Abstract

1. An underlying aim in conservation planning is to maximize the long-term persistence of biodiversity. To fulfill this aim, the ecological and evolutionary processes that sustain biodiversity must be preserved. One way to conserve such processes at the feature level (e.g. species, ecosystem) is to preserve a sample of the feature (e.g. individuals, areas) that is representative of the intrinsic or extrinsic physical attributes that underpin the process of interest. For example, by conserving a sample of populations with local adaptations—physical attributes associated with adaptation—that is representative of the range of adaptations found in the species, protected areas can maintain adaptive processes by ensuring these adaptations are not lost. Despite this, current reserve selection methods overwhelmingly focus on securing an adequate amount of area or habitat for each feature. Little attention has been directed towards capturing a representative sample of the variation within each feature.

2. To address this issue, we developed the raptr R package to help guide reserve selection. Users set “amount targets”—similar to conventional methods—to ensure that solutions secure a sufficient proportion of area or habitat for each feature. Additionally, users set “space targets” to secure a representative sample of variation in ecologically or evolutionarily relevant attributes (e.g. environmental or genetic variation). We demonstrate the functionality of this package, using simulations and two case studies. In these studies, we generated solutions using amount targets—similar to conventional methods—and compared them with solutions generated using amount and space targets.

3. Our results demonstrate that markedly different solutions emerge when targeting a representative sample of each feature. We show that using these targets is important for features that have multimodal distributions in the process-related attributes (e.g. species with multimodal niches). We also found that solutions could conserve a far more representative sample with only a slight increase in reserve system size.

4. The raptr R package provides a toolkit for making prioritizations that secure an adequate and representative sample of variation within each feature. By using solutions that secure a representative sample of each feature, prioritizations may have a greater chance of achieving long-term biodiversity persistence.

Keywords

biodiversity, conservation, mixed integer linear programming, optimization, protected areas
1 | INTRODUCTION

Perhaps the most fundamental aim of conservation is to maximize the long-term persistence of biodiversity (Margules & Pressey, 2000; McNeely, 1994). To achieve this, conservation actions must preserve biodiversity patterns (e.g. populations, species, ecosystems), but also crucially the processes that sustain them. One of the major tangible achievements of modern conservation has been the act of setting aside areas for preservation (Sanderson, Segan, & Watson, 2015; Dudley, Segnan, & Hackings, 2014). Reserve networks buffer species from gross threatening processes and set the stage for enhanced management interventions (Gaston, Jackson, Cantu-Salazar, & Cruz-Pinon, 2008). Since the resources available for conservation action are limited, protected area networks must be sited in places that satisfy conservation objectives for minimal cost (Margules & Pressey, 2000).

Today, the most widely used conservation planning tools focus on biodiversity patterns (Marxan and Zonation; Ball, Possingham, & Watts, 2009; Moilanen, 2007). Decision makers can use these tools to obtain solutions that secure a proportion of the geographic range of each biological feature (populations, species, or ecosystems) of interest, by setting targets. One method to incorporate data on the ecological and evolutionary processes that sustain such features is to conserve a representative sample of the internal or external physical attributes underpinning these processes. To achieve this, current methods typically involve partitioning features into sub-groups based on an attribute variable that relates to a biodiversity process of interest (Beger et al., 2014; Klein et al., 2009) or phylogenetic trees (Carvalho et al., 2017). For instance, by dividing species distributions into sub-groups according to habitat discontinuities and ensuring that each sub-group is conserved in the solution, conservation planners can obtain prioritizations that promote adaptive processes (Carvalho, Brito, Crespo, & Possingham, 2011). However, using these methods is challenging because biodiversity often cannot be divided into operational groups without information loss (Faith & Walker, 1996; Orians, 1993), and because the number of groups can substantially alter solutions (Pressey & Logan, 1994). This limitation has been known for quite some time, and dates back to some of the earliest reserve selection methods (e.g. Kirkpatrick, 1983).

To overcome this limitation, Faith and Walker (1994) developed the “environmental diversity” (ED) reserve selection framework to conserve a representative sample of the environmental variation found across a study area. Drawing inspiration from the p-median problem (Owen & Daskin, 1998), this method used Euclidean distances to express differences among candidate sites and maximize the representativeness of the solution (Faith, 2003; Faith & Walker, 1996; Faith, Ferrier, & Walker, 2004). Recent work has built on this method, using more advanced optimization algorithms (Engelbrecht, Robertson, Stoltz, & Joubert, 2016). However, existing environmental diversity based methods are not often used in conservation planning. This is, in part, because they seek to find the most representative solution that is within a budget, unlike the more commonly used decision support tools (e.g. Marxan) that aim to find the cheapest solution subject to a set of targets (or constraints).

Conservation planners lack a decision support tool that lets them set explicit targets to obtain solutions that (1) secure an adequate amount of habitat for each feature and (2) a representative sample of the variation in each feature. To begin to fill this gap, we unite the ideas underpinning environmental diversity and Marxan into new formulations of the reserve selection problem, and implement them in the raptR R package. This R package provides decision makers with the tools to generate prioritizations based on data that relate to biodiversity patterns and processes. Here, we aim to provide an in-depth understanding of this R package and explore its functionality.

2 | MATERIALS AND METHODS

2.1 | Problem formulation

Biodiversity features are defined as the entities that the prioritization is required to preserve (e.g. species, ecosystems). Each biodiversity feature may have one or more “attributes” which vary across its geographic range as a result of ecological or evolutionary processes. These attributes can be intrinsic (e.g. genetic or phenotypic) or extrinsic (e.g. environmental conditions) to the feature. The overarching goal of the prioritization is to conserve these biodiversity processes, and by capturing the variation in these species-specific attributes, the prioritization can make some progress towards achieving this. Thus there should be a reasonable underlying hypothesis that relates the attributes to the biodiversity processes that need to be conserved.

The variation in the attributes of a feature can be described using an n-dimensional “attribute space.” For example, a decision maker may require a prioritization that captures populations along climatic gradients. To achieve this, the decision maker might use a “climatic” attribute space with dimensions relating to mean annual temperature (°C) and precipitation (mm). Any given combination of temperature and precipitation may be conceived as a point in this climatic space. Although they exist as polygons, for simplicity, each planning unit may be thought to exist as a single point inside a given attribute space. By associating the planning units with climatic data, and calculating a descriptive statistic for each planning unit (e.g. mean), they can be mapped from geographic space to this climatic attribute space.

“Demand points” (Faith, 2003; Faith & Walker, 1994, 1996) are designated by the decision maker to indicate regions of the attribute space that they wish to conserve in the prioritization (see below for discussion on generating demand points for real-world datasets). For instance, by siting demand points throughout an attribute space, planners can obtain solutions that better capture existing variation. Alternatively, by siting demand points in specific regions of an attribute space, planners can obtain solutions that target specific samples of an attribute space. For a given set of demand points, the shorter the sum of the distances between the demand points and the planning units selected for prioritization: the better a solution is at securing the desired variation in the attribute space. Demand points are used heavily in the p-median and facility location problems (reviewed in Owen...
& Daskin, 1998). Additionally, different features and attribute spaces may require different demand points. For instance, values that might be valid in one attribute space (e.g., mean annual temperature \(-5^\circ\text{C}\)) may be invalid for another attribute space (e.g., mean annual rainfall \(-5\text{ mm}\)). Additionally, there may be some regions that are desirable for some features and undesirable for others (e.g., conditions known to be outside the physiological tolerance of certain species).

To illustrate these concepts, consider the following example: we wish to develop a prioritization for a single species that has four populations. Since we can only afford to preserve three of the four populations, our objective is to conserve the most representative sample of the intra-specific variation could be captured by securing populations A and B constitutes considerable redundancy by selecting populations that inhabit similar conditions. Instead, a more representative sample of the intra-specific variation could be captured by securing populations A (or B), C, and D. However, if the goal of the prioritization was to preserve populations living in warmer temperatures, then instead of siting demand points across the full range of conditions, we could site demand points in environmental conditions with temperatures over \(30^\circ\text{C}\) (i.e., the top two rows of demand points in Figure 1). Given this new set of demand points, populations A, B, and C would be prioritized. Since demand points can be sited and weighted in any configuration, they provide a flexible means to guide the reserve selection process.

The raptr R package utilizes two novel formulations of the reserve selection problem to generate prioritizations. These formulations are based on ideas that underpin the Marxan, environmental diversity, and uncapacitated facility location problems (Owen & Daskin, 1998). Since they are based on the unreliable and reliable facility location problems (Cui, Ouyang, & Shen, 2010), the formulations are hereafter referred to as the “unreliable” and “reliable” formulations. The difference between the two formulations is that the reliable formulation explicitly accommodates uncertainty in the probability of features occupying the planning units when calculating how well a given solution samples a feature’s attribute space. For brevity, we will define the simpler formulation—the unreliable formulation—below and define the more complex version—the reliable formulation—in Appendix S1. All mathematical terms defined hereafter are described in Table S1. For convenience, the cardinality of a given set will be denoted using the same symbol used to denote the set.

Define \(F\) to be the set of features one wishes to conserve (indexed by \(f\)). Let \(J\) be a set of planning units (indexed by \(j\)), and \(C_j\) denote the cost of preserving planning unit \(j \in J\). Also, let \(A_j\) denote the area—or some other metric such as habitat quality or amount of habitat—associated with planning unit \(j\). To assess the extent to which each feature is secured in a given prioritization, let \(q_{jf}\) denote the probability of feature \(f\) occupying planning unit \(j\). The level of fragmentation associated with a prioritization is parameterized as the total exposed boundary length (as in Marxan). Let the shared edges between each planning unit \(j \in J\) and \(k \in J\) be \(e_{jk}\).

Let \(S\) denote a set of attribute spaces (indexed by \(s\)). Each \(j \in J\) is associated with spatially explicit data that serve as coordinates in each attribute space \(s \in S\). Let \(l_{fs}\) denote a set of demand points (indexed by \(i\)) for each feature \(f \in F\) and each attribute space \(s \in S\). Let \(\lambda_{fs}\) denote the weighting for each demand point \(i \in I, f \in F\) and \(s \in S\). Let \(d_{fs}\) denote the distance between each demand point \(i \in I\) and each planning unit \(j \in J\) for each feature \(f \in F\) and attribute space \(s \in S\). To describe the inherent variation in the distribution of demand points for feature \(f\) and space \(s\), let \(\delta_{fs}\) denote the distance between each demand point \(i \in I\) and the centroid of the demand points in space \(s\).

Demand points with greater weight \(\lambda_{fs}\) are more important, and so solutions may need to select planning units closer to highly weighted demand points. Since increasing the number of demand points in the problem also increases the time required to solve it, conservation planners may find it more useful to increase of the weighting of demand points in important regions of an attribute space rather than increasing the number of demand points in the region. Decision maker will need to choose an appropriate weighting for each demand point to ensure that the resulting solution reflects their overarching goals.

Targets are used to ensure that prioritizations sufficiently conserve each feature. Amount-based targets specify the minimum amount of habitat required for each feature to be adequately conserved (similar to those used in Marxan). Let \(T_f\) denote the area or amount of habitat that needs to be preserved for each feature \(f \in F\). Space-based targets specify the minimum proportion of variation in the demand points that needs to be secured for each feature. These targets directly relate to a continuous and multi-dimensional attribute space and the set of demand points within it. Space-based targets are expressed as proportions—instead of a sum of weighted distances—by scaling the sum of weighted distances between the demand points and the selected planning units in a solution relative to the distances between demand points and the demand points’ centroid. This scaling is conceptually similar to that used in calculating the \(R^2\) statistic for \(k\)-means cluster

![Figure 1](https://example.com/figure1.png)
analyses from the within sums of squares and total sums of squares (Greenacre & Primicerio, 2014, p. 106). Let \( \tau_{fs} \) denote the space-based targets for feature \( f \in F \) and attribute space \( s \in S \). The control variables for the unreliable formulation are \( B, T_f \), and \( T_s \).

\[
T_s = \text{amount target for feature } f \\
T_f = \text{space target for feature } f \text{ in attribute space } s \\
B = \text{boundary length modifier (BLM) to penalize fragmented solutions}
\]

The decision variables are \( X_j \) and \( Y_{fsij} \):

\[
X_j = \begin{cases} 
1, & \text{if planning unit } j \text{ is selected for conservation action} \\
0, & \text{otherwise} 
\end{cases} \\
Y_{fsij} = \begin{cases} 
1, & \text{if demand point } i \text{ for feature } f \text{ in space } s \text{ is conserved by planning unit } j. \\
0, & \text{otherwise} 
\end{cases}
\]

Each demand point \( i \in I \) for feature \( f \in F \) and space \( s \in S \) is conserved by a single selected planning unit (i.e. \( a_j \in J \) where \( X_j = 1 \)). The degree to which a demand point \( i \) is conserved by a planning unit \( j \) is determined by the distance between them \( (d_{fsij}) \). Generally—unless near zero space targets are used so that the problem is effectively unconstrained by the target—demand points are conserved by their closest selected planning units. In poorer quality solutions, demand points will be conserved by planning units that are further away from them. As a consequence, the sum of the weighted distances between all of the demand points and the planning units used to conserve them will be larger, and so, the solution will capture less of the variation described by the demand points.

The unreliable formulation of the representative and adequate prioritization problem (URAP) is defined as a multi-objective optimization problem.

\[
\text{(URAP) Min } \sum_{j=0}^{J-1} (X_jC_j) + \sum_{j=0}^{J-1} \sum_{k=j}^{J-1} X_j (1 - X_k) (B_{jk}) \\
\text{s.t. } \sum_{j=0}^{J-1} A_i d_{fsij} \geq T_f, \quad \forall 0 \leq f \leq F - 1 \\
1 - \frac{\sum_{i=0}^{I-1} \sum_{j=0}^{J-1} Y_{fsij} \lambda_{fsi}}{\sum_{i=0}^{I-1} \sum_{j=0}^{J-1} \lambda_{fsi}} \geq \tau_{fs}, \quad \forall 0 \leq f \leq F - 1, \quad 0 \leq s \leq S - 1 \\
\sum_{j=0}^{J-1} Y_{fsij} = 1, \quad \forall 0 \leq f \leq F - 1, \quad 0 \leq s \leq S - 1, \quad 0 \leq i \leq J - 1 \\
Y_{fsij} \leq X_j, \quad \forall 0 \leq f \leq F - 1, \quad 0 \leq s \leq S - 1, \quad 0 \leq i \leq J - 1
\]

The objective (Equation 3a) is to minimize the total cost and fragmentation of the solution. Constraints ensure that all amount-based targets are met (Equation 3b), and all space-based targets are met for each feature and each attribute space (Equation 3c). By treating targets as constraints, rather than including them in the objective function, all feasible solutions will fulfill the targets. For each feature and attribute space, the total weighted distance between the demand points and their closest selected planning units is calculated \((\sum_{i=0}^{I-1} \sum_{j=0}^{J-1} \lambda_{fsi} \delta_{fsij} Y_{fsij})\). This total weighted distance is then scaled by the inherent variation in the demand points \((\sum_{i=0}^{I-1} \lambda_{fsi}^2 \delta_{fsij}^2 )\). The resulting fraction is used to calculate a proportion conceptually similar to the \( R^2 \) statistic used in \( k \)-means cluster analysis. Constraints (Equation 3d) ensure that only one planning unit is assigned to each demand point. Constraints (Equation 3e) ensure that demand points are only assigned to selected planning units. Constraints (Equation 3f) ensure that the \( X \) and \( Y \) variables are binary.

### 2.2 | Optimization

Although the reserve selection problems presented here are nonlinear (see Appendix S1 for the reliable formulation), they can be linearized using methods described by Beyer, Dujardin, Watts, and Possingham (2016) and Cui et al. (2010). The raptr R package provides functions to express conservation planning data as linearized versions of the optimization problems and solve them using the commercial Gurobi software suite (www.gurobi.com). Presently, academics can obtain a Gurobi license at no cost.

### 3 | EXAMPLES

To showcase the behaviour of the unreliable formulation, we conducted a simulation study and two case studies. To assess how long it would take to solve various sized problems, we also conducted a benchmark analysis (Appendix S2, Figure S2). We completed the analyses using R (version 3.3.2; R Core Team, 2016) and raptr (version 0.0.3). We solved all optimization problems to within 10% of optimality using Gurobi (version 7.0.2).

#### 3.1 | Simulation study

##### 3.1.1 | Methods

We simulated a hypothetical study area with square planning units arranged in a 10 × 10 grid (Figure 2), and three species inhabiting this study area. Firstly, we simulated a hyper-generalist species (hereafter referred to as the “uniformly distributed species”). It occupied all planning units with equal probability (Figure 2a; Equation 4a). Secondly, we simulated a species with simple habitat requirements...
and boundary length modifiers of 2.5. We generated solutions using 20% amount targets, 90% space targets, using amount and space targets as well as fragmentation penalties, we generated solutions using 20% amount targets and a boundary length modifier to reduce fragmentation (47.27% sampled; Figure 3c) because the latter solution sited more planning units closer to centre of the study area–reducing the total distance between the selected planning units and the demand points.

By explicitly targeting a representative sample of the species’ geographic spread, we obtained a solution that captured the geographic spread of the uniform species (93.39% sampled; Figure 3b). Although this solution secured an adequate amount of habitat and a representative sample of the species’ geographic spread, this solution was highly fragmented. However, by penalizing fragmentation using a boundary length modifier, we were able to obtain a well connected solution that met all of the objectives (90% sampled; Figure 3d). This solution

(hereafter referred to as the "normally distributed species"; Figure 2b). This species was most likely to be found in planning units nearest to the centre of the study area. It was simulated using the density function of a multivariate normal distribution (denoted by $\mathcal{N}$; Equation 4b). Thirdly, we simulated a species with two distinct populations (hereafter referred to as the "bimodally distributed species"; Figure 2c). It was simulated using the maximum density of two multivariate normal distributions (Equation 4c). For a given species, planning unit occupancy was calculated using the $(X, Y)$ coordinates of the units’ centroids and the relevant equation. We used a geographic attribute space to provide an intuitive visualization of the solutions. Demand points were set as the planning units’ centroids and were weighted according to the units’ probability of occupancy.

$$P(\text{uniformly distributed species}|X, Y) = 0.5 \quad (4a)$$

$$P(\text{normally distributed species}|X, Y) = \mathcal{N}\left(\begin{bmatrix} X \\ Y \end{bmatrix}\begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}, \begin{bmatrix} 12.5 & 0 \\ 0 & 12.5 \end{bmatrix}\right) \quad (4b)$$

$$P(\text{bimodally distributed species}|X, Y) = \text{Max} \mathcal{N}\left(\begin{bmatrix} X \\ Y \end{bmatrix}\begin{bmatrix} -3.75 \\ -3.75 \end{bmatrix}, \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}\right), \mathcal{N}\left(\begin{bmatrix} X \\ Y \end{bmatrix}\begin{bmatrix} 3.75 \\ 3.75 \end{bmatrix}, \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}\right) \quad (4c)$$

We generated four solutions for each species. First, to portray solutions generated using conventional methods (e.g. Marxan), we generated solutions using 20% amount targets. Second, to show how the addition of space targets can affect solutions, we generated solutions using 20% amount targets and 90% space targets to capture the geographic spread of each species. Third, to portray solutions generated using conventional planning methods that penalize for fragmentation, we generated solutions using 20% amount targets and a boundary length modifier of 2.5. Finally, to illustrate the combined effects of using amount and space targets as well as fragmentation penalties, we generated solutions using 20% amount targets, 90% space targets, and boundary length modifiers of 2.5.

We generated four solutions for each species. First, to portray solutions generated using conventional methods (e.g. Marxan), we generated solutions using 20% amount targets. Second, to show how the addition of space targets can affect solutions, we generated solutions using 20% amount targets and 90% space targets to capture the geographic spread of each species. Third, to portray solutions generated using conventional planning methods that penalize for fragmentation, we generated solutions using 20% amount targets and a boundary length modifier of 2.5. Finally, to illustrate the combined effects of using amount and space targets as well as fragmentation penalties, we generated solutions using 20% amount targets, 90% space targets, and boundary length modifiers of 2.5.

### 3.1.2 Uniformly distributed species

The solution generated for the uniformly distributed species using an amount target prioritized 20 planning units (Figure 3a). For this scenario, all solutions containing 20 units are optimal because this species has an equal chance of occurring in any given planning unit. Although it may seem odd that the solution prioritized planning units along the southern end of the study area, this behaviour occurred because there were many optimal solutions and the solver—without any other criteria to compare solutions—returned a solution that reflected the order that data were encoded in the optimization problem. As we can see in later examples, this behaviour only manifests when there are many optimal solutions. It is unlikely that conservation planners would encounter this behaviour when using real-world data, and if they did, it would suggest that they need to obtain more data to inform the reserve selection process.

The solution generated using just amount targets only captured a small proportion of the variation in the geographic attribute space (~23.64% sampled). The coverage was so poor that this “proportion” was negative because the solution captured a less representative sample than a solution containing one planning unit in the centre of the species’ distribution. In other words, the total distance between the demand points and the 20 selected planning units was larger than the total distance between the demand points and a single planning unit in the centre of the study area. Similarly, $R^2$ statistics can have negative values when they are describing models that perform worse than a null model. This solution performed worse than the solution generated using amount targets and a boundary length modifier to reduce fragmentation (47.27% sampled; Figure 3c) because the latter solution sited more planning units closer to centre of the study area–reducing the total distance between the selected planning units and the demand points.

By explicitly targeting a representative sample of the species’ geographic spread, we obtained a solution that captured the geographic spread of the uniform species (93.39% sampled; Figure 3b). Although this solution secured an adequate amount of habitat and a representative sample of the species’ geographic spread, this solution was highly fragmented. However, by penalizing fragmentation using a boundary length modifier, we were able to obtain a well connected solution that met all of the objectives (90% sampled; Figure 3d). This solution
prioritized a similar number of planning units as the previous solutions even though it is far superior. As we can see, under the simplest of circumstances, reserve selection methods may not yield solutions that secure a representative sample of features unless constraints are used to guarantee this property.

3.1.3 | Normally distributed species

The solution generated for the normally distributed species using just an amount target prioritized planning units in the species’ core distribution (Figure 3e). Although this strategy may seem cost-effective, simply conserving the places where this species is most likely to be found is a poor strategy for securing a representative sample of the species’ geographic spread (59.09% sampled) because it did not protect any peripheral habitat. By using amount and space targets, we obtained a solution that conserved an adequate amount of habitat and also secure a representative sample of the species’ geographic spread (90.02% sampled; Figure 3f). Similar to the solutions for the uniformly distributed species, this solution was highly fragmented and we were able to obtain a better connected solution by specifying fragmentation penalties (90.88% sampled; Figure 3h). However, unlike the solutions for the uniformly distributed species, the solution generated using an amount target, a space target, and fragmentation penalties required more planning units than the other solutions. These results suggest that solutions may need to select more planning units to meet additional conservation objectives.

3.1.4 | Bimodally distributed species

The solution generated for the bimodally distributed species using just an amount target conserved individuals belonging to one of the two populations (Figure 3i). As a consequence, this solution did not secure a representative sample of the species’ geographic spread (12% sampled). The addition of fragmentation penalties exacerbated this issue, and resulted in a solution that sampled even less of the species’ geographic spread (8.38% sampled; Figure 3k). However, this issue was resolved by using a space target to obtain a solution that secured both populations (90.23% sampled; Figure 3l). This finding suggests that species with large intra-specific variation could benefit the most from prioritizations generated using space-based targets.

3.2 | Case study 1

3.2.1 | Methods

We investigated how space-based targets can be used in a multi-species planning context to generate a prioritization that sufficiently preserves the species’ realized niches. By preserving the
populations in suitable habitats with different environmental conditions, conservation planners can preserve the species’ adaptive landscape and foster resilience to environmental change (Moritz, 2002). We selected Queensland, Australia as the study area, and used an equal area coordinate system for geospatial analyses (Australian Albers GDA94; EPSG:3577). We used a 50 x 50 km\(^2\) grid within the state boundary as planning units for this case study. We obtained data for 19 bioclimatic variables across the region (at 30’’ resolution from www.worldclim.org; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) and subjected them to a principal components analysis (using ArcMap 10.3.1). We used scores from the first two principal components to characterize the environmental variation across the study area (explaining 99.5% of the total variation; Figure 4).

We used four bird species in this case study: blue-winged kookaburra (Dacelo leachii), brown-backed honeyeater (Ramsayornis modestus), brown falcon (Falco berigora), and pale-headed rosella (Platycercus adscitus). These species span a range of different habitat requirements. We mapped the extent of occurrence for each species (Figure 5). To do this, we obtained occurrence records from the Atlas of Living Australia across the whole of Australia (using the ALA4R R package; Raymond, VanDerWal, & Belbin, 2015), spatially thinned the data to omit points within 10 km of each other to ameliorate the effects of sampling bias (using a modified version of the spThin R package; www.github.com/jeffreyhanson/spThin; Aiello-Lammens, Boria, Radosavljevic, Vilela, & Anderson, 2015), and fit 85% minimum convex polygons (using the adehabitatHR R package; Calenge, 2006). We used this method because it is entirely reproducible using freely available data.

We generated 500 demand points for each species (Figure S1). To ensure that the demand points reflected the core parts of the species’ realized niches, we used the following method: for each species we generated random points inside the species’ geographic range and at each point extracted the principal component values at that location. We then fitted hyperbox kernels to the distribution of principal component values to characterize the realized niche of each species using a manually chosen bandwidth of 0.2 and a 0.5 quantile to map the core parts of the species’ niches (implemented in the hyperR package; Blonder, Lamanna, Violle, & Enquist, 2014). We then generated uniformly distributed points inside the species’ distribution in environmental space, and extracted the kernel density at their locations. These uniformly distributed points and associated density estimations were used as demand point coordinates and weights (respectively).

We generated two multi-species prioritizations. The first solution was generated using 20% amount targets for each species. The second solution was generated using the same amount targets with additional 75% space targets to capture a representative sample of the realized niche for each species.

### Results

Generally, the solution generated using just amount targets preserved a representative sample of each of the four bird species’ niches (Figure 6a; left column Figure S1). This solution secured a large proportion of the realized niche of blue-winged kookaburra (90.18%), brown falcon (92%), and the pale-headed rosella (84.76%), but only a small proportion for the brown-backed honeyeater (29.09%). On the other hand, the solution generated using space targets secured a large proportion...
of the realized niche for all four of the species (Figure 6b; right column Figure S1). This result demonstrates that although conventional methods may yield solutions that conserve a representative sample of the variation in some features, only through the use of explicit targets can planners obtain cost-effective solutions that secure a representative sample of the variation in all features.

3.3 | Case study 2

3.3.1 | Methods

Here we used space-based targets to generate a prioritization securing a representative sample of a species’ intra-specific genetic variation. We used species occurrence and genetic data collected by the international IntraBioDiv project in the European Alps (Alvarez et al., 2009; Gugerli et al., 2008; Meirmans et al., 2011). Although this dataset contains multiple species, we used data for the betony-leaved rampion (Phyteuma betonicifolium) as a simple case study, because it exhibits significant inter-population genetic structure (Meirmans et al., 2011). Members of the IntraBioDiv project collected data using a 20° longitude × 12° latitude grid (c. 20 km × 22.5 km; Figure 7a). They visited every second grid cell, and if the species was detected in a cell, samples were collected from three individuals. Samples were genotyped using amplified fragment length polymorphisms (Vos et al., 1995), and used to construct matrices denoting the presence/absence of polymorphisms at loci. In total, 131 individuals were genotyped at 138 markers.

We used non-metric multi-dimensional scaling (NMDS; using Gower distances to accommodate sparsity; Gower, 1971; implemented using the cluster R package; Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2015) to ordinate the presence (or absence) of locus-specific alleles within individuals into two continuous variables (implemented in the vegan R package; Oksanen et al., 2015). These continuous variables described the main axes of genetic variation within the species. We averaged together values corresponding to samples collected in the same grid cell, and used them to create a genetic attribute space (Figure 7c,d). To assess spatial auto-correlation, we calculated Moran’s I auto-correlation index for each NMDS axis using inverse great circle distances between the grid cells’ centroids as weights (using the ape R package; Paradis, Claude, & Strimmer, 2004).

The grid cells were used as planning units for generating prioritizations, noting that data were not collected in every planning unit, and so the planning units were arranged in a checkerboard pattern. The spatially averaged ordinations were used to describe the typical genetic characteristics of individuals in each planning unit. Planning units with similar values/colours contain individuals with similar loci polymorphisms. Note that data were not collected in every grid cell, and the planning units are therefore arranged in a checkerboard pattern.
3.3.2 | Results

A two-dimensional genetic attribute space was able to capture most of the variation between individuals (stress = 0.17; Figure 7c,d). Planning units located near each other contained individuals with similar genetic characteristics (Moran’s I: NMDS axis 1, I = 0.4, p < .001; Moran’s I: NMDS axis 2, I = 0.32, p < .001). In terms of the average genetic characteristics of individuals in the planning units, they tended to cluster into two main groups, with evidence of within-group structure inside the larger group (Figure 8c). This analysis supports previous work by Alvarez et al. (2009) who also found evidence of genetic structure within this species.

The solution generated using just the amount target failed to preserve a representative portion of the species’ genetic variation (15.52% sampled; Figures 8a,c). This solution only conserved individuals in one of the two main genetic groups. In fact, this solution is just a collection of the cheapest planning units occupied by the species needed to fulfill the target. Alternatively, the solution generated using both amount and space targets conserved individuals in each of the two main genetic groups (95.2% sampled; Figure 8b,d). Although the solutions only differ by a single planning unit, by securing individuals in both main genetic groups the second solution was able to secure a much more representative sample of the species’ genetic variation for only a minor increase in cost (99.66 compared to 99.41 total cost).

4 | IMPLICATIONS AND FUTURE DIRECTIONS

The raptR R package provides a unified approach to reserve selection. Conservation planners can use this R package to generate prioritizations that secure intra- and inter-specific biodiversity patterns. Both the simulations and case studies show that explicitly targeting a representative sample of the variation within each feature in a conservation planning exercise can substantially alter prioritizations. Additionally, we found prioritizations often needed to secure more habitat to conserve a representative sample of the variation within each feature. Although it may not be practical to conserve substantially more habitat than the amount specified using amount-based targets, we did find that even small increases in reserve system size could yield large gains.

One of the key advantages of this package is that it is general enough to incorporate almost any spatially explicit variable. This package can accommodate intrinsic or extrinsic variation in the feature(s). For example, adaptation processes could be secured using environmental variation (e.g. Carvalho et al., 2011), and trophic processes could be conserved by capturing the overlapping distributions of predator and prey species (e.g. Chernomor et al., 2015; Rayfield, Moilanen, & Fortin, 2009). As is often the case with multi-objective conservation planning problems, trade-offs between different goals may be unavoidable. For example, maximizing geographic spread of sites will almost always have an impact on connectivity. Additionally, space targets can be used to secure a representative sample of features for reasons that are not related to biodiversity processes, such as obtaining a geographically representative set of reserves so there is equitable access to parks by people (e.g. Moilanen, Anderson, Arponen, Pouzols, & Thomas, 2013). As long as the variation can be described, using Euclidean distances—which could be achieved through transformation or dimension reduction—the R package can be used to obtain a representative sample.

The degree to which a prioritization truly secures a representative sample of a feature depends on (1) the attribute space(s), (2) the distribution of demand points chosen by the conservation planner, and (3) the space target used. The optimal solution will not effectively capture the decision maker’s objectives if an inappropriate set of spatial variables or demand points are used to construct the attribute space(s). Demand points should be distributed across the full range of variation in an attribute space to obtain solutions that secure a representative sample (see the make.DemandPoints function for statistical routines to achieve this). As with conventional amount-based targets, planners will need to ensure that space targets are high enough to fulfill conservation objectives (e.g. to secure genetic diversity as mandated in the convention on biological diversity). The most appropriate space target will depend on the species of conservation interest (e.g. threatened species may...
need more genetic variation to be conserved; Reed & Frankham, 2003), and how conserving more variation relates to the species’ persistence (e.g. the relationship between inbreeding and extinction has a threshold; Frankham, 1995). We encourage conservation planners to consider carefully the relevant biological information to identify suitable targets.

Attribute data need to be spatially comprehensive to map planning units to attribute spaces. However, many real-world datasets are patchy. For example, due to constrained resources, genetic data may only be available for some planning units. To use such patchy data, planners could omit the planning units that are missing data, or estimate the missing data, using spatially explicit models (e.g. generalized dissimilarity models or kriging; Ferrier, Manion, Elith, & Richardson, 2007; Oliver & Webster, 1990). This approach has been successfully applied to a range of biological datasets (e.g. Thomassen et al., 2010). We caution that inaccuracy and uncertainty in data can negatively impact prioritizations (Wilson, Westphal, Possingham, & Elith, 2005), and recommend that planners ensure that data are sufficiently reliable or use the reliable formulation (described in Appendix S1) to accommodate uncertainty into the reserve selection process.

To maximize the long-term persistence of biodiversity, decision makers need to identify prioritizations that preserve existing patterns of biodiversity and the processes that support them. It is becoming increasingly apparent that simply protecting a large area will fail to achieve this (Barnes, 2015). Here, we developed the raptr R package to provide conservation planners with the tools to deliver cost-effective prioritizations that secure an adequate amount of a representative sample of biodiversity features. By exploring the functionality of this software package, we found that explicitly targeting a representative sample of the variation in each feature can result in substantially different solutions.

ACKNOWLEDGEMENTS

J.O.H. is supported by an Australian Government Research Training Program (RTP) Scholarship. R.A.F. and H.P.P. have Australian Research Council Fellowships. This work was supported by the Centre for Biodiversity and Conservation Science (CBCS). We are also grateful to Dan Rosauer and an anonymous reviewer for their suggestions that greatly improved the manuscript.

AUTHORS’ CONTRIBUTIONS

J.O.H. performed the analysis and drafted the manuscript. All authors developed the study and edited the manuscript.

DATA ACCESSIBILITY

The raptr R package can be downloaded from The Comprehensive R Archive Network (https://CRAN.R-project.org/package=raptr). To permit replication and validation of this study, all data, code, and results are stored in an online repository (https://github.com/jeffreyhanson/raptr-manuscript) and are available from the Dryad Digital Repository https://doi.org/10.5281/zenodo.847016 (Hanson, Rhodes, Possingham, & Fuller, 2017).

ORCID

Jeffrey O. Hanson http://orcid.org/0000-0002-4716-6134
Jonathan R. Rhodes http://orcid.org/0000-0001-6746-7412
Hugh P. Possingham http://orcid.org/0000-0001-7755-996X
Richard A. Fuller http://orcid.org/0000-0001-9468-9678

REFERENCES


SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.