Bailey. 2004. Assessing the fit of site-occupancy models. Journal of Agricultural, Biological, and Environmental Statistics 9:300–318.

MacKenzie, D., J. Nichols, and G. Lachman. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248–2255.

MacKenzie, D. I., L. L. Bailey, J. E. Hines, and J. D. Nichols. 2011. An integrated model of habitat and species occurrence dynamics. Methods in Ecology and Evolution 2:612–622.

Marschak, J. T. 1950. Statistical inference in economics. Statistical inference in dynamic economic models :1–50.

Marty, J. 2005. Effects of cattle grazing on diversity in ephemeral wetlands. Conservation Biology 19:1626–1632.

Miller, Z. J. 2012. Fungal pathogen species richness: why do some plant species have more pathogens than others? American Naturalist 179:282–292.

Ospina, R., and S. L. Ferrari. 2012. A general class of zero-or-one inflated beta regression models. Computational Statistics & Data Analysis 56:1609–1623.

Pearl, J. 1998. Graphs, causality, and structural equation models. Sociological Methods & Research R-253:226–284.

Pearl, J. 2000. Causality. Cambridge, New York.

Preston, D. L., J. S. Henderson, and P. T. Johnson. 2012. Community ecology of invasions: direct and indirect effects of multiple invasive species on aquatic communities. Ecology 93:1254–1261.

R Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org

Robins, J. D., and J. E. Vollmar. 2002. Livestock grazing and vernal pools. Pages 401–430. Wildlife and rare plant ecology of eastern merced county's vernal pool grasslands. Berkeley, California.

Roche, L. M., A. M. Latimer, D. J. Eastburn, and K. W. Tate. 2012. Cattle grazing and conservation of a meadow-dependent amphibian species in the Sierra Nevada. PLoS One 7:e35734.

Rouse, J. D., C. A. Bishop, and J. Struger. 1999. Nitrogen pollution: an assessment of its threat to amphibian survival. Environmental Health Perspectives 107:799–803.

Rowe, C., and W. Dunson. 1995. Impacts of hydroperiod on growth and survival of larval amphibians in temporary ponds of central Pennsylvania, USA. Oecologia 102:397–403.

Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60:108–115.

Royle, J. A., and M. Kéry. 2007. A Bayesian state-space formulation of dynamic occupancy models. Ecology 88:1813–1823.

Rykiel, E. J. 1996. Testing ecological models: The meaning of validation. Ecological Modelling 90:229–244.

Schmutzer, A. C., M. J. Gray, E. C. Burton, and D. L. Miller. 2008. Impacts of cattle on amphibian larvae and the aquatic environment. Freshwater Biology 53:2613–2625.

Shipley, B. 2002. Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference. Cambridge University Press, Cambridge, United Kingdom.

Shipley, B. 2013. The AIC model selection method applied to path analytic models compared using a d-separation test. Ecology 94:560–564.

Stan Development Team. 2014a. Stan modeling language users guide and reference manual, Version 2.5.0. New York, New York, USA.

Stan Development Team. 2014b. RStan: the R interface to Stan, version 2.5.0. New York, New York, USA.

Stan Development Team. 2014c. Stan: A C++ library for probability and sampling, version 2.5.0. New York, New York, USA.

Stebbins, R. C. 2003. A field guide to western reptiles and amphibians. Houghton Mifflin, Boston, Massachusetts, USA.

Szuroczki, D., and J. M. L. Richardson. 2011. Palatability of the larvae of three species of *Lithobates*. Herpetologica 67:213–221.

Walters, B. 1975. Studies of interspecific predation within an amphibian community. Journal of Herpetology 9:267–279.

Welsh, H. H., K. L. Pope, and D. Boiano. 2006. Sub-alpine amphibian distributions related to species palatability to non-native salmonids in the Klamath mountains of northern California. Diversity and Distributions 12:298–309.

Zipkin, E. F., E. H. Campbell Grant, and W. F. Fagan. 2012. Evaluating the predictive abilities of community occupancy models using AUC while accounting for imperfect detection. Ecological Applications 22:1962–1972.

Supporting Information

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1890/15-0833.1/suppinfo