

Sensitivity Of Conservation Planning To Different Approaches To Using Predicted Species Distribution Data

Kerrie A. Wilson, Michael I. Westphal, Hugh P. Possingham and Jane Elith

Abstract

The main role of conservation planning is to design reserve networks to protect biodiversity in situ. Research within the field of conservation planning has focused on the development of theories and tools to design reserve networks that protect biodiversity in an efficient and representative manner. Whilst much progress has been made in this regard, there has been limited assessment of the sensitivity of conservation planning outcomes to uncertainty associated with the datasets used for conservation planning. Predicted species distribution data are commonly used for conservation planning because the alternatives (e.g. survey data) are incomplete or biased spatially. However, there may be considerable uncertainty associated with the use of predicted species distribution data, particularly given the variety of approaches available to generate a dataset from such predictions for use in conservation planning. These approaches range from using the probabilistic data directly to using a threshold identified a priori or a posteriori to convert the probabilistic data to presence/absence data. We assess the sensitivity of conservation planning outcomes to different uses of predicted species distribution data. The resulting reserve networks differed, and had different expected species representation. The choice of approach will depend on how much risk a conservation planner is willing to tolerate and how much efficiency can be sacrificed.

Keywords: conservation planning; uncertainty; reserve design; species distribution models

1. Introduction

The increasing pressures exerted on our natural environment by the human population make conservation areas crucial for the persistence of biological diversity (McNeely, 1994 and Groombridge and Jenkins, 2002). We use the term 'conservation area' to describe any area of land or sea managed for the persistence of biodiversity and natural processes in situ, through constraints on incompatible land uses. The term 'reserve network' denotes a system of conservation areas.

Selecting conservation areas in an ad hoc manner or for the protection of particular species generally results in the conservation of economically marginal land and unrepresentative reserve networks (Mark, 1985, Henderson, 1992, Pressey, 1993, Pressey, 1994, Pressey and Tully, 1994 and Rouget et al., 2003). In recognition of these problems and the limited resources available for conservation, research efforts in the field of conservation planning have focused on the development of principles and tools to configure efficient and representative reserve networks.

Scoring systems were developed in the 1980s in an attempt to provide an explicit and rational basis for selecting conservation areas (Margules and Usher, 1981, Terborgh and Winter, 1983, Purdie et al., 1986, Smith and Theberge, 1986, Smith and Theberge, 1987, Usher, 1986, Margules et al., 1988 and Pressey and Nicholls, 1989). These systems scored or rated potential conservation areas against several criteria to provide an overall indication of their conservation value. However, scoring systems did not identify efficient and representative reserve networks (Pressey and Nicholls, 1989 and Pressey, 1997). In a move away from scoring systems, systematic conservation planning techniques were developed. This commenced with Kirkpatrick's introduction of the minimum-set, which involved the identification of a network of complementary conservation areas that efficiently achieved a set of conservation targets (Kirkpatrick, 1983, Justus and Sarkar, 2002 and

Pressey, 2002). Since then, there has been increasing emphasis on meeting quantitative conservation targets within a network of complementary conservation areas.

Systematic conservation planning is an evolving discipline at the interface of biological, mathematical and social sciences (Margules and Pressey, 2000 and Possingham et al., 2000). Systematic conservation planning techniques offer significant improvements over ad hoc and scoring approaches, as they are goal-directed, transparent, defensible, flexible and aim to efficiently meet quantitative targets within a network of representative and complementary conservation areas. Systematic conservation planning has informed conservation in both terrestrial (Pressey, 1998 and Cowling and Pressey, 2003) and marine realms (Ferdana, 2002 and Great Barrier Reef Marine Park Authority, 2003).

Systematic conservation planning requires information on the distribution of biodiversity within the planning region; however, such information is often lacking. Therefore, the features used for conservation planning act as surrogates for total biodiversity (Ferrier et al., 2000). These surrogates have included particular groups of species (Richardson and Funk, 1999 and Clark and Slusher, 2000), vegetation communities (Woinarski et al., 1996) and land classifications (Kirkpatrick and Brown, 1994 and Pressey et al., 2000).

When species are used as a surrogate for biodiversity, often the information on the distribution of species is incomplete, biased spatially (e.g. towards areas that are easily accessible) or biased toward icon species (Podger et al., 1990, Austin, 1998, Keller and Scallan, 1999, Polasky et al., 2000 and Funk and Richardson, 2002). Additionally, systematic surveys for more than a few taxa on a regional scale are uncommon (Haila and Margules, 1996). Despite this, the selection of conservation areas cannot be delayed pending acquisition of complete survey data; options for conservation may be dramatically reduced in the interim. Instead, predicted species distribution data can be used (Margules and Nicholls, 1987, Margules and Stein, 1989, Scott et al., 1993, Richardson and Funk, 1999, Clark and Slusher, 2000 and Jennings, 2000). Predicted species distributions are based on the modelled relationships between species survey data and mapped environmental information (Guisan and Zimmermann, 2000). We consider approaches that predict the probability of occurrence of species in unsurveyed areas (e.g. predicted species distribution data generated using logistic regression, McCullagh and Nelder, 1989). Such predictions can be generated for most species. However, difficulties might arise when modeling the potential distribution of species with ranges that are the product of historical environmental change and complex disturbance regimes (Kirkpatrick and Brown, 1991).

Predicted species distribution data can exhibit considerable uncertainty (Elith et al., 2002). For example:

- (a) Species survey data may be inaccurately recorded, resulting in false-positive and false-negative errors;
- (b) Errors in environmental data may arise during interpretation and processing of satellite images or aerial photographs;
- (c) Uncertainty may arise from the simplifying assumptions used in the modelling process.

For example, an environmental variable might be used in a species distribution model but this variable might be correlated with another unmeasured variable that actually drives the distribution of the species.

Various approaches are available for generating a dataset from predicted species distribution data. The most common approach involves converting the probabilities of occurrence to presence/absence data using a threshold (Li et al., 1997, Manel et al., 1999, Manel et al., 2001, Fleishman et al., 2001 and Fleishman et al., 2003). An alternative approach is to use the probabilities of occurrence directly. This avoids losing information by artificially converting probabilities of occurrence into presence/absence data (Margules and Nicholls, 1987, Polasky et al., 2000 and ReVelle et al., 2002).

The different approaches to using predicted species distribution data for conservation planning can be thought of as different ways to formulate the conservation planning problem. Generally, however, only one problem formulation is assessed and the objectives and

constraints within it are regarded as the only ones applicable. Little attention has been paid to the sensitivity of conservation planning outcomes to different problem formulations. Comparatively more attention has been given to the performance of reserve selection algorithms, including their ability to generate optimal solutions and their speed of operation (for example, Pressey et al., 1996, Pressey et al., 1997, Csuti et al., 1997 and Kelley et al., 2002). The choice of approach to using predicted species distribution data might influence which areas are identified as being important for conservation and their expected representation of species. Here, the focus is on the different ways to formulate the conservation planning problem, as opposed to comparing reserve selection algorithms. Here, we assess the sensitivity of conservation planning outcomes to the different approaches to using predicted species distribution data. The similarity of reserve network solutions identified using the different datasets is evaluated. Additionally, the expected representation of species in the different reserve network solutions is used to indicate which approach may provide the best outcomes for nature conservation.

2. Methods

2.1. Study area and species survey data

The study region in Victoria, Australia, comprises portions of the Goldfields, Wimmera, Murray Mallee and Central Victorian Upland bioregions of the North Central Catchment Management Authority area: a total area of almost two million hectares. The region contains 5149 remnants of native vegetation (predominately belonging to the Box–Ironbark forest type), which cover 306,107 ha. The survey data held for four plant species: *Eucalyptus tricarpa* (a common overstorey species), *Pultenaea largiflorens* (a common understorey species), *Hibbertia exutacies* (a common ground layer species) and *Acacia ausfeldii* (a rare and threatened understorey species) was obtained (Department of Sustainability and Environment, 2002). Survey data for these four species were not available for the entire study region. Consequently, predictions of species occurrence throughout the study region were generated by extrapolating the survey data to unsurveyed areas using the modelled relationship between the known occurrence of each species and environmental variables. These relationships were derived from a logistic regression model (Agresti, 1996). The probability of occurrence of each species in unsurveyed areas were generated for each one hectare grid cell (Wilson, 2003). Grid cells in which the species were known to occur were allocated a probability of one. The one hectare grid cells are of a size that reasonably represents the variation in the landscape and vegetation, and the accuracy of the species survey data.

2.2. Conservation planning objectives and species conservation targets

The aim of systematic conservation planning, formulated in a decision theory framework, is either to minimise or to maximise the value of an objective function, subject to constraints that control the choice of areas selected to be part of the reserve network. These formulations, referred to as the minimum-set and the maximal coverage problems, differ according to the perceived constraints and objectives.

The objective of the minimum-set problem is to minimise resources expended, such as the area reserved, subject to the constraint that all features meet their conservation targets (Kirkpatrick, 1983, Margules and Nicholls, 1987, Margules et al., 1988, Possingham et al., 1993 and ReVelle et al., 2002). The objective of the maximal coverage problem is to maximise conservation of features subject to the constraint that the resources expended does not exceed a specified amount (Satersdal et al., 1993, Church et al., 1996, Haight et al., 2000, Polasky et al., 2000, Polasky et al., 2001, Arthur et al., 2002, Camm et al., 2002 and ReVelle et al., 2002). Since there are limits on the amount of land that can be set aside for nature conservation, it is prudent to select a set of conservation areas that achieves comprehensive representation or meets targets at minimum cost. We adopted the minimum-set framework; the objective was to minimise the area reserved whilst meeting areal conservation targets allocated to each of the four plant species. We recognise that a reserve network based on the requirements of only four species is unlikely to meet the requirements of all species occurring in the region. Rather, the data for these four species provide a simple data matrix to evaluate

the different approaches to using species distribution data and to assess the sensitivity of conservation planning to these approaches.

We employed the areal extent of the estimated original distribution of the Ecological Vegetation Classes in which the common plant species occur (Muir et al., 1995) as a surrogate measure of their historical distribution. A conservation target of 15% of the pre-European extent of each Ecological Vegetation Classes in which the species presently occur was allocated to each of the common plant species (Table 1). This is an arbitrary target but follows the target currently sought for Australia's forest types (JANIS, 1997). The procedure outlined by Burgman et al. (2001) was employed to set the conservation target for the rare plant species, *A. ausfeldii* (Wilson, 2003). This procedure uses expert judgement of life history parameters and the determination of a target area that results in a specified quasi-extinction risk (0.1% chance of falling below 50 individuals at least once in the next 50 years). Calculation of the target area took into account deterministic and stochastic threats, such as eucalyptus oil harvesting and land clearing that may reduce available habitat for this species within the study region (Wilson, 2003).

Table 1. Conservation targets for the four plant species

Species	Target (ha)
<i>Acacia ausfeldii</i>	1596
<i>Eucalyptus tricarpa</i>	51615
<i>Hibbertia exutacies</i>	38842
<i>Pultenaea largiflorens</i>	8862

2.3. Threshold approaches – converting probabilistic data to presence/absence data

First, we assumed that a species was absent from an area if the probability of occurrence was <0.5 and present in an area if the probability of occurrence was equal to or greater than 0.5. This threshold was chosen prior to generating the probabilities of occurrence (Li et al., 1997, Manel et al., 1999, Manel et al., 2001, Fleishman et al., 2001 and Fleishman et al., 2003) and is referred to hereafter as Threshold Method 1.

The main disadvantage of using a threshold chosen a priori is that unequal group sizes in the original survey data can influence the probabilities generated from a logistic regression model, with probabilities biased toward the larger group (Hosmer et al., 1988). This bias is sensible as the probabilities truly reflect frequencies of presences and absences in the survey data. For example, the survey data used to generate the species distribution models had unequal group sizes (that is, more absence observations than presence observations). Consequently, the predicted probabilities are biased toward zero and their mean value is low (Table 2). A threshold chosen a priori might be much greater than the probability of occurrence values predicted from the models. For example, a threshold of 0.5 may mean that nearly all predictions for rare species are realised as absent. Compared with thresholds selected a posteriori, a priori thresholds are unable to be adapted to specific datasets and user needs.

Table 2. Group sizes of the original survey data (the number of presences and absences) and the mean value of the predicted probabilities. The threshold probabilities for the four species when using Threshold Method 2 are also provided

Species	# of presences	# of absences	Mean value of predictions \pm SD	Threshold probability using Threshold Method 2
<i>Acacia ausfeldii</i>	61	662	0.09 \pm 0.12	0.60
<i>Eucalyptus tricarpa</i>	185	564	0.30 \pm 0.23	0.62
<i>Hibbertia exutacies</i>	239	531	0.36 \pm 0.26	0.77
<i>Pultenaea largiflorens</i>	202	573	0.32 \pm 0.21	0.72

We chose the second threshold a posteriori by trading off the sensitivity and specificity of the generated datasets (Zweig and Campbell, 1993, Fielding and Bell, 1997 and Manel et al., 2001). Sensitivity is the proportion of observations correctly predicted as a presence. Specificity is the proportion of observations correctly predicted as an absence. This approach is referred to hereafter as Threshold Method 2.

Using Threshold Method 2, a threshold was chosen that ensured the converted presence/absence data had a desired specificity level. In selecting areas for inclusion in a reserve network, a conservation planner might be cautious, and wish to ensure that only highly suitable areas for a species are selected. The threshold probability would therefore need to be stringent, the sensitivity of the generated dataset would be low and the specificity would be high. The sensitivity and specificity of the datasets derived using different thresholds were evaluated by converting the predicted probabilities of occurrence to presence/absence data and comparing the generated dataset with the original survey data. We selected a threshold probability so that the sensitivity of the generated dataset was 10% and the specificity was 90%. The threshold probabilities chosen using this approach are provided in Table 2.

We chose the final threshold a posteriori using information about the relative costs of false-positive and false-negative errors, whilst taking into account the prevalence of the species (the proportion of presence observations in the survey data; Zweig and Campbell, 1993). This approach is referred to hereafter as Threshold Method 3.

For Threshold Method 3, the relative costs of false-positive and false-negative errors were determined. In conservation terms, the cost of a false-positive error is that associated with selecting an area to be conserved because a species was predicted to be present in the area, but in fact was not. The cost of a false-negative error is that associated with excluding an area from the reserve network on the basis that the species was predicted to be absent from the area, when it was actually present.

The relative costs of false-positive and false-negative errors can extend from zero to infinity. For example, a cost ratio of 0.2 occurs when the cost of a false negative error is greater than the cost of a false positive error. A cost ratio of one occurs when the cost of a false negative error and the cost of a false positive error are equal. A cost ratio of five occurs when the cost of a false positive error is greater than the cost of a false negative error. The latter would be the preferred cost ratio of a cautious conservation planner who wishes to ensure that only highly suitable habitats for a species are selected for the reserve network. To choose a threshold that yields false-positive and false-negative errors in the optimal proportion, one must also incorporate information on the prevalence of the species, or the proportion of presence observations in the underlying survey data (Table 2 and Table 3). Low

prevalence generally calls for lower thresholds, as the probabilities of occurrence generated for a rare species will be biased towards zero (Swets, 1988, Zweig and Campbell, 1993 and Manel et al., 2001).

Table 3. Species prevalence measures, slope and associated threshold probabilities chosen using Threshold Method 3

Species	$(1 - p)/p$	Slope	Threshold probability using Threshold Method 3
<i>Acacia ausfeldii</i>	11.5	57.5	0.62
<i>Eucalyptus tricarpa</i>	3	15	0.64
<i>Hibbertia exutacies</i>	2.2	11	0.69
<i>Pultenaea largiflorens</i>	2.9	14.5	0.71

Combining the information on prevalence (p) and the relative costs of false-positive and false-negative errors allows the calculation of a slope (Table 3):

$$(\text{false positive cost}/\text{false negative cost}) \times ((1 - p)/p)$$

The appropriate threshold probability is found by moving a line, with the calculated slope, from the top left of a receiver operating characteristic (ROC) curve (Zweig and Campbell, 1993) to the point where the line and the curve first touch. The ROC curve relates relative proportions of correctly and incorrectly classified predictions (sensitivity and false-positive pairs) over a wide and continuous range of threshold probability values. The sensitivity is plotted on the y -axis and the false-positive rate (or the specificity rate) is plotted on the x -axis. The sensitivity and specificity values at the point on the ROC curve where the line first touches is used to determine the appropriate threshold probability, given the relative cost of false-positive and false-negative errors and the prevalence of the species (Fig. 1). In Fig. 1, the line with slope 0.75 corresponds to a cost ratio of 1:4 and a prevalence of 0.2. The line first touches the curve at the point with a sensitivity of 0.85 and specificity of 0.4, which corresponds to a threshold probability of 0.2. The threshold probabilities chosen for the four plant species using this approach are provided in Table 3. The threshold chosen for *A. ausfeldii* is slightly lower than that for the other species, due to its comparatively low prevalence.

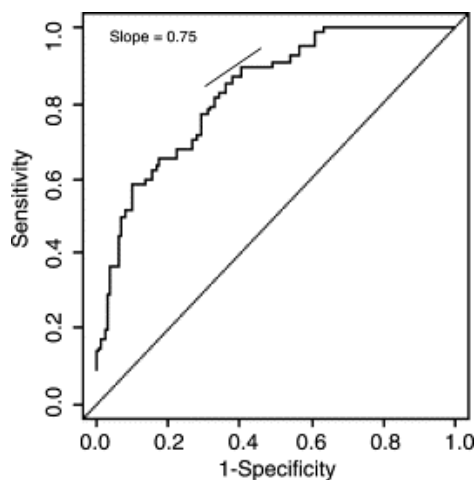


Fig. 1. The use of a ROC curve and a line with a specified slope to determine a threshold probability for use on predicted species distribution data. The slope represents the relative cost of false-positive and false-negative errors and takes into account the prevalence of the species (the proportion of presence

observations in the survey data). The point where the line with this slope first touches the curve is used to identify the appropriate threshold probability.

We used these three methods of threshold selection to convert the predicted probabilities of occurrence of the four plant species to presence/absence data. We then scaled the data from the resolution of grid cells to that of the planning units – the extant remnants of native vegetation. To achieve this, the presence/absence data were summed across the grid cells (which represent one hectare) to obtain the predicted number of occurrences of each species in each planning unit. We take this to represent the predicted area of occurrence of each species (in hectares) in each planning unit.

2.4. Probabilistic approaches – using the probabilities directly

An alternative approach to using predicted species distribution data in conservation planning is to use the probabilities of occurrence directly and not convert them to presence/absence data. We evaluated two approaches to using probabilities of occurrence directly.

First, we used the minimum expected coverage approach (referred to hereafter as Probabilistic Method 1). The expected area of occurrence (in hectares) of each species in each planning unit was calculated and the objective was to minimise the area reserved, subject to the constraint that the expected area of occurrence was greater than the pre-specified targets (r_i):

$$\sum_{j \in J} \sum_{k \in K_j} p_{ik} x_j \geq r_i \quad (2)$$

for each species (i), where p_{ik} is the probability that species i exists in grid cell k . The set of grid cells k in planning unit j is denoted by K_j and J represents the set of planning units j selected for the reserve network. The control variable x_j is equal to zero or one for $j = 1, \dots, J$ such that x_j equals one if planning unit j is in the reserve network and x_j equals zero if planning unit j is not in the reserve network. As far as we are aware, this is the first time Probabilistic Method 1 has been used for conservation planning.

Second, we modified the minimum expected coverage approach to generate the minimum expected threshold coverage approach (referred to hereafter as Probabilistic Method 2). The thresholds generated using Threshold Method 2 (Table 2), where sensitivity and specificity were traded, were used to find the grid cells with probabilities of occurrence greater than the specified threshold allocated to each species. The expected area of occurrence of each species in each planning unit was calculated, but using only the probabilities of occurrence greater than the specified thresholds.

In summary, three threshold approaches and two probabilistic approaches were used to generate a dataset from the predicted species distribution data (Table 4).

Table 4. Summary of the approaches used to generate a dataset from the predicted species distribution data. These approaches include those that use a threshold to convert the probabilities of occurrence to presence/absence data and those that use the probabilities of occurrence directly

Method of generating a dataset from predicted distribution data	Description
<i>Threshold approaches</i>	
Threshold Method 1	1. Convert all probabilities greater than 0.5 to presence/absence data
	2. Sum the converted presence/absence data
Threshold Method 2	1. Find the probability that gives converted presence/absence data that has the desired specificity
	2. Convert all probabilities greater than this threshold to presence/absence data
	3. Sum the converted presence/absence data
Threshold Method 3	1. Find the probability that gives converted presence/absence data that has the desired relative false-positive and false-negative error rates
	2. Convert all probabilities greater than this threshold to presence/absence data
	3. Sum the converted presence/absence data
<i>Probabilistic approaches</i>	
Probabilistic Method 1	1. Sum all probabilities regardless of their value
Probabilistic Method 2	1. Find the probability that gives converted presence/absence data that has the desired specificity
	2. Sum all probabilities greater than this value (without converting them to presence/absence data)

An alternative approach to using probabilities directly is the minimum threshold coverage approach. This approach has been used for selecting conservation areas (Margules and Nicholls, 1987, Williams and Araújo, 2000, Williams and Araújo, 2002, Araújo et al., 2002, Arthur et al., 2002 and ReVelle et al., 2002). The objective is to minimise the area reserved subject to the constraint that the overall probability of occurrence of each feature in the reserve network is greater than a pre-specified target:

$$1 - \prod_{j \in J} (1 - p_{ij}x_j) \geq \alpha \quad (3)$$

for all species (i), where p_{ij} is the probability of occurrence of species i in planning unit j . The objective function employed for Threshold Methods 1–3 and Probabilistic Methods 1 and 2

was to find the minimum area that meets an area target for each species. The objective function when using the minimum threshold coverage approach is to find the minimum area that represents each species with a specified probability (α).

Under this alternative approach, two simplifying assumptions are required to allow the probability of a species occurring in a reserve network to be expressed as the product of the probabilities of non-occurrence over all component planning units. The first assumption is that the probability of a species occurring in a planning unit is independent of the probability of the species occurring in other planning units. The second assumption is that the probability of a species occurring in a planning unit is independent of the probability of every other species occurring in that planning unit. These assumptions are often false because the distribution of individuals is usually strongly spatially correlated and ecological relationships often exist between many species. The objective function for the other approaches does not make these assumptions. In addition, the other approaches avoid the difficulty of setting a target based on a probability of occurrence. Consequently, we chose not to consider the minimum threshold coverage approach. Finally, optimising for the altered objective function is redundant for this dataset. This is not meant as a criticism of the approach. Rather, it illustrates that finding the minimum area that meets a specified probability of occurrence, can sometimes be achieved using a simple, exhaustive search. Planning units can simply be added to the reserve network, starting with the planning units with the highest probability of occurrence, until the probability of occurrence threshold is met. For example, when guaranteeing a stringent threshold probability (such as 0.95 for all species), the minimum area required is approximately 9 hectares based on the Box–Ironbark dataset.

2.5. Conservation planning procedure

We used the simulated annealing algorithm in the Marxan reserve selection software package (Ball and Possingham, 2000) to find the minimum area required to meet the conservation targets (Liu and Wang, 1994, Murray and Church, 1996, Csuti et al., 1997, Possingham et al., 2000 and Arthur et al., 2002). For the simulated annealing procedure, an adaptive schedule followed by iterative improvement was employed and was configured so that the number of simulated annealing iterations was 10 million and the number of ‘temperature’ decreases was 10,000. We ran Marxan 100 times to produce 100 solutions. The irreplaceability of each planning unit was calculated as the proportion of the reserve network solutions that included each planning unit. Planning units included in all of the reserve network solutions were considered irreplaceable, as these planning units will be required to meet the conservation targets for the species they contain. Existing conservation areas were taken into account by making x_j equal one for each planning unit j that was already reserved. For simplicity, we did not seek to minimise the boundary length of the reserve networks. Generally, stepwise heuristics have been used to handle probabilistic data in conservation planning (Margules and Nicholls, 1987, Polasky et al., 2000, Williams and Araújo, 2000 and Araújo et al., 2002). This is the first time we are aware that simulated annealing has been used to address this problem.

2.6. Assessing the sensitivity of the conservation planning outcomes to the different approaches to using predicted species distribution data

We investigated the sensitivity of the conservation planning outcomes to the different approaches for generating datasets from predicted species distribution data by determining:

1. The predicted area of occupancy of each species within the study region.
2. The efficiency of each reserve network solution.
3. The area of irreplaceable planning units in the reserve network solutions and the similarity of selected planning units.
4. The compactness of each reserve network solution.
5. The expected representation of each species in each reserve network solution.
6. A measure of the performance of different reserve networks is how efficiently targets are met. Efficiency (which varies from zero to 100) can be calculated via:

$$1(X/T) \times 100, \quad (4)$$

where X is the area of planning units needed to achieve the targets and T is the total area of planning units available for conservation (Pressey and Nicholls, 1989). The minimum-set of planning units needed to meet the conservation targets is the most efficient solution.

Therefore, highly efficient solutions have minimal over-achievement of targets.

The Kappa statistic (Cohen, 1960) provides a measure of the similarity of planning units selected for the different reserve networks. This statistic assesses the extent that reserve networks overlap after removing overlap due to chance. The Kappa statistic is:

$$\frac{(P_o - P_c)}{1 - P_c}, \quad (5)$$

where P_o is the proportion of planning units that overlap and P_c is the proportion of planning units that overlap due to chance. The Kappa statistic ranges from a minimum of negative one to a maximum of one, with a value of one indicating perfect overlap, zero indicating expected overlap due to chance and negative one indicating no overlap (Czaplewski, 1994).

We used a two by two classification table to calculate the Kappa statistic (Table 5); this cross-tabulated the area of planning units selected for each reserve network solution.

Table 5. A two by two classification table, where Reserve Network 1 might represent the network selected using Threshold Method 1 and Reserve Network 2 might represent the reserve network selected using Probabilistic Method 2 ($A + B + C + D = N$)

		Reserve network 1	
		Included	Not included
Reserve Network 2	Included	A	B
	Not included	C	D

The Kappa statistic was calculated (after Fielding and Bell, 1997) using the formula:

$$\frac{[(A + D) - (((A + C) \times (A + B) + (B + D) \times (C + D))/N)]}{[N - (((A + C) \times (A + B) + (B + D) \times (C + D))/N)]}. \quad (6)$$

We calculated the Kappa statistic for the subset of irreplaceable planning units selected to be part of each reserve network and for the full set of planning units selected to be part of each reserve network.

A measure of compactness or degree of spatial clustering of a reserve network is provided by the ratio of the boundary length of the reserve network and the circumference of a circle of the same area (Possingham et al., 2000):

$$\frac{\text{Boundary length}}{2\sqrt{\pi \times \text{Area}}} \quad (7)$$

A circle is the most compact shape possible, so this is the ratio of the boundary length to the theoretical minimum and is a dimensionless measure. Values approaching one resemble the shape of a circle and are highly compact.

We calculated the probability of at least one occurrence of each species in each planning unit and summed this across all planning units selected for each reserve network:

$$\sum_{j \in J} x_j \left[1 - \prod_{k \in K_j} (1 - p_{jk}) \right] \quad (8)$$

This determined the number of times each species is expected to occur in each reserve network and therefore provided the expected representation of each species in the reserve network solutions.

3. Results

3.1. The predicted area of occupancy of each species

First, a comparison is made of the predicted area of occupancy of each species. Using Threshold Methods 2 and 3, the area where the species is predicted to occur is limited to that deemed likely to contain the species, according to a chosen threshold. These thresholds are higher than the 0.5 threshold of Threshold Method 1 and hence the predicted area of occupancy of each species is less. Using Probabilistic Method 1, the probabilities were summed across all grid cells, which generally resulted in a greater predicted area of occupancy than with the other methods. Probabilistic Method 2, where only those probabilities greater than a specified threshold were summed, resulted in the least predicted area of occupancy of each species (Fig. 2). Given the small area predicted to be occupied by the species using Threshold Methods 2 and 3 and Probabilistic Method 2, many of the targets were unable to be met (Fig. 2).

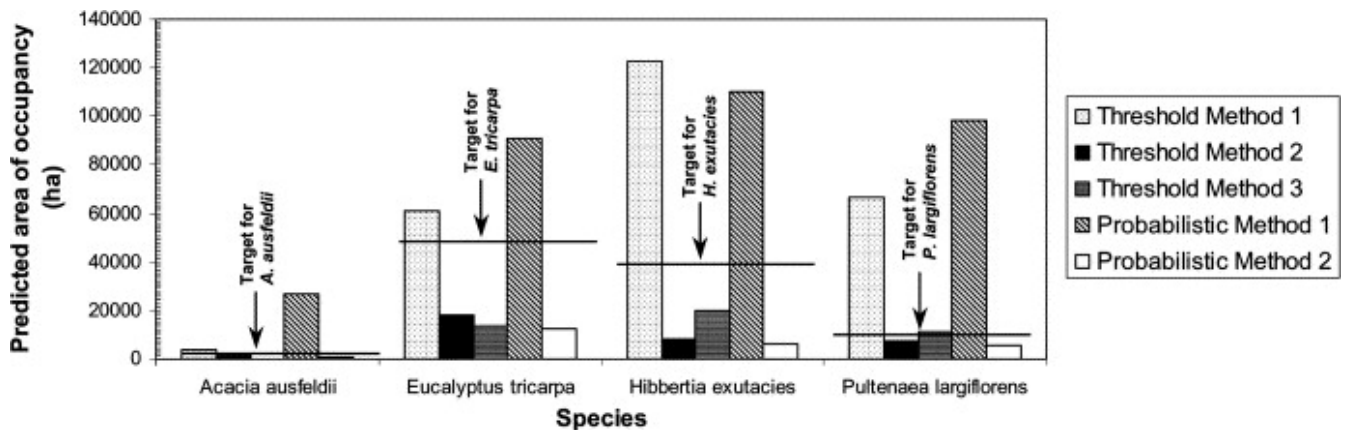


Fig. 2. The predicted area of occupancy of each species, based on the different methods for generating a dataset from predicted species distribution data. The conservation target for each species is highlighted. For many species, the original targets cannot be met, due to insufficient area being predicted as occupied.

Using Threshold Method 1, the predicted area of occupancy of *A. ausfeldii*, *E. tricarpa* and *P. largiflorens* was less than with Probabilistic Method 1. The modal probabilities of occurrence of these three species were about 0.4. When using Threshold Method 1, probabilities <0.5 were converted to zero. Therefore, the predicted area of occupancy of these three species was deflated and less than with Probabilistic Method 1 where the probabilities were summed. The modal probability of occurrence for *H. exutacies* was about 0.6. Therefore, since probabilities equal to or greater than 0.5 were converted to one when using Threshold Method 1, the predicted area of occupancy of this species was inflated and greater than with Probabilistic Method 1.

3.2. Efficiency of the reserve network solutions

The dataset generated using Threshold Method 3 resulted in the reserve network with the greatest area. Consequently, the reserve network generated on the basis of this dataset was the least efficient (Table 6). Despite its large area, the reserve network was unable to meet the original conservation targets for any of the species, due to the small area predicted to be occupied by each species.

Table 6. Number and mean size (in hectares) of the selected planning units and the area of the reserve network solutions. The efficiency of the reserve network solutions and the area of irreplaceable planning units are also detailed

Method of generating a dataset from predicted distribution data	# planning units	Mean size of planning units (ha)	Area of reserve network (ha)	Efficiency (area selected)	Area of irreplaceable planning units
Threshold Method 1	469	243	114,166	62.7%	81849 ($n = 40$)
Threshold Method 2	693	201	139,381	54.5%	139,378 ($n = 682$)
Threshold Method 3	678	220	148,820	51.4%	147,729 ($n = 640$)
Probabilistic Method 1	1197	97	116,354	62.0%	19098 ($n = 42$)
Probabilistic Method 2	668	216	144,412	52.8%	144,409 ($n = 658$)

The dataset generated using Threshold Method 1 resulted in the most efficient reserve network solution (Table 6). Probabilistic Method 1 met the conservation targets within a reserve network of comparable size to Threshold Method 1, but required almost twice the number of planning units (Table 6). Basing the analyses on datasets generated using Threshold Method 1 and Probabilistic Method 1 also allowed the original targets to be met. The analyses were repeated using targets that were attainable. The patterns exhibited using the modified targets were identical to the results obtained using the original targets. The mean size of the planning units within the study region is 62 ha; planning units range in size from <1 to 14,711 ha. The reserve network solutions generated using the threshold approaches and Probabilistic Method 2 were biased towards larger planning units (Table 6). As the probabilities of occurrence were generally <0.5, smaller planning units were less valued when a threshold was used on the component probabilities. Larger planning units were less affected by the use of a threshold because they were more likely to contain at least some high probabilities of occurrence.

3.3. Similarity of planning units in the reserve network solutions

The reserve network solutions identified using the datasets from Threshold Methods 2 and 3 and Probabilistic Method 2 contained large areas of irreplaceable planning units (Table 6). In comparison, the reserve network solution identified using Probabilistic Method 1 had the largest area of replaceable planning units, followed by Threshold Method 1. The irreplaceable planning units of Probabilistic Method 1 were largely comprised of the existing reserves, which were locked into the reserve network solutions. When the analyses were repeated using targets that were attainable, the patterns exhibited were identical.

The irreplaceable planning units selected using the datasets from Threshold Methods 2 and 3 and Probabilistic Method 2 overlapped substantially (Table 7). Those selected using Threshold Method 1 overlapped to a lesser extent and those selected using Probabilistic Method 1 overlapped the least. When the analyses were repeated using targets that were attainable, the patterns exhibited were identical, except that Threshold Method 1 and Probabilistic Method 1 exhibited perfect overlap. This is because their irreplaceable planning

units comprised only the existing conservation areas, which were locked into the reserve network solutions.

Table 7. The overlap between irreplaceable planning units in the reserve network solutions selected using the different datasets. The Kappa statistics was used to assess the overlap between reserve after removing overlap due to chance

Method of generating a dataset from predicted distribution data	Overlap in irreplaceable area			
	Threshold Method 1	Threshold Method 2	Threshold Method 3	Probabilistic Method 1
Threshold Method 2	0.55			
Threshold Method 3	0.49	0.84		
Probabilistic Method 1	0.31	0.15	0.14	
Probabilistic Method 2	0.52	0.94	0.82	0.14

None of the reserve network solutions identified on the basis of the different datasets comprised the same set of planning units (Table 8). The set of planning units selected using Probabilistic Method 1 did not overlap with those selected based on the other datasets (Table 8), with the exception of the reserve network solution identified using Threshold Method 1. The set of planning units selected using the threshold methods overlap to a moderate extent (Table 8). The set of planning units selected using Probabilistic Method 2 had considerable overlap with the reserve network identified using Threshold Methods 2 and 3 (Table 8). Similar patterns were exhibited when the analyses were repeated using targets that were attainable.

Table 8. The extent of overlap of planning units in the reserve network solutions selected using the different datasets. The Kappa statistics was used to assess the overlap between reserve after removing overlap due to chance

Method of generating a dataset from predicted distribution data	Overlap in reserve network area			
	Threshold Method 1	Threshold Method 2	Threshold Method 3	Probabilistic Method 1
Threshold Method 2	0.61			
Threshold Method 3	0.62	0.85		
Probabilistic Method 1	0.66	0.49	0.45	
Probabilistic Method 2	0.60	0.94	0.83	0.49

3.4. Compactness of the reserve network solutions

The reserve network solution identified using Threshold Method 1 was the most compact (Table 9). The reserve network solution identified using Probabilistic Method 1 was the least

compact and this is due to the selection of smaller planning units to form the reserve network solutions compared to the other approaches.

Table 9. Compactness of the reserve network solutions identified using the different datasets

Method of generating a dataset from predicted distribution data	Boundary length (km)	Area (ha)	Compactness of the reserve network
Threshold Method 1	2,901,491	114,166	24
Threshold Method 2	3,952,810	139,381	30
Threshold Method 3	4,186,473	148,820	31
Probabilistic Method 1	4,139,755	116,354	34
Probabilistic Method 2	4,057,581	144,412	30

3.5. Expected representation of each species in the reserve network solutions

The probability of at least one occurrence of each species in each planning unit was summed across all planning units selected for the different reserve network solutions. This provided the number of planning units in the reserve network solutions in which each species is expected to occur.

Initially, Probabilistic Method 2 provided the greatest expected representation of each species, closely followed by Threshold Method 2 and Threshold Method 3 (Table 10). However, these methods also resulted in the largest reserve network solutions: any approximation to a minimum-set problem that selects a larger area can include more representations of species simply by including more area. The results were therefore standardised for reserve network area. This made little difference to the results. Subsequent to standardisation, the approaches that involved the use of a threshold (Threshold Methods 2 and 3 and Probabilistic Method 2) had equally great expected species representation (Table 10). Similar patterns were exhibited when the analyses were repeated using targets that were obtainable.

Table 10. Expected species representation in the reserve network solutions identified using the different datasets

Method of generating a dataset from predicted distribution data	Mean expected species representation	Mean expected species representation (standardised for reserve network area)
Threshold Method 1	108.19	0.0008
Threshold Method 2	568.53	0.004
Threshold Method 3	564.51	0.004
Probabilistic Method 1	45.10	0.0003
Probabilistic Method 2	571.49	0.004

4. Discussion

We have shown that conservation planning outcomes are very sensitive to the uncertainty that arises from the different methods to generate a dataset from predicted species distribution data. The reserve network solutions differ in their relative efficiencies, their component planning units, and their expected representation of species. The method of generating a dataset from predicted species distribution data for use in conservation planning will reflect how much risk a planner is willing to tolerate and the amount of efficiency that can be sacrificed. A planner who is risk-averse would seek greater certainty that the target species

occur in the planning units selected for a reserve network (Terborgh and Winter, 1983, Williams, 1998 and Elith and Burgman, 2003).

The approaches used to generate a dataset from predicted species distribution data for use in conservation planning may be grouped into two broad categories: (1) those that convert the probabilities of occurrence to presence/absence data using thresholds identified a priori or a posteriori, and (2) those that use the probability of occurrence data directly. Three thresholds were compared: a threshold of 0.5 selected a priori and two thresholds selected a posteriori; one of these trades the sensitivity and specificity of the datasets, the other trades the cost of false-positive and false-negative errors.

A priori methods of threshold selection are commonly used (Li et al., 1997, Manel et al., 1999, Manel et al., 2001, Fleishman et al., 2001 and Fleishman et al., 2003), despite the fact that they take no account of the range and spread of predictions, nor of dataset characteristics and user requirements. Stringent thresholds were chosen a posteriori to ensure that only potentially suitable habitat was predicted as occupied habitat. Therefore, the a posteriori chosen thresholds were more risk-averse than the threshold chosen a priori. The approaches to using probabilistic data directly, which have not been used previously for conservation planning, employ information on the expected area of occurrence of each species in each planning unit. The second approach differs from the first in that only the probabilities of occurrence above a stringent threshold are used in determining the expected area of occurrence. Therefore, the second approach is more risk-averse.

The various methods of generating a dataset from predicted species distribution data resulted in different predictions of the area occupied by each species. When the probabilities of occurrence are summed, the predicted area of occupancy is greater than when a stringent threshold is used to first convert the probabilities to presence/absence data. This result may vary for different modelling methods and different species; nevertheless, it is likely to be partly a consequence of the datasets being biased towards zero (due to the low prevalence of the species in the survey data). The calculation of the predicted area of occupancy of a species is also dependent on the resolution of the underlying data. One hectare grid cells were used as the base unit for predictions, because this was a reasonable representation of the variation in the landscape and the accuracy of the species survey data. There is, however, a trade-off between smaller grid cells, which might deliver more accurate predictions of the area occupied by a species, but with a likely increase in computation time. Larger grid cells might result in the overestimation of the area occupied by a species by including unsuitable habitat and hence create more uncertainty in the modelled predictions. To help mitigate these problems, the sensitivity of the area of occupancy predictions to the resolution of the data could be assessed and species-specific resolutions for estimating area of occupancy could be employed (Keith, 1998 and Keith et al., 2000).

The use of a stringent threshold on the predicted species distribution data helps to ensure that only highly suitable habitat is counted towards the predicted area of occupancy of each species. When using a low threshold or summing the probabilities without using a threshold, the predicted area of occupancy can also include areas with a very low probability of occurrence and is therefore a more risk-neutral approach.

With only a small area predicted to contain each species, the datasets generated using the stringent thresholds offered less flexibility in designing a reserve network. Consequently, the reserve networks identified on the basis of these datasets covered a large area and were comparatively less efficient. The reserve network based on the summed probabilities had the largest area of replaceable planning units. This reflects the flexibility available to generate this reserve network solution.

While the reserve network solutions identified using the different datasets overlapped, they still differed from each other. Most importantly, there was limited overlap in irreplaceable planning units. Therefore, depending on how predicted species distribution data is used for conservation planning, different sets of planning units will be selected to be part of the reserve network, particularly if irreplaceability is used to determine the planning units that are the immediate priorities for conservation.

The reserve networks identified using the datasets generated with a stringent threshold (the more risk-averse approaches) had the greatest expected species representation when standardised for the area of the reserve network. The expected representation of species in reserve networks is one procedure for evaluating different reserve network solutions, in terms of how well they might conserve species (their anticipated adequacy). An alternative, more data intensive, approach to testing adequacy would be to evaluate the predicted persistence of the species in the reserve network solutions using population viability analysis. In addition to having the greatest expected species representation, the reserve network solutions identified using the more risk-averse approaches also favoured large planning units, which may provide better outcomes for nature conservation due to reduced boundary length and edge effects (Fagan et al., 1999). However, the reserve networks based on these datasets were not the most efficient. Therefore, the method of generating a dataset from predicted species distribution data for use in conservation planning will reflect a trade-off between the risk of inadequately conserving biodiversity and the amount of efficiency that can be sacrificed.

The efficiency of a reserve network has two implications. First, due to competition between incompatible land uses, limited resources for conservation and financial considerations (e.g. costs of acquisition, management and opportunity costs to other land uses) efficiency determines the likelihood of achieving conservation targets. Second, efficiency determines the degree to which reserve network proposals are defensible in the light of competing land uses. It is important to recognise, however, that 'efficiency' simply refers to the area of land required to achieve the conservation targets. The use of a lower threshold to convert probabilities of occurrence to presence/absence data might increase the predicted area of occupancy of a species and provide greater flexibility for the design of a reserve network (and hence greater efficiency). However, the use of a low threshold might also result in areas being falsely predicted as occupied and increase the uncertainty associated with the use of predicted species distribution data for conservation planning.

Predictions of species distribution are essential for conservation planning because, even in well-surveyed parts of the world, sample plots cover only a small proportion of the landscape. Conservation planning has to deal with the whole of the landscape. Used appropriately, predicted species distribution data can help decision-makers understand how best to use limited resources to achieve conservation targets. However, it is important that the uncertainty in predicted distribution data be accounted for in conservation planning given the sensitivity of the conservation planning outcomes to this source of uncertainty. Recently, the probability of a species occurring in an area has been equated with the long-term persistence of a species in that area (Williams and Araújo, 2000 and Araújo et al., 2002). Probabilities of occurrence may not indicate habitat quality for a species and may not be related to species persistence. This is because the probabilities are based on static species locality data that does not take into account population processes. Consequently, in generating these probabilities, it is assumed that the population is at equilibrium (Guisan and Zimmermann, 2000). If the population is increasing (for example, recovering) then there will be many false-negative predictions and the quality of the habitat and persistence of a species may be underestimated. Alternatively, if the population is declining (for example, confined to relict areas) many false-positive predictions may result, in which case the quality of the habitat and persistence of a species is likely to be overestimated. In either circumstance, an inaccurate assessment of habitat quality will result and estimates of persistence are likely to be unreliable.

Even if the population of a species is at equilibrium, it may not occupy, let alone persist, in habitat predicted to have a high probability of occurrence due to behaviour, intraspecific competitive exclusion, or dispersal dynamics (Van Horne, 1983). In the Box–Ironbark region, altered fire regimes will influence the distribution of plants. For example, *A. ausfeldii* is an obligate seeder that regenerates after high soil temperatures, resulting from fire, break the dormancy of its seeds (Brown et al., 2003). Therefore, if the frequency of fire is reduced, *A. ausfeldii* may not occupy habitat in which it was predicted to have high probability of occurrence. Further, meta-population theory tells us that due to colonisation-

extinction dynamics, high quality habitat may be unoccupied at the time a species survey is conducted.

This research has demonstrated that planning outcomes are very sensitive to the different methods for generating a dataset from predicted species distribution data. A natural further step is to direct effort towards producing the most reliable predictions for use in conservation planning, and to find the reserve network that is most robust to the uncertainty in the predictions. Uncertainty is not only related to the use of the probabilities generated from a model of species distribution, but also to the modelling method and the underlying data (Elith et al., 2002). Modelled predictions might be made more reliable, for example, by re-calibrating the model so it predicts more accurately to new cases (Steyerberg et al., 2001). Uncertainty might also be reduced by producing more ecologically realistic models (Austin, 2002). Nevertheless, there will always be uncertainty associated with the use of predicted species distribution data for conservation planning. As a consequence on-ground inspection of sites nominated for conservation will likely be required.

Acknowledgements

This research was supported by grants to the lead author from the University of Melbourne (Australia), the Menzies Centre for Australian Studies (Kings College, London, UK), the Australian Federation of University Women, and the Holsworth Wildlife Research Fund (Australia). We also would like to thank Mark Burgman, Adrian Newton, Daniel Wilson and Bob Pressey for earlier comments on the manuscript.

References

- Agresti, 1996 A. Agresti, An Introduction to Categorical Data Analysis, John Wiley & Sons, New York (1996).
- Araújo et al., 2002 M.B. Araújo, P.H. Williams and R.J. Fuller, Dynamics of extinction and the selection of nature reserves, *Proceedings of the Royal Society of London. Series B Biological Sciences* **269** (2002), pp. 1971–1980.
- Arthur et al., 2002 J.L. Arthur, R.G. Haight, C.A. Montgomery and S. Polasky, Analysis of the threshold and expected coverage approaches to the probabilistic reserve site selection problem, *Environmental Modeling and Assessment* **7** (2002), pp. 81–89.
- Austin, 1998 M.P. Austin, An ecological perspective on biodiversity investigations: examples from Australian eucalypts, *Annals of the Missouri Botanical Garden* **85** (1998), pp. 2–17.
- Austin, 2002 M.P. Austin, Spatial prediction of species distribution: an interface between ecological theory and statistical modeling, *Ecological Modelling* **157** (2002), pp. 101–118.
- Ball and Possingham, 2000 I.R. Ball and H.P. Possingham, Marxan (v 1.8.6): Marine reserve design using spatially explicit annealing, *A manual prepared for the Great Barrier Reef Marine Park Authority* (2000).
- Brown et al., 2003 J. Brown, N.J. Enright and B.P. Miller, Seed production and germination in two rare and three common co-occurring Acacia species from south-east Australia, *Austral Ecology* **28** (2003), pp. 271–280.
- Burgman et al., 2001 M.A. Burgman, H.P. Possingham, A.J.J. Lynch, D.A. Keith, M.A. McCarthy, S.D. Hopper, W.L. Drury, J.A. Passioura and R.J. Devries, A method for setting the size of plant conservation target areas, *Conservation Biology* **15** (2001), pp. 603–616.
- Camm et al., 2002 J. Camm, S. Norman, S. Polasky and A. Solow, Nature reserve site selection to maximize expected species covered, *Operations Research* **50** (2002), pp. 946–955.
- Church et al., 1996 R.L. Church, D.M. Stoms and F.W. Davis, Reserve selection as a maximal covering location problem, *Biological Conservation* **76** (1996), pp. 105–112.
- Clark and Slusher, 2000 F.S. Clark and R.B. Slusher, Using spatial analysis to drive reserve design: a case study of a notional wildlife refuge in Indiana and Illinois (USA), *Landscape Ecology* **15** (2000), pp. 75–84.
- Cohen, 1960 J. Cohen, A coefficient of agreement for nominal scales, *Educational and Psychological Measurement* **20** (1960), pp. 37–46.
- Cowling and Pressey, 2003 R.M. Cowling and R.L. Pressey, Introduction to systematic conservation planning in the Cape Floristic Region, *Biological Conservation* **112** (2003), pp. 1–13.
- Csuti et al., 1997 B. Csuti, S. Polasky, P.H. Williams, R.L. Pressey, J.D. Camm, M. Kershaw, R. Kiester, B. Downs, R. Hamilton, M. Huso and K. Sahr, A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon, *Biological Conservation* **80** (1997), pp. 83–97.
- Czaplewski, 1994 Czaplewski, R.L., 1994. Variance approximations for assessments of classification accuracy. Rocky Mountain Forest and Range Experiment Station, USDA Forest Service, Fort Collins.
- Department of Sustainability and Environment, 2002 Department of Sustainability and Environment, 2002. Flora Information System – May 2002. Viridans Biological Databases, Brighton East, Victoria.
- Elith and Burgman, 2003 J. Elith and M.A. Burgman, Chapter 8: Habitat models for PVA In: C.A. Brigham and M.W. Schwartz, Editors, *Population Viability in Plants*, Springer-Verlag, New York (2003).
- Elith et al., 2002 J. Elith, M.A. Burgman and H.M. Regan, Mapping epistemic uncertainties and vague concepts in predictions of species distribution, *Ecological Modelling* **157** (2002), pp. 313–329.
- Fagan et al., 1999 W.F. Fagan, R.S. Cantrell and C. Cosner, How habitat edges change species interactions, *The American Naturalist* **153** (1999), pp. 165–182.
- Ferdana, 2002 Z., Ferdana, Approaches to integrating a marine GIS into The Nature Conservancy's ecoregional planning process In: J. Breman, Editors, *Marine Geography: GIS for Oceans and Seas*, ESRI, Redlands, WA (2002), pp. 151–158.
- Ferrier et al., 2000 S. Ferrier, R.L. Pressey and T.W. Barrett, A new predictor of the irreplaceability of areas for achieving a conservation goal, its application to real-world planning, and a research agenda for further refinement, *Biological Conservation* **93** (2000), pp. 303–325.
- Fielding and Bell, 1997 A.H. Fielding and J.F. Bell, A review of methods for the assessment of prediction errors in conservation presence/absence models, *Environmental Conservation* **24** (1997), pp. 38–49.

- Fleishman et al., 2001 E. Fleishman, R. Mac Nally, J.P. Fay and D.D. Murphy, Modeling and predicting species occurrence using broad scale environmental variables: an example with butterflies of the Great Basin, *Conservation Biology* **15** (2001), pp. 1674–1685.
- Fleishman et al., 2003 E. Fleishman, R. Mac Nally and J.P. Fay, Validation tests of predictive models of butterfly occurrence based on environmental variables, *Conservation Biology* **17** (2003), pp. 806–817.
- Funk and Richardson, 2002 V.A. Funk and K.S. Richardson, Systematic data in biodiversity studies: use it or lose it, *Systematic Biology* **51** (2002), pp. 303–316.
- Great Barrier Reef Marine Park Authority, 2003 Great Barrier Reef Marine Park Authority, 2003. Representative Areas in the Marine Park. Available from: http://www.gbrmpa.gov.au/corp_site/key_issues/conversation/rep_areas/?rep_areasoverview.html.
- Groombridge and Jenkins, 2002 B. Groombridge and M.D. Jenkins, Global biodiversity: responding to the change In: B. Groombridge and M.D. Jenkins, Editors, *World Atlas of Biodiversity: Earth's Living Resources in the 21st Century*, University of California Press, Berkeley, CA (2002), pp. 195–223.
- Guisan and Zimmermann, 2000 A. Guisan and N.E. Zimmermann, Predictive habitat distribution models in ecology, *Ecological Modelling* **135** (2000), pp. 147–186.
- Haight et al., 2000 R.G. Haight, C.S. ReVelle and S.A. Snyder, An integer optimization approach to the probabilistic reserve site selection problem, *Operations Research* **48** (2000), pp. 697–708.
- Haila and Margules, 1996 Y. Haila and C.R. Margules, Survey research in conservation biology, *Ecography* **19** (1996), pp. 323–331. Abstract-GEOBASE
- Henderson, 1992 N. Henderson, Wilderness and the nature conservation ideal: Britain, Canada and the United States contrasted, *Ambio* **21** (1992), pp. 394–399.
- Hosmer et al., 1988 D.W. Hosmer, S. Lemeshow and J. Klar, Goodness-of-fit testing for the logistic regression model when the estimated probabilities are small, *Biometrical Journal* **30** (1988), pp. 911–924.
- JANIS, 1997 JANIS, 1997. Nationally agreed criteria for the establishment of a comprehensive, adequate and representative reserve system for forests in Australia. Joint ANZECC/MCFFA National Forest Policy Statement Implementation Subcommittee. National forest conservation reserves: Commonwealth proposed criteria. Commonwealth of Australia, Canberra.
- Jennings, 2000 M.D. Jennings, Gap analysis: concepts, methods, and recent results, *Landscape Ecology* **15** (2000), pp. 5–20.
- Justus and Sarkar, 2002 J. Justus and S. Sarkar, The principle of complementarity in the design of reserve networks to conserve biodiversity: a preliminary history, *Journal of Biosciences* **27** (2002), pp. 421–435.
- Keith, 1998 D.A. Keith, An Evaluation and Modification of World Conservation Union Red List Criteria for Classification of Extinction Risk in Vascular Plants, *Conservation Biology* **12** (1998), pp. 1076–1090.
- Keith et al., 2000 D.A. Keith, T.D. Auld, M.K.J. Ooi and B.D.E. Mackenzie, Sensitivity analyses of decision rules in World Conservation Union (IUCN) Red List criteria using Australian plants, *Biological Conservation* **94** (2000), pp. 311–319.
- Keller and Scallan, 1999 C.M.E. Keller and J.T. Scallan, Potential roadside biases due to habitat changes along breeding bird survey routes, *Condor* **101** (1999), pp. 50–57.
- Kelley et al., 2002 C. Kelley, J. Garson, A. Aggarwal and S. Sarkar, Place prioritization for biodiversity reserve network design: a comparison of the SITES and ResNet software packages for coverage and efficiency, *Diversity and Distributions* **8** (2002), pp. 297–306.
- Kirkpatrick, 1983 J.B. Kirkpatrick, An iterative method for establishing priorities for the selection of nature reserves: an example from Tasmania, *Biological Conservation* **25** (1983), pp. 127–134.
- Kirkpatrick and Brown, 1991 J.B. Kirkpatrick and M.J. Brown, Planning for species conservation In: D.A. Saunders, G.W. Arnold, A.A. Burbidge and A.J.M. Hopkins, Editors, *Nature Conservation: The Role of Remnants of Native Vegetation*, Surrey Beatty and Sons Pty Ltd in association with CSIRO and CALM, Sydney (1991), pp. 83–89.
- Kirkpatrick and Brown, 1994 J.B. Kirkpatrick and M.J. Brown, A comparison of direct and environmental domain approaches to planning reservation of forest higher plant communities and species in Tasmania, *Conservation Biology* **8** (1994), pp. 217–224.
- Li et al., 1997 W. Li, Z. Wang, Z. Ma and H. Tang, A regression model for the spatial distribution of red-crown crane in Yancheng Biosphere Reserve, China, *Ecological Modelling* **103** (1997), pp. 115–121.
- Liu and Wang, 1994 C.M. Liu and A.H. Wang, Solving location-allocation problems with rectilinear distances by simulated annealing, *Journal of the Operational Research Society* **45** (1994), pp. 1304–1315.
- Manel et al., 1999 S. Manel, J.M. Dias, S.T. Buckton and S.J. Ormerod, Alternative methods for predicting species distribution: an illustration with Himalayan river birds, *Journal of Applied Ecology* **36** (1999), pp. 734–747.
- Manel et al., 2001 S. Manel, H.C. Williams and S.J. Ormerod, Evaluating presence-absence models in ecology: the need to account for prevalence, *Journal of Applied Ecology* **38** (2001), pp. 921–931.
- Margules and Nicholls, 1987 C.R. Margules and A.O. Nicholls, Assessing the conservation value of remnant habitat islands: Mallee patches on the western Eyre Peninsula, South Australia In: D.A. Saunders, G.W. Arnold, A.A. Burbidge and A.J.M. Hopkins, Editors, *Nature Conservation: The Role of Remnants of Native Vegetation*, Surrey Beatty and Sons Pty Ltd in association with CSIRO and CALM, Sydney (1987), pp. 89–102.
- Margules and Pressey, 2000 C.R. Margules and R.L. Pressey, Systematic conservation planning, *Nature* **405** (2000), pp. 243–253.
- Margules and Stein, 1989 C.R. Margules and J.L. Stein, Patterns in the distribution of species and the selection of nature reserves: An example from *Eucalyptus* forest in south eastern New South Wales, *Biological Conservation* **50** (1989), pp. 219–238.
- Margules and Usher, 1981 C.R. Margules and M.B. Usher, Criteria used in assessing wildlife conservation potential: a review, *Biological Conservation* **21** (1981), pp. 79–109.
- Margules et al., 1988 C.R. Margules, A.O. Nicholls and R.L. Pressey, Selecting Networks of Reserves to Maximise Biological Diversity, *Biological Conservation* **43** (1988), pp. 63–76.
- Mark, 1985 A.F. Mark, The botanical component of conservation in New Zealand, *New Zealand Journal of Botany* **23** (1985), pp. 789–810.
- McCullagh and Nelder, 1989 P. McCullagh and J.A. Nelder, Generalized Linear Models, Chapman and Hall, London (1989).
- McNeely, 1994 J.A. McNeely, Protected areas for the 21st Century: working to provide benefits for society, *Biodiversity and Conservation* **3** (1994), pp. 3–20.
- Muir et al., 1995 A.M. Muir, S.A. Edwards and M.J. Dickins, Description and Conservation Status of the Vegetation of the Box-Ironbark Ecosystem in Victoria, Department of Conservation and Natural Resources, East Melbourne, Victoria (1995).
- Murray and Church, 1996 A.T. Murray and R.L. Church, Applying simulated annealing to location-planning models, *Journal of Heuristics* **2** (1996), pp. 31–53.

- Podger et al., 1990 F.D. Podger, D.C. Mummery, C.R. Palzer and M.J. Brown, Bioclimatic analysis of the distribution of damage to native plants in Tasmania by *Phytophthora cinnamomi*, *Australian Journal of Ecology* **15** (1990), pp. 281–289.
- Polasky et al., 2000 S. Polasky, J.D. Camm, A.R. Slow, B. Csuti, D. White and R. Ding, Choosing reserve networks with incomplete species information, *Biological Conservation* **94** (2000), pp. 1–10.
- Polasky et al., 2001 S. Polasky, J.D. Camm and B. Garber-Yonts, Selecting biological reserves cost effectively: an application to terrestrial vertebrate conservation in Oregon, *Land Economics* **77** (2001), pp. 68–78.
- Possingham et al., 1993 H. Possingham, J. Day, M. Goldfinch and F. Salzborn, The mathematics of designing a network of protected areas for conservation In: D. Sutton, E. Cousins and C. Pearce, Editors, *Decision Sciences: Tools for Today, 12th Australian Operations Research Conference*, ASOR, University of Adelaide, Adelaide (1993), pp. 536–545.
- Possingham et al., 2000 H. Possingham, I. Ball and S. Andelman, Mathematical methods for identifying representative reserve networks In: S. Ferson and M. Burgman, Editors, *Quantitative Methods for Conservation Biology*, Springer-Verlag, New York (2000), pp. 291–305.
- Pressey, 1993 R.L. Pressey, The good and the bad news for conservation planning: procedures for reserve selection are improving but we still don't know enough about birds and other fauna In: C.P. Catterall, P.V. Driscoll, K. Hulsman, D. Muir and A. Taplin, Editors, *Birds and their Habitats: Status and Conservation in Queensland*, Queensland Ornithological Society, St Lucia, Queensland (1993), pp. 146–156.
- Pressey, 1994 R.L. Pressey, Ad hoc reservations: forward or backward steps in developing representative reserve systems?, *Conservation Biology* **8** (1994), pp. 662–668.
- Pressey, 1997 R.L. Pressey, Priority conservation areas: towards an operational definition for regional assessments In: J.J. Pigram and R.C. Sundell, Editors, *National Parks and Protected Areas: Selection, Delimitation, and Management*, Centre for Water Policy Research, Armidale, New South Wales (1997), pp. 337–357.
- Pressey, 1998 R.L. Pressey, Algorithms, politics and timber: an example of the role of science in a public, political negotiation process over new conservation areas in production forests In: R.T. Wills and R.J. Hobbs, Editors, *Ecology for Everyone: Communicating Ecology to Scientists, the Public and the Politicians*, Chipping Norton, Surrey Beatty and Sons, NSW (1998), pp. 73–87.
- Pressey, 2002 R.L. Pressey, The first reserve selection algorithm – a retrospective on Jamie Kirkpatrick's 1983 paper, *Progress in Physical Geography* **26** (2002), pp. 434–441.
- Pressey and Nicholls, 1989 R.L. Pressey and A.O. Nicholls, Efficiency in conservation planning: scoring versus iterative approaches, *Biological Conservation* **50** (1989), pp. 199–218.
- Pressey and Tully, 1994 R.L. Pressey and S.L. Tully, The cost of ad hoc reservation: a case study in western New South Wales, *Australian Journal of Ecology* **19** (1994), pp. 375–384.
- Pressey et al., 1996 R.L. Pressey, H.P. Possingham and C.R. Margules, Optimality in reserve selection algorithms: when does it matter and how much?, *Biological Conservation* **76** (1996), pp. 259–267.
- Pressey et al., 1997 R.L. Pressey, H.P. Possingham and J.R. Day, Effectiveness of alternative heuristic algorithms for identifying indicative minimum requirements for conservation reserves, *Biological Conservation* **80** (1997), pp. 207–219.
- Pressey et al., 2000 R.L. Pressey, T.C. Hager, K.M. Ryan, J. Schwarz, S. Wall, S. Ferrier and P.M. Creaser, Using abiotic data for conservation assessments over extensive regions: quantitative methods applied across New South Wales, Australia, *Biological Conservation* **96** (2000), pp. 55–82.
- Purdie et al., 1986 R.W. Purdie, R. Blick and M.P. Bolton, Selection of a conservation reserve network in the mulga biogeographic region of south-western Queensland, Australia, *Biological Conservation* **38** (1986), pp. 369–384.
- ReVelle et al., 2002 C.S. ReVelle, J.C. Williams and J.J. Boland, Counterpart models in facility location science and reserve selection science, *Environmental Modeling and Assessment* **7** (2002), pp. 71–80.
- Richardson and Funk, 1999 K.S. Richardson and V.A. Funk, An approach to designing a systematic protected area system in Guyana, *Parks* **9** (1999), pp. 7–10.
- Rouget et al., 2003 M. Rouget, D.M. Richardson and R.M. Cowling, The current configuration of protected areas in the Cape Floristic Region, South Africa—reservation bias and representation of biodiversity patterns and processes, *Biological Conservation* **112** (2003), pp. 129–145.
- Satersdal et al., 1993 M. Satersdal, J.M. Line and H.J.B. Birks, How to maximise biological diversity in nature reserve selection: vascular plants and breeding birds in deciduous woodlands, western Norway, *Biological Conservation* **66** (1993), pp. 131–138.
- Scott et al., 1993 J.M. Scott, F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. d'Erchia, T.C. Edwards, J. Ulliman and R.G. Wright, Gap Analysis: a geographic approach to protection of biological diversity, *Wildlife Monograph* **123** (1993), pp. 1–41.
- Smith and Theberge, 1986 P.G.R. Smith and J.B. Theberge, A review of criteria for evaluating natural areas, *Environmental Management* **10** (1986), pp. 715–734.
- Smith and Theberge, 1987 P.G.R. Smith and J.B. Theberge, Evaluating natural areas using multiple criteria: theory and practice, *Environmental Management* **11** (1987), pp. 447–460.
- Steyerberg et al., 2001 E.W. Steyerberg, F.E. Harrell Jr., G.J.J.M. Borsboom, M.J.C. Eijkemans, Y. Vergouwe and J.D.F. Habbema, Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis, *Journal of Clinical Epidemiology* **54** (2001), pp. 774–781.
- Swets, 1988 J.A. Swets, Measuring the accuracy of diagnostic systems, *Science* **240** (1988), pp. 1285–1293.
- Terborgh and Winter, 1983 J. Terborgh and B. Winter, A method for siting parks and reserves with special reference to Columbia and Ecuador, *Biological Conservation* **27** (1983), pp. 45–58.
- Usher, 1986 M.B. Usher, *Wildlife Conservation Evaluation*, Chapman and Hall, London (1986).
- Van Horne, 1983 B. Van Horne, Density as a misleading indicator of habitat quality, *Journal of Wildlife Management* **47** (1983), pp. 893–901.
- Williams, 1998 P.H. Williams, Key sites for conservation: area-selection methods for biodiversity In: G.M. Mace, A. Balmford and J.R. Ginsberg, Editors, *Conservation in a Changing World. Integrating Processes into Priorities for Action*, Cambridge University Press, Cambridge (1998), pp. 211–240.
- Williams and Araújo, 2000 P.H. Williams and M.B. Araújo, Using probability of persistence to identify important areas for biodiversity conservation, *Proceedings of the Royal Society of London Series B Biological Sciences* **267** (2000), pp. 1959–1966.
- Williams and Araújo, 2002 P.H. Williams and M.B. Araújo, Apples, oranges, and probabilities: Integrating multiple factors into biodiversity conservation with consistency, *Environmental Modeling and Assessment Special Issue: Reserve Design Modeling* **7** (2002), pp. 139–151.
- Wilson, 2003 Wilson, K.A., 2003. Uncertainty and vulnerability in conservation planning. PhD Thesis. Institute of Land and Food Resources, The University of Melbourne.

Woinarski et al., 1996 J.C.Z. Woinarski, O. Price and D.P. Faith, Application of a taxon priority system for conservation planning by selecting areas which are most distinct from environments already reserved, *Biological Conservation* **76** (1996), pp. 147–159.

Zweig and Campbell, 1993 M.H. Zweig and G. Campbell, Receiver-Operating Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clinical Chemistry* **39** (1993), pp. 561–577.