The value of information for spatial conservation planning

William K. Morris, Heini Kujala & Peter A. Vesk

2016-07-03

Abstract

Spatial conservation plans are typically based upon uncertain inputs and may benefit from additional data to inform them. Value of informaiton analysis indicates how beneficial new information is to a decision problem. However, the toolset of spatial conservation prioritization does not yet contain a method for assessing the value of new information to a spatial conservation plan. If the value of information were to be calculated, then any given conservation plan could be more effective, benefiting from a more optimal allocation of effort to incorporating new information. Here, for the first time we demonstrate how a formal value of information analysis can be applied to a spatial conservation plan. We show how a value of information analysis can be combined with traditional conservation planning tools to map species distributions and optimize a reserve network to protect them. We incorporate uncertainty into conservation planning with Monte Carlo sampling of the planning inputs and then test the effects of uncertainty reduction to calculate the value of additional information to a conservation plan. The impact of optimally incorporating additional information into conservation plans, will be more effective plans where additional information is beneficial, and avoiding the loss of resources to unnecessary information gathering where new data has no or little benefit to the fundamental objectives of the plan.

Introduction

Spatial conservation planning is the field of science that focuses on developing methods to select candidate sites for protection and other conservation actions (Moilanen et al., 2009). Spatial conservation plans are complex and their inputs are commonly uncertain because they are generally based on few data [**cite**]. The outcomes of spatial conservation plans could be improved if they incorporated additional information. For example, additional occurrence data could be included for species of concern to the conservation planner so that the input layers of the spatial plan were more precise. But additional information comes as at cost. And that cost may have to be traded-off against implementing the conservation plan itself. Therefore, it is crucial to measure the value of potential new information to the conservation plan. If the value of new information is too low, that is, it does not significantly improve the conservation outcome, it may be better to implement a plan based on the original information alone, with remaining resources used for direct conservation actions.

Spatial conservation planning and imperfect information

Nearly all conservation actions include a spatial component: that is, decisions about where to act. Spatial conservation planning originally focused on designing networks of conservation reserves (Kirkpatrick, 1983; Margules et al., 1988), but has since then expanded to also cover other conservation actions and multi-action planning (e.g., Kujala et al., 2015; Thomson et al., 2009; Westphal et al., 2007). An aim of conservation planning is to preserve a comprehensive, adequate and representative subset of a region's biota [**cite**] by separating its constituents from threatening processes [**cite**]. States and non-state actors practising conservation do so with limited budgets and resources [**cite**]. Also, previous conservation actions need to be accounted for to target resources where they are most critically needed. Therefore, conservation planning must be systematic (Margules and Pressey, 2000).

A typical (systematic) spatial conservation plan will assess a pool of candidate locations for reservation (or some equivalent action). The planner will either find a set of locations that maximize conservation benefits within a given budget, or prioritize locations in order to meet a conservation target as cost effectively as possible. These are called maximal benefit or minimal set problems respectively (Camm et al., 1996; Cocks and Baird, 1989).

The information needed to make a spatial conservation plan

A systematic spatial conservation conservation plan is in essence a classic decision analysis requiring optimization. Like any decision analysis it first requires a comprehensive problem definition. In defining the problem the conservation planner requires five types of information to proceed (Possingham et al., 2001).

- **Objectives**: The objective may be a target (e.g., protect 20% of the habitat of a set of species) or a goal to maximize some gain or minimize some loss (e.g. protect locations so as to minimize the average loss of habitat for a set of species.)
- **Constraints**: The bounds in which the plan operates. Including but not limited to, the spatial and temporal frame the plan will operate in, and the resources (e.g., monetary) available to the planner.
- **Actions**: The actions the planner can take (e.g., protect, not protect, restore habitat, etc.) to meet their objectives within the given constraints.
- **State variables**: The components of the system in which the plan will operate that planner seeks to effect and against which, the performance of the plan will be measured. In a spatial conservation plan the state variables are typically the distribution of a set of species the plan is seeking to protect.
- **System models**: A system model links the actions the planner will take and the state variables. With a system model the planner can predict what the outcome of any given plan will be with respect to the state variables of interest (e.g., what will be the effect of protecting certain areas on the distribution of species).

Uncertainty in spatial conservation planning.

Of the components in the list above, it is the last two, state variables and system models, where the most critical uncertainty lies. By critical, here we mean that uncertainty which could, once addressed, change the decision being made and the decision outcome [**cite**]. Here we focus on the uncertainty in state variables and leave the treatment of system model uncertainty for another forum. While in some sense, there may be uncertainty in objective, constraints and actions, these cannot be critical (in the strict sense we use the term above) as these components define the decision problem itself and thus addressing them is not directly changing the decision and its outcome, it is changing the framework under which the decision maker is operating.

State variables (typically modelled species distributions) are almost always based on imperfect knowledge. Conservation planners do not know with certainty where the species they seek to protect occur and must rely on models to predict there occurrence or abundance (Burgman et al., 2005). Such models may themselves be based on uncertain and imperfect inputs (e.g., the distribution of a species may be predicted from a climate envelope that is based on uncertain climate data) (Guillera-Arroita et al., 2015). Uncertainty also arises in state variables, such as species distribution maps, because the models used to build them are trained with a sample of data points. For example, a common approach to predicting the distribution of species is to use so called presence-only species distribution models. In such models, the environment of the locations where a species is known to occur, is compared to the environment overall. Ignoring for a moment any uncertainty in the nature of the environment, the fewer locations a species is known to be present at, the more imprecise and uncertain the predictions of its overall distribution from such models will be.

State variable uncertainty matters to conservation planning because the uncertainty will propagate from inputs all the way through the planning algorithm, to the output and then to the conservation plan. This happens regardless of whether or not the uncertainty is accounted for explicitly.

Accounting for uncertainty in spatial conservation planning

Ultimately, a conservation planner, like any decision maker, can do one of two things in the face of uncertainty. They can make a decision (formulate a plan) with the uncertainty or try to reduce the uncertainty. Even if they take the second option, uncertainty is rarely completely resolved and the plan must be made with imperfect information. More often than not, spatial conservation planning is done in spite of uncertainty rather than by taking any uncertainty into account [**cite**]. Uncertainty, whether or not explicit acknowledged, is not often quantified for state variables or any of the other decision components. For example, a set of predictive species distribution maps will be produced, on top of which a spatial plan is built. However, only one map per species is typically produced and these maps will be implicitly treated as the true state of the system. In reality, the maps represent one possible, and at best, average or most likely (though typically not both and perhaps neither), version of the system state under the assumption that the data used to produce them was unbiased. If instead, a set of maps was produced that reflected the uncertainty (multiple maps for each species) then a complementary set of plans could be produced from them, that reflected the uncertainty in state variables. It is only at this point, having quantified uncertainty, that uncertainty can truly be addressed and an assessment made of whether, and/or how, to reduce it by acquiring addition additional information. To know whether or how to reduce uncertainty a conservation planner must measure the value of information.

The value of information

Value of information analysis is a tool used to quantify how much reducing the uncertainty in a predictive model is worth to a decision maker [**cite**]. The value of information is the difference between the final outcome of a decision (or plan) with or without additional information. Value, in and of itself, cannot be known in advance of any decision problem, including a spatial conservation plan, playing out. Therefore, decision theorists work with expected value. Expected value is the mean of all possible outcomes weighted by their probability of occurring. For example, if a person will earn a dollar when a fair coin toss is heads, and nothing if it is tails, the expected value of the toss is 50 cents. When making a decision with multiple alternatives, assuming the decision maker is risk neutral, it is always best to take the action that maximizes the expected value [**cite**].

The expected value with original information

The expected value (see Box 1) with original information, EVWOI, is the maximum expected value if no additional information is gathered before a decision is made. In the case of spatial conservation planning, EVWOI is the default. Typically, a conservation plan maximizes the expected value using the information at hand. For example, species occurrence data (point locations of places where each species of concern has been observed) is used to build a set of species habitat suitability maps, with a species distribution modelling algorithm such as MaxEnt (Phillips et al., 2006). A second algorithm, such as Zonation (Ball et al., 2009) or Marxan (Moilanen et al., 2009), is used to construct a plan that maximizes (or approximately maximises) the expected outcome with respect to some objective/target. The key, is that in the default case there is only one map used per species. The one-per-species map set represents the expected outcome under this level of uncertainty. Any uncertainty is averaged over and the map delineating the spatial conservation plan will indicate the expected value with original information.

The expected value with perfect information

Having perfect information is to have complete knowledge—no uncertainty. While in practice this is unlikely to ever occur, the expected value **with** perfect information (EVWPI, see Box 1 for a formal definition) is a useful construct as it allows the calculation of the expected value **of** information.

The expected value of perfect information

If the conservation planner can estimate the value of having (with) perfect information, that is knowing what it is worth to resolve all uncertainty before enacting a conservation plan, then they will have an idea of the upper bound on what they should be willing to do to reduce uncertainty. If the expected value of perfect information is relatively small then it is less likely to be worth resolving uncertainty than if the value of perfect information was relatively large.

Box 1: A simple example: value of information for a plan to protect one species given two properties

A conservation planner can afford to protect a property to help save an endangered species from extinction. Two properties are available. The planner's aim is to maximise the area of habitat protected for the endangered species. The planner is uncertain about each property's **habitat suitability**. A property's habitat suitability can be either one (suitable) or zero (unsuitable). The first property has a 50% chance of being suitable (a 50% chance that habitat suitability is one, and a 50% chance that it is zero). But the planner knows that the second property has a slightly better chance (60%) of being suitable (i.e., a 40% chance it's unsuitable). The properties are identical in all other ways. A property can be protected **with original information** (i.e., choosing a property while still ignorant of the habitat suitabilities) or **with perfect information** (i.e., knowing the actual habitat suitabilities of each property in advance). The **value** of protecting a property is its habitat suitability (i.e., the value can be zero or one). Regardless of they habitat suitability, an unprotected property has no value to the planner. The conservation planner must decide which property to protect.

Expected value

The **expected value** (EV) is the value the planner expects (but not necessarily what they get), given the outcome of their decision is uncertain. An **expected value** is a value weighted (multiplied) by a probability (or, in the case of a random, variable a probability distribution function). Before they make a decision, the planner can work out the **expected value with original information** (EVWOI) and the **expected value with perfect information** (EVWPI). The **expected value of perfect information** (EVPI) is the difference between the EVWPI and the EVWOI. If the EVPI is positive, it means it is worth (up to a point) learning before making a decision. For this example, if the EVPI is large enough, it is worth the planner working out what the habitat suitabilities of the two properties are, before they decide which one to protect. But if the EVPI is too low (when there is little difference between EVWOI and EVWPI) they should just decide which one to protect straightaway.

Expected value with original information

The EVWOI is the highest value the planner expects they could get, on average (the **maximum expected value**), by protecting a property with only the information they have on habitat suitability at hand. To work out the expected value of protecting a property, the planner weights (i.e., multiplies) the values that are possible by their probabilities and adds them together.

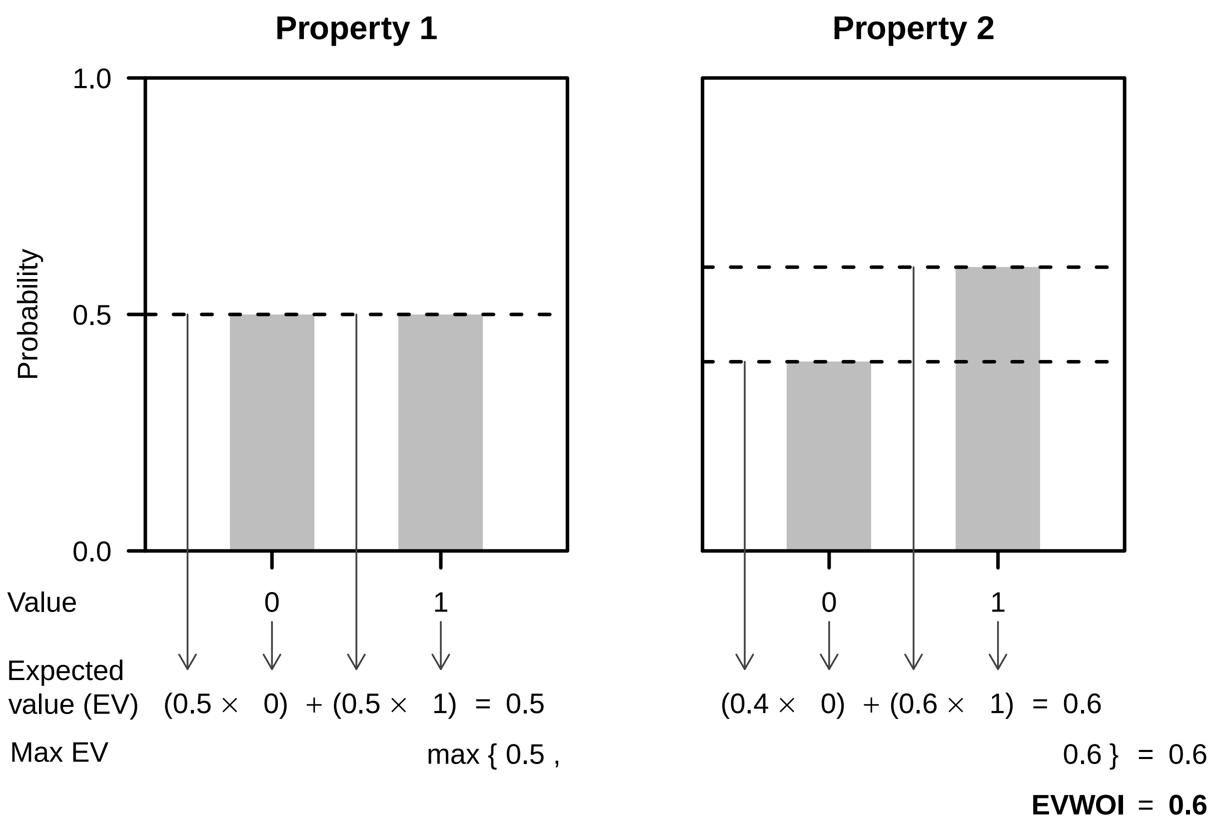


Figure 1: How to calculate EVWOI. First calculate the expected value of choosing to protect each property separately. The expected value of protecting a property is the values (the grey bars) that are possible, multiplied by their respective probabilities (the height of the grey bars) and then added together. EVWOI is the greatest (maximum) of the expected values. In this case, the EVWOI is the expected value of protecting property 2 is **0.6**.

In this case, protecting the first property has a 0.5 probability of having a value of 1, and a $1 - 0.5 = 0.5$

probability of having a value of 0. So, the expected value is:

$$1 \times 0.5 + 0 \times (1 - 0.5) = 0.5.$$

The calculation for the second property is:

$$1 \times 0.6 + 0 \times (1 - 0.6) = 0.6,$$

meaning the maximum expected value, EVWOI, is also **0.6**. The calculation can be expressed as:

$$\text{EVWOI} = Max_a[Mean_s(Value)]. \tag{1}$$

Where a represents the **action** taken by the conservation planner, s represents a **state** (or scenario) (i.e., in this case it describes the information the planner has on the properties' habitat suitabilities). Taking the mean (average) for the states gives us the expected values.

Expected value with perfect information

The EVWPI is the value the planner expects if they knew the properties' habitat suitabilities before deciding which one to protect.

In this case, there are four possible scenarios:

1. both properties are suitable,
2. the first is suitable while the second is not,
3. the second is suitable while the first is not, and
4. neither is suitable.

Because the planner is only protecting one property, the value the planner gets from any one of these scenarios is the highest of the values of the two properties for that scenario (i.e., the value for scenarios 1, 2 and 3 is

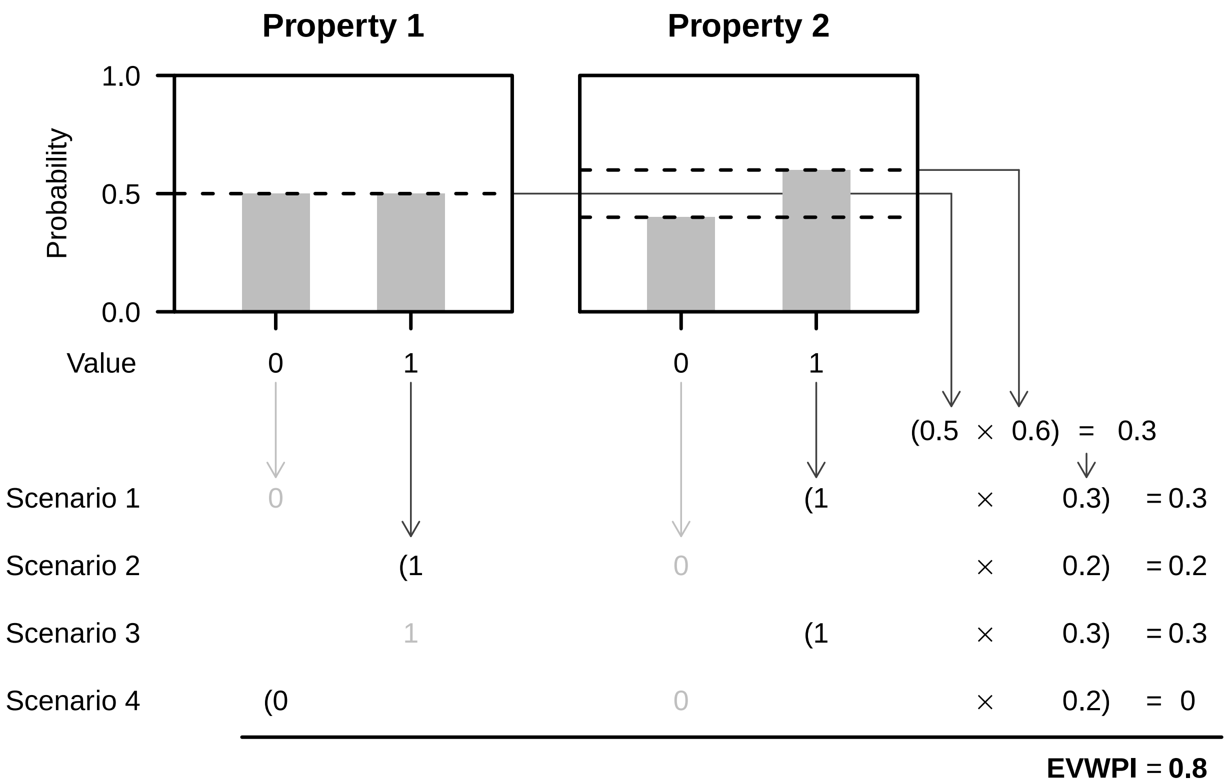


Figure 2: How to calculate EVWPI. First work out the scenarios that are possible. Then choose the property to protect in each possible scenario. When the properties have equal value it doesn't matter which one is selected (e.g., in scenarios 3 & 4). Next calculate the probability that the scenario occurs by multiplying the probabilities of the property habitat suitabilities (e.g., for scenario 1, the probability is $0.5 \times 0.6 = 0.3$, see the rightmost arrows). Next multiply the scenario probabilties by the value of the properties selected in each scenario. Finally, sum the weighted scenario values (rightmost values). In this case the EVWPI is **0.8**.

one, since there is at least one suitable property, while the value for scenario 4 is zero, since neither property is suitable). To work out the EVWPI, the planner takes the four maximum values (one for each scenario), and sums them, weighted by the scenarios' respective probabilities. The probability of a given scenario is the probability that the first property has the suitability stated in the scenario, multiplied by the probability that the second property has the suitability stated in the scenario. For scenarios 1 and 3, where the second property is suitable, the probability of the scenario is $0.5 \times 0.6 = 0.3$, irrespective of whether the first property is suitable or not. This is because the probability of the first property being suitable is the same as the probability that it is unsuitable (i.e., 0.5). Similarly, for scenarios 2 and 4, where the second property is unsuitable, the probability is $0.5 \times 0.4 = 0.2$. So, the weighted values for scenarios 1 to 4, respectively, are:

$$\begin{aligned}
&1 \times (0.5 \times 0.6) = 1 \times 0.3 = 0.3, \\
&1 \times (0.5 \times 0.4) = 1 \times 0.2 = 0.2, \\
&1 \times (0.5 \times 0.6) = 1 \times 0.3 = 0.3, \text{ and} \\
&0 \times (0.5 \times 0.4) = 0 \times 0.2 = 0.
\end{aligned} \tag{2}$$

The EVWPI is the sum of these weighted values, $0.3 + 0.2 + 0.3 + 0 = \mathbf{0.8}$. In mathematical notation this can be expressed as:

$$\text{EVWPI} = Mean_s[Max_a(Value)]. \tag{3}$$

Here the symbols have the same meaning as in equation 1. But notice that instead of calculating the **action that maximises the expected value** as in EVWOI, EVWPI is the **expected maximum value of action**. In other words, the order of maximisation and expectation has been reversed.

Expected value of perfect information

As noted above, the expected value of perfect information (EVPI) is the difference between the EVWPI and the EVWOI. Combining equations 1 and 3, the EVPI can be expressed as:

$$\text{EVPI} = \text{EVWPI} - \text{EVWOI}. \tag{4}$$

In the conservation planner's case, EVPI is $0.8 - 0.6 = \mathbf{0.2}$, meaning that if they could express habitat value as money, they should be willing to spend up to 20% of the price of a property, but no more, on learning about habitat suitability, before they decide what to protect, because beyond that, the expense of gathering the information is greater than the expected benefit from acting on the accumulated information (new information plus original).

As indicated above, spatial conservation planners rarely explicitly address uncertainty in state variables. This presents a problem, as without measuring uncertainty a conservation planner cannot know whether uncertainty is worth addressing. Without doubt there is a motive to reduce uncertainty in general, as decisions made with less uncertainty, all else being equal, will be better ones than decisions made with relatively more uncertainty [**cite**]. In light of these facts we propose that the field of spatial conservation planning should absorb the decision theoretic tools of value of information analyses. However, introducing a new tool into to an established framework is by no means trivial. As such, here we seek to incorporate the concepts of value of information in harmony with the norms of systematic spatial conservation planning. In doing so, we outline what we think is the first example of a robust and comprehensive method of calculating the value of information for a spatial conservation plan for the first time. Our approach in the following work has to been to balance simplicity with realism. While our central case study is somewhat contrived, it uses real (not simulated) data in a plan to protect species of conservation concern in a region in need of systematic planning (Whi, 2014; Kujala et al., 2015). To perform our analysis we combine the use of established software packages MaxEnt (Phillips et al., 2006; Phillips and Dudík, 2008) and Zonation (Moilanen et al., 2009), which are well known to conservation planners, with Monte Carlo resampling methods and a value of information analysis.

The rest of this work is organized as follows. We introduce the case study briefly and then demonstrate how, within the context of spatial conservation planning, the value of information can be calculated using Monte Carlo methods. To aid understanding we interweave the case study with toy, low resolution examples (such as Box 1 above) so that reader may gain a deeper understanding of the method we are proposing.

Case study: a spatial conservation plan for the Hunter region, NSW, Australia

To demonstrate how to incorporate the value of information in a systematic spatial conservation plan, we now turn to a case study on prioritizing the Hunter region for the conservation of threatened plants and animals. For the case study we make the simplifying assumptions that the entire region is an original position where no area is protected but the entire region is available for protection in a conservation plan. While this is entirely unrealistic, it would be unnecessary to complicate the demonstration with a more realistic scenario as the results are no less general having simplified the problem and added detail would only serve to distract the reader from the key components of the method we outline here.

Study area

The Hunter is a biodiverse region of north-eastern New South Wales, Australia. The region is home to many threatened species of plants and animals. There are multiple threats to biodiversity in the region. The Hunter is under active development and the area's land users utilize it's resources for mining, agriculture, transport, urban infrastructure and conservation [**cite**]. For the the analyses we present here, we consider the Hunter region to include the local government areas of Cessnock, Dungog, Gloucester, Gosford, Greater Taree, Great Lakes, Lake Macquarie, Maitland, Musselbrook, Newcastle, Port Macquarie-Hastings, Port Stephens, Singleton, Upper Hunter and Wyong, an area of 38,296 km^2.

Study species

The Hunter Valley is home to many species of national conservation significance. Here we consider six species: two birds, two mammals and two plants. For the following analyses we build conservation plans that aim to maximise the average relative carrying capacity of these six species across the hunter region. Table 1 outlines these six species and an estimate of their maximum carrying capacity (see supplement for more details) as well as their conservation status according to the *NSW Threatened Species Conservation Act 1995* (TSC), *Commonwealth Environment Protection and Biodiversity Conservation Act 1999* (EPBC) and the IUCN Red list (IUCN).

Common Name	Scientific Name	$\bar{K}^{\max}$	TSC	EPBC	IUCN
Powerful Owl	*Ninox strenua*	0.1	V	-	LC
Spotted-tailed Quoll	*Dasyurus maculatus*	0.2	V	E	NT
Squirrel Glider	*Petaurus norfolcensis*	150.0	V	-	LC
Regent Honeyeater	*Anthochaera phrygia*	200.0	CE	CE	CE
Bynoe's Wattle	*Acacia bynoeana*	250.0	E	V	-
Charmhaven Apple	*Angophora inopina*	18000.0	V	V	-

Table 1: Species used in the conservation plan for the Hunter region. LC = Least Concern; NT = Near Threatened; V = Vulnerable; E = Endangered; CE = Critically Endangered; - = Not Listed. $\bar{K}^{\max}$ = Estimated maximum carrying capacity (number of individuals per square kilometre).

Box 2: A slightly less simple example: value of information for a plan to protect one species at two properties with continuous uncertainty

In Box 1 we demonstrated how to calculate the value of information when the uncertainty in value was discrete (the value of a property could be zero or one because the value was derived from the speices being present or absent). In the following example we increase the complexity slightly and demonstrate how to calculate EVPI when uncertainty is continuous. In all other aspects, the problem remains the same. A planner has the budget to protect one property for the conservation of an endangered species and there are two properties available.

Again, the planner's aim is to maximise the area of habitat protected. And again, the planner is uncertain about the habitat suitability of both properties. This time however, the true value and uncertainty of habitat suitability is continuous. Now the planner thinks that the habitat suitability of both properties can be **any value** between zero and one, whereas before the planner thought the value could be **either** zero or one. For the continuous case the habitat suitability can be described by a continuous probability distribution function (PDF). The planner uses a different PDF to describe the uncertainty in each property's value (Figure 3). For property 1 the PDF is symmetrical indicating that planner thinks it is equally probable that the value is less than .5 as it is probable that the value is greater than .5. In contrast, the planner believes that property 2 has a value that is more likely to be greater than .5 than not. Therefore, the PDF describing the value of property 2 is shifted to the right, having a greater mass over values between .5 and 1 than over values between 0 and .5.

Expected value with original information

With these two PDFs, the planner can calculate the EVWOI. Again the planner needs to find the expected value of purchasing each property. The greater of the two, the maximum expected value, is the EVWOI. To find the expected value of a property's PDF the planner must integrate the PDF over the range of possible values. In Figure 3. we demonstrate how this is done using Monte Carlo integration. This yields expected values of .52 and .62 for properties 1 and 2 respectively. Applying equation 1 as in Box 1 we estimate the

EVWOI is .62 and the optimal action for the planner would be to purchase and protect property 2.

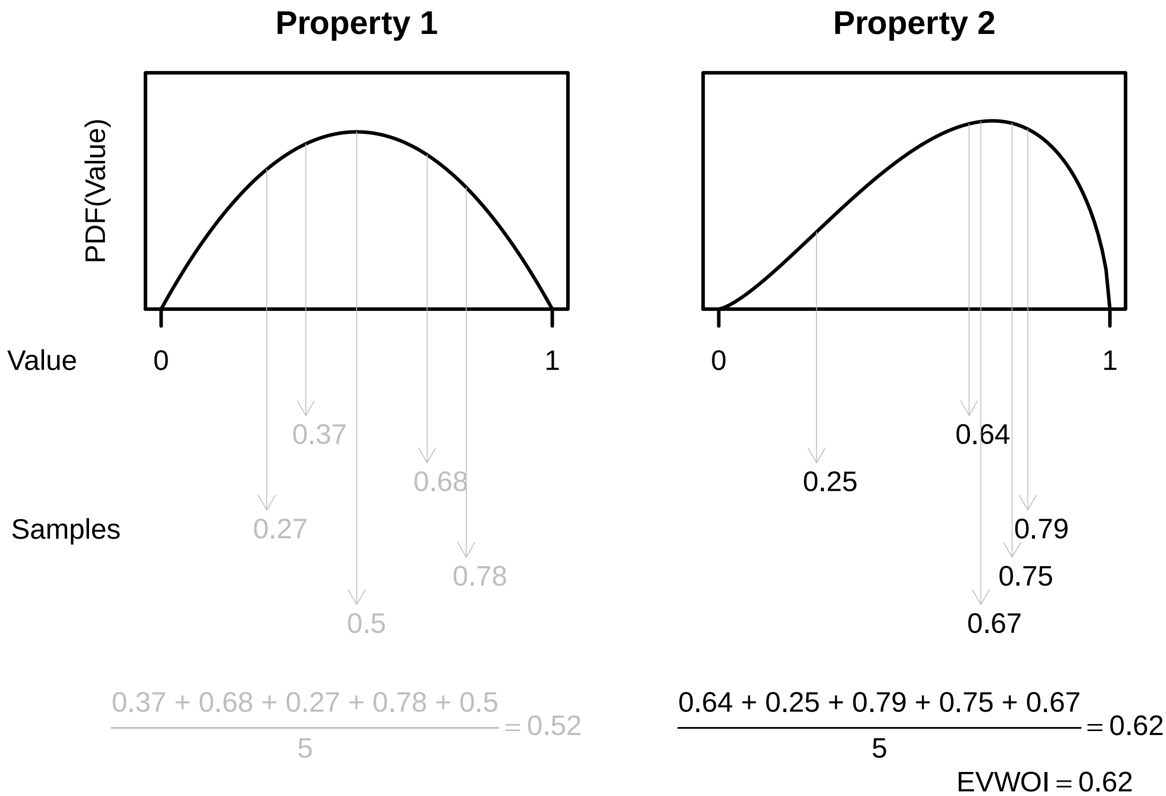


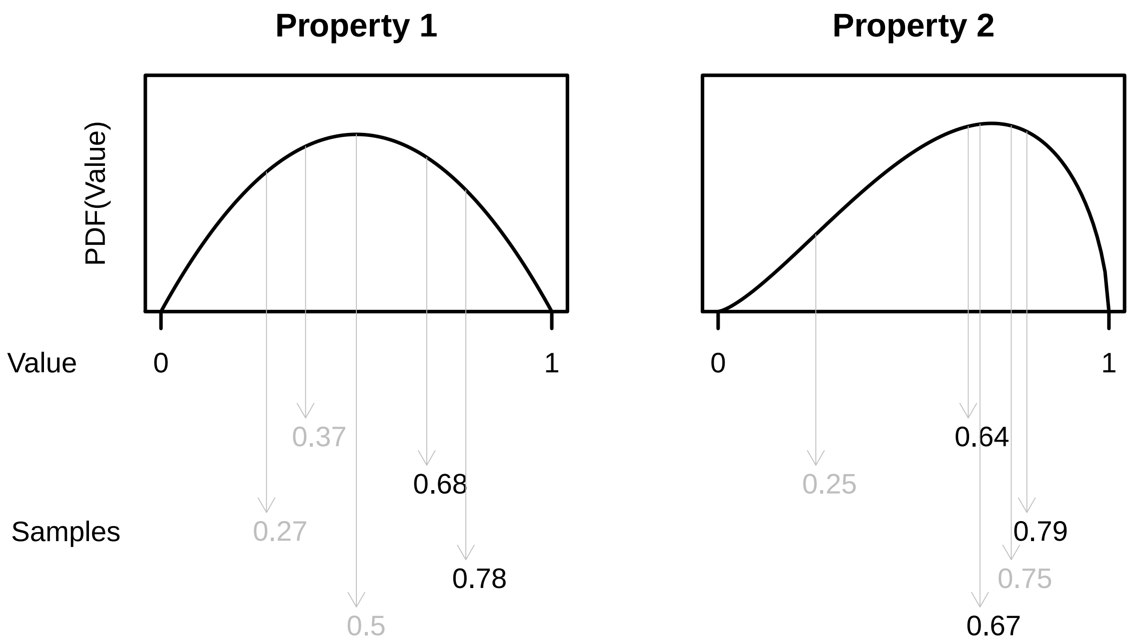
Figure 3: How to calculate EVWOI with continuous uncertainty using Monte Carlo integration. Like in Figure 1, EVWOI is the maximum expected value of the two properties. To calculate these expected values we generate samples from the respective PDFs and divide by the number of samples. By generating samples with probability according to their PDFs, as the number of samples increases, the estimate of expected value increases in accuracy. With 5 samples the estimates are 0.62 and 0.52 (note that for property 2, which has an asymmetrical distribution, the expected value, its mean, is a little lower than its most probable value, the mode). Therefore the estimate of EVWOI is 0.62, the greater of the two.

Expected value with perfect information

Figure 4 demonstrates the application of Monte Carlo integration to calculating EVWPI. Here the samples generated are the same as in figure 4 (but they need not be). However this time we generate them as pairs and as in the calculation of EVWPI in Box 1, we maximise first on each pair and then take the average (mean) at the end to arrive at the expected value, which is our estimate of EVWPI. In essence to calculate EVWPI when uncertainty is continuous the planner can calculate the expected values of sets (in this case pairs) of Monte Carlo samples from the distributions characterising uncertainty and then average them much like they do the "scenarios" of the discrete example in Box 1.

Expected value of perfect information

Once again we use equation 4 to calculate EVPI. In this case EVPI is $0.71 - 0.62 = \mathbf{0.09}$



$$\text{EVWPI} = \frac{0.64 + 0.68 + 0.79 + 0.78 + 0.67}{5}$$
$$= 0.71$$

Figure 4: How to calculate EVWPI with Monte Carlo integration. For continuous uncertainty, unlike discrete uncertainty the number of possible outcomes/scenarios are infinite. Any combination of values between 0 and 1 for each property (though some values are more likely than others). As in Figure 3, a convienent solution is to use Monte Carlo integration. We can estimate EVWPI by generating random samples in pairs, one sample from each of the property value PDFs, selecting the maximum sample in each pair (samples in black text) and then averaging by dividing by the number of samples generated. By selecting each pair in proportion to the probability distribution functions (PDFs) as we take more Monte Carlo samples the estimate of EVWPI approaches its true value. With the 5 samples above we estimate an EVWPI of 0.71 (which is close to the true value of .69, which is what we would get if we used a large number of samples.

Input data for the conservation plan

Predictors of species distributions

We summarised the environment of the Hunter region with six data layers: annual mean solar radiation, annual mean temperature, annual precipitation, precipitation seasonality (coefficient of variation), inherent soil fertility, and topographic wetness index. Each layer is a 297 by 324 grid of 1 km^2 cells. We chose this set of variables as they are publicly available (see supplement for sources), are biologically plausible drivers of the distribution of many taxa, have previously been shown to predict the distributions of the study species in the region [**cite**] and are relatively uncorrelated with one another (maximum Pearson correlation coefficient = 0.54).

Species occurence data

We obtained 30 random occurrence records within the boundaries of the Hunter region (as defined above) and collected within the date range, January 1, 1996 to May 1, 2016, for each of the six study species from the Atlas of Living Australia database [**cite**]. We chose this relatively low sample-size random subset to ensure that subsequent modelling would have an initial level of uncertainty large enough for us to demonstrate how to calculate the value of information.

Background geographic data

To fit distribution models with the occurrence records and environmental predictors, we used a set of background points selected so as to minimise sampling bias in the occurrence points. For each higher taxonomic group (birds, mammals and plants), the background points were selected by randomly sampling 10,000 occurrence records of all taxa belonging to each group from the Atlas of Living Australia database for the same spatial and temporal extent indicated above.

Modelling species distributions using MaxEnt

We used the software MaxEnt (Phillips et al., 2006; Phillips and Dudík, 2008) to create species distribution maps on which to base the conservation plan. MaxEnt can be used to describe the potential distribution of species with occurrence records alone. The algorithm does this by comparing the distribution of occurrence records in covariate space to the distribution of covariate space as a whole, also known as the background (see below). We fitted MaxEnt models to each species with only linear and quadratic features enabled.

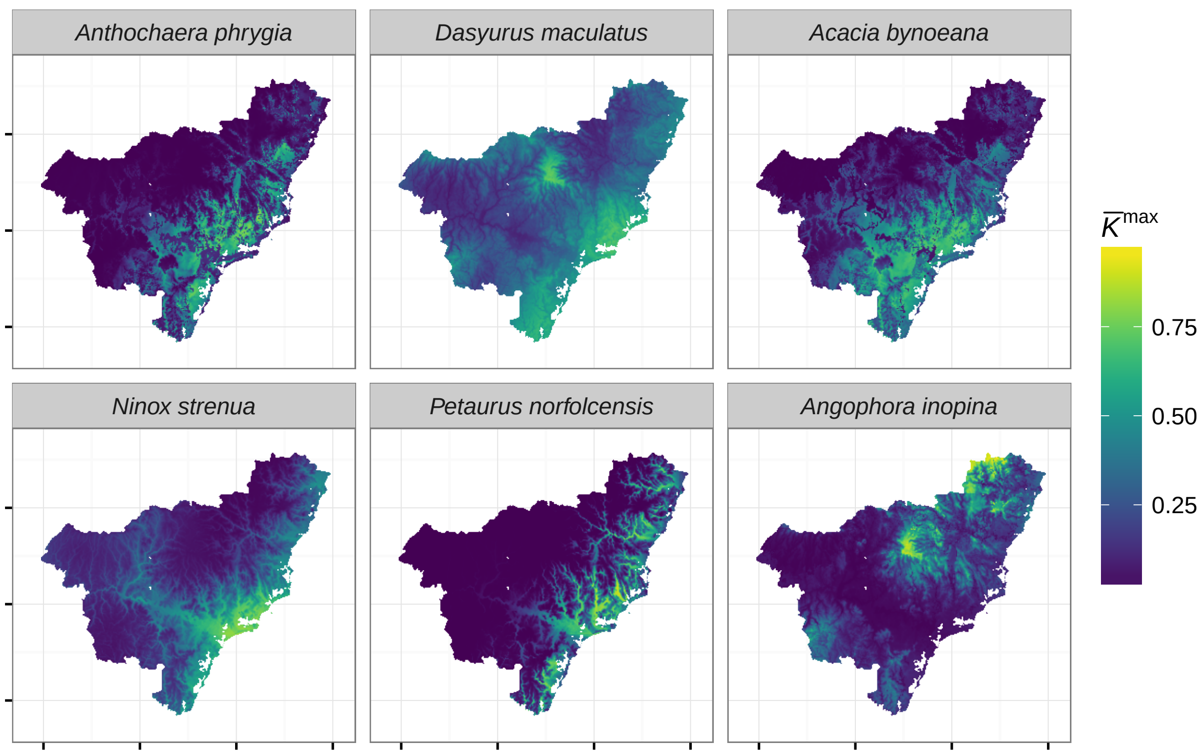


Figure 5: Distribution of study species in the Hunter region. Maps show the relative carrying capacity (logistic output) of each species estimated by fitting Maxent models. .

Intepreting the output of MaxEnt

Interpreting MaxEnt output in both it's native formats (raw and logistic) has, in the past, proved problematic and controversial. It has been shown that it while it is equally problematic to interpret MaxEnt output as either relative occurrence rate or probability of presence, it is legitimate to interpret MaxEnt output as an indicator of relative habitat suitability [**cite**]. Here we chose to interpret the logistic output of MaxEnt as a rough indicator of relative carrying capacity. Where the logistic output is zero, the carrying capacity should be zero (or close to zero) and where the output is one, the carrying capacity should be close to the maximum attainable for the species. When the output is some fraction of one, then we assume the carrying capacity is that fraction of the maximum. The important caveat here is that the relationship between MaxEnt logistic output and carrying capacity is assumed linear. While there is no prima facie reason that it should be, equally there is no prima facie reason there should be some more complicated relationship. Therefore, linearity would seem to be the most logical initial assumption, in the absence of information to the contrary. In any case, any conceivable deviation from linearity should not matter too much here, as the species of interest have carrying capacities that vary over many orders of magnitude (Table 1). The benefit of interpreting the MaxEnt output in such a way is that we can use the modelled distribution maps to calculate an estimate of the total potential population size for each species. And subsequently once we have a conservation plan, we can calculate the potential impact on that population size of the plan, or other competing plans.

Incorporating uncertainty

To calculate the value of information requires some means of quantifying the level of uncertainty in a prediction of the system state. One way to express the uncertainty in species distributions is to consider a set of distribution maps where each map is equally likely to represent the true distribution, instead having a single map that (alone) describes the distribution of a species. The multi-map description of species distributions acknowledges that species distributions are modelled based on a sample of points and those points are used to estimate the species relationship to the environment. As the points are a sample, the estimated relationship will be inherently uncertain.

Uncertainty and MaxEnt

Fitting a MaxEnt model with a set of occurrence points does not by itself account for uncertainty in a species' distribution. It provides a single possible realization (even if it is a likely, or the most likely one) of a species distribution and does not account for the fact that the distribution was estimated from a sample of occurrence points and is therefore inherently uncertain. MaxEnt does not natively account for uncertainty. That is, at the time of writing, the MaxEnt software does not provide a means of quantifying uncertainty in its output, and only provides a single map for a single sample of species occurrence points and background dataset.

The bootstrap

To overcome the shortcomings in the MaxEnt software's ability to incorporate uncertainty we used a statistical method known as the bootstrap [**cite**]. At heart the bootstrap is a simple technique that can be used to describe uncertainty in quantities that have been generated by a model. Many improvements and extensions have been made to the bootstrap since its introduction, but here, for simplicity's sake, we focus on and employ the basic bootstrap. In essence the bootstrap quantifies uncertainty by refitting a model to a dataset that has been resampled with replacement. For example, if we fit a model to a dataset consisting of three data-points labelled A, B and C (note for the purposes of the bootstrap it is unimportant how many variables they consists of and what value(s) the data-points take), there are ten bootstrap samples to fit models to (AAA, AAB, AAC, ABC, ABB, ACC, BBB, BBC, BCC and CCC), and ten possible outputs that describe the uncertainty of the model fitted to the data-set. However, as the sample-size increases the number of possible bootstrap samples increases rapidly (e.g., for the sample-size of only 30 used in this case study there are almost 6×10^{16} different possible bootstrap samples) so in practice a random (or Monte Carlo) set of n bootstrap samples is used to estimate the bootstrap quantities' uncertainty.

To bootstrap a species distribution model using MaxEnt we resampled the 30 occurrence points 1000 times, with replacement and ran the MaxEnt algorithm on each set to produce 1000 distribution maps. We performed this Monte Carlo bootstrap analysis on a MaxEnt model for each of the six focal species, to produce a total 6000 distribution maps (6 times 1000).

Prioritizing with Zonation

We used the spatial planning software Zonation (Moilanen et al., 2009) to make a conservation plan for the Hunter Region. The conservation plan we implemented had a budget large enough to protect 10% of the Hunter region (with the simplifying assumption that all of the Hunter is available to be protected, and that any 10% of region will have a cost equal to the budget). The objective of the conservation plan was to minimise the average proportional loss (maximise the average proportion remaining) of the total carrying capacity of the six species of interest. We gave equal weight to each species (no species was preferred over any other).

The Zonation algorithm

The Zonation software uses a greedy algorithm to rank grid cells in order of priority such that preserving the highest ranked 10% of grid cells is the approximately maximal solution to the objective above. The Zonation algorithm works by iteratively removing grid cells that contribute the least to the objective. The removal process is repeated until all cells have been removed and the order of removal constitutes the ranking of the cells (the last cell to be removed being the highest ranked and the first cell removed is the lowest ranked). Zonation provides a number of implementations of its greedy algorithm that differ by the way they calculate the contribution of each cell to the objective (the basis on which cell removal is determined). Here we use Zonation's "additive benefit" cell removal rule, whereby cells that contribute the least to the sum of the proportion of value remaining, are removed first.

Box 3: A simple example of calculating the value of information for a conservation plan using Zonation

In Box 2 we demonstrated how to calculate EVPI for as two site conservation plan with continuous uncertainty. Here we extend that problem to 25 sites and two species demonstrate how the problem can be solved using Zonation on random (Monte Carlo) bootstrap samples of species distributions. The same method we apply here can be used for much larger numbers of sites and species and is only limited by computing resources.

Here again, like in Boxes 1 and 2, the planner's aim is to maximise the quality of habitat protected given a limited budget that is large enough to purchase a single property. This time however, there are 25 properties to chose from and two endangered species the planner is entrusted to protect. The planner's aim is to maximise the average (over the two species) proportion of habitat quality remaining after a site has been protected and the remaining 24 have been lost.

As before, the planner is uncertain about the habitat quality of the 25 sites, and in this case they are uncertain about the habitat quality for both species. The uncertainty in site habitat quality is continuous as in Box 2. In this case habitat quality is measured as the percentage of maximum achievable carrying capacity at a site. For both species the maximum carry capacity of a site is 100, meaning that a site can at most accommodate 100 individuals and a site where the habitat quality is 50% could sustain 50 individuals.

The planner has models that predicts habitat quality for both species. With their model and the computing resources available they are able to produce two bootstrap samples (k_1 and k_2) of each species modelled distribution. These bootstrapped distribution maps represent the planner's uncertainty of the system state (Figure 6, bottom left quadrant) and with them they can calculate the EVPI.

Expected value with original information

To calculate EVWOI the planner must calculate the expected value of each site for each species. This amounts to averaging the bootstrapped maps k_1 and k_2 (Figure 6, top left quadrant). With an bootstrap average map for each species the planner can then find the maximum expected value by building a conservation plan using Zonation. Applying the Zonation algorithm to the two bootstrapped average maps, the planner finds that

E[K]

Species 1

20	76	17	34	72
38	76	47	34	64
42	64	77	74	28
50	62	36	56	20
41	12	61	88	55

Total 1244

Species 2

26	59	51	25	47
54	18	33	86	50
27	39	71	42	40
80	78	32	60	47
72	78	52	37	20

Total 1224

Rank

25	3	21	24	7
15	14	17	6	10
19	13	1	9	20
4	2	23	8	22
12	16	11	5	18

$$\text{EVWOI} = \frac{\frac{77}{1244} + \frac{71}{1224}}{2} = 0.06$$

k_1

Species 1

10	93	22	65	58
62	82	78	40	35
9	61	87	81	44
10	50	61	33	21
33	0	61	89	52

Total 1237

Species 2

39	40	35	40	49
64	26	50	92	39
0	56	82	35	74
78	98	54	40	54
52	89	79	47	25

Total 1337

Rank

24	5	23	15	14
8	13	6	7	20
25	10	1	9	11
17	2	12	22	21
18	16	3	4	19

$$\text{EVWPI} = \frac{\frac{\frac{87}{1237} + \frac{82}{1337}}{2} + \frac{\frac{90}{1246} + \frac{81}{1108}}{2}}{2} = 0.07$$

k_2

Species 1

30	59	12	3	85
13	70	16	28	94
75	66	66	67	12
90	74	10	79	20
48	23	61	87	58

Total 1246

Species 2

13	78	66	9	45
44	10	15	80	61
54	22	60	49	7
81	58	10	81	40
92	66	25	27	15

Total 1108

Rank

21	5	16	25	8
20	17	22	11	3
7	14	9	10	24
1	6	23	2	19
4	13	15	12	18

$$\begin{aligned}\text{EVPI} &= \text{EVWPI} - \text{EVWOI}\\ &= 0.07 - 0.06\\ &= 0.01\end{aligned}$$

Figure 6: How to calculate the value of information for a conservation plan for 2 species, 25 cells and a budget large enough to purchase 1 cell, using Zonation. First calculate the EVWOI (top row). With original information we rank each cell using the maps of expected (average of k_1 and k_2 carrying capacity (top left quadrant). For the highest rank cell (greyed out cell, top right quadrant) take the average proportion of carrying capacity remaining with all other unprotected cells removed (top equation). Next calculate EVWPI. For each set of bootstrapped species carrying capacity maps k_1 and k_2 rank each cell and calculate the average proportion of carrying capacity remaining. The EVWPI is the average of the average proportion of carrying capacity remaining for each bootstrap sample k_1 and k_2 (middle equation). Finally we calculate EVPI which is the difference between the EVWOI and EVWPI (bottom equation).

the top ranked cell (the cell that best meets the criteria of the objective above) has habitat quality scores of 77 and 71 percent for species 1 and 2 respectively. To calculate the EVWOI the planner divides these expected values by the total carrying capacity of all sites (to arrive at the proportion of carrying capacity remaining) for each species and then averages that value across species by dividing by two (Figure 6, top equation). In this case the EVWOI is 0.06, that is, relying on the original information alone, the planner would expected to preserve an average of 6% of the initial carrying capacity for both species if they protected the best site available and lost the remaining 24.

Expected value with perfect information

As in Box 2, the planner can apply the same method to the bootstrap sample maps they generated here. Again the process of taking the mean (averaging) and maximising (using Zonation) is reversed when calculating the EVWPI as opposed to EVWOI. This time the planner creates two conservation plans, one for each set of bootstrap sample distribution maps (Figure 6, bottom right quadrant). Each conservation plan gives the planner a different average proportion of total carrying capacity remaining and average over the two gives an EVWPI of 0.7.

Expected value of perfect information

As in Boxes 1 and 2 the EVPI is the difference between EVWPI and the EVWOI in this case a value of 0.01. This means that the planner would expect on average to increase the average proportion of carrying capacity remaining by 17% if they resolved all their initial uncertainty in their knowledge of habitat suitability of the two species across the 25 sites.

The expected value of information for the Hunter conservation plan

Using the methods outlined in Boxes 1, 2 and 3 and the data highlighted above, we calculated EVPI and it's components (EVWOI and EVWPI) for our conservation plan for the Hunter region.

The expected value with original information

First we calculated the EVWOI, that is, how well do we expect the conservation plan to perform (on average) when basing it only on the 30 observations per species. To calculate EVWOI we first averaged the bootstrap species distributions to produce one map per species. Using these six maps (Figure 5) we applied the Zonation algorithm to rank the grid cells of the Hunter region in order of priority. Figure 7 shows the grid cells of the Hunter ranked in their order of priority. According to Zonation, the grid cells in the top 10% are located in the hinterland of the south-east and central coast. Protecting this region of the Hunter, on average, protects 25.5% of the total carrying capacity of the six species of interest. This value is the EVWOI.

The expected value with perfect information

To calculate the EVWPI we applied the method outlined in Box 3 reversing the two steps of averaging over the bootstrap samples and applying the Zonation algorithm to find the solution that maximised (approximately) the objective of the conservation plan. This time, we first applied the Zonation algorithm to each set of bootstrap distribution maps. This resulted in 1000 conservation plans (instead of the single plan arrived at when calculating EVWOI). In each of these maps the ranking of the grid cells differs. Figure 8 shows the range of rankings each grid cell has across the bootstrapped conservation plans and highlights those grid cells that are ranked in the top 10% at least once in the bootstrapped plans. For each bootstrapped conservation plan we calculate the average proportion of total carrying capacity remaining after protecting the top 10% of cells. These 1000 values are the expected performances if we were certain of the distribution of the six species across the Hunter (with each set of bootstrapped maps and plan representing a separate realization of that certainty). Averaging these 1000 values is the final step and gives the EVWPI or the performance we

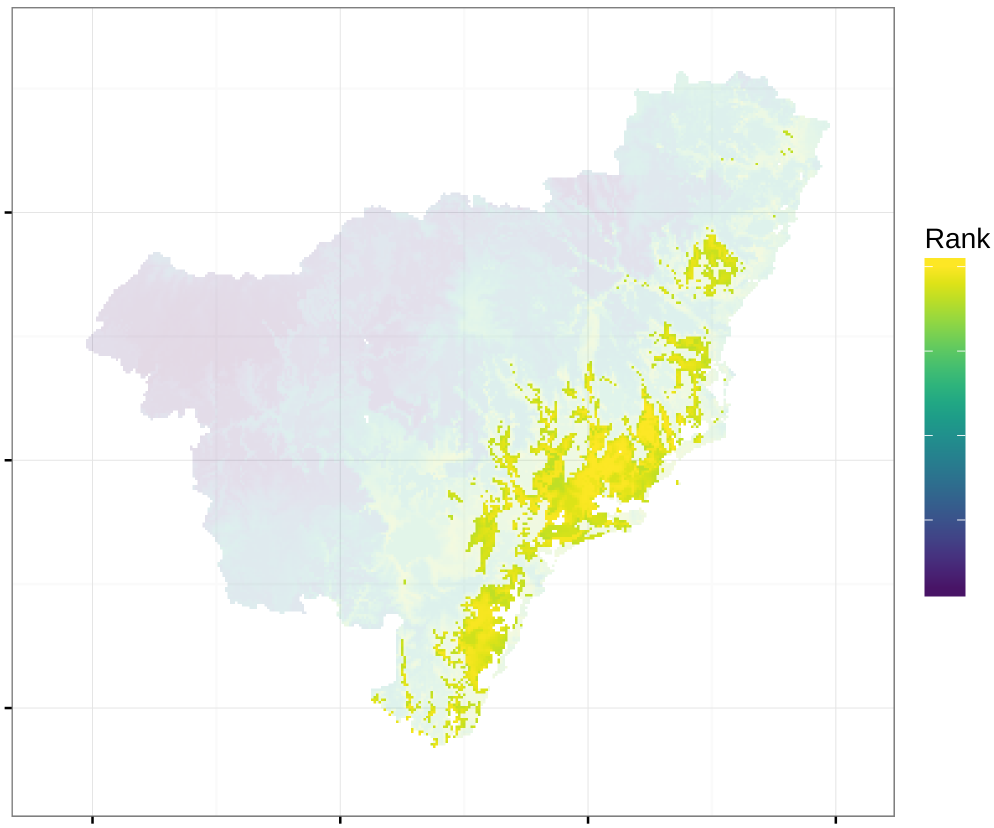


Figure 7: Priority map for the Hunter region. Grid cells have been ranked from lowest priority (blue) to highest priority (yellow) using the Zonation greedy algorithm. Grid cells ranked in the top 10% have been highlighted. The additive benefit function was used to iteratively remove cells with the lowest marginal benefit. In calculating marginal benefit each species was weighted equally. The map can be used to calculate the expected value with original information for a fixed budget (number of cells protected) with value measured as the average proportion of species total carrying capacity remaining if all unprotected cells were removed.

would expect to conservation plan to achieve on average if we had no uncertainty about the distribution of each of the six species. In this case the EVWPI is 26.5%.

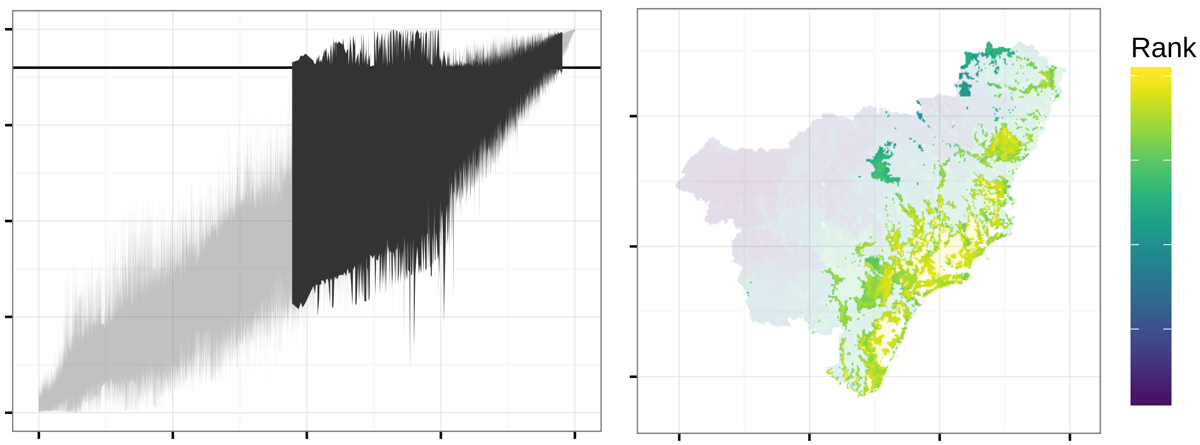


Figure 8: Uncertainty in grid cell ranks. In the left pane, grid cells have been ordered along the x-axis according to their mean bootstrap ranking. The grey shaded region indicated the range of bootstrap rankings (along the y-axis). The black horizontal line marks the top 10^{th} percentile, cells that have bootstrap rankings that span this line are shaded a darker grey. The right panel shows the same map displayed in Figure 7 however this time the cells that have bootstrap ranks spanning the top 10^{th} percentile have been highlighted.

The expected value of perfect information

Finally we arrive at the EVPI, which is simply the difference between EVWPI and EVWOI. Here EVPI is 1. This is the increase in performance we would expect to see on average if we learned everything there was to know about the distribution of the species of conservation concern in our conservation plan. To express this value in different terms, if we consider a species with a carrying capacity of 1 individual per km^2 and that species had a total carrying capacity in the Hunter of 10,000 individuals, enacting a conservation plan with the information at hand (based on 30 observations) we would expect (on average) to preserve enough habitat to accomodate 2,552 individuals. Were we to then enact a plan with complete certainty the we would expect to preserve enough habitat (though the area) to accomodate 97 more individuals.

Expected value	Average percentage of carrying capacity remaining
EVWOI	25.5
EVWPI	26.5
EVPI	1.0

Table 2: The expected value of perfect information and its components for a conservation plan for the Hunter region

References

(2014). 28(4):992–1003.

Ball, I. R., Possingham, H. P., and Watts, M. (2009). Marxan and relatives: Software for spatial conservation prioritisation. In Moilanen, A., Wilson, K. A., and Possingham, H. P., editors, *Spatial Conservation Prioritization: Quantitative Methods and Computational Tools*, chapter 14, pages 185–195. Oxford University Press, 1 edition.

Burgman, M. A., Lindenmayer, D. B., and Elith, J. (2005). MANAGING LANDSCAPES FOR CONSERVATION UNDER UNCERTAINTY. *Ecology*, 86(8):2007–2017.

Camm, J. D., Polasky, S., Solow, A., and Csuti, B. (1996). A note on optimal algorithms for reserve site selection. *Biological Conservation*, 78(3):353–355.

Cocks, K. and Baird, I. (1989). Using mathematical programming to address the multiple reserve selection problem: An example from the eyre peninsula, south australia. *Biological Conservation*, 49(2):113–130.

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., and Wintle, B. A. (2015). Is my species distribution model fit for purpose? matching data and models to applications. *Global Ecology and Biogeography*, 24(3):276–292.

Kirkpatrick, J. (1983). An iterative method for establishing priorities for the selection of nature reserves: An example from tasmania. *Biological Conservation*, 25(2):127–134.

Kujala, H., Whitehead, A., Morris, W., and Wintle, B. (2015). Towards strategic offsetting of biodiversity loss using spatial prioritization concepts and tools: A case study on mining impacts in australia. *Biological Conservation*, 192:513–521.

Margules, C., Nicholls, A., and Pressey, R. (1988). Selecting networks of reserves to maximise biological diversity. *Biological Conservation*, 43(1):63–76.

Margules, C. R. and Pressey, R. L. (2000). Systematic conservation planning. *Nature*, 405(6783):243–253.

Moilanen, A., Wilson, K. A., and Possingham, H., editors (2009). *Spatial Conservation Prioritization: Quantitative Methods and Computational Tools*. Oxford University Press, 1 edition.

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4):231–259.

Phillips, S. J. and Dudík, M. (2008). Modeling of species distributions with maxent: new extensions and a comprehensive evaluation. *Ecography*, 31(2):161–175.

Possingham, H., Andelman, S., Noon, B., Trombulak, S., and Pulliam, H. (2001). Making smart conservation decisions. In Soulé, M. E., Orians, G., and Boersma, P. D., editors, *Conservation Biology: Research Priorities For The Next Decade*, chapter 10, pages 225–44. Island Press, 1 edition.

Thomson, J. R., Moilanen, A. J., Vesk, P. A., Bennett, A. F., and Nally, R. M. (2009). Where and when to revegetate: a quantitative method for scheduling landscape reconstruction. *Ecological Applications*, 19(4):817–828.

Westphal, M. I., Field, S. A., and Possingham, H. P. (2007). Optimizing landscape configuration: A case study of woodland birds in the mount lofty ranges, south australia. *Landscape and Urban Planning*, 81(1-2):56–66.