Predicting avian patch occupancy in a fragmented landscape: do we know more than we think?

Danielle F. Shanahan* and Hugh P. Possingham

The Ecology Center, School of Biological Sciences, University of Queensland, Brisbane, QLD 4072, Australia

Summary
1. A recent and controversial topic in landscape ecology is whether populations of species respond to habitat fragmentation in a general fashion. Empirical research has provided mixed support, resulting in controversy about the use of general rules in landscape management. Rather than simply assessing post hoc whether individual species follow such rules, a priori testing could shed light on their accuracy and utility for predicting species response to landscape change.

2. We aim to create an a priori model that predicts the presence or absence of multiple species in habitat patches. Our goal is to balance general theory with relevant species life-history traits to obtain high prediction accuracy. To increase the utility of this work, we aim to use accessible methods that can be applied using readily available inexpensive resources.

3. The classification tree patch-occupancy model we create for birds is based on habitat suitability, minimum area requirements, dispersal potential of each species and overall landscape connectivity.

4. To test our model we apply it to the South East Queensland region, Australia, for 17 bird species with varying dispersal potential and habitat specialization. We test the accuracy of our predictions using presence–absence information for 55 vegetation patches.

5. Overall we achieve Cohen’s kappa of 0.33, or ‘fair’ agreement between the model predictions and test data sets, and generally a very high level of absence prediction accuracy. Habitat specialization appeared to influence the accuracy of the model for different species.

6. We also compare the a priori model to the statistically derived model for each species. Although this ‘optimal model’ generally differed from our original predictive model, the process revealed ways in which it could be improved for future attempts.

7. Synthesis and applications. Our study demonstrates that ecological generalizations alongside basic resources (a vegetation map and some species-specific information) can provide conservative accuracy for predicting species occupancy in remnant vegetation patches. We show that the process of testing and developing models based on general rules could provide basic tools for conservation managers to understand the impact of current or planned landscape change on wildlife populations.

Key-words: a priori model, classification tree, patch occupancy model, rainforest birds, South East Queensland, species generalizations, threshold responses

Introduction
A primary goal in landscape ecology is to identify the factors that influence the persistence or extinction of populations in fragmented habitat patches (Rodriguez et al. 2007). Recently research has increasingly focused on generalizing species’ responses to landscape structure, as the ability to predict how different species will respond to human or climate driven change remains critical for effective pro-active management. There has been particular focus on examining whether ‘ecological thresholds’ or rules-of-thumb exist in landscape ecology (With & Crist 1995; Bennett & Ford 1997; Radford & Bennett 2004). An ecological threshold is the point where small changes in an environmental factor produce sudden, rapid responses in the ecosystem itself (Groffman et al. 2006); an example of such a threshold is the ‘30% rule’, where connectivity is thought to be broken below 30% vegetative cover and species extinctions are more likely to occur (Andren 1994; McAlpine, Fensham & Temple-Smith 2002). Empirical evidence for the existence of

*Correspondence author. E-mail: d.shanahan@uq.edu.au
such thresholds is inconclusive and different values are often suggested (Lindenmayer & Luck 2005). These inconsistencies are attributed to the high level of variability in systems and species-specific traits that also have a significant role in determining extinction probabilities (Mac Nally, Bennett & Horrocks 2000; Vos et al. 2001; Sekercioglu & Sodhi 2007). Consequently, there has been little success in translating such advances into real-world prediction solutions for land managers (Mac Nally et al. 2000; Namsauer & Opdam 2008). There is nevertheless a paucity of studies directly testing such broad generalizations in an a priori fashion, with consideration of important species-specific traits.

MacNally & Bennett (1997) published a seminal paper in which they noted that taking an a priori approach in landscape ecology (i.e. making quantifiable predictions of what we expect to see based on our current knowledge) would enable us to understand the relevance and application of the knowledge we have already gained, and translate general rules into readily useable tools for wildlife and land managers. They went on to predict occurrence of several bird species across a fragmented landscape based on broad generalizations. They tested these predictions in a later research (Mac Nally et al. 2000). The model was unsuccessful in predicting the occurrence of many bird species, and the authors suggested that it is probably better to consider each species in isolation taking into account more life-history details. Perhaps because of this lack of predictive success few similar studies directly testing generalizations have been carried out since. This approach is important, however, as it provides a means to examine why general rules may not work in every instance, and thus is also a means to improve and develop better general models.

More recent research by Vos et al. (2001) took into account varying life-history traits of species when predicting the fraction of habitat patches likely to be occupied in a fragmented landscape. This approach primarily used area requirements and dispersal distance of a species to generate ecological species profiles, or ‘ecologically scaled landscape indices’. The index correlates well with the fraction of patches occupied by a species in a fragmented landscape, which is thought to reflect a species’ overall susceptibility to fragmentation (Hanski 1994a; Vos et al. 2001). Considering each species in isolation, but still within the context of general habitat fragmentation theory and rules, may therefore provide the key to achieving the right balance between complexity and generality when predicting population response to landscape change. Incorporating these ideas into a priori predictive models will not only allow them to be tested but will also take the next step towards translating them into a cost effective practical tool.

Our primary aim in this study is to examine how much we really know about habitat fragmentation by using general rules to predict occupancy of bird species in remnant vegetation patches. We intend to create an a priori predictive model broadly based on theory and factors generally known to influence the persistence of populations in fragmented landscapes. While aiming to minimize model complexity, within the model framework we also intend to consider important species specific traits that are thought to influence the persistence of populations. In this way we aim to strike a balance between model complexity and generality, potentially creating a more successful model than past predictive attempts. In this paper we will also explore the model’s success and assess its failings. We consider this a critical step in continuing to improve predictive models to ensure high accuracy for practical application. Our secondary goal is to ensure the predictive model is presented in a readily understandable manner, and to make our predictions using only readily available, inexpensive data resources.

**Materials and methods**

**MODEL DEVELOPMENT**

Our basic predictive model is broadly based on two principles that are considered important determinants of the persistence of a population in a habitat fragment: extinction probability and colonization rate (Levins 1970; Hanski 1994b). In fragmented systems extinction probability can be considered a function of patch area, which simply assumes that population size is limited by the area available to inhabit (Levins 1970; Shaffer 1981; Hanski & Gilpin 1991). Colonization rate can be considered a function of patch isolation, where increased isolation means a lower probability of colonization in the event of extinction, and decreased migration between populations also reducing persistence (Levins 1970; Shaffer 1981; Hanski & Gilpin 1991). Hanski (1994b) presents the relationship as:

\[
J = C/(C + E)
\]

where \(J\) is the probability a patch will be occupied, \(C\) is colonization rate and \(E\) is extinction rate.

Vos et al. (2001) developed ‘ecologically scaled landscape indices’ to adjust the expected probability of patch occupancy for life-history traits of the species in question, incorporating the dispersal potential of a species (as a predictor for colonization rate) and the size of the population (a predictor for extinction rate). In our study we extend this idea and assume population occupancy in any one patch is determined by landscape context, habitat availability, area requirements and dispersal capabilities of the species.

We used a classification tree model (which can be applied to each patch, for each species of interest) to assimilate these principles into a common framework (Fig. 1). This model can be used to make binary predictions of species’ presence–absence, which is particularly useful for data where threshold responses to variables potentially exist. Though classification trees do not necessarily provide the best species distribution models when working with training data sets [compared with, for example, logistic multiple regression (Munoz & Feliscisimo 2004)], they are easy to develop and interpret (Vayssieres, Plant & Allen-Diaz 2000), an important factor to ensure accessibility for land managers. We used the tree structure in a non-traditional manner by assigning variable values to each node without actual presence–absence data (hence an a priori approach). The difficulty with this theoretical model is numerically defining such rules in a broad fashion so they can be adapted and applied to different species.

**MODEL DEFINITION AND PARAMETERIZATION**

We break the problem down into four simple ‘node rules’. Each rule serves to predict presence or absence of a species in each vegetation patch (Fig. 1). We used a combination of threshold theory and species-specific information to determine values for the cut-off points,
above or below which the presence–absence predictions were made. These rules are described below, and are adapted for the landscape of interest in Fig. 1.

**Rule 1: Habitat suitability.** Is there suitable habitat within the vegetation patch? Suitable habitat is required for a population to be present. Suitable habitat descriptions can be derived for many species from the scientific literature or from expert opinion. A vegetation map is required to apply this information in a practical manner. Suitable habitat connected together with no discernable breaks can be considered a ‘patch’ for the purposes of this model.

**Rule 2: Minimum patch size.** Is there sufficient suitable habitat to sustain the population (even with complete isolation) for the period it has been fragmented for? No accurate number can be given for the minimum viable population size that will survive in a habitat fragment over time because of environmental variation and the probabilistic nature of extinction (Thomas 1990). Minimum viable population size literature has, however, provided very general rules-of-thumb that can be used to estimate minimum patch size for survival over x years: for birds, in areas with approx. 80–100 years of habitat fragmentation, an area sufficient for 200 individuals or 100 pairs should provide a high probability of population persistence (Jones & Diamond 1976) (see Fig. 1 for threshold prediction). For other species groups or in areas fragmented for longer periods of time, a more conservative estimate for persistence is 1000 individuals, and 10 000 for a very conservative estimate (Thomas 1990). For areas where fragmentation time is unknown, 1000 individuals could be considered a moderate estimate. To translate these estimates into minimum areas, territory size estimates can be used (Vos et al. 2001).

**Rule 3: Isolation.** If a patch does not meet the criteria for Rule 2, is it within dispersal distance of one that does fit Rule 2? We anticipate that small habitat patches within dispersal distance of larger self-sustaining populations (populations that fulfill Rule 2) can be supported by the larger patch through migration, and thus extinction risk is reduced (source–sink dynamics; Pulliam 1988). Hence, for the purposes of this model we can predict a population will be present (Fig. 1). Bowman (2003) provides a simple ‘guesstimate’ for dispersal distance based on previous empirical dispersal studies:

$$\text{Median dispersal distance} = 12\sqrt{\text{Territory size}}$$

The estimate is not data intensive, requiring only territory size to generate dispersal distance values.

### Classification tree

![Classification tree](image)

**Rule 4: Landscape connectivity.** Is there sufficient vegetation in the landscape to ensure connectivity, and hence migration between populations? Theoretical and some anecdotal research suggests that 30% or more forest cover is the threshold value above which landscapes provide sufficient habitat and connectivity, allowing population survival even in small patches (Andren 1994; McAlpine et al. 2002). We therefore consider a landscape with 30% or more forested vegetation cover sufficient to ensure connectivity between populations (see Fig. 1 for threshold diagram). As such, we can predict a population will be present in a small patch if it is connected through a sufficiently forested landscape to a large patch that fulfills Rule 2. A landscape is defined in this study as remnant vegetation patches separated by cleared land of a maximum width of the species dispersal distance or less – this in effect connects ‘stepping stones’ of vegetation. Therefore, a primary assumption of this rule is that a population will not persist in suitable habitat unless there is at least one population in the landscape large enough to at least be self-sustaining.

### Applying the model to an existing landscape

We tested our model in South East Queensland, Australia. The area once had substantial subtropical rainforest and notophyll vine forest cover which was largely cleared post-European settlement (80–150 years; Taylor 1994). Using our model we predicted occupancy of 17 forest bird species in each vegetation patch in the landscape. These species represent varying dispersal potential and habitat specialization: life-history traits that are deemed important in determining population response to habitat fragmentation (Wiens 1985; MacNally & Bennett 1997; Vos et al. 2001) (Table 1). We used a simple Habitat Specialization Index: the number out of eight possible vegetation types the species is commonly known to inhabit. The ‘dispersal potential’ measurement was separate from the median

![Fig. 1. An a priori classification tree model for predicting presence or absence of bird species in South East Queensland, fragmented for approx. 100 years. The figure also provides a description of each node rule and theoretical threshold values used to make those predictions.](image)
dispersal distance described earlier, and simply a relative estimate of low, medium or high; this information was generated from descriptions in HANZAB (1991–2007). We also estimated predicted resilience to habitat fragmentation: high habitat specialization and low dispersal potential species are predicted to be susceptible, the converse are predicted to be resilient (MacNally & Bennett 1997; Vos et al. 2001) (Table 1).

Data set descriptions

To define predicted values for the model variables presented in Fig. 1. We used data sets readily available to land managers in the region:

Vegetation map. We used a vegetation map created by local government which is widely available to land managers in the region.

Table 1. Relevant life-history information (and relevant sources) of the Australian bird species for which the classification tree presence–absence prediction model was applied

<table>
<thead>
<tr>
<th>Species</th>
<th>Habitat preference (Queensland)</th>
<th>Habitat specialization (1, high and 8, low)*</th>
<th>Mobility estimate (low, moderate and high)</th>
<th>Predicted resilience to fragmentation †</th>
<th>References (additional to HANZAB 1993–2006) or naming authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black-faced monarch (Monarcha melanopsis)</td>
<td>Rainforests; wet sclerophyll; vine thickets</td>
<td>3</td>
<td>H</td>
<td>Migratory</td>
<td>Susceptible</td>
</tr>
<tr>
<td>Brown cuckoo-dove (Macropygia phasianella)</td>
<td>Rainforests; wet sclerophyll</td>
<td>3</td>
<td>M</td>
<td>?</td>
<td>Somewhat resilient</td>
</tr>
<tr>
<td>Brown gerygone (Gerygone moaki)</td>
<td>Rainforest; wet forest and gully</td>
<td>3</td>
<td>M</td>
<td>✓</td>
<td>Somewhat susceptible</td>
</tr>
<tr>
<td>Brown thorntail (Canthiza pusilla)</td>
<td>Dense shrub layer</td>
<td>5</td>
<td>H</td>
<td>?</td>
<td>Resilient</td>
</tr>
<tr>
<td>Eastern whipbird (Psephotus olivaceus)</td>
<td>Dense understorey</td>
<td>6</td>
<td>L</td>
<td>✓</td>
<td>Somewhat resilient</td>
</tr>
<tr>
<td>Eastern yellow-robin (Eopsaltria australis)</td>
<td>Most habitat types so long as there are vertical stems to rest on</td>
<td>7</td>
<td>L</td>
<td>✓</td>
<td>Resilient</td>
</tr>
<tr>
<td>Emerald dove (Chalocopsis indica)</td>
<td>Dense forests</td>
<td>2</td>
<td>M</td>
<td>✓ ?</td>
<td>Resilient</td>
</tr>
<tr>
<td>Golden whistler (Pachycephala pectoralis)</td>
<td>Rainforest; enclosed eucalypt woodland</td>
<td>4</td>
<td>H</td>
<td>Partly migratory</td>
<td>Resilient</td>
</tr>
<tr>
<td>Green catbird (Alloraeus crassirostris)</td>
<td>Rainforest</td>
<td>1</td>
<td>L</td>
<td>✓</td>
<td>Susceptible</td>
</tr>
<tr>
<td>Grey shrike-thrush (Collaricincla harmonica)</td>
<td>Wet eucalypt; eucalypt woodland</td>
<td>4</td>
<td>H</td>
<td>✓</td>
<td>Resilient</td>
</tr>
<tr>
<td>Large-billed scrubwren (Sericornis magnirostris)</td>
<td>Rainforest; wet sclerophyll forest</td>
<td>3</td>
<td>H</td>
<td>✓</td>
<td>Somewhat resilient</td>
</tr>
<tr>
<td>Logrunner (Orthonyx temmincki)</td>
<td>Rainforests; adjacent wet sclerophyll forest</td>
<td>1</td>
<td>L</td>
<td>✓</td>
<td>Susceptible</td>
</tr>
<tr>
<td>Spectacled monarch (Monarcha trivirgatus)</td>
<td>Dense; low vegetation; rainforest; wet sclerophyll forest</td>
<td>3</td>
<td>H</td>
<td>Partly migratory</td>
<td>Somewhat resilient</td>
</tr>
<tr>
<td>White-browed scrubwren (Sericornis frontalis)</td>
<td>Dense shrub layer</td>
<td>7</td>
<td>L</td>
<td>✓</td>
<td>Resilient</td>
</tr>
<tr>
<td>White-throated treecreeper (Cormobates leucophaeus)</td>
<td>Rainforest; wet eucalypt; eucalypt woodland</td>
<td>6</td>
<td>L</td>
<td>✓</td>
<td>Resilient</td>
</tr>
<tr>
<td>Wompoo fruit-dove (Ptilinopus magnificus)</td>
<td>Rainforest</td>
<td>1</td>
<td>L</td>
<td>✓ ?</td>
<td>Susceptible</td>
</tr>
<tr>
<td>Yellow-throated scrubwren (Sericornis citreogularis)</td>
<td>Rainforest</td>
<td>2</td>
<td>L</td>
<td>✓</td>
<td>Susceptible</td>
</tr>
</tbody>
</table>


*1–8 refers to the number of habitat types considered suitable for that species.
†Susceptible, high habitat specialization and low dispersal potential; resilient, habitat generalist and high dispersal potential.
region (Regional Ecosystems Version 5.2; Environmental Protection Agency 2007). The vegetation mapping was originally carried out using satellite imagery and aerial photograph interpretation, and was later ground-truthed and classified. We simplified the original 151 vegetation classes into eight vegetation types relevant to the bird species studied: Rainforest, Vine forest, Wet sclerophyll forest, dry eucalypt woodland (grassy understory), Woodland with shrubby understory, Heathland, shrubs, non-remnant vegetation. Each site we visited was ground-truthed, and the resulting User's accuracy (probability that the actual vegetation type is what the map says it is) of this more broadly classified map was 95%.

Applying the model

These data sets were used to generate a vegetation cover map, and for each species a suitable habitat map (with associated 'patch area' values), minimum area requirement estimates, and median dispersal distance estimates. This information was then used to define variable values for the a priori model (Fig. 1). Each suitable habitat patch was assessed against the classification tree and assigned a final 'population present' or 'population absent' prediction. This process was carried out in ArcGIS 9.1 (ESRI, California, USA).

TEST OF MODEL ACCURACY

Field data collection

We tested the model accuracy using an independent presence-absence data set. Fifty-five rainforest/vine-forest patches were randomly chosen for survey, stratified across the region to avoid spatial autocorrelation. Sites were surveyed four times (randomized across the field season, between October 2006 and January 2007, from 5.00 h to 9.00 h) by a single observer using the '20-minute 2-hectare' approach (Barrett et al. 2003). The 2 ha was a randomly positioned 300 m × 66 m transect that was crossed in the 20-min period. All birds heard or seen on the transect line were recorded, and a patch was considered occupied if that species was recorded on any survey. The probability of detecting each species was calculated to ensure there was a reasonable chance of detection within four repeat surveys. This probability was simply the total proportion of surveys where the species was heard or seen in patches where it was eventually determined to be present.

We examined the spatial pattern of presence-absence records for each species using Moran's I to determine if records were clustered, uniform or random across the landscape. Moran's I was estimated using Euclidean distance between survey sites, and the inverse distance relationship method (Legendre & Fortin 1989) in the software program R (R Development Core Team 2005). Moran's I values close to 1 indicate a uniform pattern, values close to -1 indicate a clustered pattern. A z-score was also estimated to determine whether the difference between the estimated and expected Moran's I was statistically significant.

We used the Habitat Hectares approach (Parkes, Newell & Cheal 2003) to rank the quality of each vegetation patch (considering vegetation structure and predominant plant species present) against a pristine or undisturbed vegetation patch of the same type. Patches were scored from 100 (pristine) to 1 (destroyed). The same value was given to the entire patch, assuming a heterogeneous level of disturbance.

Percentage accuracy of model

We calculated standard model assessment measures that adjust for species prevalence to determine the overall accuracy of the model: sensitivity (the number of observed presences divided by the total number of species presence records) and specificity (the number of observed absences correctly predicted as absent divided by the total number of recorded species absences). Low sensitivity indicates erroneous predictions of absence, and low specificity erroneous predictions of presence (Fielding & Bell 1997). We also calculated the absolute prediction accuracy for each species (percentage correct 'present' and 'absent' predictions) as an indicator of performance if applied in a conservation context. Although this measure does not account for species prevalence, it provided a much clearer indication of which life-history traits influence model success. We also calculated Cohen’s kappa statistic, which records overall agreement between predictions and observations, corrected for agreement expected to occur by chance: 1 indicates perfect agreement; zero or less suggests performance no better than random (Cohen 1960). We assessed whether habitat specialization and dispersal potential (from Table 1) influenced model accuracy. We also estimated the Cohen’s kappa statistic across all species to determine the overall agreement between the model and the test data, and the probability that this value was different from random following the calculation of variance among samples (following Fleiss, Cohen & Everitt 1969) in the program R.

IDENTIFYING THE ERRORS IN THE PREDICTIVE MODEL

Inaccuracies within this model could arise from two main sources: the model itself (through omission of important variables), or the values used to define node splits (from both the general rules and species specific information). We compared the predictive model, both the values used to define splits and the variables themselves, against the statistically derived classification tree model for each species.

To create these post hoc classification trees, recursive partitioning (splitting) was used, where a data set was split into increasingly similar groups through the use of single predictor variables. The categorical variable was the presence or absence of the species and the landscape or patch measurements were the predictor (or continuous) variables. We used Gini goodness-of-fit measures (Jammalamadaka & Goria 2004) to determine optimum splits, with a minimum of five observations required to split the node. A maximum of 10 splits was set to avoid over-fitting the tree. V-fold cross-validation of trees was used to check the accuracy of the model and determine optimal length (Sherrod 2003). This process splits the data set into v subsamples (here v = 10) of the presence-absence data, producing v trees. In each run, one subsample is left out of the training set, and this sample is then compared back to the resulting tree as a cross-validation. The classification error is calculated each time this is done. The error rate is averaged for a particular tree size, providing a function from which the optimal tree size (where the error cost is the lowest) can be calculated. Trees were created in the program dtree (Sherrod 2003). Backward pruning was carried out in a stepwise fashion following the cost-complexity measure described by Breiman et al. (1984); this process fits the model created from the entire data set to the optimal length as determined from the cross-validation process. Tree topography was allowed to vary for both sets of trees generated (described below), as this process allowed us to determine optimal variable values, tree size and variable order.
Testing for accuracy of variable selection

The first set of post hoc trees was generated for each species using additional variables that we did not consider in our predictive model. The variables were chosen because they represent factors that have been found to influence the survival of bird populations in other fragmentation studies. These variables were:

- Total patch area. This is comparable to the ‘minimum patch size’ variable from Rule 2 in the predictive model.
- Isolation (the Fragstats ‘proximity index’; this is the sum of patch area (m²) divided by the nearest edge-to-edge distance squared (m²) between the focal patch and other patches of the same habitat type within 2000 m of that patch; McGarigal et al. 2002). This variable is comparable with the ‘isolation’ variable from Rule 3.
- Total percentage forested in 100 km² landscape (100 km² method as used by Radford & Bennett 2004). This is comparable to the ‘landscape connectivity’ variable from Rule 4.
- Percentage suitable habitat in the 100 km² landscape.
- Vegetation quality score.
- Patch shape (the Fragstats Shape Index; this is patch perimeter divided by the minimum perimeter possible for a maximally compact patch of the same patch area; McGarigal et al. 2002).

Because the classification tree method does not provide a measure of importance for each variable and different variables can occur at the same level in a tree, we simply used the occurrence of a variable in a tree as an indication of its importance (thus recording the percentage of all possible nodes that each variable occurred in). To assess the influence of life-history on which variables were most important, we separated the species into groups: high and low dispersal potential, habitat specialists and generalists, and ‘susceptible’ or ‘resilient’ species (Table 1). Species that fell between these extreme groups were not used in this analysis. Again, the percentage of all possible nodes each variable occurred in for the entire group was used as an indication of variable importance for that group.

Testing for ‘node-split’ value accuracy

The second set of classification trees were created using only the variables considered in the original predictive model to obtain ‘optimal’ threshold values for each species:

- Patch size.
- Isolation (straight line distance to the nearest 100 territory patch).
- Landscape connectivity (percentage vegetation cover in the landscape).

For each species, we recorded the threshold value for each variable where it occurred at the highest branch in the tree (some variables did not improve model accuracy at all, and hence no value was obtained). We compared these statistically derived values to those we predicted for each species. For variables that did display linear trends, we used linear regression to examine this.

Results

MODE ACCURACY

The probability of detection for all bird species used in this study was above 40% (Table S1, Supporting Information), hence the chance of detecting all species in patches where they

Fig. 2. The accuracy of the predictive patch occupancy model [percentage correct ‘present’ predictions, ‘absent’ predictions and overall agreement between model and test data set (Cohen’s kappa)] for each species against: (a) habitat specialization (measured as the number of habitat categories that were considered suitable habitat for that species); and (b) the estimated dispersal potential of the species (measured as high, low or moderate, as described in Table 1).
were present in at least one of the four repeat surveys was high. Moran’s I analysis indicated that presence–absence records for all species at the scale sampled in this study were spatially random rather than clustered or uniform (values provided in Table S1, Supporting Information). We therefore considered the test data set to be sufficiently accurate and unbiased for model testing.

Sensitivity values were high, all above 0.65, however, model specificity was generally low, between 0 and 0.57 (see Table S1, Supporting Information for values). This suggests a low accuracy in ‘present’ predictions for most species, but high accuracy in ‘absent’ predictions. The accuracy percentage of ‘present’ and ‘absent’ predictions (Fig. 2) appeared to be influenced by the habitat specialization of the species concerned (Fig. 2a), and a particularly low ‘present’ prediction accuracy was achieved for three habitat specialists; the logrunner Orthonyx temminckii (Ranzani 1825), green catbird Ailuroedus crassirostris (Paykull 1815) and emerald dove Chalcopsitta indica L. Low ‘absent’ prediction accuracy was achieved for three habitat generalist species [white-browed scrubwren Sericornis frontalis (Vigors & Horsfield 1827), eastern yellow robin Eopsaltria australis (White 1790), and eastern whipbird Psophodes olivaceus (Latham 1801)]. No obvious trends were evident when the presence–absence accuracy data was compared with dispersal potential of the species concerned (Fig. 2b). Cohen’s kappa showed that the model performed reasonably well for a number of species, generally classified as ‘fair’ to ‘moderate’ agreement between model predictions and the test data set (Fig. 2a,b). The overall Cohen’s kappa value across all species was 0.33 (‘fair’ agreement), and this was significantly different from random with $P < 0.01$.

**IDENTIFYING THE ERRORS IN THE PREDICTIVE MODEL**

Figure 3 provides an example of a statistically derived classification tree for a single species, the yellow-throated scrubwren Sericornis citreogularis. Table 2 shows the total percentage occurrence of each variable (out of all possible node positions) in the statistically derived classification trees for each group of species. This percentage represents the relative importance of each variable in enabling persistence of each species group (higher percentage means more important). This analysis showed that for habitat specialists, susceptible species and resilient species, vegetation quality was a particularly important determinant of species presence. Interestingly, this is a variable that was not considered in the original predictive model. For both habitat generalists and species with high dispersal potential, the percentage suitable habitat remaining in the landscape was the most important factor. For species considered to have a low dispersal potential, no clear patterns of variable importance were observed.

We were able to obtain statistically derived values for the three variables used in the original predictive patch occupancy model for only some species: 13 species for ‘minimum patch size’ (variable for Rule 2) and ‘landscape connectivity’ (Rule 4, required % vegetation cover for connectivity); and six species for ‘isolation’ (Rule 3). For the remainder of species none of the three variables provided any predictive power, and thus no optimal value could be derived. Our a priori estimates of ‘minimum patch size’ were linearly related to the statistically derived value for this variable (Fig. 4a; F-statistic: 14.52, $r^2 = 0.65$, $P = 0.003$). There was, however, no evident relationship between the predicted and statistically derived values of the variable ‘isolation’ (Fig. 4b). The statistically derived value for ‘landscape connectivity’ was between 15% and 20% for most species, and all but one obtained optimal values below the predicted value of 30% (Fig. 4c).

**Discussion**

A relatively recent and controversial topic in landscape ecology is whether species respond to habitat fragmentation in a general way (Mac Nally *et al*. 2000; Lindenmayer & Luck 2005). The need for simple generalizations is obvious: under-resourced conservation projects generally lack the funds and time for intense localized research. In this study we have attempted to strike a balance between the species specific traits that need to be incorporated into a predictive patch occupancy model.
MacNally & Bennett 1997; Vos et al. 2001), and general theory that can be applied across all species, providing a skeleton theoretical model which can be adapted on a species by species basis. The result is a simple and reasonably accurate model that is accessible to non-experts. As such, our research provides some advantage that although species do respond differently to landscape change, some generalizations could be useful for decision making until better approaches and further local research becomes available.

Our predictive model achieved a conservative level of success when tested with an independent data set: if implemented for conservation prioritization purposes, for most species we would be more likely to protect areas that we thought contained a population but did not, rather than ignore areas where a population was present. This high ‘absence’ prediction accuracy featured across a large range of species with differing life-history traits including three migratory species. It was, however, compromised for habitat generalists. Our post hoc analysis of the presence-absence data indicated that the amount of suitable habitat remaining in the landscape was more important for these species, suggesting there may be more suitable habitat in the landscape (such as heavily infested weedy areas) than was recorded by the vegetation maps we used. For habitat specialists we achieved markedly low accuracy in predicting population presence. Vegetation quality and forest cover was much more important for these species than patch size or isolation. The model inaccuracy for both specialists and generalists was therefore driven by the inability to consider vegetation quality in the predictive model, as the vegetation map used had no vegetation quality rating. This post hoc analysis suggests that an accurate vegetation map with associated degradation information should be considered one of the most important resources for land managers and decision makers to make informed conservation decisions.

Fig. 3. An example of a statistically derived classification tree generated for the yellow-throated scrubwren Sericornis citreogularis from field survey data of vegetation patches.

Fig. 4. The threshold values for three landscape variables which were used to predict presence or absence of bird species in South East Queensland. The graphs show our a priori estimate of the threshold point for each variable (details in Fig. 1.), against the actual threshold point as determined from classification tree analysis on actual presence absence data. The variables these graphs represent are: a) minimum patch size for a population to survive independently of other patches; b) distance between patches above which isolation occurs; c) per cent vegetation cover above which connectivity between remnants was maintained.
The most challenging part of model building was determining, a priori, values for the key components of the classification tree model: the minimum patch size for population persistence, dispersal distances of species, and connectivity levels required in the landscape. We did demonstrate some success, with a strong linear relationship between our predicted ‘minimum patch size’ and that determined as statistically optimal from our post hoc analysis. We also found that the 30% vegetation cover rule was a conservative estimate when compared with the statistically derived data values, and as such this research supports studies proposing 30% as a target for conservation purposes (Andren 1994; McAlpine et al. 2002). Our median dispersal distance estimates were less successful, demonstrating no relationship between predicted and statistically derived values. This was a consistent pattern through our post hoc and accuracy analysis, where dispersal distance was not a particularly useful way to group species. This could have implications for management, suggesting that habitat specialization or resilience may be a more effective way to group species response to habitat fragmentation.

Conclusions and management implications

The model presented in this paper is conservative in its estimates of species absence. Its simple nature and limited data requirements mean it is likely to be accessible to land managers. This model can certainly be improved, and perhaps the next step is to include vegetation quality as a major component within the predictive framework. However, with continued testing and improvement, models such as ours could provide a useful starting point for making speedy estimates as to where species of interest occur, and predictions as to how these species might respond to planned landscape alterations.

The process of creating and testing models such as ours has been identified as a critical step in bridging the gap between theoretical notion and practical conservation tool (Mac Nally & Bennett 1997). Although we did not develop a perfect model, the process of creating and testing it has allowed us to lend support to general rules such as protecting the largest most connected remnants first to ensure persistence of a wide range of species. However, our study also emphasizes the need for protecting high quality remnants or restoring quality in already protected remnants to preserve specialist species. In general, this study demonstrates a use of the a priori approach for testing general rules in landscape ecology (MacNally & Bennett 1997; Martin & Possingham 2005) which allowed us to examine fragmentation as a mechanism behind extinctions of species with different life-history traits.

Acknowledgements

The author would like to acknowledge CSIRO Sustainable Ecosystems, particularly Craig Miller, for financial and logistical support. The authors would also like to thank Wildlife Protection Society of Australia and Birds Australia for funding, and two anonymous referees for invaluable suggestions to improve this manuscript.

References


Received 22 December 2009; accepted 1 July 2009
Handling Editor: Jack Lennon

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1. Detectability of each species during field surveys of 55 vegetation patches, spatial autocorrelation estimates (and the significance of the deviation from random spatial pattern) of presence–absence data, and accuracy values of the predictive patch occupancy model, including presence and absence prediction accuracy, model sensitivity and specificity, and the overall accuracy statistic Cohen’s kappa

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.