

When do conservation planning methods deliver? Quantifying the consequences of uncertainty

William T. Langford^a, Ascelin Gordon^a, Lucy Bastin^b

^a RMIT University, Melbourne, Australia ^b Aston University, Birmingham, UK

Abstract

The rapid global loss of biodiversity has led to a proliferation of systematic conservation planning methods. In spite of their utility and mathematical sophistication, these methods only provide approximate solutions to real-world problems where there is uncertainty and temporal change. The consequences of errors in these solutions are seldom characterized or addressed. We propose a conceptual structure for exploring the consequences of input uncertainty and oversimplified approximations to real-world processes for any conservation planning tool or strategy. We then present a computational framework based on this structure to quantitatively model species representation and persistence outcomes across a range of uncertainties. These include factors such as land costs, landscape structure, species composition and distribution, and temporal changes in habitat. We demonstrate the utility of the framework using several reserve selection methods including simple rules of thumb and more sophisticated tools such as Marxan and Zonation. We present new results showing how outcomes can be strongly affected by variation in problem characteristics that are seldom compared across multiple studies. These characteristics include number of species prioritized, distribution of species richness and rarity, and uncertainties in the amount and quality of habitat patches. We also demonstrate how the framework allows comparisons between conservation planning strategies and their response to error under a range of conditions. Using the approach presented here will improve conservation outcomes and resource allocation by making it easier to predict and quantify the consequences of many different uncertainties and assumptions simultaneously. Our results show that without more rigorously generalizable results, it is very difficult to predict the amount of error in any conservation plan. These results imply the need for standard practice to include evaluating the effects of multiple real-world complications on the behavior of any conservation planning method.

1. Introduction

The rapid global loss of habitat and biodiversity has led governments and non-governmental organizations to reserve and manage habitat for conservation purposes. Optimally balancing financial and ecological constraints to select sets of land parcels to acquire, preserve and rehabilitate is a complex problem. In reaction to this, there has been exponential growth in the literature on systematic reserve selection, design and optimization methods since the field first developed in the early 80s (Kirkpatrick, 1983; Pressey et al., 1993; Margules and Pressey, 2000; Sarkar et al., 2006). These techniques (from here on referred to as “conservation planning methods”) have been used to design and maintain reserve systems in ecosystems around the world and have a significant role to play as constraints on undertaking conservation actions become more complex (Margules and Pressey, 2000; Pierce et al., 2005; Oetting et al., 2006).

While many conservation planning methods are both useful and mathematically sophisticated, they currently only provide approximate solutions to real-world planning problems that have complications such as uncertainty in the inputs and conditions that change over time. The sensitivity of methods to a small number of input uncertainties is sometimes measured, but in terms of the resulting variation in solutions, not in terms of error. A detailed analysis of error resulting from mismatches to the structure of real-world conservation processes is almost never considered. In this paper, we present a structured approach for measuring and examining the consequences of these errors and the critical effect that they have on generalization and evaluation of method performance and utility. We demonstrate a unified modelling framework for improving conservation outcomes through predicting and characterizing error, as well as testing outcomes under a variety of conditions. We then control for attributes of problem structure and use the framework to highlight previously under-explored effects of factors such as how species are distributed across the landscape.

Numerous studies raise questions about the real-world complications and uncertainties faced in reserve design. Some studies investigate uncertainty in the species maps used in the reserve selection (Gaston and Rodrigues, 2003; Burgman et al., 2005; Wilson et al., 2005; Halpern et al., 2006; Ray and Burgman, 2006; Rondinini et al., 2006; van Teeffelen et al., 2006). Other studies examine sequential decisions and issues associated with loss and availability of habitat over time as opposed to immediately implementing an entire reserve network (Costello and Polasky, 2004; Meir et al., 2004; Pressey et al., 2004; Drechsler, 2005; Sarkar et al., 2006; Turner and Wilcove, 2006). Rodrigues et al. (2000) and Araujo et al. (2004) address temporal issues associated with variability in population locations due to climate change. Strange et al. (2006) look at the option to sell reserved land on which biodiversity value has decreased. O’Hanley et al. (2007), Cabeza (2003) and

Cabeza and Moilanen (2003) examine unreserved habitat and how its existence or loss affects reserved areas. Both Gaston and Rodrigues (2003) and Turpie (1995) investigate the use of species abundance instead of just presence/absence data. While most studies address only one or two complications at a time, in a single location, there are exceptions with a somewhat larger scope (Pressey et al., 1999; Drechsler, 2005; Turner and Wilcove, 2006). In spite of all these papers and more that warn about the consequences of uncertainties, it is still not standard practice to evaluate methods or plans under uncertainty, nor do commonly used methods address the bulk of these uncertainties in their algorithms and objective functions. Our aim here is to build on all of these works by addressing many real-world uncertainties simultaneously in a systematic way that will enable both better generalization and more reliable estimates of the error in particular outcomes.

We believe that the primary obstacles to quantifying and generalizing estimates of error in methods and solutions are the difficulty and lack of standardization in doing it. It requires the use, sharing, programming, and understanding of many different kinds of models related to landscape structure, species persistence, economics, land use change, as well as error and uncertainty in species modelling. Given that most of these tasks are tedious and lie outside the primary focus of an author’s original interest and/or expertise, it is not surprising that they are left undone. We address these issues by describing a single conceptual framework that can be wrapped around any given conservation planning method or rule of thumb to measure and deal with the effects of multiple uncertainties.

When implemented, this framework is designed to allow the modelling of different conservation planning methods using performance measures such as species representation and persistence across a range of different initial and time-varying conditions. The range of conditions includes landscape structure, species distribution patterns and temporal changes in populations and habitat condition. Most importantly, this framework allows us to investigate the effects of uncertainty in all of the above conditions. It is intended to evolve and have modules that are easily replaced by users, making as few assumptions as possible. Finally, it also allows users and researchers to address issues of generalization, by providing a mechanism for measuring outcomes for large numbers of examples.

In the remainder of this paper, we describe the framework and demonstrate its use with four example experiments that highlight the critical importance of quantitatively evaluating the performance of conservation planning methods and scenarios. These experiments would be much more difficult without the framework and can address many useful questions. These include comparing conservation planning methods in different situations, as well as how resources should be spread between activities such as reducing error in existing habitat distribution models or generating habitat models for new species.

2. Materials and methods

The evaluation framework that we present measures the consequences of a conservation planning method's interactions with four classes of uncertainties. The first is uncertainty in the inputs. The second is uncertainty in the planning process itself, that is, uncertainty in the match of the planning method to the real-world process. The third is uncertainty in our knowledge of how the real world will evolve over time, such as what habitat will be lost and when. Fourth and last, is uncertainty in the ability of the planning method to find the best outcome even when given correct inputs. We address these uncertainties through a framework consisting of six modular steps:

- 1) define landscapes,
- 2) define species distributions over landscapes,
- 3) define costs over landscapes,
- 4) determine and undertake conservation actions,
- 5) model system dynamics,
- 6) measure consequences.

The first three steps relate to input uncertainty, the fourth to planning process uncertainty, the fifth to uncertainty in real-world dynamics. The sixth step addresses the collection and presentation of results as well as the method's ability to find the best outcome when it is possible to solve for the true optimal outcome. Each of the self-contained steps that use input data must allow the use of either real or synthetic data. This enables specific real scenarios to be studied, as well as allowing for the more general questions to be addressed under controlled conditions. Uncertainties and sensitivities in data and parameters are injected into or wrapped around each step. Details of each module are discussed below. We have implemented this framework by integrating a number of existing models and writing other missing pieces ourselves in R (R Development Core Team, 2008) and Java. Currently the source code is available from the authors on request, with plans to publicly release an open source version of our framework in the future. Screen shots of the graphical user interface that controls most of the processes in the uncorrelated and richness is uniformly distributed across the landscape. Given the number of species and the number of patches in the landscape, occupancy maps can be generated for each species by randomly sampling from the rarity and richness distributions to designate a set of patches for each species to occupy.

In what follows, the real or synthesized input maps are considered to be the "correct" occupancy maps which designate the true locations of a species. Uncertainty in these maps can then be modelled by creating an erroneous version of each Correct map through addition and/or subtraction of habitat patches. (Throughout this paper, words such as "correct" and "apparent" are capitalized when they are used as the names of maps or names of reserve selection methods rather than as generic adjectives.) The erroneous maps are intended to represent the map that a user would actually have available to them, for example as the result of running a habitat model which over-predicted occupancy and/or quality, or doing a field survey that recorded false absences. These erroneous versions of the Correct maps are referred to as the Apparent maps. In all the results presented in this paper, the conservation planning methods only have access to the Apparent maps.

2.3. Define costs over landscapes

Define costs over landscapes assigns costs to actions taken on planning units, for example, the cost of reserving a planning unit. As with defining habitat maps, uncertainty can be applied to the costs to give Correct and Apparent costs.

2.4. Select conservation actions

Select conservation actions allows the user to plug in the conservation planning method of their choice. In the experiments described here, we concentrate on five reserve selection methods (though the analysis of actions

framework can be seen in the [Supplemental Fig. S1](#).

2.1. Define landscapes

Define landscapes creates or identifies the landscape where actions will be evaluated. This may be based on real data provided by the user, or it can be generated by a program that synthesizes landscapes, such as RULE (Gardner, 1999). The landscape definition consists of two types of base maps: habitat and planning units. The base habitat map shows the union of all habitat patches for all species in the system. Habitat maps for individual species are subsets of this map. The planning unit base map shows the discrete areas over which conservation actions such as reserve selection are constrained to act and need not bear any relationship to habitat patches. This addresses the fact that the unit of action for conservation is often at the level of land ownership parcels or administrative boundaries rather than habitat patches, though planning units can also be defined to be the habitat patches themselves.

2.2. Define species distributions over landscapes

Define species distributions over landscapes initializes a set of species-specific habitat maps showing the extent and condition of each species' habitat. Sets of real maps may be used, such as those from field surveys or habitat distribution models, or simulated maps may be used to test how well a conservation planning outcome generalizes to different distributions. For example, these maps could be generated using one of the methods provided in our implementation of the framework, which allow the user to specify a "rarity distribution" and a "richness distribution" over all species. We use the term "rarity distribution" to mean the description of the number of species occupying a given number of patches (for example, 5 species occupy 1 patch each, 8 species occupy 2 patches each, etc.). We use the term "richness distribution" to mean the number of species on each of the patches, the species richness, sorted in decreasing order of number of species. It represents the extent to which species co-occur on the same patches. At one extreme, species occurrence on patches is highly correlated, forming species richness "hotspots", while at the other extreme, a flat richness distribution, the locations of species are

beyond just reserve selection is permitted within the framework). We test two heuristic rules which we call Simple Richness and Unprotected Richness. The Simple Richness heuristic selects patches in descending order of the number of species on the patch with ties broken randomly. The Unprotected Richness heuristic is similar to that used in Meir et al. (2004); patches are chosen in descending order of the number of species on the patch whose representation targets have not yet been met. Ties are first broken by species richness, then randomly. The third conservation planning method is simply selecting reserves at Random. This method exists to provide a baseline against which other methods can be compared.

The last two methods employed are the software packages Marxan (Ball and Possingham, 1999; Possingham et al., 2000) and Zonation (Moilanen et al., 2005; Moilanen and Kujala, 2006). These are chosen because they are well-known and commonly used. Marxan uses a stochastic search algorithm (simulated annealing) to determine sets of conservation reserves that meet specified targets for reservation (e.g., 30% of the habitat for all species) while minimising other constraints such as economic cost and the reserve boundary length. The Zonation package uses a reverse stepwise heuristic that iteratively removes cells from the landscape in an order that minimizes marginal loss while maintaining connectivity (Moilanen et al., 2005). The cell removal order can be used as a priority rank, with cells remaining last receiving the highest priority. The lowest scored cells can be used to identify areas where further habitat loss causes the smallest loss of conservation value. When using the Zonation package in this paper, the core-area analysis method was selected out of the several variants available (Moilanen et al., 2005).

2.5. Model system dynamics

Model system dynamics models the processes that control system outcomes; for example, models of changing landscape condition, habitat loss, disturbance, metapopulation dynamics, and economics. Each of these may

also be modelled either statically or over time. Because there is no one “correct” model of these processes, this module must allow for the use of different models ranging from very simple constant value or random models to more sophisticated modelling packages like the dynamic landscape metapopulation package RAMAS Landscape (Akçakaya et al., 2003) or any other user-supplied model. It is important to note that unlike the reserve selectors, models used for system dynamics are to be run over the Correct landscapes rather than the Apparent landscapes, since it is the Correct species distributions and habitat configurations that will determine the true outcome of a conservation planning exercise.

2.6. Measure consequences

Measure consequences allows the summarizing and quantifying of the consequences of the modelled actions, using any method that can operate over the output from the modelling of system dynamics. This can involve summarizing biodiversity outcomes such as the estimated minimum population size or extinction risk derived from RAMAS Landscape, or the difference between the Apparent and Correct representations of species achieved through reserve selections. If other forms of models are attached, they can also quantify outcomes unrelated to biodiversity such as the economic consequences of a strategy or the amount of carbon sequestered.

2.7. Four experiments

We conducted four experiments, designed to illustrate analyses which are currently difficult to do without this framework and to demonstrate the conservation consequences of uncertainties and real-world complications. These experiments represent just a small proportion of the scenarios that can be investigated with a framework such as this. Many of the experiments in the current literature which relate to uncertainties in conservation planning and reserve selection can be easily replicated, extended, and combined within this framework.

In the first experiment, we compared the efficiency and robustness to uncertainty of several different conservation planning methods on a given landscape and initial species distribution. In the second experiment, we again looked at robustness and efficiency but this time we considered the effects of different richness distributions of species across the same landscape. In the third experiment, we examined the performance of a single conservation planning method in terms of persistence instead of just representation. It also shows how the framework allows us to measure the effects of combining many different real-world complications simultaneously. Again, we emphasize the ability to consider effects of uncertainty. Finally, we demonstrate how the framework can be used to investigate the consequences of tradeoffs, in this case, spending resources to improve existing species maps versus acquiring more low quality species maps. We also demonstrate how users can examine the side effects of including or excluding different rare and common species in a conservation planning method. Below, we give an overview of the parameters and models used for the experiments. Further details of each experiment are given in the [Supplementary data](#).

2.7.1. Landscapes

In the experiments described in this paper, we used two different landscapes. In the first case, we performed the analysis over a synthesized landscape containing 1600 habitat patches. This number was chosen as being small enough to keep processing times within reasonable limits, but large enough to provide a complex problem for optimization. The experiments carried out using this landscape made no use of the spatial location or absolute area of the patches, so any map with 1600 separate patches would have given the same result. The planning units in experiments using this landscape were set to be the patches themselves. In the second case, we used a real landscape with 212 habitat patches that is a 25 km×25 km subset of the Melbourne Open Space Map (ARCUE, 2002). This area is on the eastern side of the city of Melbourne, in the state of Victoria in south eastern Australia, and has varying patch sizes ranging from 7 ha to 846 ha (median of 12 ha).

2.7.2. Species distributions

All but one of the experiments described in this paper used distributions based on a species rarity distribution derived from data gathered by the Victorian Government's Department of Sustainability and Environment. That data indicates the presence/absence of 4080 flora species on 36,787 30×30 m quadrats across the state of Victoria. Some of the experiments were also based on the richness distribution of species from this survey. Throughout this paper, these distributions will be referred to as the Victorian distributions. For these analyses, realistic richness and rarity distributions were simulated by randomly sampling from the Victorian distributions using the “sample” command in R. Fig. 1(A) shows the rarity distribution using 200 species and 1600 patches drawn using the Victorian distributions as the base, and Fig. 1(B) shows the corresponding richness distribution drawn using the same base. Fig. 1(C) shows the richness distribution of just the 50 rarest species in the distribution shown in Fig. 1(B).

Three different types of species richness distributions were used in these experiments: flat, clumpy, and Victorian. The Victorian distribution, as described above, was intended to represent a realistic distribution of species across patches. The clumpy distribution was intended to simulate species “hotspots” on the landscape, while the flat distribution, where species were more uniformly distributed across the landscape, was intended to reflect a more challenging problem for conservation planning methods. That is, if most of the species occur on a few patches, then selecting patches that contain a representation of each species is relatively simple, while a more homogeneous distribution makes it more difficult to choose patches that capture diversity.

2.7.3. Habitat map error model

The example experiments described here only examined Apparent maps where there were more patches of habitat than in the Correct map. This was intended to represent both erroneous overestimation (Hurlbert and Jetz, 2007) and the case where it is easier and cheaper to identify potential habitat (for example by modelling and/or remote sensing) than habitat which is actually occupied. This is particularly important for species that are poor dispersers and in situations where there are large barriers to dispersal, such as roads in peri-urban settings. For the purposes of this paper, our point is simply to demonstrate a single example error model which bears a useful relation to common conservation situations. In some contexts it may be more appropriate to model underestimation of species representation, or a mix of both, to capture the phenomenon where species go undetected in a field survey. This is also easy to do using the framework.

2.7.4. Loss model

Experiments in this paper assumed that reserved areas were implemented completely at the start of a run, but this is not a requirement. Unreserved habitat was either all lost when the reserves are implemented or lost gradually over time based on a simple model where a fixed proportion of unreserved planning units are randomly selected for development every 10 years over a 100 year period. When planning units were lost, they were considered to no longer contain habitat for any species that existed on them prior to development.

2.7.5. Economic model

In these experiments, economic costs were incorporated using a range of simple static cost models. In some experiments, all patches were assumed to have the same cost. In others, they were assigned a randomly-selected cost ranging between 10 and 500 units. While random costs are not likely to occur in reality, the examples here exist only to demonstrate the techniques and the value of manipulating and modelling economic costs within the framework.

2.7.6. Representation and persistence measures

In our example experiments we measured estimates of two quantities, representation and persistence. The goal of conservation planning methods is often phrased in terms of attaining a given or maximal level of representation of each species. After running various planning methods over the Apparent habitat maps discussed above, we computed the number of patches that each method reserved for each species given a specified budget and costs. This computation was done on both the Apparent habitat map used to make the selections and on the Correct map which is not seen by the planning method. We then compared the relative effectiveness of the mechanisms in achieving representation and

compared their robustness with respect to the uncertainties in their inputs. That is, we assessed how much representation was actually achieved (on the Correct map) compared to the level that appeared to be achieved using the Apparent map as a model input.

Since representation alone does not guarantee persistence of species, we also investigated more direct estimates of persistence within the selected reserve sets. In particular, we measured the expected minimum population size (McCarthy and Thompson, 2001) of some species using the RAMAS

Landscape metapopulation modelling package. As RAMAS is a complicated modelling system that requires a large number of parameters, we only ran these models for a handful of species whose requirements and characteristic distributions are well surveyed and documented. Details are given in the description of Experiment 3 below.

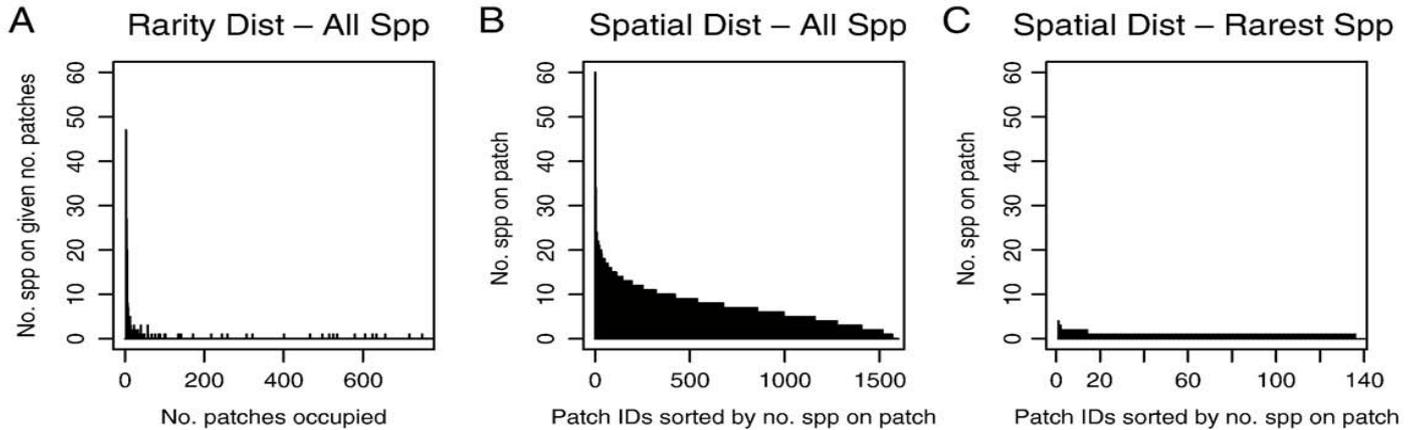


Fig. 1. Species distributions for 200 species sampled using Victorian distributions as base. Panel (A) shows the rarity distribution, i.e., for each possible number of patches that a species could occupy (shown on the x-axis), the y-axis indicates the total number of species in the sampled distribution that occupy the given number of patches. (B) and (C) show the species richness distributions. In (B) the number of species on each of the 1600 patches used in Experiment 1 is shown. Values are sorted in decreasing order of number of species. (C) shows the same information for just the 50 rarest species from (B).

2.8. Experiment 1 — comparing performance of multiple methods under uncertainty

For this experiment we investigated a scenario with 200 species and 1600 patches and assessed the budget required to purchase a set of patches that would contain at least one representation of each species. The four conservation planning methods tested were Zonation, Unprotected Richness, Simple Richness, and Random. Each method except Random was run approximately 130 times, starting with a budget of zero and gradually incrementing the budget until it reached the amount required to buy all 1600 patches in the landscape. The Random method was run 1000 times at each budget level and the mean value at each budget level was computed, along with the 95% confidence interval derived from the 1000 runs.

Before any of the tests were run, the landscape was initialized as follows: costs between 10 and 500 units were randomly assigned to each patch, and species were randomly assigned to patches using both the Victorian rarity and richness distributions (Fig. 1(A) and (B)). For this experiment, Apparent maps were generated by adding extra patches at random until the resulting maps had 30% more patches than the corresponding Correct maps. To determine the difference in apparent and correct performance, percent error was calculated as:

$$\text{PercentErr} = \times 100 \frac{p_a - p_c}{p_c}$$

where p_a and p_c are the Apparent and Correct proportions of species habitats represented in a given reserve network, respectively.

2.9. Experiment 2 — comparing effects of different species richness distributions

Here we extended the results of Experiment 1 to consider different richness distributions of species, by adding runs which used the flat and clumpy richness distributions while still using the same rarity distribution. The aim for each of the four conservation planning methods was to identify the minimum budget (as a proportion of the sum required to purchase all patches in the landscape) required to acquire patches containing at least one representation of each species. As in Experiment 1, the conservation planning methods only had access to the Apparent maps. To genuinely cover each species, a greater number of patches would always need to be purchased,

increasing the required budget by differing amounts depending on the richness distribution of species. Using the Correct maps, the relative performance of each method, for each species richness distribution, could be compared to the actual representation goals achieved.

2.10. Experiment 3 — measuring persistence and combining many complications at once

In this experiment we demonstrated the framework's ability to incorporate multiple real-world uncertainties simultaneously, and to assess selected reserves in terms of species persistence instead of representation. Here we briefly describe the experiment and give further details in the [Supplementary data](#). We use real species and a real landscape, based on the Melbourne Open Space Map (ARCUE, 2002) containing spatially realistic patch size variability and connectivity (Fig. 4). This experiment used 7 species, for which well-parameterised metapopulation models were available and employs the framework to generate habitat distributions for each species. Fig. 4 shows the resulting habitat maps for two of the species. To illustrate a realistic situation where planning units do not align with habitat, we tiled the study area with 16 ha square planning units and selected a range of reserves using the Marxan reserve selection tool (Ball and Possingham, 1999; Possingham et al., 2000). Once reserves had been selected, two loss models were used to simulate landscape change over 100 years where habitat is destroyed due to development. Loss Model 1 removed planning units at a rate that ensured all unreserved planning units were lost within the 100 years. In Loss Model 2, the loss rate was set so that 25% of unreserved planning units remained after 100 years, potentially leaving some habitat outside the reserved areas selected by Marxan.

The performance of the reserve system was determined by running metapopulation models in RAMAS Landscape for each of the seven species, over the 100-year time series of maps produced by the Loss Models. Species densities used to estimate the carrying capacity of ideal habitat were set at realistic values (see [Supplemental Table S8](#)). Expected minimum population size (McCarthy and Thompson, 2001) was used as the persistence measure as it provides a good indication of the propensity for population decline and is less sensitive to model assumptions than risks of decline or extinction.

We used four error models to create the Apparent maps from the Correct maps. The first model assumed no error (Apparent and Correct maps were

identical). The second assumed that the Apparent map overestimated the number of patches in the Correct map by 50%. The value of 50% was chosen to produce a demonstrable effect, although the overestimation of potential habitat may be larger in some situations (see for example, [Hurlbert and Jetz \(2007\)](#)). The third model assumed the carrying capacity of each patch was 50% lower than shown on the Apparent maps. The fourth model combined models 2 and 3 and in addition, set the initial abundance of each patch to be a uniform random proportion of the Apparent carrying capacity.

2.11. Experiment 4 — effects of varying uncertainty, number of species, and choice of species included

In this experiment, we combined several sources of variation, showing how we can evaluate the performance of a particular method (in this case Marxan) as input map uncertainty increases and the species of interest change. In all of the experiments described so far, we have evaluated each method's performance in achieving representation over the same set of species that it used to select reserves. However, we may also want to know what the effects of that selection are on other species not included in the original set or on a particular subset of the original set. For example, how effective is the planning method for species with a similar or a different distribution than are steps of 25 species until we reached 200. This gave a total of 6 [error levels] × 8 [species levels] = 48 different reserve selection maps. We did this whole process once for a small budget (determined by the cost to purchase 3% of the whole landscape) and again for a large budget (set to 10% of the cost to purchase the whole landscape). The cost of each patch was allocated from a uniform random distribution, ranging from 10 to 500 units. We then calculated the correct and apparent number of species represented by the reserve selection in each case, using the normal matched evaluation method. At this point, the experiment was similar to Experiments 1 and 2 except for the use of a different conservation planning method and exploring a range of levels of map uncertainty and numbers of species. These values could then be used to determine how much reduction in map error would be necessary to achieve a given level of correct performance at each number of species.

We then went one step further and ran unmatched evaluations on these reserve selections to examine the impact of including and excluding different sets of species when running Marxan. Because the problem is comparatively easier when there is a large budget, we did the unmatched evaluations using only the small budget. To keep the example simple, we did all of the unmatched evaluations against the 50 rarest species in the original set of 200. For each of the 48 different reserve selections in the matched evaluations, we computed the proportion of the 50 rarest species that had at least one representation. For example, we took the results of reserve selection using the 25 rarest species at each uncertainty level and computed how many of the 50 rarest species had at least one patch reserved for them by that selection. We then asked the same question when the 75 rarest were used to do the reserve selection instead of 25, and so on. Doing these evaluations at 25 species begins to look at the question of how well species not included in the optimization fare under that reserve selection. Doing the evaluation for 75 or more species in the optimization begins to look at a different question: how well does a focal set of species do as more and more species are included in the optimization?

3. Results

3.1. Experiment 1 — comparing performance of multiple methods under uncertainty

Out of the four conservation planning methods tested under uncertainty, Zonation performed best on the Apparent maps, followed by Unprotected Richness, then Simple Richness and finally, Random (Fig. 2A). Zonation achieved at least one representation of each species at 0.027 of the total landscape cost, Unprotected Richness at more than 1.5 times that cost, 0.043, and Simple Richness at more than twelve times the cost, 0.328. Random never achieved a representation for every species within the given budget.

For all reserve selection methods, the level of representation achieved on the Apparent maps was higher than on the corresponding Correct maps. Fig.

not included in the original set (e.g., less charismatic species)? Or, how does the method do on just the rarest species in its original set, particularly when it is unable to attain its goals for some of the species in the original set? We will refer to evaluating on the original set as normal or matched evaluation. Experiments 1, 2, and 3 all use matched evaluation. We will refer to evaluation on any set other than exactly the original species used in the conservation planning method as unmatched evaluation. In this experiment, we used both matched and unmatched evaluation so that we could highlight the side effects of choosing particular species to use in the conservation planning method.

We created habitat maps by distributing habitat for 200 species over 1600 patches drawn from the Victorian distributions, as in Experiments 1 and 2. For each of these 200 Correct maps, we then created 6 erroneous maps simulating a range of uncertainty levels between 0% and 50% overestimation of habitat, in steps of 10%. This allowed us to plot the change in true and apparent performance as error in the Apparent maps increases. We then ran Marxan at each error level using only the maps of the 25 rarest species, which gave us six different reserve selections with the goal of reserving at least one patch for each species. We then repeated this process for the 50 rarest species, the 75 rarest and so on in

2(B) shows the percent error between the species representation calculated using the Apparent and Correct maps over the budget range. The point at which the percent error curve reached zero coincides with the budget at which a given method actually achieved its goal of representing all species. The ranking of the performance of the four methods was the same for both the Apparent and the Correct results. Initially Zonation had the largest difference between Correct and Apparent performance, but this rapidly decreased to levels below the other methods. At the lowest cost, where each method achieved one representation on the Correct maps, Zonation's true cost was 0.077, Unprotected Richness was 0.314 and Simple Richness was 0.341. Thus, Zonation outperformed Unprotected Richness both in efficiency of selecting patches and in robustness to error. Simple Richness was less efficient in selecting patches but the difference between Correct and Apparent representation was small (3.96%). Unprotected Richness had the largest overall difference and was therefore, the least robust to error in the habitat maps.

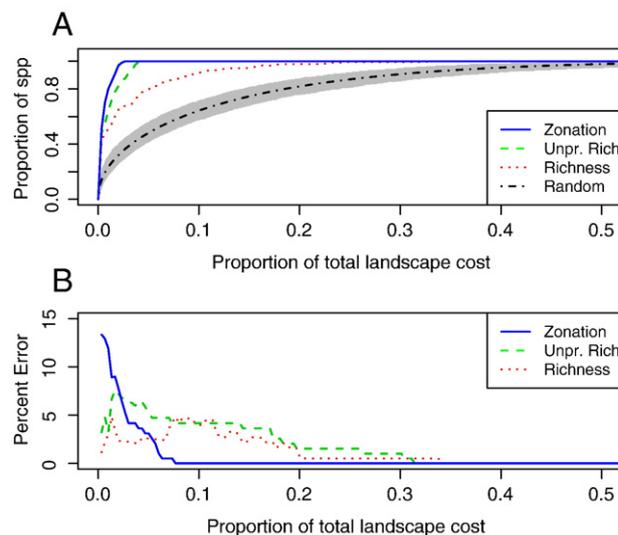


Fig. 2. Comparison of conservation planning methods across budgets. The efficiency and robustness to uncertainty are shown for three conservation planning methods. In this experiment habitat for 200 species is distributed over 1600 patches. The Victorian distributions are used for species rarity and richness and 30% overestimate of habitat is used when generating the Apparent maps. In (A), the y-axis shows the proportion of species with at least one patch in the reserve set, while the x-axis records each solution's cost as a proportion of the cost of buying every patch in the landscape. The grey region for Random shows the 95% confidence interval for 1000 runs at each cost level. In (B), the y-axis shows percentage error in the Apparent representation with respect to the representation attained on the corresponding Correct maps (i.e., $100 \times (\text{apparent} - \text{correct}) / \text{correct}$).

3.2. Experiment 2 — comparing effects of different species richness distributions

The rankings of all methods remained the same as in Experiment 1, regardless of the richness distribution of species. Each method, however, exhibited a different pattern of performance and robustness to uncertainty across richness distributions. Results for Experiment 2 are shown in Fig. 3. The exact values used to create the plots are presented in Supplemental Table S1.

Zonation achieved at least one representation of each species in the Apparent maps at 0.023 of the total landscape cost on the clumpy distribution, 0.027 on the Victorian distribution, and 0.033 on the flat distribution. In the corresponding Correct maps, Zonation did not reach a true single representation of each species until the cost was

2.74 times higher on the clumpy (0.063), 2.85 times as high on Victorian (0.077), and 2.64 times higher on flat (0.087). Zonation outperformed all of the others on all three problems in both Apparent and Correct performance, but in each problem, its true performance was more than 2.5 times worse than its Apparent performance.

Like Zonation, Unprotected Richness required larger budgets as the richness distribution became flatter. Unlike Zonation, the difference between Correct

and Apparent performance increased dramatically with increasing flatness of the richness distribution. The true performance was 15 times worse than the Apparent performance for the flat distribution. Unprotected Richness had much better Apparent performance than Simple Richness and Random. However, its true performance was similar to that of Simple Richness and completely unlike that of Random. In fact, it was worse than Random on the flat distribution. Simple Richness performed well on the Apparent clumpy distribution but like Unprotected Richness, had deteriorating performance as the richness distribution flattened, to the extent where it performed no better than Random on the flat distribution. For Simple Richness, the performance based on the Apparent maps was a good indication of the true performance.

Random choice of reserves gave nearly identical Apparent mean performance across richness distributions (requiring a budget of approximately 0.67). Similarly, for the Correct maps, mean performance was nearly identical but roughly 1.15 times the values for the corresponding Apparent maps. All three distributions had similar size confidence intervals around the mean, both for Apparent and Correct maps. Like Simple Richness, the Apparent performance was a reasonable indication of true performance.

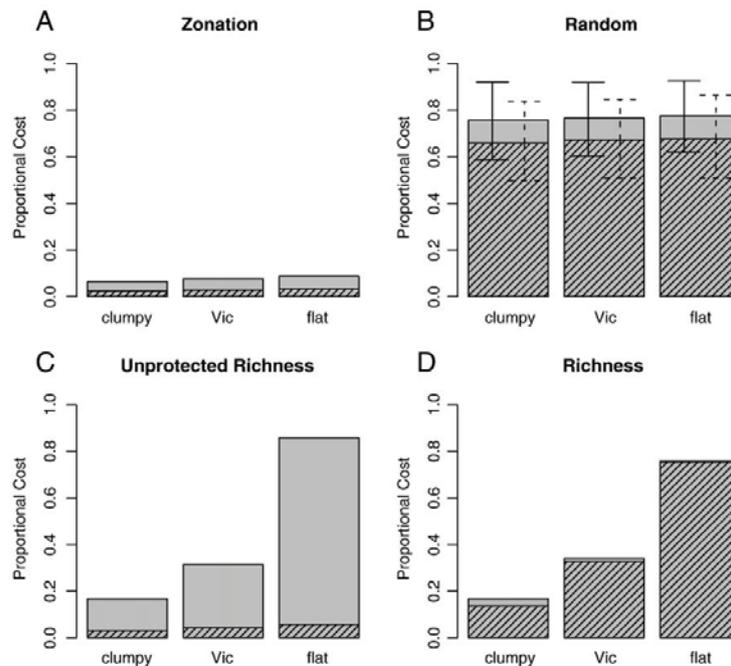


Fig. 3. Comparison of Apparent and Correct cost to purchase a conservation network with at least one representation of each species under different richness distributions and 30% overestimate of habitat. Bar charts are shown for the different conservation planning methods and the three richness distributions (clumpy, Vic and flat; see text). The y-axis shows the cost, which is measured as a proportion of total landscape cost. The hatched areas show the cost based on the Apparent maps and the solid grey shows the extra budget required to represent each species at least once based on the Correct maps. Random has error bars for the 95% confidence interval around the mean; solid bars are for the Correct result's confidence interval and dashed bars for the Apparent result's confidence interval.

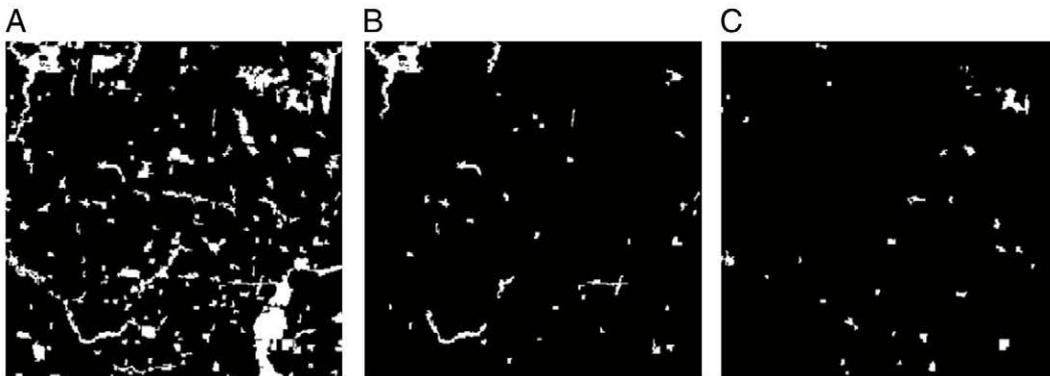


Fig. 4. Melbourne Open Space habitat maps. White patches are habitat and black patches are non-habitat. (A) shows the master habitat map obtained from the Melbourne Open Space Map. This map shows all habitat patches in the study area and has patch sizes ranging from 7 ha to 846 ha. Panels (B) and (C) show the subset of the full habitat that is assigned as habitat for the legless lizard and the ringtail possum, respectively, in Experiment 3.

3.3. Experiment 3 — measuring persistence and combining many complications at once

Experiment 3 demonstrates how reserves selected with Marxan can be evaluated in terms of species persistence in the presence of several real-world uncertainties and complications. In this experiment there were no cost constraints and in every case Marxan was able to find reserves that satisfied the representation goals.

Fig. 5 shows the results of selecting conservation areas with Marxan and then running metapopulation models with Ramas Landscape for two of the seven species: the striped legless lizard (*Delma impar*) and the ringtail possum (*Pseudocheirus peregrinus*). These two species vary in their dispersal abilities; the ringtail possum is a medium disperser and the legless lizard is a short disperser (see [Supplementary data](#) for details of the metapopulation models). The estimated minimum population sizes (EMPs) shown in Fig. 5 tend to increase as the species representation goals are increased, but the increase is dependant on the loss and error models used.

An integrated way to understand the results in Fig. 5 is to view them with respect to achieving a particular EMP for a species. For example, suppose all unreserved habitat was lost and that the minimum acceptable EMP was 20% of the EMP value for the species when there was no habitat loss (i.e. 0.2 on the vertical axis in Fig. 5). In that case, with no error in the habitat maps, appropriate representation goals would have been 300 ha for the legless lizard and 100 ha for the ringtail possum.

If carrying capacity was reduced by 50%, these goals should have been increased to 500 ha and 300 ha, respectively. If the Correct habitat was overestimated by 50%, then the possum would have needed a 600 ha representation goal and the lizard would not have been able to achieve its persistence goal within 600 ha. If both errors were present, along with

variable initial abundance, then neither species could have achieved their persistence goals within 600 ha. Relaxing the assumption that all unreserved habitat was lost results in the lizard meeting its persistence goal within 600 ha, even with all errors present, while the possum fails to meet its persistence goal within 600 ha.

When all unreserved habitat was lost (Loss Model 1), overestimating the habitat by 50% resulted in a greater reduction of EMP values than overestimating the carrying capacity. When some unreserved habitat was retained in the landscape (Loss Model 2), then the relative impact of the two error models was reversed. Combining overestimation of carrying capacity and habitat with variation in initial abundance resulted in greater reductions in EMP values under both loss models, and the EMP values improved relatively little when more Apparent habitat was added to the reserve.

Thus, the impact of an error model was determined by the loss model. Overestimation of habitat resulted in planning units being chosen which did not constitute any genuine habitat, whereas overestimating carrying capacity only reduced the number of individuals a selected planning unit could support. When unreserved habitat was present, the impact of overestimation of habitat was reduced, as selected planning units that did not constitute habitat could be balanced by remaining unreserved habitat for the species. The impact of overestimating the carrying capacity was also reduced by unreserved habitat, but to a much lesser extent due to the carrying capacity of the unreserved habitat also being overestimated.

The extra variability in the EMP values for Loss Model 2 resulted from the presence of unreserved habitat. Loss Model 2 was parameterised so that 25% of all planning units remained (including all reserved planning units). As the planning units selected to be lost were chosen randomly, the amount of unreserved habitat left for a given species could vary. There was also an interaction between the reserve selection and the loss model as the selected planning units determined the planning units available to be lost.

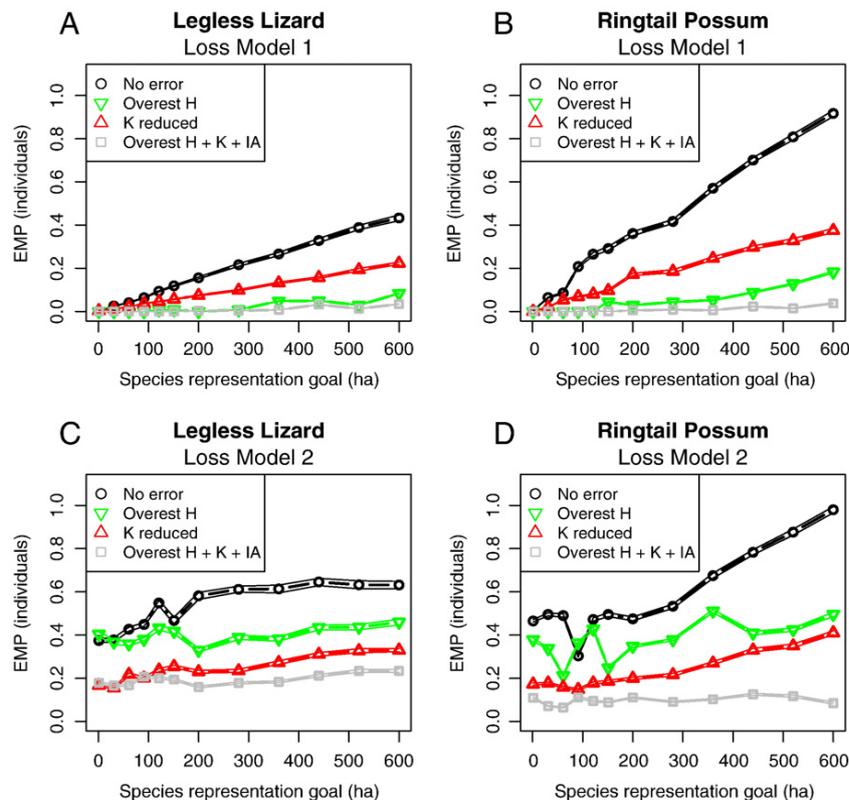


Fig. 5. Selecting reserves using Marxan and evaluating the reserves using persistence measures under various real-world complications. Plots show the increase in persistence with increasing reserved habitat for two species, the striped legless lizard (*Delma impar*) and the ringtail possum (*Pseudocheirus peregrinus*), under different loss and error models. The x-axis shows the representation goal used in Marxan for each species and the y-axis shows the expected minimum population (EMP) size resulting from the metapopulation model as a proportion of the EMP obtained using the habitat maps with no error and with no habitat loss occurring. (A) and (B) show results for Loss Model 1, where all unreserved planning units are lost by the end of the 100 years. (C) and (D) show results for Loss Model 2, where only 75% of all unreserved planning units are lost by the end of the 100 years. Each plot has a line showing the results of running Marxan over a range of representation goals (0 to 600 ha) for each of four different error models: no error (open black circles), 50% overestimation of habitat (inverted green triangles), 50% reduction in carrying capacity (red triangles), and a combination of the carrying capacity and habitat errors plus a 50% reduction in initial abundance (grey squares). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. Experiment 4 — effects of varying uncertainty, number of species, and choice of species included.

In this experiment, we measured the effects of the number and choice of species included in the reserve design over a range of uncertainty levels in the habitat maps. The results are shown in Fig. 6 and in Supplemental Tables S2–S7. A key point to remember is that in some cases, the optimization and evaluation were done on different sets of species. In all cases in Fig. 6, the objective function inside Marxan was optimized using a different number of species at each point along any contour of uncertainty, but in Fig. 6(E) and

(F), its performance was evaluated on a different set of species. For example, corresponding point in Fig. 6(F) shows the score for the 50 rarest in Fig. 6(D), the point (uncertainty = 0.5, number of species = 75) has species when Marxan was optimizing for the 75 rarest. This means that 60% of the 75 rarest species actually that plot has a value of 0.66, which means that 66% of the 50 rarest had at least one representation in Marxan's result when it was given species actually had at least one representation in the solution for Apparent input maps for the 75 rarest species. On the other hand, the optimizing over the 75 rarest species.

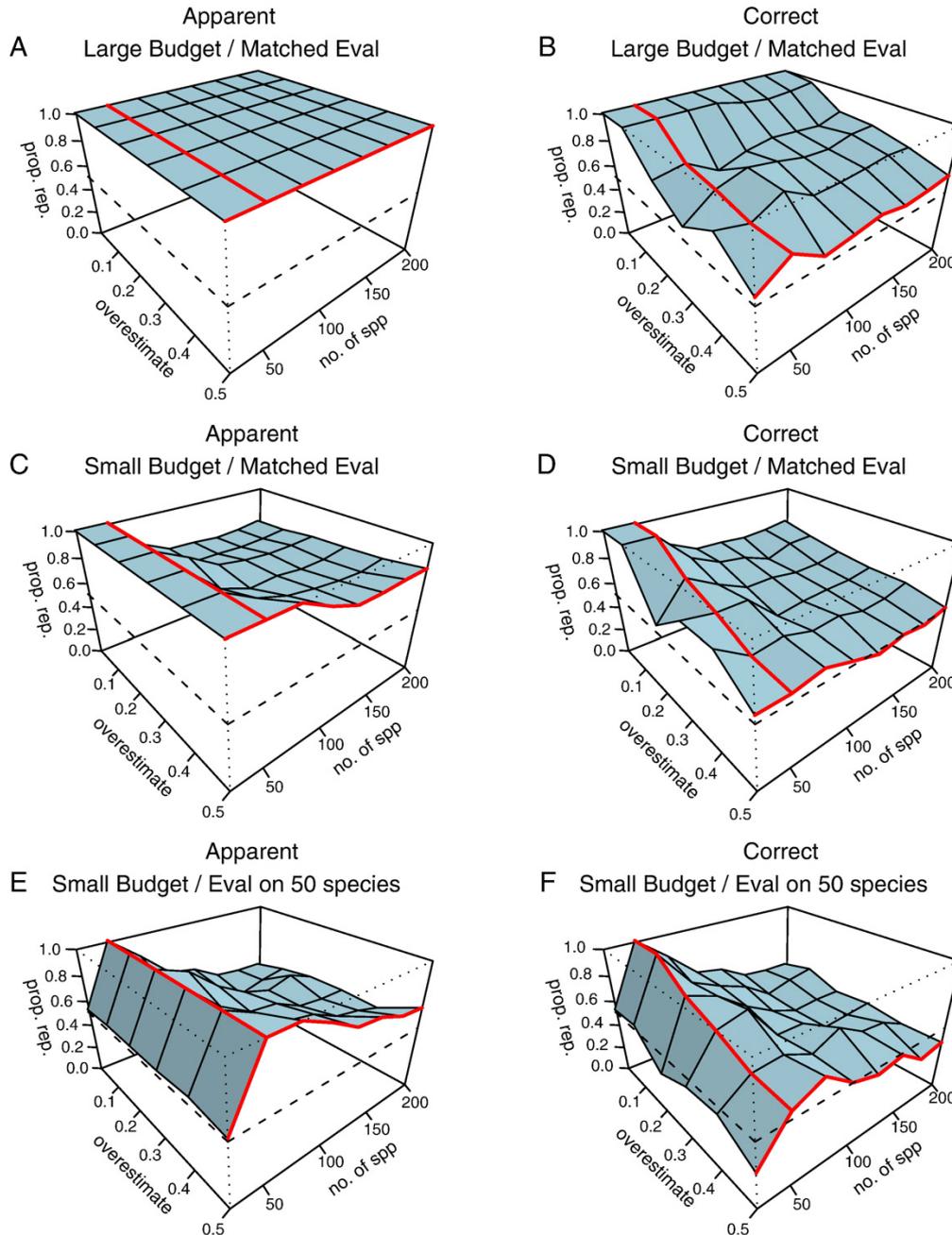


Fig. 6. Apparent and Correct proportion of species with at least one patch reserved using Marxan as a function of error in species habitat maps and number of species used in the optimization. In all plots the two axes in the horizontal plane are i) the number of species habitat maps used to generate Marxan's input data, and ii) the amount of uncertainty used in the error model to overestimate the true habitat. The vertical axis represents the proportion of species with at least one representation in the reserve network selected by Marxan. The left column shows the Apparent scores for each reserve selection and the right column shows the corresponding scores using the Correct maps for the selection made on the Apparent maps. The first row is for reserves selected using the large budget. The second and third rows are for reserves selected using the small budget. The first two rows show the scores using the normal/matched evaluation method. The last row shows the scores for an unmatched evaluation of the same reserve selections shown in the second row. For the unmatched scores the y-axis represents the proportion of just the 50 rarest species that have at least one representation in the set reserved by Marxan. Note that in all of the figures, a thick red line traces the contours for 50 species and 0.5 error as a reference. A dashed line is also shown for reference at the 0.5 representation level above both axes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The case of normal evaluation with a large budget is shown in Fig. 6(A) and (B). Results using the same Apparent maps for optimization and evaluation are shown in Fig. 6(A). In this case, Marxan found at least one representation for each of the species. However, in the corresponding evaluations on the Correct maps shown in Fig. 6(B), this perfect representation only held for error levels of 0.0 and for cases of error level 0.1 where Marxan saw less than 125 species. Performance fell off steeply for all species counts as overestimation in habitat maps went from 0.1 to 0.2. In fact, the large budget case illustrated in Fig. 6(B) was the only case in Fig. 6 where reducing the error in the input maps had a substantial effect on the Correct score. Even there, the map error had to be reduced to a level of 0.1 to get a large increase in performance.

Fig. 6(C) and (D) shows results with a restricted budget, still using normal evaluation. In Fig. 6(C), Marxan still appeared to perform perfectly for 25 and 50 species across all tested levels of uncertainty, but it no longer had enough budget to find a perfect solution for more than 50 species. In the corresponding results for the Correct maps, Fig. 6(D), the only area where Marxan got perfect performance was where the 25 or 50 rarest species were used and map error was 0.0 or 0.1. All other levels of uncertainty and species count rapidly fell off to a nearly 50% loss in Correct performance.

Finally, in Fig. 6(E) and (F) we considered this same restricted budget case but looked at the effect of evaluating performance with only the 50 rarest species in all cases. Since these evaluations were all done on the same reserve selections as the previous row of the figure, the Apparent scores along the contour for the 50 rarest species were identical in both rows. For anything other than 50 species though, this was not the case. When optimizing on the 25 rarest species, evaluating on the 50 rarest species led to roughly only 50% of the 50 rarest species getting at least one representation. Given that Marxan had no knowledge of the extra 25 species that it was now being evaluated on, that is not surprising. As in our other experiments, the Correct performance was even lower. When optimizing on more than the 50 rarest species, we found similar behavior to the normal evaluation case with limited budget. That is, there was not enough budget to get representation for all species. Interestingly, when we optimized for 75 species we actually got a better Correct result on the 0.5 uncertainty level than we got optimizing for the 50 species, even though we did worse on the Apparent scores. For uncertainty=0.5, the Apparent and Correct scores for optimizing and evaluating on 50 species were 1.0 and 0.56 while at that uncertainty level, optimizing for the 75 rarest and evaluating on the 50 rarest gave an Apparent score of 1.0 and a Correct score of 0.66.

4. Discussion

Conservation planning is inherently an exercise in tradeoffs. It can mean trading one species for another, trading biodiversity values for economic benefit, or trading more abstract concepts such as risk or transparency for apparent efficiency. It can also mean deciding between allocating resources to gather more data, improve existing data, improve models, or to purchase more land for conservation. Nearly all of these issues are excluded from systematic conservation planning exercises that do not provide for an evaluation of their outcomes in the context of the real-world issues that envelope them.

Our experiments demonstrate the utility of an evaluation framework for parameterising and comparing the behavior of conservation planning methods under many types of uncertainties to allow users to better understand and choose among the tradeoffs they are making. Experiments 1 and 2 show how a user can easily explore the difference in performance and robustness to error of multiple different methods, across a range of budget conditions. They show how a user can address the ever-present questions about the relative effectiveness of complex optimization methods and various simple heuristics (Pressey et al., 1996; Csuti et al., 1997; Rodrigues and Gaston, 2002; van Teeffelen and Moilanen, 2008). They also demonstrate how uncertainty complicates defining the “optimality” of solutions as even an algorithm that is guaranteed to find the optimal solution on Correct data is likely to be deceived by uncertain data. All four experiments show how a user can be assisted in developing empirical support for their choice of parameter values for any conservation planning method, for example, the boundary length penalty in

Marxan. All of these tasks can be performed without this evaluation framework if the user has the programming desire, but our framework significantly reduces the effort required, and therefore makes these tasks more likely to be done.

4.1. Generalization

The most important outcome from all of our experiments relates to the generalizability of results for conservation planning methods and solutions. The results in Fig. 3 show why it is so crucial to report the results of method performance on many different problems and for users not to base their actions and beliefs about a method's performance solely on small numbers of studies using specific locations and datasets. Had we generalized from the results in Fig. 2 that used only the Victorian richness distribution, we would not have predicted the performance variation shown in Fig. 3. In that example, only the richness distribution of the species was changed between problems, yet the four reserve selection methods changed behavior in completely different ways. They differed with respect to how their true performance changed as the richness distribution flattened and in how much their Apparent performance reflected their true performance. Worse yet, as all of these measurements were over just three instances of a single scenario, we cannot even generalize these relative performance values to other scenarios (different landscapes, costs, species counts, etc.).

4.2. Importance of synthetic data

One important aspect of our framework is that it allows for the use of both real and artificial data. While the ability to use real data is a necessity, the ability to use synthetic data is equally important. With synthetic data, the correct values can be defined and therefore known, giving a standard against which to measure error. Without synthetic data, we can only do the typical sensitivity analysis, that is, measure the variation in response to perturbations in inputs over what we believe to be the bounds of possible error. While extreme sensitivity of a system is not a good sign, a lack of variation does not necessarily tell us anything about error. As with typical statistical measures, we are interested in both the variance and the bias of our estimates of system performance. An excellent summary of the necessity and use of artificial data in the context of species distribution modelling can be found in Austin et al. (2006).

Rather than just comparing the apparent optimality of any of the methods, synthetic data also allows us to ask questions concerning to what extent the methods do achieve their goals with respect to their objective functions or on externally specified goals, such as carbon sequestration. The relatively simple experiments shown in Fig. 2 only reflected one kind of complication, error in the species maps, but even there we found that a method's Apparent performance can be very deceptive. An error of 0.3 in the input led to much larger errors in estimates of the necessary budget. The two best performing methods required 2.85 and 7.30 times as much budget to achieve a single Correct representation of each species as was required for a single Apparent representation. That experiment also showed that methods can have similar Apparent performance but very different Correct performances. For the Victorian distribution in experiment 1, Zonation and the Unprotected Richness heuristic reached at least one Apparent representation of every species at fairly similar budgets (0.027 and 0.043), but required very different budgets to reach the same true representation level (0.077 and 0.314). In this case, the true difference between methods for achieving one representation for each species (0.237) was nearly 15 times the Apparent difference (0.016).

The uncertainties that we have examined here fall into two different categories: uncertainties in observation and uncertainties in process. Understanding the consequences of uncertainties in observation can be achieved relatively easily by stochastically modifying existing data or by drawing data from theoretically postulated distributions. Consequently, observation uncertainties are the more frequently examined of the two types of uncertainty. Uncertainties in process however, are rarely examined, if only because the conservation process or context is not included in typical reserve selection mechanisms and therefore, difficult to vary in sensitivity tests. Examining the consequences of uncertainty in the process requires the

addition of models of the processes surrounding the implementation and lifespan of reserves (e.g., our models of land use change and metapopulation dynamics). There are situations where process uncertainty could easily be more important than observational uncertainty and ignoring process uncertainty is likely to lead to underestimation of error in the outputs of reserve selection. The importance of both observation and process uncertainty can be measured in the same way if they are incorporated in a unified framework such as we have proposed here. An even broader spectrum of epistemic and linguistic uncertainties to consider is discussed in Regan et al. (2002). Even though they are not all modelled in this framework, they should also be considered by anyone evaluating a conservation plan.

4.3. Effects of richness distribution on error and performance

Another important reason for using artificial data is that the generative models used to produce the data can control for problem characteristics that are difficult to control in real data. In Experiment 2, we showed the explanatory utility of controlling for the richness distribution of species. In that experiment, the number and relative rarities of species did not change, but as the species were more evenly distributed across the landscape, each conservation planning method's performance decreased considerably, as one might expect. Each method also had approximately the same degree of error regardless of the richness distribution, with the exception of Unprotected Richness, which had between 2 and 3 times as much error on the flat distribution as on the clumpier distributions. Based on the Apparent data alone, there was a comparatively small difference in the performance of Zonation and Unprotected Richness on any of the richness distributions, while the Correct data shows a large difference, particularly on the flat distribution. Thus, we can conclude that there are circumstances where the species richness distribution can affect both the level of performance and the degree of error. Again, it should be pointed out that we cannot generalize from this experiment and predict each method's behavior across different distributions and landscapes, as this experiment is based on a small number of examples. However, it suggests that it is an area that needs further investigation. Other authors have also demonstrated consequences of the distribution of species in different ways, particularly with respect to the nestedness of the distribution (Pressey et al., 1999). To our knowledge though, these studies have not considered the effects of uncertainty in the distributions or in their nestedness. This suggests another important direction for further research.

Even though real distributions of all species on a landscape are unlikely to be flat, there are important reasons for evaluating the performance of a reserve selection method using a flat richness distribution. The primary reason is that the set of species a user chooses to include in the optimization has no restrictions on it. Because a user may choose any subset of species to include, that set can be approximately flat, even though the set of all species is not flat. As an example, the Victorian richness distribution shown in Fig. 1(B) is far from flat, but the distribution of its 50 rarest species shown in Fig. 1(C) is almost completely flat (89.7% of the patches have the same number of species on them, that is, one species on each patch). This is another demonstration of why it is important to examine behavior across a wide range of conditions rather than implicitly generalizing from a single scenario.

4.4. Side effects of choosing which species to include in the reserve selection process

The flattening of the richness distribution is not the only example of side effects of choosing a particular set of species to include in the conservation planning process. The results shown in Fig. 6 detail the effects on just the 50 rarest species when the number of species included in the optimization was increased or reduced. In this example we see that regardless of the amount of error in the species maps, optimizing for either more or less species than the target value (50 in this case) reduced performance on the target set when there was a constrained budget. On the other hand, with the larger budget, increasing the number of species in the optimization for powerful routines like simulated annealing had less effect on the target set of species, because the method could find a solution that achieved its goal. However, this behavior depends on the level of input map error because as before, the selection method is subject to optimizing the wrong values in the Apparent

maps.

Perhaps the most important question raised by Experiment 4 is, "what are the effects of a strategy on species not included in the optimization?" Every study has its own reasons for choosing which species to include. Some typical reasons might be legislative mandate, charisma, surrogacy for other species, or just being the species with the best or only data (Lambeck, 1997; Andelman and Fagan, 2000; Lindenmayer et al., 2002; Roberge and Angelstam, 2004). The choice of a particular set of species necessarily trades their well-being for that of other species not included in the optimization. In response to this, the evaluation framework allows us to examine questions such as: What is the effect of this on other important but less charismatic species such as fungi? What is the effect on species with similar distributions to those being reserved and on those with very different distributions? The framework's ability to handle experiments across many different types of distributions will be useful in drawing general results for these types of questions.

4.5. Evaluation issues

In some sense, it is unfair to evaluate conservation planning methods on criteria such as these that are not included in their objective function. However, they are still important to evaluate because they are often the things that determine whether a solution will be useful. In cases like persistence, these criteria may be left out of the objective function because of the processing time required to estimate the persistence measures on every step of an optimization process. Similarly, estimating robustness to uncertainty and to violations of assumptions is computationally expensive and difficult to phrase concisely within an objective function. In other cases, goals may be completely outside the original intent of the planning method or be added after the plan is completed. In particular, preserving biodiversity is seldom the only goal of land managers. In a method whose objective function uses species representation, it may be important to know the solution's effect on ecosystem services such as water quality and carbon sequestration (Chan et al., 2006; Nelson et al., 2008). Few of these important extra measures and goals will ever be included in the conservation planning tool's objective function, although they may be among the most important criteria for determining whether a solution is actually used. The key point here is that there will never be just one correct way to evaluate the performance of a method or solution and consequently, no objective function can ever be completely correct. This makes it even more important to use an evaluation framework that can apply arbitrary objective functions outside the optimization process.

4.6. Dynamics

An important capability of the evaluation framework not discussed so far is in dealing with dynamic systems and temporal change. Various works have shown the consequences of mistakenly assuming that the reserve selection occurs all at once (Costello and Polasky, 2004; Meir et al., 2004; Pressey et al., 2004; Drechsler, 2005) or that habitat outside reserves is lost immediately (Cabeza and Moilanen, 2003) but these consequences are generally ignored in conservation planning methods. In Experiment 3 (Fig. 5), we have demonstrated how different kinds of dynamics can be incorporated into the evaluation of a method. For example, we have incorporated two different habitat loss models and shown the consequences of assuming all unreserved land is lost over time compared to the case where some unreserved habitat remains. The ease with which aspects such as the impact of unreserved land over time can be incorporated in the framework can increase the frequency with which they are explored by analysts and planners.

4.7. Persistence

These kinds of temporal change in the environment have significant effects on species persistence. In Experiment 3, we used species persistence estimates derived from metapopulation models as a performance measure instead of representation. Representation is often used as a surrogate for persistence, but it does not guarantee persistence, as there are many other

influential factors, such as habitat quality, habitat arrangement, and species interactions (Gaston et al., 2002; Baskett et al., 2007; Sabo, 2008). An important example of this is given in Salomon et al. (2006), where they found that in their study area, sites with high species richness did not correspond with sites that had high productivity of keystone species. Their results suggest that in some systems, using representation as the single driver for conservation may lead to poorer conservation outcomes than using measures more directly tied to ecological processes and viability.

Some reserve selection methods do attempt to directly include persistence in the problem statement and objective function to attempt to optimize it directly (Haight et al., 2002; Moilanen and Cabeza, 2002; Westphal et al., 2003; Drechsler and Burgman, 2004; Nicholson et al., 2006). We believe that this is a step in the right direction; however, for computational reasons, these approaches necessarily make use of fast methods for approximating persistence and usually ignore complexities like barriers to movement such as roads and ocean currents. The question is, does this matter? No rigorous attempt has yet been made to generalize beyond a single location and situation or to use multiple models of species persistence and real-world complications. It is important to have a framework for determining under what circumstances it is acceptable to ignore these complexities.

4.8. Future directions

The framework described here is a first step toward evaluating conservation plans but it can be enhanced in a number of ways. For example, useful improvements include the ability to use pixels instead of polygons, extension to marine models, as well as better models of error, land use change, economics, landscape structure, and species viability. Even without these enhancements there are many experiments that still need to be done with larger scale studies aimed at generalizing results about the behavior of different methods.

5. Conclusions

We have presented a structure for exploring the consequences of many simultaneous uncertainties and evaluating the outcomes of quantitative conservation planning tools. Our results clearly show that it is very difficult to reliably predict the difference between the apparent and the true performance of any conservation plan or method without testing. Taken together, the experiments and factors discussed in this paper imply the need for standard practice to incorporate evaluation of the effects of real-world complications on the behavior of a method or plan. Currently, the consequences of complications and uncertainties are difficult to evaluate and explore. We believe that one step toward making evaluation a standard practice is to have a systematic framework for performing any arbitrarily specified evaluation on a conservation plan or planning tool. Such a framework needs to be able to evaluate a method or plan with respect to both error and sensitivity and with respect to robustness to input data uncertainty, assumption violation, and mismatch to real-world processes. It also needs to be able to evaluate performance with respect to arbitrary criteria specified by the user through mechanisms such as models of a plan's effects on persistence, economics, or ecosystem services. Finally, it needs to be able to do these things in specific cases with real data and to contribute to deriving more generalizable results. The evaluation framework that we have presented here provides a first step toward making it easier to do this kind of testing.

Acknowledgments

We would like to thank the following people who provided comments on the manuscript and feedback during the development of this research: Sarah Bekessy, Simon Jones, Alex Lechner, Mick McCarthy, Leeza Pachepsky, Jai Ranganathan, Piero Visconti, Matt White, Brendan Wintle, and an anonymous reviewer. This research was funded by the Australian Research Council through Linkage Projects LP0454979 and LP0882780, and the Applied Environmental Decision Analysis research hub (through the Australian Commonwealth Environment Research Facilities programme).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecoconf.2009.04.002.

References

- Akçakaya, H., Mladenoff, D., Hong, S., 2003. RAMAS landscape: integrating metapopulation viability with LANDIS forest dynamics model. Applied Biomathematics, Setauket, NY.
- Andelman, S.J., Fagan, W.F., 2000. Umbrellas and flagships: efficient conservation surrogates or expensive mistakes? Proceedings of the National Academy of Sciences of the United States of America 97, 5954–5959.
- Araujo, M.B., Cabeza, M., Thuiller, W., Hannah, L., Williams, P.H., 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. Global Change Biology 10, 1618–1626.
- ARCUE, 2002. Public open space database for the greater Melbourne area. Australian Research Centre for Urban Ecology, Melbourne. http://www.rbg.vic.gov.au/research_and_conservation/arcue/databases/references_on_the_ecology_and_natural_resources [accessed 2nd July 2008].
- Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D., Luoto, M., 2006. Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. Ecological Modelling 199, 197–216.
- Ball, I., Possingham, H., 1999. MARXAN — a reserve system selection tool. The Ecology Centre. The University of Queensland, Brisbane. <http://www.ecology.uq.edu.au/index.html?page=27710> [accessed 2nd April 2008].
- Baskett, M.L., Micheli, F., Levin, S.A., 2007. Designing marine reserves for interacting species: insights from theory. Biological Conservation 137, 163–179. Burgman, M.A., Lindenmayer, D.B., Elith, J., 2005. Managing landscapes for conservation under uncertainty. Ecology 86, 2007–2017.
- Cabeza, M., 2003. Habitat loss and connectivity of reserve networks in probability approaches to reserve design. Ecology Letters 6, 665–672. Cabeza, M., Moilanen, A., 2003. Site-selection algorithms and habitat loss. Conservation Biology 1402–1413. Chan, K.M.A., Shaw, M.R., Cameron, D.R., Underwood, E.C., Daily, G.C., 2006. Conservation planning for ecosystem services. PLoS Biology 4, e379. Costello, C., Polasky, S., 2004. Dynamic reserve site selection. Resource and Energy Economics 26, 157–174.
- Csuti, B., Polasky, S., Williams, P.H., Pressey, R.L., Camm, J.D., Kershaw, M., Kiester, A.R., Downs, B., Hamilton, R., Huso, M., Sahr, K., 1997. A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon. Biological Conservation 80, 83–97.
- Drechsler, M., 2005. Probabilistic approaches to scheduling reserve selection. Biological Conservation 122, 253–262. Drechsler, M., Burgman, M.A., 2004. Combining population viability analysis with decision analysis. Biodiversity and Conservation 13, 115–139.
- Gardner, R.H., 1999. RULE: map generation and a spatial analysis program. In: Klopatek, J.M., Gardner, R.H. (Eds.), Landscape Ecological Analysis: Issues and Applications. Springer-Verlag, New York, pp. 280–303.
- Gaston, K.J., Pressey, R.L., Margules, C.R., 2002. Persistence and vulnerability: retaining biodiversity in the landscape and in protected areas. Journal of Biosciences 27, 361–384.
- Gaston, K.J., Rodrigues, A.S.L., 2003. Reserve selection in regions with poor biological data. Conservation Biology 17, 188–195.
- Haight, R.G., Cypher, B., Kelly, P.A., Phillips, S., Possingham, H.P., Ralls, K., Starfield, A.M., White, P.J., Williams, D., 2002. Optimizing habitat protection using demographic models of population viability. Conservation Biology 16, 1386–1397.
- Halpern, B.S., Regan, H.M., Possingham, H.P., McCarthy, M.A., 2006. Accounting for uncertainty in marine reserve design. Ecology Letters 9, 2–11.
- Hurlbert, A.H., Jetz, W., 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. Proceedings of the National Academy of Sciences 104, 13384–13389.
- Kirkpatrick, J.B., 1983. An iterative method for establishing priorities for the selection of nature reserves: an example from Tasmania. Biological Conservation 25, 127–134.
- Lambeck, R.J., 1997. Focal species: a multi-species umbrella for nature conservation. Conservation Biology 11, 849–856.
- Lindenmayer, D.B., Manning, A.D., Smith, P.L., Possingham, H.P., Fischer, J., Oliver, I., McCarthy, M.A., 2002. The focal-species approach and landscape restoration: a critique. Conservation Biology 16, 338–345.
- Margules, C.R., Pressey, R.L., 2000. Systematic conservation planning. Nature 405, 243–253.
- McCarthy, M.A., Thompson, C., 2001. Expected minimum population size as a measure of threat. Animal Conservation 4, 351–355.
- Meir, E., Andelman, S., Possingham, H.P., 2004. Does conservation planning matter in a dynamic and uncertain world? Ecology Letters 7, 615–622.
- Moilanen, A., Cabeza, M., 2002. Single-species dynamic site selection. Ecological Applications 12, 913–926.
- Moilanen, A., Franco, A.M.A., Eary, R.I., Fox, R., Wintle, B., Thomas, C.D., 2005. Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. Proceedings of the Royal Society B 272, 1885–1891.
- Moilanen, A., Kujala, H., 2006. Zonation Spatial Conservation Planning Framework and Software v. 1.0, User Manual, Helsinki. <http://www.helsinki.fi/bioscience/consplan/> [accessed 4th June 2008].
- Nelson, E., Polasky, S., Lewis, D.J., Plantinga, A.J., Lonsdorf, E., White, D., Bael, D., Lawler, J.J., 2008. Efficiency of incentives to jointly increase carbon sequestration and species conservation on a landscape. Proceedings of the National Academy of Sciences 105, 9471–9476.
- Nicholson, E., Westphal, M.I., Frank, K., Rochester, W.A., Pressey, R.L., Lindenmayer, D.B., Possingham, H.P., 2006. A new method for conservation planning for the persistence of multiple species. Ecology Letters 9, 1049–1060.
- O'Hanley, J.R., Church, R.L., Keith Gilles, J., 2007. Locating and protecting critical reserve sites to minimize expected and worst-case losses. Biological Conservation 134, 130–141.
- Oetting, J.B., Knight, A.L., Knight, G.R., 2006. Systematic reserve design as a dynamic process: F-TRAC and the Florida Forever program. Biological Conservation 128, 37–46.
- Pierce, S.M., Cowling, R.M., Knight, A.T., Lombard, A.T., Rouget, M., Wolf, T., 2005. Systematic conservation planning products for land-use planning: interpretation for implementation.

- Biological Conservation 125, 441–458.
- Possingham, H., Ball, I., Andelman, S., 2000. Mathematical methods for identifying representative reserve networks. In: Ferson, S., Burgman, M.A. (Eds.), *Quantitative Methods for Conservation Biology*. Springer-Verlag, New York, pp. 291–306.
- Pressey, R.L., Humphries, C.J., Margules, C.R., Vane-Wright, R.I., Williams, P.H., 1993. Beyond opportunism: key principles for systematic reserve selection. *Trends in Ecology & Evolution* 8, 124–128.
- Pressey, R.L., Possingham, H.P., Logan, V.S., Day, J.R., Williams, P.H., 1999. Effects of data characteristics on the results of reserve selection algorithms. *Journal of Biogeography* 26, 179–191.
- Pressey, R.L., Possingham, H.P., Margules, C.R., 1996. Optimality in reserve selection algorithms: when does it matter and how much? *Biological Conservation* 76, 259–267.
- Pressey, R.L., Watts, M.E., Barrett, T.W., 2004. Is maximizing protection the same as minimizing loss? Efficiency and retention as alternative measures of the effectiveness of proposed reserves. *Ecology Letters* 7, 1035–1046.
- R Development Core Team, 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Ray, N., Burgman, M.A., 2006. Subjective uncertainties in habitat suitability maps. *Ecological Modelling* 195, 172–186.
- Regan, H.M., Colyvan, M., Burgman, M.A., 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications*, 12, 618–628.
- Roberge, J., Angelstam, P., 2004. Usefulness of the umbrella species concept as a conservation tool. *Conservation Biology* 18, 76–85.
- Rodrigues, A.S.L., Gaston, K.J., 2002. Optimisation in reserve selection procedures—why not? *Biological Conservation* 107, 123–129.
- Rodrigues, A.S.L., Gregory, R.D., Gaston, K.J., 2000. Robustness of reserve selection procedures under temporal species turnover. *Proceedings of the Royal Society B* 267, 49–55.
- Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H., Possingham, H.P., 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters* 9, 1136–1145.
- Sabo, J.L., 2008. Population viability and species interactions: life outside the single-species vacuum. *Biological Conservation* 141, 276–286.
- Salomon, A.K., Ruesink, J.L., DeWreede, R.E., 2006. Population viability, ecological processes and biodiversity: valuing sites for reserve selection. *Biological Conservation* 128, 79–92.
- Sarkar, S., Pressey, R.L., Faith, D.P., Margules, C.R., Fuller, T., Stoms, D.M., Moffett, A., Wilson, K.A., Williams, K.J., Williams, P.H., Andelman, S., 2006. Biodiversity conservation planning tools: present status and challenges for the future. *Annual Review of Environment and Resources* 31, 123–159.
- Strange, N., Thorsen, B.J., Bladt, J., 2006. Optimal reserve selection in a dynamic world. *Biological Conservation* 131, 33–41.
- Turner, W.R., Wilcove, D.S., 2006. Adaptive decision rules for the acquisition of nature reserves. *Conservation Biology* 20, 527–537.
- Turpie, J., 1995. Prioritizing South African estuaries for conservation: a practical example using waterbirds. *Biological Conservation* 74, 175–185.
- vanTeuffelen, A., Cabeza, M., Moilanen, A., 2006. Connectivity, probabilities and persistence: comparing reserve selection strategies. *Biodiversity and Conservation* 15, 899–919.
- van Teuffelen, A., Moilanen, A., 2008. Where and how to manage: optimal selection of conservation actions for multiple species. *Biodiversity Informatics* 5, 1–13.
- Westphal, M.I., Pickett, M., Getz, W.M., Possingham, H.P., 2003. The use of stochastic dynamic programming in optimal landscape reconstruction for metapopulations. *Ecological Applications* 13, 543–555.
- Wilson, K.A., Westphal, M.I., Possingham, H.P., Elith, J., 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation* 122, 99–112.