Predicting patch occupancy in fragmented landscapes at the rangewide scale for an endangered species: an example of an American warbler

Bret A. Collier1*, Julie E. Groce2, Michael L. Morrison3, John C. Newnam4, Andrew J. Campomizzi3, Shannon L. Farrell1, Heather A. Mathewson1, Robert T. Snelgrove1, Raymond J. Carroll5 and Robert N. Wilkins1

ABSTRACT

Aim Our objective was to identify the distribution of the endangered golden-cheeked warbler (*Setophaga chrysoparia*) in fragmented oak–juniper woodlands by applying a geoadditive semiparametric occupancy model to better assist decision-makers in identifying suitable habitat across the species breeding range on which conservation or mitigation activities can be focused and thus prioritize management and conservation planning.

Location Texas, USA.

Methods We used repeated double-observer detection/non-detection surveys of randomly selected (n = 287) patches of potential habitat to evaluate warbler patch-scale presence across the species breeding range. We used a geoadditive semiparametric occupancy model with remotely sensed habitat metrics (patch size and landscape composition) to predict patch-scale occupancy of golden-cheeked warblers in the fragmented oak–juniper woodlands of central Texas, USA.

Results Our spatially explicit model indicated that golden-cheeked warbler patch occupancy declined from south to north within the breeding range concomitant with reductions in the availability of large habitat patches. We found that 59% of woodland patches, primarily in the northern and central portions of the warbler’s range, were predicted to have occupancy probabilities ≤0.10 with only 3% of patches predicted to have occupancy probabilities >0.90. Our model exhibited high prediction accuracy (area under curve = 0.91) when validated using independently collected warbler occurrence data.

Main conclusions We have identified a distinct spatial occurrence gradient for golden-cheeked warblers as well as a relationship between two measurable landscape characteristics. Because habitat-occupancy relationships were key drivers of our model, our results can be used to identify potential areas where conservation actions supporting habitat mitigation can occur and identify areas where conservation of future potential habitat is possible. Additionally, our results can be used to focus resources on maintenance and creation of patches that are more likely to harbour viable local warbler populations.

Keywords Bayesian inference, golden-cheeked warbler, habitat conservation, occupancy, semiparametric regression, *Setophaga chrysoparia*.

INTRODUCTION

Species inhabiting human-dominated environments often exist in locations where habitat loss and fragmentation have reduced patch contiguity, patch size, and increased edge and isolation effects (Marzluff, 2001; Bolger, 2002). Such changes in structural features at the local scale also influence dynamics in surrounding areas (Forman, 1995; Saab, 1999). Moreover,
fragmentation can create ecological thresholds for demographic values such as population size (Homan et al., 2004; Betts et al., 2007), survival (Ruiz-Gutiérrez et al., 2008), dispersal (Bayne & Hobson, 2002) and reproduction (Robinson et al., 1995; Lloyd et al., 2005). Such thresholds can potentially increase local extinction risk for species in areas under intensive land use pressures (Lande, 1998). Whether or not a species is structured as a metapopulation, knowledge of how habitat is occupied through space allows predictions to be made on colonization and extinction probabilities (Mills, 2007:265).

A core issue of conservation biology is the distribution of individuals through space, typically with a focus on the relationship between availability, size and proximity of potential habitats to one another. Information on how environmental metrics predict habitat use across space and time provides the foundation for population management and conservation strategies (Sagarin et al., 2006; Brotons et al., 2007). As such, environmental metrics from locations where presence-absence surveys have been conducted represent the basis for predictive modelling of species distributions and for monitoring changes in distributions as environmental conditions change (Moore & Swihart, 2005; Elith et al., 2006; MacKenzie, 2006; Syphard & Franklin, 2009). Species inhabiting human-dominated environments present challenges for habitat modelling because they often exist in locations where habitat loss and fragmentation has reduced patch contiguity and patch size and increased edge and isolation effects (Bolger, 2002). Thus, it is likely that the underlying distribution process varies nonlinearly in space wherein we would expect, for instance, that conditions will become less similar as spatial proximity declines (e.g. spatial autocorrelation; Augustin et al., 1996; Royle et al., 2007). In general, the impact of incorporating spatial relationships into predictive models is an attempt to create a proxy for addressing unmeasurable or unidentifiable environmental metrics, which otherwise would not be incorporated into model structure (Tognelli & Kelt, 2004).

Our goal was to predict patch occupancy in a fragmented landscape for the golden-cheeked warbler (Setophaga chrysoparia) across the entirety of its breeding range. The golden-cheeked warbler was listed as federally endangered in the United States in 1990 owing to concerns about habitat loss within the warbler’s restricted breeding range (88,878 km² in central Texas, USA; Fig. 1a; U.S. Fish and Wildlife Service 1992). The warblers’ endemism to central Texas during breeding is driven by its relationship to the oak (Quercus spp.)–Ashe juniper (Juniperus ashei) woodland communities that provide foraging habitat, nesting cover, and shredded bark from Ashe juniper for nest construction (Pulich, 1976; Ladd & Gass, 1999). The breeding range of the warbler in central Texas has seen an increase in human populations of approximately 50% since species listing (Groce et al., 2010). Previous studies of the warbler in Texas have focused on public lands (Anders & Dearborn, 2004; Peak, 2007; Reidy et al., 2008), which represent <5% of land within the warblers range; hence, few data exist for the accurate assessment of rangewide warbler distribution and factors affecting distribution (DeBoer & Diamond, 2006; Collier et al., 2010). Several recent attempts to map the distribution of potential warbler habitat within the breeding range (Diamond, 2007; SCWA 2007, Loomis Austin 2008) used remotely sensed vegetative conditions that were deemed appropriate for warblers and classified patches into qualitative categories representing habitat quality (e.g. high or low quality) based on expert opinion and limited field data. However, these qualitative assessments of potential habitat were unable to quantitatively estimate likelihood of warbler presence across the species range, thus limiting their usefulness for conservation planning.

We developed a geoadditive semiparametric occupancy model (Ruppert et al., 2003; Crainiceanu et al., 2005; Gimenez et al., 2006) for repeated detection/non-detection survey data to predict patch-specific occupancy of the golden-cheeked warbler across the breeding range in Texas. We corrected for imperfect detection (Royle & Kéry, 2007; Royle et al., 2007) and allowed the spatial relationship between patch occupancy probabilities to be determined as a function of a nonparametric interaction (Ruppert et al., 2003; Gimenez et al., 2006; Grosbois et al., 2009). The geoadditive component of our model represents the merging of the underlying semiparametric model with the spatial aspect provided by the spline basis

Figure 1 Distribution (a) of woodland patches (n = 63,616) and posterior predicted spatial process (b) centred at the regression means for the geoadditive semiparametric model containing patch size, landscape composition, X and Y location, and the interaction between patch size and landscape composition within the 35-county breeding range of the golden-cheeked warbler in Texas, USA.
functions as detailed in Kammann & Wand (2003) and Ruppert et al. (2003). Our objective was to predict those areas of high and low occurrence probability to better assist decision-makers in prioritizing conservation, management or mitigation activities across the golden-cheeked warbler’s breeding range. We also detail how our sampling and analytical approach can be broadly applied to many other species in habitating fragmented landscapes.

METHODS

Study area and biological covariates

Golden-cheeked warblers are endemic to oak–juniper woodlands (Pulich, 1976; Ladd & Gass, 1999); thus, we defined woodland patches as the sampling unit on which we measured occurrence (Collier et al., 2010). We delineated woodland patches based on spring 2007 and 2008 LANDSAT five imagery (30 m pixel) that maximized vegetative spectral differences after leaf emergence in spring and prior to the stresses associated with heat and drought in the summer (Collier et al., 2010). We conducted an unsupervised classification of woodlands across central Texas, aggregating land cover types into two classes (oak–juniper woodland and other) and used the 2001 National Land Cover Dataset (NLCD) to mask any notable areas (e.g. croplands, wetlands) that were misclassified as woodlands by our unsupervised classification. We created breaks between patches traversed by paved roads using the Texas strategic mapping program (STRATMAP) by intersecting road data with our woodland classification and deleting any woodland classified pixels that intersected the road layer. We assigned each patch of woodland habitat to an administrative unit based on US Fish and Wildlife Service Recovery Regions (USFWS 1992). For each woodland patch, we calculated patch size (Collier et al., 2010) and landscape composition (Magness et al., 2006) using ESRI ArcGIS 10 as both are known to influence warbler presence (see Appendices S1 and S2 in Supporting Information). We estimated landscape composition for each patch as the mean percentage of woodlands within a 400-m radius circle surrounding a given pixel as this radius was determined to capture landscape variation relevant to warbler presence at the patch scale (Magness et al., 2006). The mean value for all pixels within a patch was used as the landscape composition estimate of the patch for analysis.

Sample size and survey methodology

We determined minimum sample size following MacKenzie and Royle (2005). We used a probability proportional to size sampling design (PPS; Thompson 2002) and selected patches randomly in proportion to size for surveying so not to over- or under-weight our sample frame because of the non-normal distribution of patches sizes across the range. We focused our survey efforts on patches ≤200 ha owing to our knowledge of size-based threshold effects in species presence (He & Gaston, 2000; Butcher et al., 2010; Collier et al., 2010). Because we anticipated access restrictions to patches on private lands (Hilty & Merenlender, 2003), we created a randomly selected sampling frame five times greater than the minimum sample size. If we were unable to obtain access to the selected patch, we contacted landowners of the next randomly selected patch akin to assuming that missing or inaccessible properties were missing completely at random (Stevens & Jensen, 2007). We supplemented our random sampling with patches on public and private properties that fell within the bounds of our sampling design on which we currently had access. We assumed throughout our study that access to property was not influenced by known or perceived warbler presence or absence and that local management and access were unrelated, and we made no assumptions that any one patch surveyed would have a greater or lesser likelihood of warbler presence. We used standardized protocols for surveying regardless of patch ownership. We attempted to survey the entirety of every accessed patch, and a patch was contained on multiple properties where some access was restricted, and we assumed that the probability of detecting warblers did not differ between accessible and inaccessible areas of the patch.

We conducted auditory and visual surveys for warblers between mid-March and late May 2009 to determine patch occupancy (Collier et al., 2010). Two simultaneous independent observers surveyed each patch systematically for warbler presence–absence (MacKenzie, 2006). If at least one observer detected a warbler within a patch during a survey period, we did not revisit the patch (e.g. a removal approach, MacKenzie, 2006). Based on previous research (Collier et al., 2010), if no warblers were detected during a survey, it was necessary to resurvey patches up to a maximum of six times (three double-observer surveys) in an attempt to detect warblers. While most patches (>95%) were surveyed ≤6 times, several were surveyed one additional time by double observers, so we used a maximum number of survey occasions of 8 for our analysis. Thus, our data represented repeated detection/non-detection surveys of \(i = 1, 2, \ldots, R\) patches of warbler habitat where each site was surveyed \(j = 1, 2, \ldots, J\) times to determine warbler detection/non-detection during the 2009 survey season. Data resulting from our repeated presence–absence surveys represent a capture history (e.g. 0011\(\ldots\)) wherein a positive detection is given a 1, no detection is given a 0, and no survey conducted is designated using ‘.’ notation (White and Burnham 1999).

Analysis

Although there are many statistical methods used to predict species distributions (Guisan & Zimmermann, 2000; Elith et al., 2006; Guisan et al., 2006; MacKenzie, 2006), generalized linear (GLM) and additive (GAMs) models represent two popular approaches for presence–absence data (Guisan & Zimmermann, 2000; Guisan et al., 2006; Syphard & Franklin, 2009). Models such as these are useful for modelling linear and nonlinear effects of biological covariates (Kammann & Wand,
modelled the detection process (Royle et al., 2006; Royle et al., 2007). Thus, we used geoaditive semiparametric regression (Ruppert et al., 2003; Gimenez et al., 2006, 2009) to model golden-cheeked warbler patch occupancy and associated detection probabilities across the species breeding range in Texas.

The dependent variable in our analysis was the presence or absence of golden-cheeked warblers in patches of oak–juniper woodland. We modelled patch-specific occupancy probability with patch size, landscape composition and their interaction entering the model linearly with spatial predictors (latitude and longitude) incorporated into the model as a nonparametric interaction (Ruppert et al., 2003; Crainiceanu et al., 2005). We predicted occupancy probability ($\psi_i$) as a function of those covariate data and spatial location within the range expressed as

$$
\log_2(\psi_i) = \beta_0 + \beta_1 X_1 + \sum_{k=1}^{20} u_k (\text{Location}_i - \kappa_k) + e_i
$$

where the $\beta X_i$ represents a vector of $l$ predictor variables (patch size, landscape composition, patch size–landscape composition interaction, and patch-specific UTM) entering the model linearly and $(\text{Location}_i - \kappa_k)$ represents the spatial effect for each surveyed habitat patch (Gimenez et al., 2009). Spatial relationships in our model used radial basis penalized splines (Ruppert et al., 2003), which can be efficiently modelled in a generalized linear mixed model framework and have the added benefit of being rotationally invariant when used for geographical smoothing (Ruppert et al., 2003). We fit the above model using $k = 20$ knots (Ruppert et al., 2003; Crainiceanu et al., 2005), which ensured adequate flexibility and used the space-filling algorithm of Nychka & Saltzman (1998) and the R package fields (Fields Development Team 2006) to select knot locations within our landscape. Our model was adapted for use as a single-season occupancy model (MacKenzie, 2006) and relied on the penalized spline structure detailed by Ruppert et al. (2003), Crainiceanu et al. (2005) and Gimenez et al. (2006, 2009). We have provided annotated WinBUGS code adapted to our specific study as an Appendix S3.

Detection model

Many species distribution modelling approaches (Knapp et al., 2003; and reviews by Elith et al., 2006; Guisan et al., 2006) do not mention detection rates when discussing modelling methods even though variable detection rates can have significant impacts on distribution model predictions (MacKenzie, 2006; Royle & Kéry, 2007; Kéry et al., 2010). Because our model was hierarchical in nature, we accounted for the impact of imperfect detection on our occupancy predictions by modelling the detection process (Royle et al., 2007). We used a temporal covariate representing survey date for detection modelling as date of survey has been shown to adequately predict detection rates of warblers at the patch scale (Collie et al., 2010). Thus, we addressed issues associated with imperfect detection using the linear logistic relationship (Kéry, 2008; Royle & Dorazio, 2008)

$$
\log_2(p_0) = \alpha_0 + \alpha_1 \text{Day},
$$

where Day represents the numeric day since 15 March 2009 and continuing through the end of the breeding season.

Bayesian inference

We adopted a Bayesian approach that has been shown to be computationally efficient for hierarchical generalized linear mixed models with radial basis splines (Wood et al., 2002; Ruppert et al., 2003; Crainiceanu et al., 2005; Gimenez et al., 2009; King et al., 2010). We provided a set of prior distributions for all model parameters to fully specify our model (Royle & Dorazio, 2008; King et al., 2010). We used normal prior's $N(0,10^0)$ on the $\beta$'s that enter linearly into our model and specified independent, normal priors on random effect parameters $u_k \sim (0, \tau_k)$ where $\tau_0 = \tau(0.1, 0.1)$ (Ruppert et al., 2003; Crainiceanu et al., 2005). We used normal priors $N(0,100)$ for the intercept and slope of the detection sub-model. We standardized covariates prior to analysis to assist with model convergence (Crainiceanu et al., 2005).

We performed all analysis using WinBUGS v. 1.4 (Spiegelhalter et al. 2003) and R (R Core Development Team, 2009) using R package R2WinBUGS (Sturz et al., 2005) for the Markov chain Monte Carlo (MCMC) simulations. Annotated R and WinBUGS code for running the semiparametric occupancy model and for predicting patch-specific occupancy is available as Appendix S3. We ran our MCMC algorithm for $1 \times 10^4$ iterations after a 50,000 iteration burn-in. We thinned every 100th iteration for model diagnostics and inference. We assessed model convergence based on the Gelman and Rubin statistic (Gelman & Rubin, 1992) and through residual evaluation (Ruppert et al., 2003) using R packages boa (Smith, 2007) and coda (Plummer et al., 2006).

Model evaluation

To evaluate our semiparametric model, we used independent survey data on golden-cheeked warblers in Texas. During 2003–2007 and 2010, we conducted observational studies in woodland patches across the warbler’s breeding range where presence–absence data at the patch scale was collected using those methods detailed in Sample size and survey methodology, which represented an optimal, data-driven approach for model validation (Guisan & Zimmerman, 2000). We followed advice of Guisan & Zimmermann (2000) and Elith et al. (2006) and developed receiver operating curves (ROC) and sensitivity/specificity comparisons using the independent survey data to evaluate predictive accuracy of our modelling approach (Sing et al., 2005).
RESULTS
We identified and assigned biological metrics to 63,616 patches of woodlands across our study area in Texas (Fig. 1a). Our remotely sensed habitat layer included approximately 1.678 million ha of woodlands. Approximately 70% of the patches were ≤10 ha in area and encompassed about 11% of total available habitat. In 2009, we surveyed 287 patches for warbler presence and had positive detections in 150 of the 287 patches, providing a naïve estimate of patch occupancy of 52%. Surveyed patches ranged from 2 to >11,000 ha in area, and we surveyed patches in 34 of the 35 counties in the warbler’s range.

Semiparametric model results
We provide summaries of the posterior distribution of model parameters in Table 1. Brooks–Gelman–Rubin diagnostics indicated model convergence with scale reduction factors for all parameters between 1.00 and 1.10 and a multivariate scale reduction factor of 1.08 (Gelman & Rubin, 1992). We developed a predictive surface of the spatial effect by predicting occupancy probability for locations on a 100-m grid distributed across our study area while holding all model parameters at their mean covariate value. There was substantial evidence of spatial variability affecting warbler distribution, as we found areas of higher relative occurrence probability on both the eastern border of the region roughly between 30°50’N and 31°55’N latitude, as well as in the south-western third of the study region south of 30°10’N and west of 99°00’W (Fig. 1b). Habitat patches towards the northern and western edge of the species range showed a much lower predicted occurrence probability of warblers.

Detection and occupancy predictions
As expected, the effect of sample survey date on detection probability was negative, indicating that detection probability declines as the breeding season progresses. Using the mean date (43 days since 15 March) for all detections, the posterior mean detection probability was 0.701. Based on this value, we estimated that the probability of not detecting a warbler when one was present within a patch of warbler habitat when surveyed by the mean date would be approximately (1–0.701)^8 = 0.00006, or effectively 0.

Our model predicted that 86% of the patches (see Methods: Study area and biological covariates) had occupancy probabilities ≤50%, and 59% had a predicted occupancy ≤10% (Table 2). Only 2.9% of patches were predicted to have occupancy probabilities >90% (Table 2), most of which were large patches in the south-western portion of the species breeding range (Fig. 1a). Patches with occupancy probabilities

Table 1 Posterior parameter estimates for the geoadditive semiparametric occupancy model applied to the golden-cheeked warbler survey data collected in Texas during 2009. The β terms represent the model intercept (β0), latitudinal spatial location (β1), longitudinal location (β2), patch size (β3), landscape composition (β4) and patch size-landscape composition interaction (β5). The µ terms represent the random effect terms for the 20 knot locations relative to each patch, and σ0 and σ1 represent the intercept and slope for the linear-logistic detection model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>β0</td>
<td>−11.56</td>
<td>3.530</td>
</tr>
<tr>
<td>β1</td>
<td>0.044</td>
<td>3.440</td>
</tr>
<tr>
<td>β2</td>
<td>−1.107</td>
<td>3.085</td>
</tr>
<tr>
<td>β3</td>
<td>−0.338</td>
<td>5.961</td>
</tr>
<tr>
<td>β4</td>
<td>0.676</td>
<td>6.004</td>
</tr>
<tr>
<td>β5</td>
<td>1.152</td>
<td>5.957</td>
</tr>
<tr>
<td>µ1</td>
<td>0.077</td>
<td>0.517</td>
</tr>
<tr>
<td>µ2</td>
<td>0.268</td>
<td>0.799</td>
</tr>
<tr>
<td>µ3</td>
<td>0.086</td>
<td>0.965</td>
</tr>
<tr>
<td>µ4</td>
<td>1.469</td>
<td>1.205</td>
</tr>
<tr>
<td>µ5</td>
<td>0.007</td>
<td>1.151</td>
</tr>
<tr>
<td>µ6</td>
<td>−0.891</td>
<td>1.154</td>
</tr>
<tr>
<td>µ7</td>
<td>−0.321</td>
<td>1.346</td>
</tr>
<tr>
<td>µ8</td>
<td>−0.725</td>
<td>1.352</td>
</tr>
<tr>
<td>µ9</td>
<td>1.430</td>
<td>1.439</td>
</tr>
<tr>
<td>µ10</td>
<td>−0.253</td>
<td>0.965</td>
</tr>
<tr>
<td>µ11</td>
<td>0.099</td>
<td>1.231</td>
</tr>
<tr>
<td>µ12</td>
<td>0.877</td>
<td>1.390</td>
</tr>
<tr>
<td>µ13</td>
<td>−0.341</td>
<td>1.258</td>
</tr>
<tr>
<td>µ14</td>
<td>1.546</td>
<td>1.344</td>
</tr>
<tr>
<td>µ15</td>
<td>0.116</td>
<td>1.219</td>
</tr>
<tr>
<td>µ16</td>
<td>−1.683</td>
<td>1.446</td>
</tr>
<tr>
<td>µ17</td>
<td>0.026</td>
<td>0.494</td>
</tr>
<tr>
<td>µ18</td>
<td>0.919</td>
<td>1.232</td>
</tr>
<tr>
<td>µ19</td>
<td>0.225</td>
<td>1.219</td>
</tr>
<tr>
<td>µ20</td>
<td>−0.816</td>
<td>1.236</td>
</tr>
<tr>
<td>σ0</td>
<td>1.101</td>
<td>0.141</td>
</tr>
<tr>
<td>σ1</td>
<td>−0.145</td>
<td>0.137</td>
</tr>
</tbody>
</table>

Table 2 Number of classified habitat patches, patch area (ha), and percentage of total potential patches and total area by predicted occupancy category based on a geoadditive semiparametric occupancy model for golden-cheeked warblers in Texas during 2009.

<table>
<thead>
<tr>
<th>Predicted occupancy</th>
<th>No. of patches</th>
<th>Total ha</th>
<th>Percentage of total patches</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.10</td>
<td>37,717</td>
<td>200,654</td>
<td>59.3</td>
<td>12.0</td>
</tr>
<tr>
<td>0.10–0.20</td>
<td>7341</td>
<td>103,439</td>
<td>11.5</td>
<td>6.2</td>
</tr>
<tr>
<td>0.20–0.30</td>
<td>4129</td>
<td>85,883</td>
<td>6.5</td>
<td>5.1</td>
</tr>
<tr>
<td>0.30–0.40</td>
<td>3022</td>
<td>77,761</td>
<td>4.8</td>
<td>4.6</td>
</tr>
<tr>
<td>0.40–0.50</td>
<td>2600</td>
<td>86,655</td>
<td>4.1</td>
<td>5.2</td>
</tr>
<tr>
<td>0.50–0.60</td>
<td>2107</td>
<td>82,346</td>
<td>3.3</td>
<td>4.9</td>
</tr>
<tr>
<td>0.60–0.70</td>
<td>1841</td>
<td>101,792</td>
<td>2.9</td>
<td>6.1</td>
</tr>
<tr>
<td>0.70–0.80</td>
<td>1710</td>
<td>134,357</td>
<td>2.7</td>
<td>8.0</td>
</tr>
<tr>
<td>0.80–0.90</td>
<td>1447</td>
<td>175,870</td>
<td>2.3</td>
<td>10.5</td>
</tr>
<tr>
<td>&gt;0.90</td>
<td>1702</td>
<td>629,941</td>
<td>2.7</td>
<td>37.5</td>
</tr>
<tr>
<td>Total</td>
<td>63,616</td>
<td>1,678,698</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
between 0.50 and 0.90 were more widely distributed across the range although occupancy estimates declined in the northern regions. Overall, the expected occurrence distribution at the rangewide scale indicated significant variation in occurrence probabilities dependent upon both spatial location and patch-specific metrics (Fig. 2).

**Model validation**

Using survey data collected during 2003–2007 and 2010 in woodland patches (n = 143), we classified patch detection/non-detection for comparative purposes with our occupancy predictions. We used a scoring classifier (Sing *et al.*, 2005) to visualize a ROC graph showing model prediction accuracy versus false-positive rate and sensitivity versus specificity (Fig. 3a,b). Our area under curve (AUC) estimate was high (0.91), indicating our model predicted reality based on our field survey data.

**DISCUSSION**

Population distribution is often limited by the amount of appropriate habitat available (Hanski & Gilpin, 1991) and the proximity or isolation of habitat patches (Shanahan & Possingham, 2006). Identification of potential habitat distribution and the ability to distinguish among patches of varying occupancy probabilities are important for driving conservation actions for species (Guisan & Thuiller, 2005). Our general approach to study design and analysis is most directly applicable to species where individuals are difficult to follow, such as smaller-bodied animals that cannot be readily sampled by telemetry and other marking methods (Fahrig, 2007:74–75).

We found that the majority (59%) of woodland patches within the breeding range of the golden-cheeked warbler were predicted to have ≤0.10 of being occupied. Thus, as it is infeasible to maintain all current habitat for warblers in Texas in perpetuity, our results can be used to (1) focus resources on maintenance of those patches with higher occupancy estimates that may be likely to harbour viable local warbler populations (He & Gaston, 2000; Collier *et al.*, 2010) and (2) identify locations where habitat management actions can assist in creating, maintaining or linking available habitat.

Our results indicate that warbler occurrence declined from south to north across the breeding range, which corresponded with a decrease in the proportion of large patches from south to north. The decrease in patch size was correlated with an overall decrease in environmental conditions supporting large patches with high canopy cover in the northernmost portion of the species range, and additionally, greater residential and
commercial development in the south-east portion of the range (Groce et al., 2010). Our data indicate that with an overall shift to smaller and more fragmented patches within the northern portions of the range, the probability of warbler occurrence declines significantly, even for large patches of woodland habitats.

Species distribution models incorporating imperfect detection and spatial relatedness are known to outperform standard GLMs in predictive accuracy (Royle & Kéry, 2007). Using our data as an example, consider two patches where one was in the northern edge of the warblers range and the other was in the south-western region of range. Both patches had similar patch size (25.4 and 25.9 ha) and landscape composition (57.1 and 57.2) but differed significantly in predicted occupancy from our model (northern patch occupancy, 0.10; south-western patch occupancy, 0.72). Based on our values, we can infer obvious differences in patch occupancy and hence use that information for focusing conservation at the landscape scale. However, if one were to use the same biological metrics in a standard GLM, occurrence probabilities for both patches would be 0.63, with a potential consequence of conservation effort focused naively, and possibly ineffectively, on the northern patch.

We adopted a set of biological covariates (patch size and landscape composition) that previous research had indicated was useful for predicting warbler occurrence at the patch scale (Magnness et al., 2006; Collier et al., 2010). Although we are confident that the covariates we used were appropriate for determining presence–absence across the range, it is likely that additional environmental metrics, such as within-patch tree communities, juniper density, relative age of oak–juniper woodland, or other metrics may help refine occupancy in small patches as well as provide information for better prediction of warbler abundance and fecundity. However, model evaluation indicated that our model’s predictive accuracy was more than adequate overall; thus, we are comfortable with the environmental metrics used. We assumed a fairly simple process for species detection rates, assuming that detection was solely a function of observation date (Collier et al., 2010). However, it is plausible that more detailed determination of the detection function could be accomplished via incorporation of additional variables for factors such as observers (e.g. differing abilities), patch metrics or potential interactions between spatial location and survey date. However, given the high detection rates of warblers during our surveys, we thought that additional metrics would add unnecessary complexity to our modelling approach.

Our approach used variables relevant to a landscape scale, which translated into a model that was insensitive to fine-scale variability in habitat composition. While we acknowledge the fact that site-specific (e.g. within patch) variation in habitat characteristics can influence how warblers distribute themselves within patches of habitat (Ladd & Gass, 1999) and could potentially affect between-patch distribution, we stress that our work was not focused on evaluating mechanisms driving within-patch differences in local selection (Manly et al., 2002). Rather, our approach shows how robust models of species distributions using coarse-scale metrics can be developed for supporting conservation decisions at a scale not easily attainable with local-scale models.

Our application of a geoadditive semiparametric regression occupancy model to golden-cheeked warbler breeding range survey data provides a flexible framework for predicting warbler distributions while addressing latent spatial variation and issues associated with imperfect species detection. Thus, our work builds upon others who have incorporated spatial relationships into models (Elith et al., 2006), while accounting for imperfect detection (Royle et al., 2007), and provides an additional framework for model-based prediction of species distributions. We applied our geoadditive model as a single-season occupancy model, where occurrence does not change; thus, our example provides a snapshot prediction of golden-cheeked warbler patch occupancy distribution across the species range. Our model can be useful for other questions linking space to demography (Grosbois et al., 2009) as the general structure allows for direct incorporation of spatial relationships into future distribution models. Thus, our approach could easily be applied to dynamic models to address temporal variation in golden-cheeked warbler habitat patch occupancy state (MacKenzie et al., 2002, Royle & Kéry, 2007), or given our model’s structure, it could easily be modified to fit a variety of additional hierarchical models focusing on golden-cheeked warbler abundance estimation across the species breeding range (Royle, 2004; Kéry et al., 2005; Royle et al., 2005; Thogmartin et al., 2006; Conroy et al., 2008).

As reviewed by Mills (2007:265), one limitation of patch-occupancy models is they ignore local population dynamics largely because they are data-intensive. Although we acknowledge this limitation, a core issue of conservation biology is the distribution of individuals through space, with a focus on the size of potential habitat patches and their proximity to one another. Whether or not a species is structured as a metapopulation, knowledge of how habitat is occupied through space allows predictions to be made on colonization and extinction probabilities. Thus, our broad-scale approach serves as a template for addressing modelling of colonization and extinction based on the occupancy of habitat, hence a more thorough understanding of population viability. For example, we are using our resulting model to develop more regionally focused studies of patch abundance, productivity and dispersal. If properly designed, regionally focused studies can then be expanded more broadly to the population of interest.

In summary, our study provides an approach for developing a broad-scale assessment of the potential distribution of a species. By using detection/non-detection surveys and remotely sensed habitat metrics within a spatially explicit modelling context, we identified a spatial gradient of occurrence for golden-cheeked warblers as well as relationships between two measurable landscape characteristics that can be used for further conservation planning. As opposed to methods
typically used to predict species distributions where habitats are predicted to be either usable or unusable for a species (e.g. Elith et al., 2006), our approach allows for a probabilistic prediction of the likelihood that a patch would harbour the target species. Our model was accurate when evaluated using an independent dataset, suggesting that our predictions were both robust and applicable rangewide.

ACKNOWLEDGMENTS

We thank J. Andrew Royle, Marc Kéry, Frank Thompson and two anonymous reviewers for constructive comments on earlier versions of this manuscript. We appreciate discussions with J. Andrew Royle, Ciprian Crainiceanu, Marc Kéry and Marc Genton during model development. Support for our research was provided by Texas Department of Transportation. Additionally, we appreciate support from Texas Parks and Wildlife Department. B.A. Collier and R.J. Carroll acknowledge partial support for their work from Award No. KUS-C1-016-01. We thank J. Andrew Royle, Marc Kéry, Frank Thompson and B.A. Collier and R.J. Carroll acknowledges partial support for their work from Award No. KUS-C1-016-01, made by the King Abdullah University of Science and Technology (KAUST). We gratefully acknowledge the private landowners who allowed us access to their lands during the course of this project as well as managers of public lands surveyed during this study.

REFERENCES


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

**Appendix S1** Distribution of woodland patches \((n = 63,616)\) by patch size (ha) within the 35-county breeding range of the golden-cheeked warbler in Texas, USA.

**Appendix S2** Distribution of woodland patches \((n = 63,616)\) by landscape composition within the 35-county breeding range of the golden-cheeked warbler in Texas, USA.

**Appendix S3** R and WinBUGS code for the semiparametric regression occupancy model.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

**BIOSKETCH**

**Bret A. Collier** is a research ecologist with the Institute of Renewable Natural Resources at Texas A&M University where his research focuses on quantitative ecology and wildlife population dynamics. Work on this project was conducted by the Research and Management System for Endangered Species (RAMSES) research group at the Institute of Renewable Natural Resources at Texas A&M University (irnr.tamu.edu/RAMSES/).

Author Contributions: All authors contributed to the development of the study ideas and design presented in this work. J.E.G., S.L.F., A.J.C. and H.A.M. directed field operations; R.T.S. directed spatial data analysis, B.A.C. and R.J.C. analysed the data; B.A.C., M.L.M. and R.N.W. led the writing.

Editor: Ralph Mac Nally