Using habitat extent and composition to predict the occurrence of woodland birds in fragmented landscapes.

Maksym Polyakov[1], Alexei D. Rowles[2, 6], James Q. Radford[3], Andrew F. Bennett[4], Geoff Park[5], Anna Roberts[2] and David Pannell[1].

Abstract: The removal, alteration and fragmentation of habitat are key causes of biodiversity decline worldwide. In Australia, temperate woodlands have been disproportionately cleared following European settlement. Biodiversity decline in such systems may be reversed by restoration of native vegetation on agricultural land. However, rebuilding functioning habitat will require understanding the determinants of species distributions in existing habitat. We used logistic regression of bird occurrence data from 240 sites across northern Victoria, to determine the probability of occurrence of 29 woodland-dependent bird species. We modelled occurrence as a function of habitat variables that characterise both the extent (amount) and composition of native vegetation surrounding sites. Our specific goal was to determine whether the predictive performance of models is improved by accounting for both extent and composition of native vegetation compared with models that characterise native vegetation by extent alone. For nearly all species, accounting for vegetation composition in addition to extent and weighting habitat variables by distance improved the explanatory power of models, explaining on average 5.4% (range 0 – 27.6%) of the residual uncertainty in models that accounted for extent alone. Models that incorporate variation in vegetation composition can not only provide more accurate predictions of species occurrence, but also guide more appropriate restoration. They highlight the need for restoration to incorporate sites with fertile soils that support productive vegetation types. These models of woodland birds will be used to inform a spatially-explicit optimisation model for restoring native vegetation cover on agricultural land in this region, with the goal of achieving biodiversity gains while minimizing loss to production.

[1] Centre for Environmental Economics and Policy, School of Agricultural and Resource Economics, University of Western Australia, Crawley, WA, 6009

[2] Department of Primary Industries, Rutherglen, RMB 1145 Chiltern Valley Rd, Rutherglen, Vic., 3685

[3] Bush Heritage Australia, P.O. Box 329, Flinders Lane, Melbourne, Vic., 8009

[4] School of Life and Environmental Sciences, Deakin University, Burwood, Vic., 3125

[5] North Central Catchment Management Authority, PO Box 18, Huntly, Vic., 3551

[6] Department of Environmental Management and Ecology, La Trobe University, PO Box 821, Wodonga, Vic., 3689

Introduction

The loss and modification of habitat through intensive human land-use has contributed greatly to biodiversity decline throughout the world (Tilman et al. 2001; Foley et al. 2005; Green et al. 2005; Fischer et al. 2010). With much of the terrestrial environment committed to agriculture and other anthropogenic land-uses, formally protected reserves (e.g. National Parks) will, on their own, be inadequate to sustain biodiversity (Brooks et al. 2004). The need for conservation strategies that can maintain biota within agricultural landscapes is increasingly recognised as a global conservation priority (Polasky et al. 2005; Vandermeer and Perfecto 2007; Foley et al. 2011; Balmford et al. 2012).

One solution to the decline of biodiversity is the active restoration of native vegetation and rebuilding of functioning landscapes in agricultural areas (Thomson et al. 2009; Gibb and Cunningham 2010). In the sheep-wheatbelt region of southern Australia, for example, where native vegetation has been massively cleared for agricultural production (Hobbs and Yates 2000; Lindenmayer et al. 2010), much attention has been given to the benefits of revegetation and restoration for maintaining native biota (Munro et al. 2007; Lindenmayer et al. 2010; Mac Nally et al. 2010). To be most effective, such restoration must be optimised to achieve desired gains in species abundance and richness, while also minimising losses to production (Stoneham et al. 2003; Polasky et al. 2005). Optimisation requires being able to accurately predict the probability of species' occurrence based on relevant habitat and landscape attributes (Thomson et al. 2007, 2009; Montague-Drake et al. 2011).

Knowledge of the value of faunal habitats in agricultural mosaics has typically been gained from studies of species at individual sites within habitat patches of varying sizes and types. Species occurrence is commonly assessed as a function of attributes that characterise the habitat at the site, together with attributes that quantify its context in the surrounding landscape (Montague-Drake et al. 2009; Brown et al. 2010). The context of a site, represented

2

by attributes such as the extent of nearby habitat and measures of isolation and connectivity, has consistently been found to have an important influence on the occurrence of biota (McGarigal and Cushman 2002; Radford and Bennett 2004; Lindenmayer et al. 2010; Boscolo and Metzger 2011). The extent of habitat influences the population size and potential persistence of a species, while the spatial configuration and isolation of habitat relative to surrounding patches affect a species' distribution and dispersal (Fahrig 2003; Bennett et al. 2006).

Less attention has been given to variation in the relative composition of vegetation in the surrounding landscape, which often is represented as a single suitable habitat type (Debinski and Holt 2000). Accounting for the composition of the surroundings by differentiating between vegetation types (e.g. species composition) or structures (e.g. overstorey height) may enable more accurate predictions of species distributions. Species may favour specialised habitats such as riparian vegetation (Mac Nally et al. 2000) or scattered trees in grassland (Gove et al. 2009); and heterogeneity of vegetation types may provide species with a wider array of resources (Tews et al. 2004; Lindenmayer et al. 2012).

In this study, we assess how well models based on combinations of variables that characterise the context of a site can explain the distribution of woodland bird species in remnant vegetation in north-central Victoria, Australia. Woodland-dependent birds typical of this region are sensitive to habitat variation at both site and landscape scales (Radford and Bennett 2007; Montague-Drake et al. 2009; Haslem and Bennett 2011). Further, approximately two-thirds of the species have shown recent decline associated with sustained dry conditions over a decade (Mac Nally et al. 2009); with this decline being attributed to reduced habitat quality and food availability.

Here, our aim is to improve the predictive capacity of models of the distribution of species in remnant habitats in agricultural environments. First, we derive a set of variables

that describe the extent of different vegetation types within a 2 km radius around sampling sites, which provide a measure of the composition of vegetation types in the surrounding landscape. Second, we apply weights to these site-context variables to account for their varying influence on species distribution models at different scales. Our specific aim is to determine to what degree accounting for both the extent *and* composition of native vegetation types surrounding sampling sites improves the predictive performance of incidence models for woodland bird species, compared with models that characterise the surrounding landscape by the extent of wooded habitat alone.

Materials and Methods

Study sites and bird occurrence data

The study region is a fragmented, agricultural landscape in northern Victoria, Australia (Fig. 1). It extends from the alluvial plains of the Victorian Riverina in the north (elevation: < 150 m) to the inland slopes of the Great Dividing Range (elevation: 150—700 m). Native vegetation has been extensively modified by agriculture, mining and forestry since European settlement and remaining tree cover (~17%) is highly fragmented and degraded (ECC 1997). The region has a Mediterranean climate with hot, dry summers and cool winters. Most rainfall is received in winter and spring, with mean annual rainfall (400-670 mm) increasing from the north-west to the south-east.

Bird surveys were conducted in 24 10×10 km 'landscapes' during two 12-month study periods in 2002-2003 and 2006-2007 (Radford et al. 2005; Radford and Bennett 2007). Each landscape contains ten 2 ha sites, 240 sites in total. Four surveys were conducted in each study period. The survey design is detailed in Radford et al. (2005). The total number of surveys was 24 landscapes × 10 sites × 2 periods × 4 surveys = 1920 surveys. We treated the

4

four surveys on each site in a particular study period as an observation (480 observations). For the analysis, we use only woodland-dependent bird species that had more than 5% non-zero observations (29 species total).

The model

We assume that the probability of sighting a species in each survey depends on the site location and associated site and landscape characteristics, as well as on species-specific characteristics such as behaviour, detectability and song. The observed sighting of species s on site i is the realisation of the probability of sighting species s on site i ($p_{s,i}$). The response variable for each species is incidence (number of surveys in which the species was detected) per site and period. The incidence for species s on site i over 4 surveys, incidence $c_{s,i}$, is a realisation of a multiple-trial binomial distribution in a fixed number (4) of surveys. This can be modelled using logistic regression:

$$c_{s,i} \sim \text{Binomial}\left(4, p_{s,i}\right); \log\left(p_{s,i}/\left(1-p_{s,i}\right)\right) = \mathbf{x}_i\,'\boldsymbol{\beta}_s \,, \tag{1}$$

where $\mathbf{x}_i$ is a vector of site-specific explanatory variables and $\boldsymbol{\beta}_s$ is a vector of species-specific parameters.

Since sites within each landscape are located close to each other, there may be factors specific to landscapes which cannot be observed or cannot be incorporated into the model due to the limitation of degrees of freedom. In other words, the errors of observations within a landscape could be correlated. Logistic models that ignore spatial autocorrelation tend to overestimate habitat effects (Betts et al. 2006). To account for possible spatial correlation, we used a random effects logistic model. We introduced a random effect: that part of the error that is specific to landscape l and is normally distributed:

$$c_{s,i} \mid l \sim \text{Binomial}\left(4, p_{s,i}\right); \log\left(p_{s,i} / \left(1 - p_{s,i}\right)\right) = \mathbf{x}_i ' \boldsymbol{\beta}_s + u_l, \tag{2}$$

where u_l is a normally distributed landscape-specific random effect.

Site variables

Sites were characterised by a set of variables that describe the extent of native vegetation types within a 2 km radius (Table 1). We classified native vegetation types based upon a combination of ecological characteristics and tree density. Ecological characteristics were derived from a GIS dataset that estimates the extent of plant communities classified by Ecological Vegetation Class (EVC) (DSE 2007). Ecological Vegetation Classes incorporate community composition, structure and associated environmental factors (DSE 2002). Tree density was based on a 1:25,000 scale GIS dataset describing extent and density of existing tree cover. Three categories: 'Dense', 'Medium' and 'Scattered' are differentiated by crown cover densities of > 80%, 50-80% and 10-50%, respectively (DSE 2004). The Ecological Vegetation Classes represented within the study landscapes (DSE 2007) were aggregated into four broad vegetation types according to related ecological characteristics: 'Dry-infertile', 'Fertile', 'Plains' and 'Riparian' (Table 1). These classifications aggregate dense and medium density vegetation, except for 'Riparian' vegetation which is classified irrespective of its tree density. An additional vegetation type, 'Scattered', was classified instead by density: it captures scattered tree cover regardless of ecological characteristics (Table 1). To retain a manageable number of classes and allow a sensible model to be built, we excluded 'Riparian' and 'Scattered' vegetation classes from interactions between vegetation type and density groups due to their comparatively smaller area and lower likelihood of being influential.

The proportions of each of these five vegetation types characterise extent and their combination defines the composition of habitat surrounding a site. In order to model possible additive effects of explanatory variables, where certain bird species had the same preference

for two or more vegetation types, we calculated variables that represent combinations of vegetation types. For example, variable DF represents the sum of the proportion of 'Dry-infertile' and 'Fertile' vegetation types, and DFPRS includes all vegetation types implying that a particular bird species is indifferent to vegetation type. In order to approximate a normal distribution, we applied a natural log transformation to variables representing vegetation type or combination of vegetation types. A multiplicative effect of explanatory variables is implied when two or more log transformed variables are included in logistic models. For example, if a model contains D and F variables, the probability of occurrence depends on the level of variable D conditional on the level of variable F (or vice versa), implying that a particular bird species requires both 'Dry-infertile' and 'Fertile' vegetation types.

We assumed landscape features in close proximity to a site have greater effects on the suitability of habitat for bird species than those further away (Fahrig 2003). When measuring site-context, the extent or proportion of habitat usually is measured within a fixed radius of sample plots (Westphal et al. 2003; Betts et al. 2006). To accommodate the assumption about the diminishing effect of landscape features on site suitability, we derived weighted proportions of native vegetation types by calculating the area of all vegetation types within three bands (0 to 450 m, 450 to 1200 m, and 1200 to 2000 m radius) from the survey site and applying weights inverse to the squared distance between the survey site and median of the respective band (Hoover 1971; Polyakov et al. 2008).

The relative occurrence of some species may vary among bioregions. To take this into account, we introduced a binary variable representing the Riverina and Goldfields bioregions. A second binary variable was used to account for temporal variation across the two survey periods (2002/03 and 2006/07). The complete list of explanatory variables is presented in Table 1.

Model selection and validation

We estimated random effect logistic models of probability of occurrence with all possible combinations of explanatory variables characterising the extent of native vegetation types in the surrounding landscape as well as the temporal and bioregion dummy variables. To facilitate selection of the best-fitting model, we employed the Akaike Information Criterion (AIC):

$$\text{AIC} = 2k - 2\ln(L)$$

where k is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model. This criterion selects models with better log-likelihood value, but penalises explanatory variables that do not add explanatory power. A lower AIC indicates a better fitting model (Quinn and Keough 2002).

In developing species occurrence models, there is a risk of correlation among site variables for extent and vegetation type, particularly as the more extensive areas of vegetation were frequently 'Dry-infertile'. However, we did not use 'overlapping' combinations of vegetation types as candidate variables in the models. Furthermore, using AIC for model selection also prevented inclusion of highly correlated variables in the models. Consequently, the selected models did not contain highly correlated explanatory variables (highest correlation between variables in selected models r = -0.65.

To test for spatial autocorrelation, we calculated Moran's I in Pearson residuals of the selected logistic models with and without the random landscape effects. The weight matrix for Moran's I is inverse distance within 6 km to ensure every observation has at least one neighbour. As the data consist of two panels (time periods 2002-03 and 2006-07), the weight matrix is a block-diagonal, implying that spatial dependencies are tested only between observations within the same period.

To evaluate predictive performance of the selected final models, we used *k*-fold cross-validation (Boyce et al. 2002; Fielding and Bell 1997) by partitioning the entire dataset into 10 subsets, each containing an equal number of randomly selected observations. We then trained our models on nine of the 10 data sets using logistic regression and used estimated coefficients to predict probabilities of occurrence for the validation set. This procedure was repeated so that each of 10 subsets was used once as validation data and every observation received an out of sample prediction.

The predictive performance of the models was evaluated by applying information-theoretic statistics developed by Hauser (1978) to the out of sample predictions obtained by 10-fold cross-validation. These statistics (equations presented in Appendix 1) allow estimation of the amount of information provided by the model relative to prior knowledge, comparison with the maximum amount of information achievable, and testing of the accuracy and significance of the model. Given a set of possible outcomes $\mathbf{A}$ and a matrix of explanatory variables $\mathbf{X}$, the information index $I(\mathbf{A};\mathbf{X})$ quantifies the information provided by the model $p(a \mid \mathbf{x}_n)$ relative to the null or prior knowledge model $p(a)$.

The value of the information index varies between zero (when the model does not explain any additional uncertainty) and the value of the corresponding index of the prior entropy $H(\mathbf{A})$, which defines the uncertainty inherent in the system. This relationship permits testing the *usefulness* (U^2) of the model by measuring the proportion of uncertainty explained by the estimated model.

Information index, prior entropy, and usefulness can be calculated using different prior knowledge or null models. Two of the often-used prior knowledge models are $p_0(a)$, which is an equal probability of all possible outcomes (observing and not observing the species in this case), and $p_1(a)$, which is a probability proportional to the occurrence of

outcomes in the sample. Another prior knowledge or null $p_2(a)$ is the set of probabilities predicted by an alternative (inferior) model (Hauser 1978). When calculated using $p_0(a)$, the information index $I_0(\mathbf{A};\mathbf{X})$ indicates the total amount of information explained by the model, the entropy $H_0(\mathbf{A})$ indicates the total amount of information in the system , and usefulness U_0^2 indicates the proportion of total uncertainty explained by the model. When another null is used, for example $p_2(a)$, the information index $I_2(\mathbf{A};\mathbf{X})$ indicates the additional amount of information provided by the model, the measure of entropy $H_2(\mathbf{A})$ indicates the amount of residual uncertainty unexplained by null $p_2(a)$, and usefulness U_2^2 indicates the proportion of residual uncertainty (unexplained by model $p_2(a)$) that the model in question explains.

Furthermore, the information index $I(\mathbf{A};\mathbf{X})$ is normally distributed with a mean of $EI(\mathbf{A};\mathbf{X})$ and a variance of $V(\mathbf{A};\mathbf{X})$ which provides a *test of the accuracy* of the model. Lastly, the log–likelihood ratio $LLR = 2n \times I(A;\mathbf{X})$ is χ^2 distributed with degrees of freedom equal to the number of coefficients in the model and allows testing the *significance* of the empirical model—i.e., the null hypothesis that the model provides no additional information compared to the alternative model.

Results

Models of the probability of occurrence for 29 woodland-dependent bird species are presented in Table 2, with the selected model being that which had the lowest AIC value. In order to evaluate the effect of accounting for both the extent *and* composition of vegetation types in the surrounding landscape, compared with extent only of native vegetation, we

estimated models for each species that included a variable characterising extent of vegetation (DFPRS) and possibly regional and/or temporal dummies. Further, we also estimated the same models (i.e. extent and composition versus extent only) for every species by using variables that were calculated *without* applying weight proportional to inverse squared distance. The AIC values for these four alternative models for each species are presented in (Table 2).

For all 29 species, models that accounted for both extent and composition of vegetation types in the surrounding landscape outperformed (i.e. lower AIC value) models that included extent of vegetation only (Table 2). Incorporating distance weighting into calculation of the vegetation variables improved the performance of models for 86% (25/29) of species (Table 2).

We detected positive spatial autocorrelation (Moran's *I* significant at 95% level) for 19 species. By incorporating landscape-specific random effects into logistic models, we corrected or decreased the severity of the spatial autocorrelation for most species, with the exception of one species (Fuscous Honeyeater) where severity of autocorrelation was not substantially decreased, and two species (Dusky Woodswallow; Eastern Yellow Robin) where negative spatial autocorrelation was introduced. The models for these species were not amended further as the initial problem was not severe for our species set and had been improved by the inclusion of random effects.

Information-theoretic statistics were used to assess the improvement in accuracy of the predicted probabilities of cross-validated data relative to the equal probability null model (Table 3). For all species, Hauser's significance test based on the log-likelihood ratio test (LLR) indicates rejection of the null hypothesis that the models provide no improvement over the null model at the 99.9% significance level (Table 3). The Hauser's usefulness measure (U^2_0) indicates that these species models explain from 25 to 88% of total uncertainty (Table

3), suggesting a reasonably good fit for most models. The accuracy test, which compares the information index against the confidence interval around the expected information index, indicates that the empirical models are accurate: in all models, the information index is within 95% confidence interval of expected information (Table 3).

In addition, we calculated Hauser's usefulness measure based on $p_1(a)$ and $p_2(a)$ prior knowledge models, where $p_2(a)$ is the probability predicted by the model with random effects, dummies, and extent of woody vegetation. Respective Hauser's Usefulness measures U^2_0, U^2_1, and U^2_2 are presented in Table 4. The values of U^2_2 suggest that inclusion of vegetation composition in the models of species occurrence increased their explanatory power by an average of 5.4% (range 0.0 - 27.6%) of the residual uncertainty not explained by the models without composition (Table 4).

Regional and temporal explanatory variables were prominent in the probability of occurrence models with at least one being featured in 83% (24/29) of models (Table 2). The regional variable was included in 62% (17/29) of species models, with a positive 'Goldfields' value (10/17) indicating a higher probability of occurrence at sites in the Goldfields bioregion and a negative value indicating higher occurrence in the Riverina bioregion. Similarly, 62% (18/29) of species were influenced by 'Year' but only two had positive values suggesting increased abundance in the 2006/07 survey period.

Discussion

The probability of occurrence of 29 woodland-dependent bird species in northern Victorian landscapes was successfully modelled using explanatory variables that characterised both the extent and composition of the wooded vegetation surrounding survey sites. For all species, these models provided a stronger fit to the data, explaining up to 27.6 %

of the residual uncertainty of models based on the extent of wooded vegetation alone (i.e. U^2_2

values, Table 4). This implies that for species occurrence to be increased or maximised,

accounting for context and vegetation type is important when considering both protection of

remnant vegetation and re-planting vegetation in restoration projects.

Including landscape composition enables the habitat requirements of individual

species to be more effectively represented in models and leads to more targeted habitat

management (Fuller et al. 1997; Hinsley and Bellamy 2000; Montague-Drake et al. 2009). For

example, including vegetation composition in models for the Diamond Firetail

(*Stagonopleura guttata*) and the Superb Fairy-wren (*Malurus cyaneus*) accounted for 8.0 and

27.6% of the residual uncertainty of their respective occurrence models including habitat

extent only. Importantly, these models identify the vegetation most likely to benefit these

species: the Diamond Firetail requires both plains and woodlands of fertile areas (i.e. fertile,

plains/scattered trees, and riparian); and the Superb Fairy-wren displays a preference for

riparian vegetation. Mesic riparian vegetation more consistently provides low shrubby

vegetation that this species requires for shelter and foraging, than dry forest and woodlands

(Higgins et al. 2006). In contrast, other species were associated with landscapes comprising

any combination of several vegetation types as expressed within models where effects of

variables describing vegetation type are additive. For instance, the Mistletoebird (*Dicaeum*

hirundinaceum) is indifferent to Dry-infertile or Fertile or Plains or Riparian vegetation.

Inclusion of several vegetation types in a model suggests that species respond

positively to a site context of greater habitat diversity, while the arrangement of vegetation

types in the model (additive or multiplicative) indicates their relative importance. Species

may require such heterogeneity to provide resources for different activities such as breeding,

foraging and nesting, or to meet changing food requirements across seasons or life stages

(Law and Dickman 1998; Benton et al. 2003). Dry-infertile, Fertile and Riparian (DFR) or

Plains (DFP) were the most common variables included in the additive models for most species (Table 2), suggesting that any combination of this vegetation can provide suitable habitat. However, many species models included additional variables. 'Fertile', 'Plains' or 'Riparian' vegetation were included as an independent variable in the selected model for 15 of the 29 species (Table 2). Thus, for more than half of the species considered, simply protecting or restoring vegetation on the dry-infertile hills will not be adequate to secure or maximise their occurrence. They also require vegetation that is generally found in more fertile, productive parts of the landscape (*viz* fertile, riparian or plains vegetation), thus establishing potential trade-offs with other land uses incompatible with nature conservation, such as high-intensity agriculture.

Regional and temporal binary variables were included in the models of many species. A positive "Goldfields" value indicates that the occurrence of this species is influenced by the differences in habitat conditions of the Goldfields region relative to the Riverina (DSE 2002). Regional differences in edaphic, topographic and climatic conditions influence species distributions (Thomson et al. 2007) and drive differences in vegetation types. Open grassy box woodlands (e.g. Grey Box *Eucalyptus microcarpa*) are typical of the Riverina (though many areas are now reduced to scattered trees in paddocks) and dry forests and box-ironbark woodlands are more common in the Goldfields region. Also critical to woodland birds, is that remnant tree cover is much greater in the Goldfields (25% cover) compared with just 2.2% cover in the Riverina (Lowe et al. 2002; Wierzbowski et al. 2002).

The temporal binary variable 'Year' was also frequent in species' models, but was predominately negative, such as for nectar-feeding species (e.g. honeyeaters, lorikeets, Red Wattlebird *Anthochaera carunculata*). This indicates a decrease in the abundance of these species in the 2006/2007 survey data. This widespread regional decline has been attributed to

prolonged drought and increased temperatures in the last decade affecting food supplies such as nectar and invertebrates (Mac Nally et al. 2009).

Applying a distance weighting to the calculation of habitat variables improved the performance of occurrence models for most species (Table 2). This supports our assumption that the influence of landscape variables such as habitat extent and composition diminishes with distance. As our measure of vegetation extent and composition included the sampling site itself, species may be partly responding to characteristics of the site rather than only the vegetation in the surrounding landscape. Part of the unexplained variation in our models is undoubtedly due to such site-level habitat differences.

Studies such as this, that are based on cross-sectional data, have implicit limitations that may affect the accuracy and predictive capacity of models. For example, sites may differ in the land-use history and the time since clearing of their surroundings, potentially leading to different time lags and 'extinction debts'. Similarly, regional variation in the character of landscapes, including soils and dominant vegetation types, means that current high cover landscapes are not necessarily representative of low cover landscapes before they were cleared.

The increased probability of species occurrence with the increased availability of proximal vegetation, provides an opportunity to benefit biodiversity by incorporating the restoration and revegetation of contiguous agricultural land into conservation strategies (Bennett and Mac Nally 2004; Thomson et al. 2009; Munro et al. 2011). Although these local gains in extent of habitat and possible gains in connectivity will help sustain species presence, such land-use change will need to be widespread to counter impacts of habitat loss (Dorrough et al. 2004; Vesk and Mac Nally 2006). The species models described here confirm a need for conservation to incorporate more than just dry-infertile land in order to meet species habitat requirements. An important task is to retain and restore more productive parts of the

landscape, not only to provide for the habitat requirements of species but because such sites may also support more productive and resilient bird communities (Watson 2011). However, such change potentially introduces greater trade-offs with agriculture..

In order to increase their likelihood of implementation or adoption (Pannell 2008; Polasky et al. 2005; Stoneham et al. 2003), actions to restore biodiversity on agricultural land must be optimised to minimise loss to production. The predictive models of bird occurrence identified here will be used to inform a spatially explicit, bio-economic optimisation model for the restoration of native vegetation cover on agricultural land (Polyakov et al. 2011). Although constructed for a select suite of woodland-dependent bird species, such models could be derived for other taxa where data are available. The habitat requirements identified in these models ensure that restoration can be optimised for individual species. The question is then how best to optimise habitat restoration across all species (Thomson et al. 2009; Polyakov et al. 2011)? Applied in this way, these models predicting the probability of species occurrence will help address the ongoing challenge in conservation to create a landscape mosaic that can adequately meet targets for both biodiversity conservation and agricultural production.

Acknowledgments

Tables

Table 1. Description of explanatory variables relating to survey sites and surveys

Variable	Description
a) Weighted proportion of vegetation types within 2 km of site:	
D	Dry infertile
F	Fertile
P	Plains
R	Riparian
S	Scattered
b) Spatial and temporal binary variables (dummies):	
G	"1" if site is located in Goldfields region "0" otherwise
Y	"1" for data collected in 2006-2007 study period "0" otherwise

Table 2. Results of model selection for the probability of occurrence of woodland birds at survey sites in northern Victoria. The selected model is the model with lowest AIC value, incorporating both vegetation extent and composition variables. AIC values for alternative models that incorporate vegetation extent alone (i.e. without composition) are also given. For both types of model, AIC values are also presented for an alternative in which variables are not weighted by distance (i.e. without weight). Boldface indicates the model with lowest AIC value. The explanatory variables included in models relate to vegetation type (D dry, F fertile, P plains, R riparian, S scattered trees), bioregion (G) and Year (Y). Sightings refers to the number of surveys (n=1920) on which a species was recorded across both study periods.

Species	Sightings	Selected model	AIC of the models			
			With weight		Without weight	
			With composition	Without composition	With composition	Without composition
Black-chinned Honeyeater (*Melithreptus gularis*)	157	DFR, -S, G, -Y,	**560.9**	570.4	586.8	586.3
Brown Treecreeper (*Climacteris picumnus*)	709	DFR, P, -S,	**1557.3**	1783.7	1747.1	1809.8
Brown-headed Honeyeater (*Melithreptus brevirostris*)	172	DPS, -Y,	**648.0**	664.3	682.0	690.9
Buff-rumped Thornbill (*Acanthiza reguloides*)	58	DP, -G,	**262.1**	265.4	281.9	289.8
Common Bronzewing (*Phaps chalcoptera*)	162	DFR, -S,	**729.3**	739.9	739.2	741.0
Diamond Firetail (*Stagonopleura guttata*)	41	F, PS, R, G,	**255.8**	287.5	271.1	287.7
Dusky Woodswallow (*Artamus cyanopterus*)	188	DFR, G,	**750.4**	758.9	779.0	781.7
Eastern Yellow Robin (*Eopsaltria australis*)	94	DF, P, -Y,	**375.4**	392.6	400.0	421.4
Fuscous Honeyeater (*Lichenostomus fuscus*)	210	D, F, -PR, -S, G, -Y,	**541.6**	558.9	586.2	627.3
Grey Fantail (*Rhipidura fuliginosa*)	96	D, R, -G, -Y,	**437.7**	479.2	454.1	487.5
Grey Shrike-thrush (*Colluricincla harmonica*)	500	DFR, P, -S, G,	**1269.3**	1330.6	1334.1	1337.7
Jacky Winter (*Microeca fascinans*)	143	DFP, R, S,	596.2	653.4	**591.6**	653.1
Mistletoebird (*Dicaeum hirundinaceum*)	67	DFPR, -Y,	**351.4**	359.2	352.9	354.5
Musk Lorikeet (*Glossopsitta concinna*)	427	-DPS, R, G, -Y,	**956.2**	975.8	971.8	983.6
Peaceful Dove (*Geopelia striata*)	72	DPS, F, R,	**389.0**	428.4	402.2	427.7
Purple-crowned Lorikeet (*Glossopsitta porphyrocephala*)	67	-DR, -P, G, -Y,	**356.9**	359.2	357.2	359.5
Red Wattlebird (*Anthochaera carunculata*)	536	-P, R, G, -Y,	1002.4	1006.8	**1002.3**	1005.8
Red-capped Robin (*Petroica goodenovii*)	52	DFP, -G,	**288.3**	296.2	306.0	312.7
Rufous Whistler (*Pachycephala rufiventris*)	142	DFP, -G, -Y,	**537.3**	542.5	552.3	560.4
Spotted Pardalote (*Pardalotus punctatus*)	164	DFP, R, -Y,	**571.9**	611.7	595.5	625.2
Superb Fairy-wren (*Malurus cyaneus*)	259	-FP, R, -Y,	**786.5**	1126.8	915.7	1126.3

Species	Count	Codes				
Varied Sittella (*Daphoenositta chrysoptera*)	36	DS, P,	**252.2**	255.5	259.6	260.0
Weebill (*Smicrornis brevirostris*)	274	DFP, -G, Y,	**855.3**	871.9	882.7	898.0
White-browed Babbler (*Pomatostomus superciliosus*)	113	DF, P, R, S, G, -Y,	**467.1**	479.2	481.0	489.4
White-naped Honeyeater (*Melithreptus lunatus*)	37	DF, -Y,	213.1	218.5	**211.2**	217.9
White-throated Treecreeper (*Cormobates leucophaeus*)	181	DPR, -G,	**600.1**	627.8	641.2	668.6
White-winged Chough (*Corcorax melanorhamphos*)	343	F, P, -RS, Y,	**1105.3**	1120.2	1125.8	1124.9
Yellow Thornbill (*Acanthiza nana*)	87	D, -FS, -G, -Y,	485.7	496.2	**482.8**	500.5
Yellow-tufted Honeyeater (*Lichenostomus melanops*)	254	DFP, -RS, G, -Y,	**660.2**	679.5	733.6	735.4

Table 3. Results of cross-validation of the models of probability of occurrence of woodland bird species. Model diagnostic terms are described in Appendix 1.

Species	Expected information $EI_0(\mathbf{A};\mathbf{X})$	95% confidence interval of the expected information	Information $I_0(\mathbf{A};\mathbf{X})$	Usefulness U_0^2	Log likelihood ratio LLR	99.9% critical value of LLR
Black-chinned Honeyeater	0.476	(0.428, 0.524)	0.467	67.4%	448.6	20.5
Brown Treecreeper	0.197	(0.155, 0.239)	0.170	24.5%	163.1	18.5
Brown-headed Honeyeater	0.440	(0.388, 0.492)	0.431	62.2%	413.7	16.3
Buff-rumped Thornbill	0.611	(0.578, 0.644)	0.600	86.6%	576.0	16.3
Common Bronzewing	0.418	(0.360, 0.476)	0.412	59.4%	395.5	16.3
Diamond Firetail	0.610	(0.567, 0.652)	0.601	86.8%	577.3	20.5
Dusky Woodswallow	0.406	(0.351, 0.460)	0.400	57.7%	383.8	16.3
Eastern Yellow Robin	0.557	(0.519, 0.595)	0.548	79.0%	525.9	18.5
Fuscous Honeyeater	0.497	(0.460, 0.533)	0.482	69.6%	463.1	22.5
Grey Fantail	0.548	(0.500, 0.595)	0.537	77.5%	464.2	20.5
Grey Shrike-thrush	0.187	(0.144, 0.230)	0.174	25.0%	166.6	20.5
Jacky Winter	0.496	(0.449, 0.543)	0.481	69.4%	461.9	18.5
Mistletoebird	0.573	(0.528, 0.618)	0.565	81.5%	542.1	16.3
Musk Lorikeet	0.314	(0.272, 0.356)	0.305	44.0%	293.0	20.5
Peaceful Dove	0.561	(0.513, 0.609)	0.549	79.1%	526.7	18.5
Purple-crowned Lorikeet	0.567	(0.520, 0.613)	0.557	80.4%	534.9	20.5
Red Wattlebird	0.252	(0.209, 0.296)	0.247	35.6%	236.8	20.5
Red-capped Robin	0.596	(0.555, 0.638)	0.588	84.8%	564.6	16.3
Rufous Whistler	0.482	(0.433, 0.532)	0.476	68.6%	456.7	18.5
Spotted Pardalote	0.464	(0.416, 0.512)	0.457	65.9%	438.6	18.5
Superb Fairy-wren	0.452	(0.407, 0.497)	0.426	61.5%	409.4	18.5
Varied Sittella	0.609	(0.564, 0.654)	0.603	87.0%	578.9	16.3
Weebill	0.381	(0.334, 0.428)	0.368	53.1%	353.6	18.5

White-browed Babbler	0.535	(0.493, 0.577)	0.517	74.7%	496.8	24.3
White-naped Honeyeater	0.616	(0.576, 0.656)	0.612	88.2%	587.2	16.3
White-throated Treecreeper	0.500	(0.458, 0.541)	0.483	69.6%	463.2	16.3
White-winged Chough	0.243	(0.190, 0.295)	0.233	33.5%	223.2	20.5
Yellow Thornbill	0.537	(0.488, 0.587)	0.521	75.2%	500.6	18.5
Yellow-tufted Honeyeater	0.479	(0.442, 0.516)	0.458	66.0%	439.4	18.5

Table 4. Proportion of residual uncertainty explained by the models with both distance weighting and vegetation composition in comparison with the equal probability model (U^2_0), the probability proportional to the occurrence of outcomes in the sample (U^2_1), and the model with intercept, random effects, dummies and vegetation extent (U^2_2)

| Species	U^2_0 (%)	U^2_1 (%)	U^2_2 (%)
Black-chinned Honeyeater	67.4	20.8	3.4
Brown Treecreeper	24.5	20.6	11.4
Brown-headed Honeyeater	61.1	13.2	4.3
Buff-rumped Thornbill	86.6	32.1	8.5
Common Bronzewing	59.4	3.1	0.7
Diamond Firetail	86.8	11.9	8.0
Dusky Woodswallow	57.7	8.9	2.9
Eastern Yellow Robin	79.0	26.3	7.4
Fuscous Honeyeater	69.6	39.7	10.6
Grey Fantail	77.1	20.8	8.8
Grey Shrike-thrush	25.0	9.6	3.1
Jacky Winter	69.4	20.7	6.8
Mistletoebird	81.5	15.2	0.5
Musk Lorikeet	43.9	27.2	1.8
Peaceful Dove	79.1	10.3	6.9
Purple-crowned Lorikeet	80.4	10.6	0.0
Red Wattlebird	35.6	24.6	0.1
Red-capped Robin	84.8	16.2	6.1
Rufous Whistler	68.6	17.7	3.2
Spotted Pardalote	65.9	19.1	5.9
Superb Fairy-wren	61.5	33.2	27.6
Varied Sittella	87.0	4.2	2.2
Weebill	53.1	20.9	4.1
White-browed Babbler	74.7	22.0	1.8
White-naped Honeyeater	88.2	14.8	1.2
White-throated Treecreeper	69.6	32.7	8.5
White-winged Chough	33.5	1.9	1.1
Yellow Thornbill	75.2	7.5	2.5
Yellow-tufted Honeyeater	66.0	40.6	8.5
Mean	**66.0**	**18.8**	**5.4**
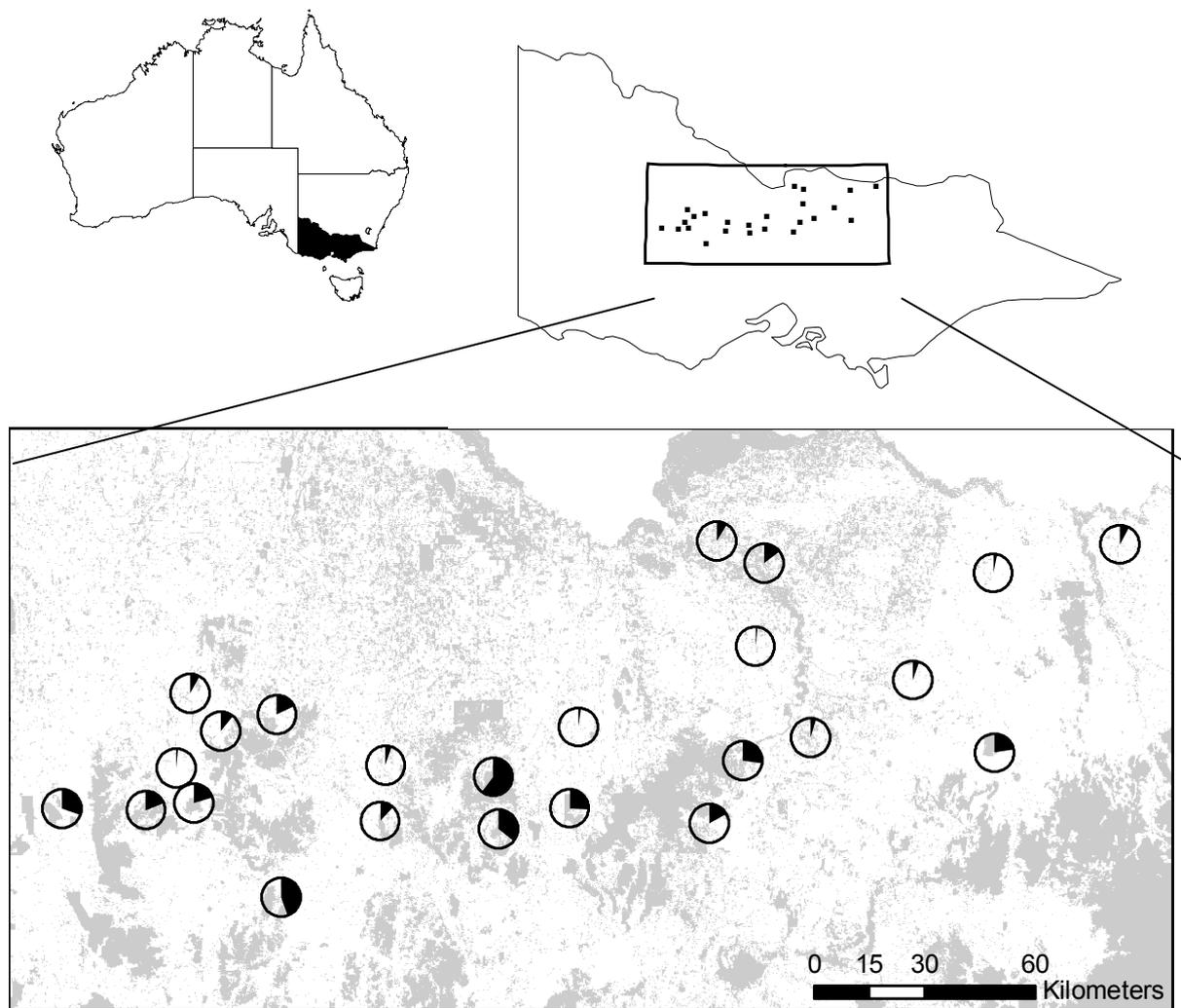
Figure 1: Location of the study landscapes. Grey indicates native vegetation; pie charts represent the proportion of tree cover in each landscape.

References

Balmford A, Green R, Phalan B (2012) What conservationists need to know about farming. Proceedings of the Royal Society B (doi:10.1098/rspb.2012.0515)

Bennett AF, Mac Nally R (2004) Identifying priority areas for conservation action in agricultural landscapes. Pacific Conservation Biology 10:106-123

Bennett AF, Radford JQ, Haslem A (2006) Properties of land mosaics: Implications for nature conservation in agricultural environments. Biological Conservation 133:250-264

Benton TG, Vickery JA, Wilson JD (2003) Farmland biodiversity: is habitat heterogeneity the key? Trends in Ecology & Evolution 18(4):182-188

Betts MG, Diamond AW, Forbes GJ, Villard MA, Gunn JS (2006) The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. Ecol Model 191:197-224

Boscolo D, Metzger JP (2011) Isolation determines patterns of species presence in highly fragmented landscapes. Ecography 34:1018-1029

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecol Model 157:281-300

Brooks TM et al. (2004) Coverage provided by the global protected-area system: Is it enough? BioScience 54:1081-1091

Brown GW, Dorrough JW, Ramsey DSL (2010) Landscape and local influences on patterns of reptile occurrence in grazed temperate woodlands of southern Australia. Landscape and Urban Planning 103:277-288

Debinski DM, Holt RD (2000) A survey and overview of habitat fragmentation experiments. Conservation Biology 14:342-355

Dorrough J, Yen A, Turner V, Clark SG, Crosthwaite J, Hirth JR (2004) Livestock grazing management and biodiversity conservation in Australian temperate grassy landscapes. Australian Journal of Agricultural Research 55:279-295

DSE (2002) Victoria's native vegetation management: a framework for action. In. Department of Sustainability and Environment, Melbourne, Victoria

DSE (2004) Corporate Geospatial Data Library: TREEDEN25. In. Land Victoria, Melbourne, Australia

DSE (2007) Overview of native vegetation datasets, Information sheet No. 1. In. Department of Sustainability and Environment, Melbourne, Victoria

Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Annu Rev Ecol Evol Syst 34:487-515

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence / absencce models. Environmental Conservation 24:38-49

Fischer J, Zerger A, Gibbons P, Stott J, Law BS (2010) Tree decline and the future of Australian farmland biodiversity. PNAS 107:19597-19602

Foley JA et al. (2005) Global consequences of land use. Science 309:570-574

Foley JA et al. (2011) Solutions for a cultivated planet. Nature 478:337-342

Fuller RJ, Trevelyan RJ, Hudson RW (1997) Landscape composition models for breeding bird populations in lowland English farmland over a 20 year period. Ecography 20:295-307

Gibb H, Cunningham SA (2010) Revegetation of farmland restores function and composition of epigaeic beetle assemblages. Biological Conservation 143:677-687

Gove AD, Majer JD, Rico-Gray V (2009) Ant assemblages in isolated trees are more sensitive to species loss and replacement than their woodland counterparts. Basic and Applied Ecology 10:187-195

Green RE, Cornell SJ, Scharlemann JPW, Balmford A (2005) Farming and the fate of wild nature. Science 307:550-555

Haslem A, Bennett AF (2008a) Birds in agricultural mosaics: The influence of landscape pattern and countryside heterogeneity. Ecological Applications 18:185-196

Haslem A, Bennett AF (2008b) Countryside elements and the conservation of birds in agricultural environments. Agriculture, Ecosystems and Environment 125:191-203

Hauser JR (1978) Testing Accuracy, Usefulness, and Significance of Probabilistic Choice Models - Information-Theoretic Approach. Oper Res 26:406-421

Higgins PJ, Peter JM, Cowling SJ (2006) Handbook of Australian, New Zealand and Antarctic Birds, Volume 7 (Dunnock to Starlings) Part B. Oxford University Press, Melbourne

Hinsley SA, Bellamy PE (2000) The influence of hedge structure, management and landscape context on the value of hedgerows to birds: A review. Journal of Environmental Management 60(1):33-49

Hobbs RJ, Yates CJ (eds) (2000) Temperate Eucalypt Woodlands in Australia. Biology, Conservation, Management and Restoration. Surrey Beatty & Sons, Chipping Norton, New South Wales

Hoover EM (1971) An Introduction to Regional Economics. Alfred A. Knopf, New York

Law BS, Dickman CR (1998) The use of habitat mosaics by terrestrial vertebrate fauna: implcations for conservation and management. Biodiversity and Conservation 7:323-333

Lindenmayer D, Bennett A, Hobbs R (eds) (2010a) Temperate Woodland Conservation and Management. CSIRO Publishing, Melbourne

Lindenmayer DB, Bennett AF, Hobbs RJ (2010b) An overview of the ecology, management and conservation of Australia's temperate woodlands. Ecological Management and Restoration 11:201-209

Lindenmayer DB, Knight EJ, Crane MJ, Montague-Drake R, Michael DR, MacGregor CI (2010c) What makes an effective restoration planting for woodland birds? Biological Conservation 143:289-301

Lowe KW, Ahern LD, Park G, Moorrees A, Price R (2002) Biodiversity Action Planning: Strategic Overview for the Goldfields Bioregion Victoria. In. Department of Natural Resources and Environment, Melbourne, Victoria

Mac Nally R, De Vries L, Thomson JR (2010) Are replanted floodplain forests in south-eastern Australia providing bird biodiversity benefits? Restoration Ecology 18:85-94

Mac Nally R, Bennett AF, Horrocks G (2000) Forecasting the impacts of habitat fragmentation. Evaluation of species-specific predictions of the impact of habitat fragmentation on birds in the box-ironbark forests of central Victoria, Australia. Biological Conservation 95:7-29

Mac Nally R et al. (2009) Collapse of an avifauna: climate change appears to exacerbate habitat loss and degradation. Diversity and Distributions 15:720-730

McGarigal K, Cushman SA (2002) Comparative evaluation of experimental approaches to the study of habitat fragmentation effects. Ecol Appl 12:335-345

Munro NT, Lindenmayer DB, Fischer J (2007) Faunal response to revegetation in agricultural areas of Australia: A review. Ecological Management & Restoration 8(3):199-207

Munro NT, Fischer J, Barrett G, Wood J, Leavesley A, Lindenmayer DB (2011) Bird's response to revegetation of different structure and floristics—are "restoration plantings" restoring bird communities? Restoration Ecology 19:223-235

Pannell DJ (2008) Public benefits, private benefits, and policy mechanism choice for land-use change for environmental benefits. Land Economics 84:225-240.

Polasky S, Nelson E, Lonsdorf E, Fackler P, Starfield A (2005) Conserving species in a working landscape: Land use with biological and economic objectives. Ecologial Applications 15:1387-1401

Polyakov M, Majumdar I, Teeter L (2008) Spatial and temporal analysis of the anthropogenic effects on local diversity of forest trees. Forest Ecology and Management 255:1379-1387

Polyakov M, Pannell DJ, Rowles AD, Park G, Roberts A (2011) Optimising the spatial pattern of landscape revegetation. In: Australian Agricultural and Resource Economics Society Conference, Melbourne, Victoria

Quinn GP, Keough MJ (2002) Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge, United Kingdom

Radford JQ, Bennett AF (2007) The relative importance of landscape properties for woodland birds in agricultural environments. Journal of Applied Ecology 44:737-747

Radford JQ, Bennett AF, Cheers GJ (2005) Landscape-level thresholds of habitat cover for woodland-dependent birds. Biological Conservation 124:317-337

Stoneham G, Chaudhri V, Ha A, Strappazzon L (2003) Auctions for conservation contracts: An empirical examination of Victoria's BushTender trial. Australian Journal of Agricultural and Resource Economics 47:477-500

Tews J et al. (2004) Animal species diversity driven by habitat heterogeneity/diversity: The importance of keystone structures. Journal of Biogeography 31:79-92

Thomson JR, Mac Nally R, Fleishman E, Horrocks G (2007) Predicting bird species distributions in reconstructed landscapes. Conservation Biology 21:752-766

Thomson JR, Moilanen AJ, Vesk PA, Bennett AF, Mac Nally R (2009) Where and when to revegetate: a quantitative method for scheduling landscape reconstruction. Ecological Applications 19:817-828.

Tilman D, Fargione J, Wolff B et al (2001) Forecasting agriculturally driven global environmental change. Science 292:281-284

Vandermeer J, Perfecto I (2007) The agricultural matrix and a future paradigm for conservation. Conservation Biology 21:274-277

Vesk PA, Mac Nally R (2006) The clock is ticking - Revegetation and habitat for birds and arboreal mammals in rural landscapes of southern Australia. Agriculture, Ecosystems and Environment 112:356-366

Watson DM (2011) A productivity-based explanation for woodland bird declines: poorer soils yield less food. Emu 111:10-18.

Westphal MI, Field SA, Tyre AJ, Paton D, Possingham HP (2003) Effects of landscape pattern on bird species distribution in the Mt. Lofty Ranges, South Australia. Landscape Ecology 18:413-426

Wierzbowski P, Lowe KW, Handley K, Berwick S, Robinson D, Ahern LD (2002) Biodiversity Action Planning: Strategic Overview for the Victorian Riverina Bioregion. In. Department of Natural Resources and Environment, Melbourne, Victoria

Appendix 1.

The set of information-theoretic statistics (Hauser 1978) used for evaluating the predictive performance of models.

1. The information index $I(\mathbf{A};\mathbf{X})$, quantifies the information provided by the model:

$$I(\mathbf{A};\mathbf{X}) = \frac{1}{N}\sum_{n=1}^{N}\sum_{j\in(0,1)} (y_n = j)\ln\left(\frac{p(y_n = j\mid\mathbf{x}_n)}{p(y_n = j)}\right)$$

where y_n is the outcome for observation n, which takes value 0 if species was not observed and 1 if species was observed, $p(a)$ is the prior probability of an outcome, and $p(a\mid\mathbf{x}_n)$ is the probability of outcome predicted by the model given vector of explanatory variables $\mathbf{x}_n$.

2. The index of the prior entropy:

$$H(\mathbf{A}) = \sum_{j\in(0,1)} p(y_n = j)\ln\left(\frac{1}{p(y_n = j)}\right)$$

defines the uncertainty inherent in the null model. Because this expression is equivalent to $I(\mathbf{A};\text{Perfect knowledge})$, it quantifies the maximum uncertainty that could be explained by the estimated model.

3. The index of prior entropy permits a *usefulness test* by measuring the proportion of uncertainty explained by the estimated model:

$$U^2 = \frac{I(\mathbf{A};\mathbf{X})}{H(\mathbf{A})}.$$

4. The expected information provided by the estimated model can be expressed as:

$$EI(\mathbf{A};\mathbf{X}) = \frac{1}{N}\sum_{n=1}^{N}\sum_{j\in(0,1)} p(y_n = j \mid \mathbf{x}_n)\ln\left(\frac{p(y_n = j \mid \mathbf{x}_n)}{p(y_n = j)}\right).$$

5. The variance of $I(\mathbf{A};\mathbf{X})$ is:

$$V(\mathbf{A};\mathbf{X}) = \frac{1}{N^2}\sum_{n=1}^{N}\left\{\left[\sum_{j\in(0,1)} p(y_n = j \mid \mathbf{x}_n)\left[\ln\left(\frac{p(y_n = j \mid \mathbf{x}_n)}{p(y_n = j)}\right)\right]\right]^2 \right.$$
$$\left. - \left[\sum_{j\in(0,1)} p(y_n = j \mid \mathbf{x}_n)\ln\left(\frac{p(y_n = j \mid \mathbf{x}_n)}{p(y_n = j)}\right)\right]^2\right\}.$$

6. The log–likelihood ratio:

$$LLR = 2n \times I(\mathbf{A};\mathbf{X})$$