Modelling Habitat Suitability for the White-throated Treecreeper (Cormobates leucophaeus); a GIS Approach

Alizadeh Shabani, A., L. McArthur, and M. Abdollahian

School of Mathematics and Geospatial Sciences, RMIT University, Melbourne
E-Mail: s3005184@student.rmit.edu.au

Keywords: *Habitat modelling; Cormobates leucophaeus; habitat suitability; binary logistic regression; GIS.*

EXTENDED ABSTRACT

The habitats of plants change through time, and the plants change with them. Plants, in fact, are able themselves to change various aspects of the environment they inhabit. Succession occurs because altered environmental conditions favour certain species, which therefore can compete more successfully than before for nutrients, light, space and many other factors. As a result, populations of well-adapted species replace earlier ones now less well equipped to compete in the altered conditions. Thus, during the course of succession, individual populations come and go, giving rise to a gradual progressive change in the community. This can be considered as the natural process of the vegetation change. But most of the time human presence is inexorably changing the vegetation conditions. Human influence, through farming, deforestation, and urbanization, fragments the natural plant communities into smaller and smaller units. These activities affect not only the plant communities, but the animal communities as well.

This fact leads us to search for a reliable method, which can be used widely on a regional scale, to determine the level of ecosystem changes that may affect the biodiversity as a consequence. In most of the cases, fragmentation of terrestrial habitats is a warning that they will be destroyed soon. On the other hand, measuring and investigating the cause of such losses may be costly and/or time-consuming. One of the methods used to investigate the complexity of an ecosystem is the use of indicators. To determine indicators of ecosystem conditions among wildlife, first one must know about the ecological needs of a species and the factors that affect them.

In this study, the White-throated Treecreeper was selected to be used as an indicator of the quality of forest patches in the Mornington Peninsula and Western Port Biosphere Reserve. However, at the first instance the distributional pattern of the species must be identified. Thus, a probabilistic model of its presence was developed, based on recorded data through a four year period.

Species data were obtained from Birds Australia. This dataset indicates the coordinates of each individual within the study area during four years. For determining habitat patches, the Ecological Vegetation Class layer was used. An ArcView extension called Landscape Context Tool was used to identify and rate habitat patches in both patch and landscape scales. In order to calculate the distance of each habitat patch to built-up areas and coast line, the centroid of patches were determined.

By overlaying the species data on the habitat patch layer, the presence of individuals in particular patches was demonstrated. Since the species data is showing only the presence, and the absence is not recorded, Binary Logistic Regression was performed to develop the distribution model of the White-throated Treecreeper according to four habitat variables.

The results show a very significant correlation ($P=0.00$) between the presence of the Treecreepers and considered environmental gradients. To test the model, three Goodness of Fit tests were carried out.

To validate the model, suitable habitat patches from the model were created by using GIS software. Then, locations that the model showed as suitable habitats of the White-throated Treecreeper, but where there were no observations recorded, were checked in the field. The field test showed that the model was successful in predicting the probability of the presence of the White-throated Treecreepers in the Mornington Peninsula and Western Port Biosphere Reserve.

1. INTRODUCTION

For many years, ecologists have been arguing that the diversity of life on earth (biodiversity) provides the essential support for sustainability. These arguments have been focussed considerably over the past 5 years as a way to get scientists, economists and the rest of our human communities thinking about and discussing the relationships between people and their natural environments and the values and opportunities that come from those relationships.

Loss of habitat is widely recognized as the greatest threat to wildlife (as one of the main elements of the natural environments) today. But even on land that has been permanently protected within our preserves, habitat does not remain constant over time. It is normal for changes in vegetation to occur as a result of long periods of wet or dry weather, fire, and other natural causes. Species have adapted to these natural processes, sometimes known as succession. Evaluating the degree of association between wildlife and their habitat, and to determine the relative importance of specific habitat characteristics for conservation purposes, requires procedures that are relatively inexpensive and that yield fast accurate results.

Since fragmentation is among the important known threats to natural wildlife habitats, it must be detected at the early stages for better decision making. The detection method should be reliable, applicable, cost-effective, and time-saving. One of the methods used to investigate the complexity of an ecosystem is the use of indicators. Indicators are variables or indices that represent, integrate, and characterize information embodied in comprehensive data sets, which are often not measurable directly (Muller et al., 1999). Indicators are suitable tools whenever the primary information of an object is too complex to be handled without aggregations. Using biological indicators for management purposes is of great value, especially for developing countries in terms of being cost effective and time saving. But to identify indicators, one must have useful data from both the group of species from which indicators are to be selected, and environmental gradients that represent the ecosystem and biodiversity status. These datasets can be obtained in developed countries, but they are not available in most developing countries. Therefore, a method, in which the species and environmental factors have basic criteria for indicator identification will be needed.

These primary criteria on identifying indicators led to the choice of birds as the species group. Birds select habitats based on their suitability, which makes them very useful as biomonitors indicating the environmental changes taking place. Among the most important factors affecting habitat selection are foraging sites, nesting sites, protection from the elements and predators, and competition. However, with the earth constantly changing due to climatic variations, pollution, human activity, and development, habitat selection is becoming increasingly difficult. Each stressor can take a serious toll on individuals and communities of birds. For this reason, birds can be very useful as biomonitors indicating the environmental changes taking place, even sometimes before apparent and visual changes can be detected. For this purpose, one should know about the ecological requirements of a species, and be aware of how changes in habitat conditions affect the distribution and habitat selection of the individuals. Once known, the species presence represents the ecosystem characteristics.

During recent years there has been growing attention to the need to consider models as an integral part of GIS, and to improve understanding and application of models. When models are applied to the environment, it is expected that insights about the physical, biological, or socio-economic system may be derived. They may also allow prediction and simulation of future conditions. The reasons for building models are to understand, and ultimately manage, a sustainable system.

A variety of analytical techniques have been used to investigate the relationship of animal distribution and environment. These include logistic regression (Pereira and Itami, 1991; Buckland and Elston, 1993; Osborne and Tigar, 1992; Walker, 1990; Rodriguez, 1997), discriminant analysis (Haworth and Thompson, 1990) classification and regression trees (Walker and Moore, 1988; Skidmore et al., 1996), canonical correlation analysis (Andries et al., 1994), supervised non-parametric classifiers (Skidmore, 1998; Skidmore et al., 1996) and neural networks (Skidmore et al., 1997).

Presence-only models are derived from datasets in which only the locations of known presences of a species are recorded. No record is kept of those areas that were surveyed without detecting the species, which are absent sites. Because of this, presence-only datasets are particularly biased in geographical or environmental coverage. However, these datasets can be analysed to predict the relative likelihood of the presence sample sites and then compare those sites with areas in which no observations were recorded, according to the studied environmental metrics.

This study aims to show how the presence of the White-throated Treecreeper is correlated to known environmental variables and habitat metrics, and to investigate the possibility of building a probabilistic model for the presence (and/or absence) of this bird species.

2. METHODS AND STUDY AREA

Habitat selection among avifauna operates through a series of behavioural decisions at several spatial scales, that's why studying their distributional pattern is difficult. Based on our knowledge of their natural history and ecological factors, some of the most relevant biophysical and anthropogenic factors that most affect their habitat suitability are studied in order to develop a distribution model of focal bird species within the study area.

The distribution of a species may be related to many independent variables identified by GIS. Many layers may be irrelevant and the knowledge about the ecology of the species could reduce the number of the unnecessary independent variables included in the analysis. Thus, the most important factors that seemed to have profound effect on the distributional pattern of the species were identified according to the existing dataset and literature (e.g. Simpson and Day, 2004, and Frith, 1976).

The bird data for this study was obtained from Birds Australia. This dataset indicates the geographical coordinates of bird observation records during a four year period from 1998 to 2002. This information has been converted into point data to be used in GIS software.

Ecological Vegetation Classes (EVC) (NRE, 1997) were used to identify habitat patches, according to each species' ecological need, and the spatial records within each unit. After identifying 'habitat patches', those patches in which observations of the species were recorded were discriminated with other patches.

Since the "patch rate" and "landscape context rate", which describe the size, shape, connectivity, and abundance of potential habitats, were the best determinant factors for the presence of the species, they were plotted with each other to find out the probable limits of the species presence. From the outcome, patches with suitable characteristics for the species where selected using GIS software. Thus, a set of potential suitable habitat patches were identified according to the habitat metrics.

Other important factors, in terms of habitat selection in White-throated Treecreepers that were included in the analysis are the distance from the sea shore and built-up areas. In other words, it seemed that habitat patches located at a specific distance from the coast line and human settlement, were preferred.

Since identified habitat patches had different shapes, in order to calculate their distance from shore and built-up areas, the centre of mass (centroid) of patches were determined.

The most common analysis to define habitat suitability in the case, where records show only the presence, is logistic regression. Since other factors rather than vegetation seemed to be important for at least some of the species, and factor interactions might be significant, a species distribution model is required, based on environmental factors that determine habitat suitability.

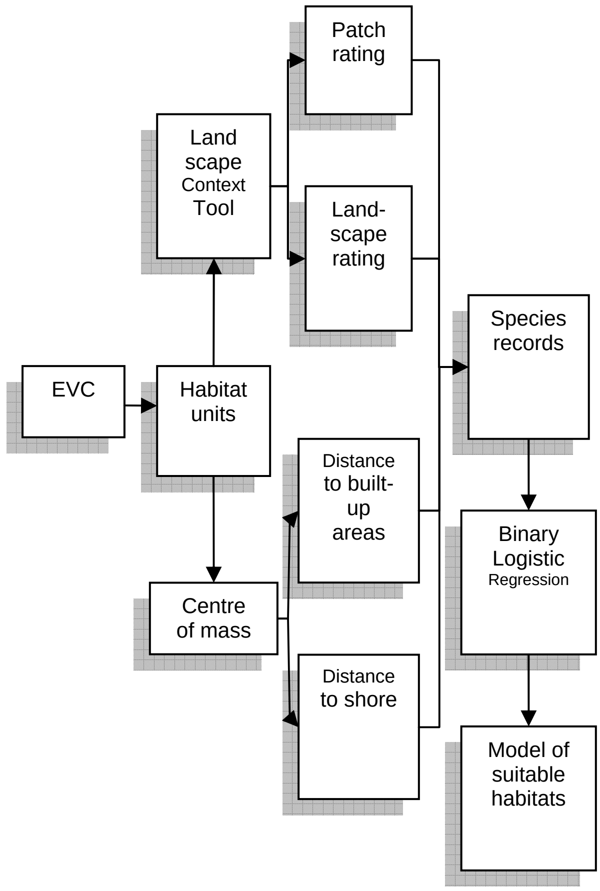

Figure 1- Building habitat suitability model process

For this purpose, a General Linear Model was developed by using binary logistic Regression design. This statistical analysis is especially helpful when presences have to be compared with absences (binary). At the same time, the outcomes show us how much a factor is significant in terms of bird distribution. Binary logistic regression has also been used to classify observations into one of two categories, and it may give fewer classification errors than discriminant analysis for some cases. Minitab (2003) was used to run the binary logistic regression for the White-Throated Treecreeper data. The presence of the species in some habitat patches was considered as the response of the model. Four environmental factors (patch rates, landscape rates, distance of patches to built-up areas, and distance of patches to sea shores) were the model predictors. In addition, three goodness-of-fit tests (Pearson, Deviance, and Hosmer-Lemeshow) were performed; a table of observed and expected frequencies, and measures of association were produced. Finally, to validate the model, a field inspection was carried out to ensure the appropriateness of the built models.

In Figure 1, the process of building the model of habitat suitability is demonstrated.

3. RESULTS

Since the logistic coefficients (Table 1) are measures of the change in log odds associated with a one unit change in the explanatory variable if it is a continuous numerical variable, they are used as coefficients of variables to produce the model.

Thus the model for the predicted presence of species "*i*" can be written as in Equation 1.

Equation 1- Linear predictor model of species distribution

$$Y_i = \beta_{0i} + \beta_{1i}x_{1i} + \beta_{2i}x_{2i} + \beta_{3i}x_{3i} + \beta_{4i}x_{4i}$$

Where Y_i is the value of the linear predictor for species *i*, β_{0i} is the constant coefficient, β_{1i} to β_{4i} are variable coefficients, and x_{1i} to x_{4i} are variable values.

These equations give the value of the linear predictor for each species. Therefore, to calculate the predicted probability of the presence of a bird species in a habitat, with a known patch rate, landscape rate, nearest distance to built-up area, and nearest distance to sea shore, Equation 2 is used, where "Y_i" indicates the outcome result of the above:

Equation 2-The probability of the presence of species *i* according to the linear model

$$P_{(presence)} = \frac{1}{1+\exp(-Y_i)}$$

The result will be a number between 0 and one, the closer to one; the higher is the probability of the species presence.

The log-likelihood of the model was calculated along with the G statistic. The G value tests that all the coefficients associated with model variables equal zero versus these coefficient not all being equal to zero. This value is especially useful when the P-value is greater than 0.05. In this study the P-value of the logistic regression analysis for the White-throated Treecreeper is equal to zero. In other words, this shows that there is sufficient evidence that at least one of the coefficients is different from zero, given that the accepted α level is less than 0.05.

Normally, in a modelling process using binary logistic regression, at this stage it is assumed that the data fits the model at a very significant level (p=0.000), and there will be no need for further analysis. However, since the aim of this study is to identify indicators among bird species to represent the conditions of (or possible changes in) the environment, more detailed investigation is needed to make sure that this model can be used to predict the habitat suitability for management purposes. Therefore, other tests such as goodness-of-fit, groupings, and measures of association should be carried out to ensure the validity of the model for indicator determination.

Table 1 - Logistic Regression Table for White-Throated Treecreeper: Log-Likelihood = -2631.065 (Test that all slopes are zero: G = 757.139, DF = 4, P-Value = 0.000)

Predictor	Coef	SE Coef	Z	P	Odds Ratio	Lower	Upper
Constant	-5.45271	0.198098	-27.53	0.000			
Patch Rate	-0.080288	0.0051079	-15.72	0.000	0.92	0.91	0.93
Landscape Rate	0.176747	0.0079219	22.31	0.000	1.19	1.17	1.21
Distance to bld.	-0.000119	0.0000081	-14.81	0.000	1.00	1.00	1.00
Distance to shore	-0.000006	0.0000035	-1.72	0.085	1.00	1.00	1.00

3.1 Goodness-Of-Fit Tests

A variety of statistical tests can be applied in order to assess how well the model describes the data. In this study three tests (Pearson, Deviance, and Hosmer-Lemenshow) were used. The Hosmer-Lemenshow test assesses the fit of the model by comparing the observed and expected frequencies. The estimated probabilities are grouped from lowest to highest, and then the Chi Square statistic is calculated to determine if the observed and expected frequencies are significantly different. When the Hosmer-Lemenshow test is significant, it means that the observed counts and those predicted by the model are not close and the model does not describe the data well, and vice versa. The results show a significant *P*-value of the Hosmer-Lemenshow test for all seven selected species. This is due to the great difference between presence and absence data, so using this test is not a reliable method to test the data fit in the model. On the other hand, the Deviance and Pearson tests, which are types of residuals, show a large *P*-value (>0.05) that indicates that there is a good fit of the data in the model. Table 2 shows a summary of goodness-of-fit tests for the White-throated Treecreeper.

Table 2-Summary of applied goodness-of-fit tests

Pearson P	*Deviance P*	*Hosmer-Lemenshow P*
1.000	1.000	0.000

3.2 Diagnostic Plot

The delta Chi-Square versus probability plots helps to identify patterns that did not fit well. Delta Chi-Square measures changes in the Pearson Goodness-of-fit statistics due to deleting a particular factor. Points that are away from the main cluster correspond to low predicted probability. In this study, according to this plot, there are only a few points away from the main cluster, and these represent low probability of presence (Figure 2).

3.3 Model Validation

To validate the model, suitable habitat patches from the model were created by using GIS software. According to the model, habitat patches with higher patch ratings and landscape ratings, also those units located further from built-up areas were determined. Then, locations that the model showed as suitable habitats of the White-throated Treecreeper, but where there were no observations recorded, were determined and checked in the field. The field observation at the predicted site, by using a portable GPS (Geographical Positioning System) and determination on the map, led to find a colony of White-throated Treecreepers at the site.

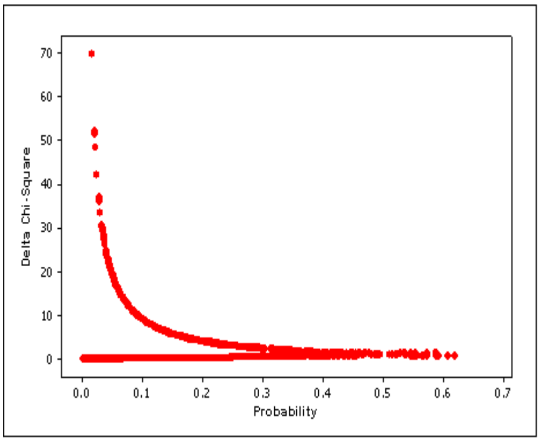

Figure 2- Plotted delta Chi Square versus probability for White-throated Treecreeper (the points forming a horizontal cluster represent the absence data)

4. DISCUSSION AND CONCLUSION

Wildlife distribution and abundance patterns depend on many environmental factors. This study was a simplification of the process of habitat selection, and identifying the most affecting factors in habitat preference by the White-throated Treecreeper. The study clearly showed that specific vegetation patterns were playing the most important determinant role for this species.

Two variables (distance to shores and distance to built-up areas) had close correlation for some locations, since they are broadened to the same extension. However, for this species that have farther observations from the built-up areas, because of the peninsula this is not happening. In this case, these two factors have split effects.

The model produced in this study successfully predicts the probability of the presence of the White-throated Treecreeper by knowing the value of the studied environmental factors. Checking the presence of selected bird species in the field was performed to validate the model. Where the model indicated high probability of species occurrences was the location to be inspected in the field. This study showed that the model is effectively operating (within the study area).

Since the final goal of such study, is to apply the method in developing countries as a helping tool for ecosystem managers, it is suggested that for better understanding of the ecological processes, other case studies with similar focus on specialist species be undertaken. This could be a shortcut for habitat managers in terms of saving time and cost, and also a big help for nature conservation.

REFERENCES

Andries A. M., H. Gulinck (1994). Spatial modelling of the barn owl *Tyto alba* habitat using landscape characteristics derived from SPOT data. *Ecography* (17): 278-287.

Buckland, S. T., and D. A. Elston (1993). Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology* (30): 478-495.

Frith, h. j. (1976). *Complete book of Australian Birds*. Sydney, Readers Digest Services PTY LTD.

Haworth, P. F., and D. B. A. Thompson (1990). Factors associated with the breeding distribution of upland birds in the south pennines, England. *Journal of Applied Ecology* (27): 562-577.

Minitab Inc. (2003). MINITAB Statistical Software, Release 14 for Windows, State College, Pennsylvania.

Muller, F., R. Hoffmann-Kroll, and H. Wiggering. (2000). Indicating ecosystem integrity, theoretical concepts and environmental requirements. *Ecological Modelling* 130: 13-23.

NRE, (1997). Victoria's Biodiversity Strategy. Department of Natural Resources and Environment, Victoria, Australia.

Osborne, P. E. and B. J. Tigar (1992). Interpreting bird atlas data using logistic models: an example from Lesotho, Southern Africa. *Journal of applied ecology* (29): 55-62.

Pereira, J. M. C. and R. M. Itami (1991). GIS-based habitat modelling using logistic multiple regression: a study of the Mt. Graham red squirrel. *Photogrammetric Engineering and Remote Sensing* (57): 1475-1486.

Rodriguez, D. J. (1997). A method to study competition dynamics using de Wit replacement series experiments. *Oikos* (78): 411.

Simpson, K. and N. Day (2004). *Field guide to the birds of Australia*. Penguin.

Skidmore, A. K. and A. Gauld (1996). Classification of kangaroo habitat distribution using three GIS models. *International Journal of Geographical Information Systems* (10): 441-454.

Skidmore, A. K., and B. J. Turner (1997). Performance of a neural network: Mapping Forest Using GIS and Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing* (63): 501-514.

Skidmore, A. K. (1998). Nonparametric Classifier for GIS data Applied to kangaroo distribution mapping. *Photogrammetric Engineering and Remote Sensing* (64): 217-226.

Walker, P. A. (1990). Modelling wildlife distribution using a geographic information system: kangaroos in relation to climate. *Journal of biogeography* (17): 279-289.

Walker, P. A. and D. M. Moore (1988). Simple, an inductive modelling and mapping tool for spatially-oriented data. *International Journal of Geographical Information Systems* (2): 347-363.