Using species distribution and occupancy modeling to guide survey efforts and assess species status

William E. Peterman [a,*], John A. Crawford [b], Andrew R. Kuhns [c]

[a] *University of Missouri-Columbia, 110 Tucker Hall, Division of Biological Sciences, Columbia, MO 65211, USA*
[b] *Lindenwood University, Department of Biology, Saint Charles, MO 63301, USA*
[c] *University of Illinois at Urbana-Champaign, Prairie Research Institute, Illinois Natural History Survey, Champaign, IL 61820, USA*

ABSTRACT

Habitat loss and fragmentation continue to be major issues affecting the persistence and conservation of species, but identification of critical habitat remains a challenge. Species distribution modeling and occupancy modeling are both approaches that have been used to predict species distributions and can identify critical habitat characteristics associated with species occurrence. Additionally, occupancy sampling can provide measures of detectability, increasing the confidence that a species is truly absent when not detected. While increasingly popular, these methods are infrequently used in synergy, and rarely at fine spatial scales. We provide a case study of using distribution and occupancy modeling in unison to direct survey efforts, provide estimates of species presence/absence, and to identify local and landscape features important for species occurrence. The focal species for our study was *Ambystoma jeffersonianum*, a threatened salamander in the state of Illinois, U.S.A. We found that fine-scale distribution models accurately discriminated occupied from unoccupied breeding ponds (78–91% accuracy), and surveys could be effectively guided using a well-fit model. We achieved a high detection rate (0.774) through occupancy sampling, and determined that *A. jeffersonianum* never used ponds inhabited by fish, and the probability of a pond being used for breeding increased as canopy cover increased. When faced with limited resources, combining fine-scale distribution modeling with a robust occupancy sampling design can expedite survey efforts, confidently designate species occupancy status, prioritise habitat for future surveys and/or restoration, and identify critical habitat features. This approach is broadly applicable to other taxa that have specific habitat requirements.

© 2012 Elsevier GmbH. All rights reserved.

Introduction

Habitat loss and fragmentation have come to the forefront of conservation and species management, and pose significant threats to biodiversity and species persistence (Brooks et al. 2002; Stuart et al. 2004; Vitousek, Mooney, Lubchenco, & Melillo 1997). Populations of species at the edge of their natural range limit may be particularly susceptible due to limited opportunities for recolonisation following local extinction (Brown & Kodric-Brown 1977). With increased land use continuing to fragment habitats and isolate populations, critical habitat and biodiversity are in jeopardy. In order to dampen the effects of these losses, quick and accurate habitat assessments are needed to prioritise the preservation of remaining habitat, but a major impediment to the assessment of biodiversity is identification of suitable habitat. To this end, modeling of species' distribution and occupancy have become critical

conservation tools (e.g. De Wan et al. 2009; Raxworthy et al. 2003). These approaches can direct survey efforts (Guisan et al. 2006; Raxworthy et al. 2003; Rebelo & Jones 2010; Williams et al. 2009), give a high degree of confidence concerning species detection/non-detection (Andelt, White, Schnurr, & Navo 2009; Bailey, Simons, & Pollock 2004b), and identify vital environmental covariates for species persistence (Hamer & Mahony 2010; Kroll et al. 2008; Schmidt & Pellet 2005). While both distribution and occupancy modeling approaches are receiving increased attention as tools to facilitate surveys, they are not readily utilised together (but see Gormley et al. 2011; Newbold et al. 2010). The focus of our study is to test the ability of distribution models to provide fine-scale discrimination between suitable and unsuitable habitat and to conduct robust sampling of sites to provide estimates of species detectability, rates of occupancy, and covariates critical to occupancy.

Numerous modeling approaches exist for predicting species distributions (see Elith et al. 2006 for a comparison of several methods). Generally, these models predict a species realised niche (Elith & Leathwick 2009), and describe the range of biological and/or

---
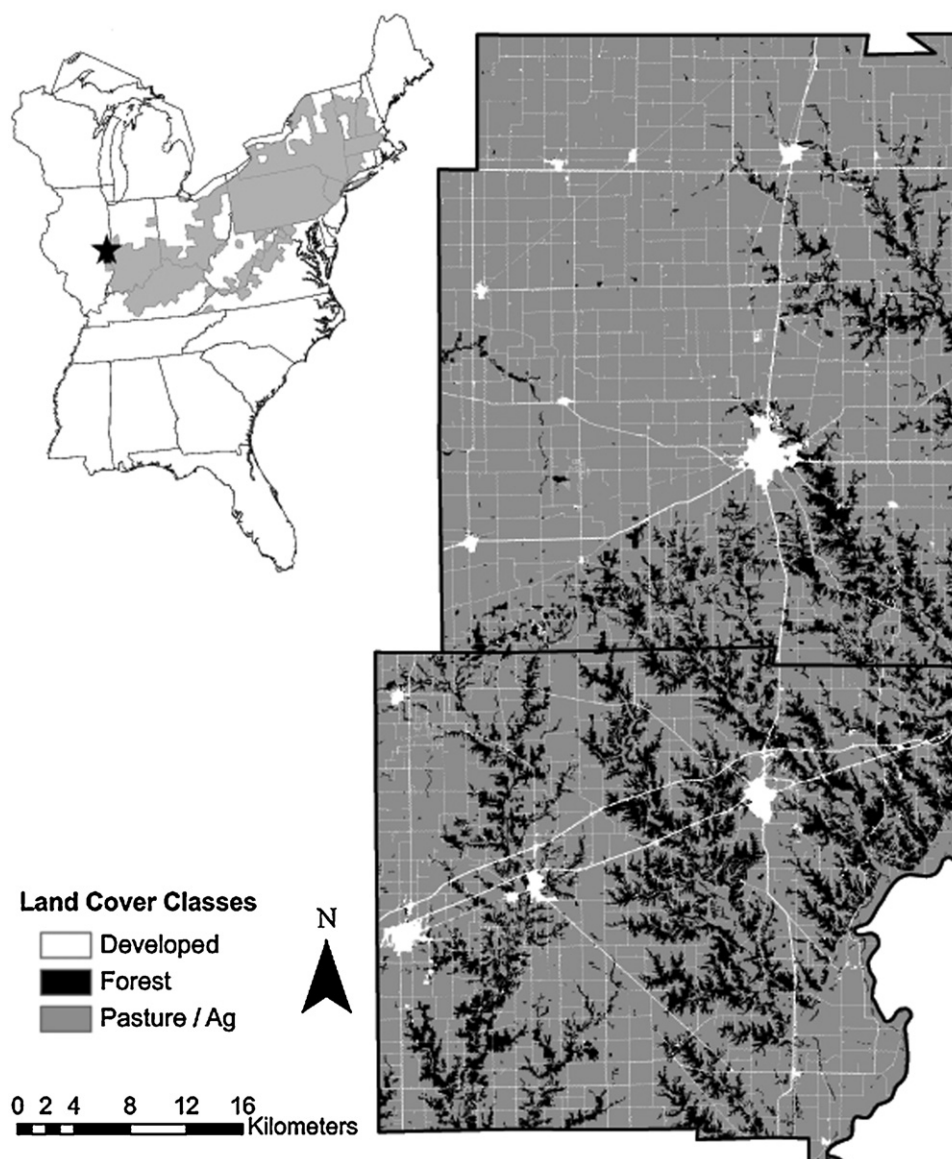
Fig. 1. Map of study area showing range of *Ambystoma jeffersonianum* in the eastern U.S.A. in light gray shading. The star in the regional map indicates the 2-county study area. Edgar County is to the north and Clark County is to the south. Approximately 50% of these two counties are forested, with the remainder being equally divided into pasture/agriculture or developed land. Developed land consists of roadways and urban areas.

physical conditions that an organism has been found to exist in. Such models are being used to direct surveys for the discovery of new populations (Guisan et al. 2006; Williams et al. 2009) and species (Raxworthy et al. 2003), conservation prioritisation (Urbina-Cardona & Flores-Villela 2010), and to predict impacts of global climate change (Milanovich, Peterman, Nibbelink, & Maerz 2010). In addition to advances in distribution modeling, the field of occupancy modeling has grown rapidly over the last decade (MacKenzie et al. 2006), with estimates of occupancy now corrected for the imperfect detection of species (Bailey, Hines, Nichols, & MacKenzie 2007; MacKenzie 2006). Using detection/non-detection data, this approach models species detection and occupancy as a function of covariates, and through model selection provides information about critical aspects of sampling efficacy and necessary habitat associations (MacKenzie et al. 2006).

With the strengths and versatility of these two different fields apparent, it is of relevance to determine how they interplay, and in particular, if there are advantages to using them in combination. To this end, we designed a study to utilise fine-scale distribution models to predict species occurrence, and then carry out robust sampling of sites in an occupancy framework. Specifically, we conducted surveys for *Ambystoma jeffersonianum* at breeding ponds in east-central Illinois, U.S.A. where it is listed as a threatened species. Illinois is the western-most extent of *A. jeffersonianum*, where they were first discovered in 1990 (Fig. 1; Phillips, Brandon, & Moll 1999), and prior to 2008 were known to breed in only six ponds. Our objectives in conducting this study were to: (1) develop a spatial model of suitable habitat for a threatened species to direct surveys and prioritise management and restoration efforts; (2) validate a robust sampling protocol for determining species presence; (3) determine the local and landscape-scale characteristics associated with occurrence; and (4) assess the generality of our approach to other taxa.

Methods

Study species

Ambystoma jeffersonianum are pond breeding salamanders that range from eastern Illinois, USA to New Hampshire, USA (Lannoo

2005). These salamanders require fishless ponds situated within deciduous forest habitat for reproduction (Petranka 1998). Following metamorphosis, salamanders disperse up to 1600 m away from ponds (Williams 1973) where they are predominantly subterranean (Rasic & Keyghobadi 2012). Salamanders return to ponds to breed in late winter and early spring (Rasic & Keyghobadi 2012; Williams 1973).

Study area

To assess the distribution and occupancy of *A. jeffersonianum*, we sampled 57 ponds in Clark and Edgar Counties in east-central Illinois (U.S.A.; 39.41°N, −87.64°W). This region is part of the Wabash Border Division (Schwegman 1973), which encompasses 7123 km² and is characterised as having forested ravines and uplands, with agricultural lowlands and terraces. Approximately 50% of our study region is forested, with the remainder being equally composed of pasture/row crop agriculture and other developed land (Fig. 1). Ponds were identified in ArcGIS 9.3 (ESRI, Redlands, CA) from 2005 leaf-off digital orthophoto quadrangles as well as through the Illinois national wetland inventory data layer (http://www.isgs.illinois.edu/nsdihome/). Only ponds located within forest patches or at forest edges were sampled in this study. Thirty-one ponds were sampled in 2008 and 46 ponds (26 new ponds, 20 re-sampled ponds from 2008) were sampled in 2009. Each pond was sampled for four nights during the spring breeding season with 6–8 double-throated minnow traps.

Data analysis

Distribution modeling

We developed distribution models for *A. jeffersonianum* using Maxent version 3.2.19 using the default settings (Phillips, Anderson, & Schapire 2006; Phillips, Dudík, & Schapire 2004). Maxent is a machine learning method that utilises the principle of maximum entropy to model species distributions using presence-only data coupled with environmental data (Phillips et al. 2006). This approach finds a probability distribution of maximum entropy using a set of environmental variables to estimate a species' ecological niche using the defined Maxent probability distribution. Maxent utilises presence-only data to discriminate species occurrences from background points. We used 2500 background points to make our models, and these points were randomly generated within a 10 km² polygon that encompassed all sampled ponds. This approach is a modification of the target-group approach (Phillips 2008), which utilises occurrence records of related species to reduce model bias. Since this region has been poorly sampled, a true target-group background could not be used. In ArcGIS, we measured the elevation and percent canopy cover at each pond and background point from a 30 m digital elevation layer (DEM) and percent canopy layer, respectively. We also derived slope and position layers from the 30 m DEM. Therefore, each of our 2500 background points were a 30 m × 30 m pixel on the map layers. The position layer identified whether points were on ridges, in valleys, or mid-slope, and was calculated using the Topographic Position Index (TPI; Jenness 2006), as implemented in the Topography Tools 9.3 ArcGIS toolbox. Topographic position was calculated based on a 90 m × 90 m neighbourhood setting.

We made two distribution models: one from the 2008 data alone to direct 2009 sampling; and, one model from the combined dataset of 2008–2009. Model fit was assessed using area under the receiver operator curve (AUC) and omission plots; we sought to maximise AUC and minimise omission. Model significance (i.e., the model performed better than random) was determined using the null model approach of Raes and ter Steege (2007), whereby we generated 1000 random data sets containing 33 samples, corresponding to the number of ponds where *A. jeffersonianum* were present. These 33 sample points were drawn without replacement from our 2500 background points. *Ambystoma jeffersonianum* model AUC values were then compared to the 95 percentile of the null AUC frequency distribution. This approach to model validation was chosen over other methods (e.g. cross validation, bootstrapping, or subsampling) as we had relatively few occurrence records, and wanted to preserve all available information in our model. Since AUC has come under criticism for its ability to accurately assess model fit (Jiménez-Valverde 2012; Lobo, Jimenez-Valverde, & Real 2008), we also calculated the true skill statistic (TSS; Allouche, Tsoar, & Kadmon 2006). The TSS is analogous to kappa, but is independent of prevalence, therefore making it a less biased evaluation statistic than kappa (Allouche et al. 2006). We further tested the ability of our distribution models to accurately discriminate between occupied and unoccupied breeding ponds (as determined by field occupancy surveys) by calculating the percent ponds correctly classified by the Maxent model, presence/absence AUC, sensitivity (correct occupied designations), specificity (correct absence designation), and Cohen's Kappa (level agreement between occupancy and Maxent model predictions) in ROC_AUC version 1.3 (http://lec.wzw.tum.de/index.php?id=67&L=1). To convert our continuous Maxent probability surface into a binary presence-absence surface, we used the P_fair threshold as implemented in ROC_AUC. This threshold minimises the difference between sensitivity and specificity, as determined from field occupancy surveys (Campagne, Smouse, Varouchas, Silvain, & Leru 2012). The P_fair threshold was also used in the calculation of TSS.

Occupancy modeling

Our occupancy sampling spanned two years with 20 ponds being sampled in both years. Because occupancy status did not change for any of the resampled ponds, only the 2009 data were included in the occupancy analysis. Four site covariates were collected at each pond: (1) FISH – presence/absence of predatory fish; (2) COV – canopy cover (%); (3) ELEV – elevation (m); and (4) AG – distance to nearest agricultural field (m). Ponds were designated as containing predatory fish if fish were captured in minnow traps during amphibian sampling. Data layers for land use/land cover (2001 data), canopy cover (2001 data), and elevation were obtained from the USGS Seamless Server (http://seamless.usgs.gov/) and canopy cover, elevation, and distance to nearest agricultural field were calculated for each GPS-recorded pond location. Additionally, data were collected on four sampling covariates: (1) PRECIP – amount of precipitation in the 24 h preceding sampling; (2) AVGTEMP – average temperature in the 24 h preceding sampling; (3) MAXTEMP – maximum temperature in the 24 h preceding sampling; and (4) TIME – day of survey. Temperature and rainfall data were obtained from the National Climatic Data Center (http://www.ncdc.noaa.gov).

To investigate the relationships between amphibian occupancy and environmental variables, we used an information-theoretic approach to model selection (Burnham & Anderson 2002), and followed methods developed by MacKenzie and Kendall (2002) and MacKenzie et al. (2006) to estimate occupancy probability of breeding ponds by *A. jeffersonianum* as a function of measured environmental covariates. All analyses were conducted with program PRESENCE 3.0 (http://www.mbr-pwrc.usgs.gov/software/presence.html).

We used a two-step process to address a priori hypotheses on *A. jeffersonianum* occupancy of breeding ponds. First, we modeled sampling covariates that we thought would affect detection

Table 1
Environmental covariates[a] at ponds sampled for *Ambystoma jeffersonianum* in east-central Illinois.

Occurrence	Fish present	Fish absent	Canopy cover	Elevation	Agriculture
Detected	0	33	72.4 (3.0)	174.0 (2.2)	94.2 (13.6)
Not detected	14	10	33.3 (6.6)	172.0 (3.2)	56.1 (12.3)

[a] Mean and standard error values for canopy cover (%), elevation (m), and distance to nearest agricultural field (m).

probabilities (precipitation, average daily temperature, maximum daily temperature, time of survey, or constant), while holding site occupancy constant. Second, we tested the a priori occupancy models simultaneously with the best model for detection probability incorporated (Bailey et al. 2004b; MacKenzie et al. 2006). All continuous environmental covariates were standardised, which was necessary to avoid problems with parameter estimation in program PRESENCE. Environmental covariates were tested for multicollinearity, but no covariates had $r > 0.5$, so no covariates were excluded from the same model (Graham 2003).

Using four environmental site covariates, we built 12 a priori models to test hypotheses predicting occupancy of breeding ponds by *A. jeffersonianum*. The environmental variables we selected (and subsequent models we built) were based on previous studies detailing their importance to amphibians. The predictive models are: (1) Null (intercept only); (2) FISH (fish presence/absence); (3) COV (percent canopy cover); (4) ELEV (altitude of site); (5) AG (distance to nearest agricultural field); (6) FISH-COV; (7) FISH-ELEV; (8) FISH-AG; (9) COV-AG; (10) FISH-COV-ELEV; (11) FISH-COV-AG; and (12) Global (FISH-COV-ELEV-AG).

For each model, we calculated the AIC value corrected for small sample sizes (AIC_c), based on the number of ponds we sampled (57). $QAIC_c$ is a measure that corrects for overdispersion ($\hat{c} > 1.0$), which was assessed using goodness-of-fit tests on the global occupancy model and simulated using 10,000 bootstrap samples. All candidate models were ranked according to their $QAIC_c$ values, and the best model had the smallest value. We also calculated $\Delta QAIC_c$, which is the difference in QAIC between each model and the best model in the set. We then calculated Akaike weights (ω_i) to determine the weight of evidence in favour of each model and we examined the model averaged point estimates and confidence intervals to determine the importance of each parameter in the model set (Burnham & Anderson 2002). To determine the probability of breeding pond use, we used model averaging of models with $\Delta QAIC_c < 7$ (Hamer & Mahony 2010). For model-averaged estimates of probability of breeding pond use, we used a logistic regression approach and assessed model fit with the D^2 value (Guisan & Zimmermann 2000).

Results

Occupancy modeling

From 2008 to 2009, we captured *A. jeffersonianum* at 33 of 57 ponds sampled (Table 1). The global model fit the *A. jeffersonianum* data well ($\hat{c} = 1.66$), so we proceeded with analyses. Our detection rate for *A. jeffersonianum* was 0.774 (SE = 0.0395; 95% CI = 0.687–0.842), giving us 99.7% confidence in occupancy status following four survey periods. The naïve occupancy of breeding ponds in our study area was 0.509, and model-averaged estimated occupancy was 0.511. Since our sampling was not completely random, but was directed toward ponds located in forests or at forest edges, these occupancy estimates can be viewed as the proportion of forested ponds occupied in our study region. We held the detection probability (p) constant when assessing site occupancy for *A. jeffersonianum* as it was 2.3 times more likely than the next best detection model tested (Table 2). Of the 12 a priori models, the FISH-COV model best predicted the occupancy of breeding ponds (Table 3). This model included only fish (detection/non-detection)

Table 2
Model selection results of *Ambystoma jeffersonianum* detection probability in breeding ponds in east-central Illinois.

Model	−2 Log-likelihood	K^a	AIC_c	ΔAIC_c	ω_i [b]
Ψ(.), p(.)	202.30	2	206.52	0.00	0.29
Ψ(.), p(MAXTEMP-TIME)	193.86	6	207.54	1.02	0.17
Ψ(.), p(PRECIP)	201.26	3	207.71	1.19	0.16
Ψ(.), p(AVGTEMP-TIME)	194.16	6	207.84	1.32	0.15
Ψ(.), p(AVGTEMP)	201.81	3	208.26	1.74	0.12
Ψ(.), p(MAXTEMP)	202.13	3	208.58	2.06	0.10
Ψ(.), p(PRECIP-TIME)	199.06	6	212.74	6.22	0.04

[a] Number of estimable parameters in approximating model.
[b] Probability that the current model (i) is the best approximating model among those tested.

Table 3
Model selection results of *Ambystoma jeffersonianum* occupancy of breeding ponds in east-central Illinois.

Model	−2 Log-likelihood	K^a	$QAIC_c$	$\Delta QAIC_c$	ω_i [b]
Ψ(FISH-COV), p(.)	156.60	4	103.03	0.00	0.36
Ψ(FISH-COV-AG), p(.)	153.20	5	103.39	0.36	0.30
Ψ(GLOBAL), p(.)	151.33	6	104.77	1.74	0.15
Ψ(FISH-COV-ELEV), p(.)	155.75	5	104.92	1.89	0.14
Ψ(FISH-AG), p(.)	163.63	4	107.26	4.23	0.12

[a] Number of estimable parameters in approximating model.
[b] Probability that the current model (i) is the best approximating model among those tested.

and canopy cover (%) covariates, which were the only variables from the model averaged point estimates for which the estimate of beta had a confidence interval that did not overlap zero (Table 4). The probability of a breeding pond being used by *A. jeffersonianum* declined significantly in the presence of fish (Fig. 2A; Table 1). The probability of use was >50% at sites with >70% canopy cover and probability of use increased to >75% at sites with canopy cover >80% (Fig. 2B and Table 1).

Distribution modeling

Our Maxent distribution models had good discrimination with high TSS and significantly better than the random AUC estimates than null models for both 2008 (TSS = 0.722; AUC = 0.910; $p < 0.001$) and 2008–2009 combined (TSS = 0.754; AUC = 0.928; $p < 0.001$). Percent canopy cover was the most influential parameter in these models (85.3% and 88.6% contribution, respectively) with TPI, slope, and elevation making smaller contributions in this order. When assessing the 2008 model's ability to discriminate occupied from unoccupied localities (based on our occupancy sampling), our model had moderate discrimination (AUC = 0.765; 95% CI = 0.614–0.957; Kappa = 0.541) with 78% of ponds being

Table 4
β estimates and 95% confidence intervals for parameters in the a priori global model for *Ambystoma jeffersonianum*.

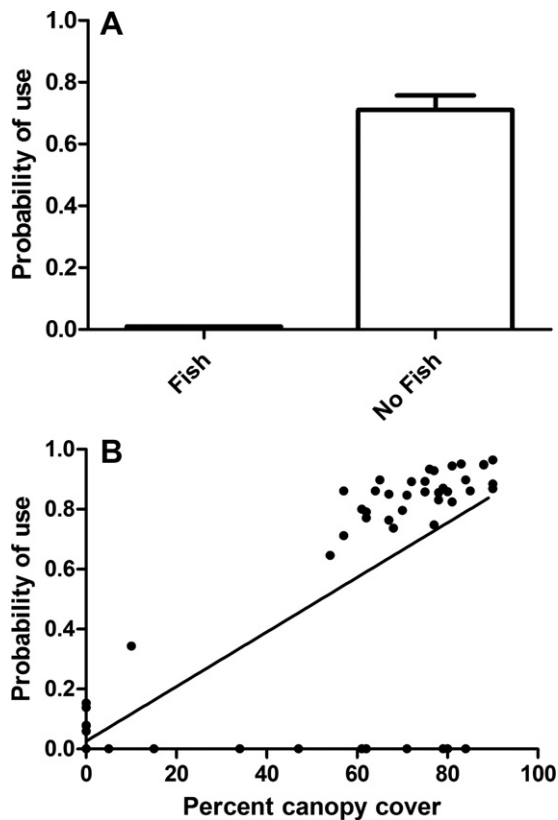| Parameter	β	Lower 95% CI	Upper 95% CI
Fish	−2.925	−3.872	−1.978
Canopy cover	1.726	0.534	2.918
Elevation	0.427	−0.658	1.512
Agriculture	1.354	−0.434	3.141

Fig. 2. Model-averaged estimates of probability of breeding pond use for *Ambystoma jeffersonianum* in east-central Illinois: (A) presence/absence of fish and (B) canopy cover ($R^2 = 0.543$; $SS_{res} = 4.311$, $p < 0.001$).

Fig. 3. Relationship between Maxent model probability predictions and occupancy model probability of use estimates ($R^2 = 0.400$; $SS_{res} = 3.323$; $p < 0.001$).

correctly classified by the model. We found that model discrimination was hindered by the presence of predatory fish in some ponds with high predicted probability values. When we removed ponds with fish from the analysis, the discriminatory ability of our model was substantially improved (AUC = 0.902; 95% CI = 0.819–1.00; Kappa = 0.812) with 90% of ponds being correctly classified. The ability of our 2008 model to discriminate suitable ponds gave us confidence to use it as a guide for 2009 surveys. Of the 26 new ponds surveyed in 2009, our model predicted that 13 ponds would be occupied and 13 ponds would be unoccupied. We had 80% accuracy in assigning these ponds correctly (AUC = 0.821; 95% CI = 0.606–0.968; Kappa = 0.615). When ponds with fish were removed, our accuracy increased to 95% (AUC = 0.957; 95% CI = 0.821–1.00; Kappa = 0.873).

Our full model including all data from 2008 and 2009 surveys had very good discrimination (AUC = 0.878; 95% CI = 0.772–0.983; Kappa = 0.709) with 86% of ponds classified correctly by the model. Once again, when ponds containing fish were removed, the model improved (AUC = 0.9457; CI = 0.863-1.00; Kappa = 0.782) with 91% of ponds being correctly classified. Our full model had good sensitivity and specificity, which are the conditional probabilities that occupied or unoccupied ponds, respectively, are correctly classified as such (Fielding & Bell 1997). The Maxent model was accurate in correctly assigning ponds unoccupied by *A. jeffersonianum* for both data sets (fish and no fish; specificity = 0.909), but our model had greater ability to correctly assign occupied ponds after removing ponds containing fish (sensitivity = 0.800 with fish, 0.909 without fish). Of the 57 ponds surveyed, 13 were found to have fish. Three ponds were incorrectly predicted to have salamanders in the model that included fish (commission error), but only one pond was incorrectly predicted when ponds with fish were excluded. There were five ponds that were occupied by salamanders, but
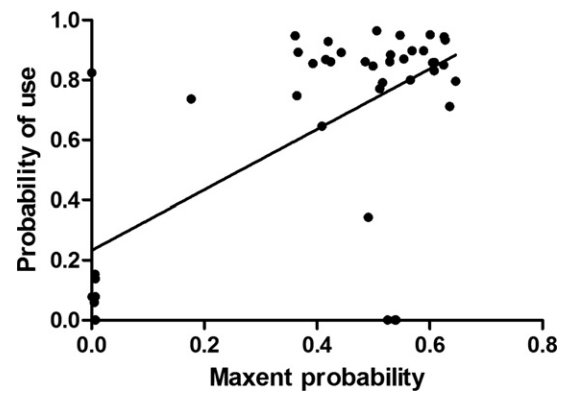
were incorrectly predicted to be unoccupied in the model with fish ponds included; only three ponds that were occupied were incorrectly predicted to be unoccupied when ponds containing fish were removed. Overall, there was a significant correlation between Maxent model probability values and predicted pond occupancy values from the occupancy modeling (Fig. 3; $R^2 = 0.400$, $p < 0.001$). From the sensitivity-specificity analysis, a critical Maxent-model probability value was estimated (i.e., the threshold at which Maxent model predictions correctly predicted site occupancy). Applying this critical threshold for correct classification to our continuous probability Maxent model, we found that approximately 59% of the forested habitat in our study region is conducive to supporting *A. jeffersonianum*.

Discussion

Using easily obtained GIS layers, we successfully developed a distribution model capable of identifying forested areas likely to support *A. jeffersonianum*, while accurately discriminating unlikely forested areas. Although our model performed well with the full detection/non-detection data set, accuracy was hindered by the presence of fish in ponds with a high probability of occurrence. This represents a limitation of our distribution model as parameterised. Distribution models such as Maxent are largely unable to incorporate biotic or physiological interactions (Guisan & Thuiller 2005). Presence of *A. jeffersonianum* was heavily influenced by the presence of fish (Table 2 and Fig. 2a). Amphibian eggs and larvae are readily depredated by fish (Petranka, Kats, & Sih 1987; Semlitsch & Gibbons 1988; Semlitsch 1988), leading many pool-breeding amphibian species to exhibit a strong aversion to water bodies containing fish (Binckley & Resetarits 2008; Hopey & Petranka 1994). It should be pointed out that our distribution model identified the fundamental niche encompassing the terrestrial habitat conditions conducive to harbouring breeding ponds of *A. jeffersonianum*, but no intra-pond characteristics were incorporated in the regional distribution model. Since we do not have detailed, regional information on the presence of fish in all ponds across the landscape, this will remain a constraint of distribution modeling of pond use by *A. jeffersonianum*, and further emphasises the need for a multi-faceted approach to link the fundamental niche with the realised niche.

Within suitable forested areas, ponds used by *A. jeffersonianum* for reproduction had two features in common: no fish and high canopy cover. The presence of predatory fish negatively influenced the presence of *A. jeffersonianum*, and there was a significant increase in the probability of ponds being occupied as the percentage of canopy cover increased (Table 3 and Fig. 2b). As a forest-dependent species that is most often associated with upland forest habitats (Petranka 1998), this association with canopy cover

is not unexpected. Even though agricultural practices can have negative effects on amphibians (e.g. Williams & Semlitsch 2010), distance to the nearest agricultural field was not a significant predictor in our models (Table 2). Although not significant, the average distance from agriculture of ponds occupied by *A. jeffersonianum* was nearly twice the distance of unoccupied ponds (94.2 m vs. 56.1 m). Despite a lack of local effects on pond use, the agricultural matrix severely fragments the forested landscape (Fig. 1), which may have implications for population connectivity and rescue-colonisation processes found in amphibian metapopulations (Greenwald, Purrenhage, & Savage 2009; Marsh & Trenham 2001).

Although we estimate that more than 50% of the forested habitat is potentially suitable for *A. jeffersonianum*, successful reproduction is limited by availability of fish-free ponds. There was a notable loss of such suitable breeding pond habitat in our study region. In addition to the presence of fish, we found that many potential ponds in our study region have been lost or converted since the development of the 1987 wetland data layer. In trying to guide our surveys in 2009, we identified the 2008 Maxent model probability of occurrence for each pond in the Illinois wetland layer. We then systematically began our 2009 surveys using the top-ranked ponds. Of the top 50 ponds, 16 no longer existed (i.e., were drained), and the majority of the remaining 34 had been altered or modified such that they were no longer suitable (e.g. filled, dredged, inadequate hydrology, stocked with fish). This finding highlights two important points for our study. First, there appears to be a high rate of potential breeding habitat loss in our study area, which may affect local population persistence and overall regional connectivity. Second, because we spatially modeled the fundamental niche of *A. jeffersonianum*, the lack of accurate data to identify breeding ponds hinders our ability to fully assess the extent and status of *A. jeffersonianum*, with or without a predictive model to help direct surveys.

Our surveys for *A. jeffersonianum* in the poorly sampled east-central region of Illinois increased the number of known breeding ponds from 6 to 33. Our multi-trap, multi-night sampling scheme produced a high detection rate (0.774), which when combined with multiple surveys, led to a high degree of confidence that ponds where salamanders were not captured truly were unoccupied. Without this measure of detection, the credibility of sampling and ultimately our occupancy model validation would be severely hindered (MacKenzie et al. 2006). Additionally, analysis of our data in an occupancy framework with model selection afforded us the additional benefit of being able to identify critical environmental variables significant to the presence and detection of *A. jeffersonianum*, which are needed for any future recovery plans.

Even though we focused our modeling and sampling efforts on a single species, this approach is broadly applicable to any species that has specific habitat requirements, particularly those with ties to microclimate that varies topographically. In our modeling, we used 30 m resolution GIS layers of topographic and landscape variables, giving us a relatively fine resolution across our study region. The majority of distribution models published to date use climate or climate-derived data (e.g. http://www.worldclim.org/) or soils data, which are generally not available at resolutions less than 1 km. The main limitation of models made from coarse resolution data are that they may not be appropriately scaled to the data used in model construction, may not appropriately match the spatial extent of the study region, or may not match the scale of the process being modeled (Elith & Leathwick 2009; Pearson, Dawson, & Liu 2004). Whenever possible, it is important to match scales in model construction and application. Depending on the characteristics of the species and the scale at which sampling is to occur, 1 km resolution may be appropriate, but for generally philopatric species like *A. jeffersonianum*, models with coarse resolution will

not suffice to accurately discriminate at a local scale. Future fine-scale distribution modeling can make use of the many abiotic data layers that can be easily derived from digital elevation models (DEM) within a GIS (TauDEM Tools: Tarboton 2010; Topography Tools: Dilts 2010; DEM Surface Tools: Jenness 2011). The resolution of these data layers is only constrained by the resolution of the DEM, which for many regions is now 10 m or less (for U.S.A. data see http://seamless.usgs.gov). Any organism that has microhabitat associations with moisture or temperature may be readily modeled using these data.

Maxent and other distribution models have seen extensive and increasing use, and their application to presence-only data continues to be evaluated (e.g. Bean, Stafford, & Brashares 2012; Elith & Graham 2009; Elith et al. 2006; Hijmans 2012; Phillips & Dudík 2008). Many researchers turn to these methods because of their ease of use, and relatively basic, obtainable data requirements. It should be noted that spatial distribution or abundance maps can be made from detection/non-detection data (Royle, Kéry, Gautier, & Schmid 2007; Wilson, Odei, & Hooten 2010), and methods to conduct such analyses have recently been implemented in program 'unmarked' (Fiske & Chandler 2011), which is an R package for conducting analysis of hierarchical models. With repeated observations of sites to account for imperfect detection, and spatially indexed covariates, maps depicting occupancy or abundance can be produced. Nonetheless, data sets capable of utilising methods to correct for imperfect detection may be financially or logistically infeasible.

Future application of this multi-model or other spatial modeling approaches can be carried out in a few different ways. First, any known locality records (e.g. museum or herbarium records) can be used to develop a distribution model for the region of interest, but these models may have inadequate discriminatory resolution or accuracy due to a scarcity or imprecision of localities. Alternatively, a small-scale pilot survey can be conducted to obtain accurate locality information and refine sampling protocols. From the pilot data, a distribution model for the region of interest can be made to direct survey efforts. It is also important to use a rigorous sampling protocol to gain a high degree of confidence that the target species is truly absent from a locality. Many species, especially species of conservation concern, have low or variable detection rates (e.g. Bailey, Simons, & Pollock 2004a; De Wan et al. 2009), and thus require multiple periods and/or methods to confirm presence/absence (Platts et al. 2010). Distribution models can then be iteratively updated as novel localities are confirmed to incorporate more information into the model. Such an iterative approach has been shown to increase new population identification by up to 4-fold over random sampling in a similar geographic region (Guisan et al. 2006). If enough repeated observations are made in space and time, a shift from a presence-only (e.g. Maxent) distribution model to a distribution or abundance model that corrects for imperfect detection could then be instituted. This iterative process is invaluable for the identification of new populations, identification of local and landscape factors related to species presence and persistence, environmental characteristics associated with high detection rates (important for optimising sampling protocols), and finally an adaptable model that can be used to guide future surveys and/or identify highly suitable regions for habitat restoration, preservation, or species translocation.

Our natural landscapes are continuing to be altered at an unprecedented rate, resulting in the loss of species diversity. In order to manage, preserve, or restore regional biodiversity, it is necessary to have an accurate assessment of species distributions and critical habitat requirements. With limited resources for such surveys, focus and direction are of utmost importance. Our case study with *A. jeffersonianum* outlines an approach that is broadly applicable across taxa, and can be used to direct survey efforts, confirm

occupancy status, and determine critical habitat features associated with species occurrence.

Acknowledgments

References

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*, 1223–1232. http://dx.doi.org/10.1111/j.1365-2664.2006.01214x

Andelt, W. F., White, G. C., Schnurr, P. M., & Navo, K. W. (2009). Occupancy of random plots by white-tailed and Gunnison's prairie dogs. *Journal of Wildlife Management*, *73*, 35–44. http://dx.doi.org/10.2193/2007-312

Bailey, L. L., Hines, J. E., Nichols, J. D., & MacKenzie, D. I. (2007). Sampling design trade-offs in occupancy studies with imperfect detection: Examples and software. *Ecological Applications*, *17*, 281–290.

Bailey, L. L., Simons, T. R., & Pollock, K. H. (2004a). Estimating detection probability parameters for plethodon salamanders using the robust capture-recapture design. *Journal of Wildlife Management*, *68*, 1–13.

Bailey, L. L., Simons, T. R., & Pollock, K. H. (2004b). Estimating site occupancy and species detection probability parameters for terrestrial salamanders. *Ecological Applications*, *14*, 692–702.

Bean, W. T., Stafford, R., & Brashares, J. S. (2012). The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography*, *35*, 250–258. http://dx.doi.org/10.1111/j.1600-0587.2011.06545.x

Binckley, C. A., & Resetarits, W. J. (2008). Oviposition behavior partitions aquatic landscapes along predation and nutrient gradients. *Behavioral Ecology*, *19*, 552–557.

Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., Rylands, A. B., Konstant, W. R., et al. (2002). Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, *16*, 909–923.

Brown, J. H., & Kodric-Brown, A. (1977). Turnover rates in insular biogeography: Effect of immigration on extinction. *Ecology*, *58*, 445–449. http://dx.doi.org/10.2307/1935620

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, USA: Springer-Verlag.

Campagne, P., Smouse, P. E., Varouchas, G., Silvain, J. F., & Leru, B. (2012). Comparing the van Oosterhout and Chybicki-Burczyk methods of estimating null allele frequencies for inbred populations. *Molecular Ecology Resources*, *12*, 975–982. http://dx.doi.org/10.1111/1755-0998.12015

De Wan, A. A., Sullivan, P. J., Lembo, A. J., Smith, C. R., Maerz, J. C., Lassoie, J. P., et al. (2009). Using occupancy models of forest breeding birds to prioritise conservation planning. *Biological Conservation*, *142*, 982–991. http://dx.doi.org/10.1016/j.biocon.2008.12.032

Dilts, T. (2010). *Topography Tools for ArcGIS v. 9.3.* <http://arcscriptsesricom/detailsasp?dbid=15996>.

Elith, J., & Graham, C. H. (2009). Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, *32*, 66–77.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*, 129–151.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and predictions across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677–697. http://dx.doi.org/10.1146/annurev.ecolsys.110308.120159

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*, 38–49.

Fiske, I. J., & Chandler, R. B. (2011). Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, *43*.

Gormley, A. M., Forsyth, D. M., Griffioen, P., Lindeman, M., Ramsey, D. S., Scroggie, M. P., et al. (2011). Using presence-only and presence–absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology*, *48*, 25–34.

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, *84*, 2809–2815. http://dx.doi.org/10.1890/02-3114

Greenwald, K. R., Purrenhage, J. L., & Savage, W. K. (2009). Landcover predicts isolation in *Ambystoma* salamanders across region and species. *Biological Conservation*, *142*, 2493–2500.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A., et al. (2006). Using niche-based models to improve the sampling of rare species. *Conservation Biology*, *20*, 501–511. http://dx.doi.org/10.1111/j.1523-1739.2006.00354.x

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

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135.

Hamer, A., & Mahony, M. (2010). Rapid turnover in site occupancy of a pond-breeding frog demonstrates the need for landscape-level management. *Wetlands*, *30*, 287–299.

Hijmans, R. J. (2012). Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null model. *Ecology*, *93*, 679–688. http://dx.doi.org/10.1890/11-0826.1

Hopey, M. E., & Petranka, J. W. (1994). Restriction of wood frogs to fish-free habitats: How important is adult choice? *Copeia*, 1023–1025.

Jenness, J. (2006). *Topographic Position Index (tpi_jen.avx) extension for ArcView 3.x, v. 1.3a.* Jenness Enterprises Available at: <http://wwwjennessentcom/arcview/tpihtm>.

Jenness, J. (2011). *DEM Surface Tools v. 2.1.292.* Jenness Enterprises: <http://www.jennessent.com/arcgis/surface_area.htm>.

Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, *21*, 498–507. http://dx.doi.org/10.1111/j.1466-8238.2011.00683.x

Kroll, A. J., Risenhoover, K., McBride, T., Beach, E., Kernohan, B. J., Light, J., et al. (2008). Factors influencing stream occupancy and detection probability parameters of stream-associated amphibians in commercial forests or Oregon and Washington, USA. *Forest Ecology and Management*, *255*, 3726–3735.

Lannoo, M. J. (Ed.). (2005). *Amphibian declines: The conservation status of United States species.* Berkeley: University of California Press.

Lobo, J. M., Jimenez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*, 145–151.

MacKenzie, D. I. (2006). Modeling the probability of resource use: The effect of, and dealing with, detecting a species imperfectly. *Journal of Wildlife Management*, *70*, 367–374.

MacKenzie, D. I., & Kendall, W. L. (2002). How should detection probability be incorporated into estimates of relative abundance? *Ecology*, *83*, 2387–2393.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2006). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence.* Boston, USA: Academic Press.

Marsh, D. M., & Trenham, P. C. (2001). Metapopulation dynamics and amphibian conservation. *Conservation Biology*, *15*, 40–49.

Milanovich, J. R., Peterman, W. E., Nibbelink, N. P., & Maerz, J. C. (2010). Projected loss of a salamander diversity hotspot as a consequence of projected global climate change. *PLoS One*, *5*, e12189. http://dx.doi.org/10.1371/journal.pone.0012189

Newbold, T., Reader, T., El-Gabbas, A., Berg, W., Shohdi, W. M., Zalat, S., et al. (2010). Testing the accuracy of species distribution models using species records from a new field survey. *Oikos*, *119*, 1326–1334. http://dx.doi.org/10.1111/j.1600-0706.2009.18295.x

Pearson, R. G., Dawson, T. P., & Liu, C. (2004). Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, *27*, 285–298.

Petranka, J. W. (1998). *Salamanders of the United States and Canada.* Washington, DC: Smithsonian Institution Press.

Petranka, J. W., Kats, L. B., & Sih, A. (1987). Predator–prey interactions among fish and larval amphibians: Use of chemical cues to detect predatory fish. *Animal Behaviour*, *35*, 420–425.

Phillips, C. A., Brandon, R. A., & Moll, E. O. (1999). *Field guide to amphibians and reptiles of Illinois, Illinois Natural History Manual 8.* Illinois Natural History Manual 8.

Phillips, S. J. (2008). Transferability, sample selection bias and background data in presence-only modelling: A response to Peterson et al. (2007). *Ecography*, *31*, 272–278.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259.

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, *31*, 161–175.

Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. In *Proceedings of the twenty-first international conference on machine learning* 69, (pp. 83–99).

Platts, P. J., Ahrends, A., Gereau, R. E., McClean, C. J., Lovett, J. C., Marshall, A. R., et al. (2010). Can distribution models help refine inventory-based estimates of conservation priority? A case study in the Eastern Arc forests of Tanzania and Kenya. *Diversity and Distributions*, *16*, 628–642. http://dx.doi.org/10.1111/j.1472-4642.2010.00668.x

Raes, N., & ter Steege, H. (2007). A null-model for significance testing of presence-only species distribution models. *Ecography*, *30*, 727–736.

Rasic, G., & Keyghobadi, N. (2012). The pitcher plant flesh fly exhibits a mixture of patchy and metapopulation attributes. *Journal of Heredity*, *103*, 703–710. http://dx.doi.org/10.1093/jhered/ess040

Raxworthy, C. J., Martinez-Meyer, E., Horning, N., Nussbaum, R. A., Schneider, G. E., Ortega-Huerta, M. A., et al. (2003). Predicting distributions of known and unknown reptile species in Madagascar. *Nature, 426*, 837–841.

Rebelo, H., & Jones, G. (2010). Ground validation of presence-only modelling with rare species: A case study on barbastelles Barbastella barbastellus (Chiroptera: Vespertilionidae). *Journal of Applied Ecology, 47*, 410–420.

Royle, J. A., Kéry, M., Gautier, R., & Schmid, H. (2007). Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs, 77*, 465–481.

Schmidt, B. R., & Pellet, J. (2005). Relative importance of population processes and habitat characteristics in determining site occupancy of two anurans. *Journal of Wildlife Management, 69*, 884–893.

Schwegman, J. E. (1973). *Comprehensive plan for the Illinois Nature Preserves System, Part 2*. Rockford, IL: Illinois Nature Preserves Commission.

Semlitsch, R. D. (1988). Allotopic distribution of two salamanders: Effects of fish predation and competitive interactions. *Copeia*, 290–298.

Semlitsch, R. D., & Gibbons, J. W. (1988). Fish predation in size-structured populations of treefrog tadpoles. *Oecologia, 75*, 321–326.

Stuart, S. N., Chanson, J. S., Cox, N. A., Young, B. E., Rodrigues, A. S. L., Fischman, D. L., et al. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science, 306*, 1783–1786.

Tarboton, D. G. (2010). *TauDEM Tools 5.0*. <http://hydrologyusuedu/taudem/taudem50/>.

Urbina-Cardona, J. N., & Flores-Villela, O. (2010). Ecological-niche modeling and prioritisation of conservation-area networks for mexican herpetofauna. *Conservation Biology, 24*, 1031–1041.

Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science, 277*, 494–499.

Williams, B. K., & Semlitsch, R. D. (2010). Larval responses of three Midwestern anurans to chronic, low-dose exposures of four herbicides. *Archives of Environment Contamination and Toxicology, 58*, 819–827.

Williams, J. N., Changwan, S., Thorne, J., Nelson, J. K., Erwin, S., O'Brien, J. M., et al. (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions, 15*, 565–576. http://dx.doi.org/10.1111/j.1472-4642.2009.00567.x

Williams, P. K. (1973). *Seasonal movements and population dynamics of four sympatric mole salamanders, genus Ambystoma*. Bloomington, IN: Indiana University.

Wilson, T. L., Odei, J. B., & Hooten, M. B. T. C. E., Jr. (2010). Hierarchical spatial models for predicting pygmy rabbit distribution and relative abundance. *Journal of Applied Ecology, 47*, 401–409.