

rapr: Representative and Adequate Prioritisations in R

Jeffrey O. Hanson¹, Jonathan R. Rhodes², Hugh P. Possingham¹, Richard A. Fuller¹

¹*School of Biological Sciences, The University of Queensland, Brisbane, QLD, Australia*

²*School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane, QLD, Australia*

Correspondance should be addressed to jeffrey.hanson@uqconnect.edu.au

18 July 2016

Summary

1. An underlying aim in spatial conservation planning is to maximise the long-term persistence of biodiversity. Yet current approaches overwhelming focus on “adequacy”, the concept ensuring a target proportion of the distribution of a biodiversity feature (species or ecosystem) is reserved. Ultimately, to achieve long-term persistence, the processes the sustain biodiversity must also be conserved. Many biodiversity processes can be preserved by conserving a representative sample of each biodiversity feature. This idea has for decades been encapsulated in principle of “representativeness”. For example, representing the full range of genetic diversity or ability to persist in a range of climatic conditions might be crucial for the long -term persistence of a species in a world where environmental change is accelerating. However, fully operationalising the principle of repressiveness in spatial conservation planning has proven methodologically challenging.
2. To address this issue, we developed the **rapr: Representative and Adequate Pioritisations in R**, a toolkit to guide reserve selection using explicit targets for representing biodiversity processes as well as meeting area protection targets. Using a novel formulation of the reserve selection problem, users set “space targets” to secure a representative sample of each species within an attribute space reflecting any biodiversity process (eg. an attribute space expressing variation in genetic characteristics between individuals in an area, or variation in climatic conditions between areas occupied by indivduals). We explored the functionality of this R package using a simulation study and two case studies. In each study, we generated prioritisations that aimed to preserve an amount of habitat for each species using amount target—to represent conventional reserve selection methods—and compared them with prioritisations generated using both habitat and space targets—to represent this method.

3. We show that explicitly considering biodiversity processes in reserve selection can substantially change the configuration of the resulting solution. Simulations suggest that including representativeness in conservation planning is particularly important where the biodiversity feature has multi-model distributions in an attribute space. Results from initial case studies show that explicitly considering space-based targets for biodiversity processes can result in prioritisations that do not necessarily require much more area than traditional approaches, but that could be much more effective in achieving long term biodiversity persistence.
4. The **rapr** R package provides a unified framework achieving spatial conservation prioritisations that are both adequate and representative, resulting in a greater chance of long term biodiversity persistence than area-based planning alone.

Short running title: Representative and adequate prioritisations

Abstract word count: 370 / 350

Total word count: XXXXX / 8000

Number of references: 34

Number of figures: 8

Number of tables: 0

Number of textboxes: 0

Number of equations: 8

49 Introduction

50 Perhaps the most fundamental aim of conservation is to maximise the long-term persistence of
51 biodiversity at organisational levels (McNeely 1994; Margules & Pressey 2000). To achieve this,
52 conservation actions must preserve biodiversity patterns (eg. populations, species, ecosystems), but
53 also crucially the processes that sustain them. One of the major tangible achievements of modern
54 conservation has been the act of setting aside areas for preservation (Sanderson *et al.* 2015). Reserve
55 networks buffer species from gross threatening processes and set the stage for enhanced management
56 interventions (Gaston *et al.* {2008}). However, the resources available for conservation action are
57 limited, and so reserve networks must be sited in places that satisfy conservation objectives for
58 minimum cost (Margules & Pressey 2000). To achieve this, reserve selection is often formulated as
59 an optimisation problem and then solved to identify cost-effective candidate reserve systems (termed
60 prioritisations; Margules & Pressey 2000). Many decision support tools have been developed to help
61 identify cost-effective solutions (eg. **ConsNet**, Ciarleglio *et al.* 2009; **Marxan**, Ball *et al.* 2009;
62 **Zonation**, Moilanen 2007). Typically, these tools are used to deliver a prioritisation that preserves
63 a set of species of interest to the conservation planner (termed target species or features). Given a
64 set of species, decision makers can use these tools to preserve biodiversity patterns by generating a
65 prioritisation that secures an adequate proportion of each species' range.

66 However, to achieve effective conservation, reserve networks must do more than cover enough
67 area. They must capture the processes that sustain biodiversity over the long term; they must
68 capture both ecological and evolutionary processes (Crandall *et al.* 2000; Margules & Pressey
69 2000). Ecological processes, such as predator-prey interactions, pollination, and decomposition,
70 are required for biodiversity to persist over short time-scales. Typically, they operate over small
71 geographic domains—with exceptions such as migration and refugial habitats—and can be preserved
72 using suitably large planning units (Ciarleglio *et al.* 2009) that each contain a discrete unit of habitat
73 (Klein *et al.* 2009). On the other hand, evolutionary processes are required for biodiversity to persist
74 over long time-scales, and they typically operate over large geographic domains. Protected areas
75 must represent adaptive evolutionary processes to foster resilience against environmental change
76 (eg. climate change; Pyke & Fischer 2005). Protected areas must also preserve neutral evolutionary

processes (Moritz 1994, 1999, 2002), arising from restriction of gene flow between populations. They are important for maintaining genetic diversity, and avoiding inbreeding depression. In recent decades, a wealth of data relevant to evolutionary processes has become available to conservation planners (eg. bioclimatic, genetic, and trait data; Hijmans *et al.* 2005; Raymond *et al.* 2015; Jones *et al.* 2009). Yet such data are rarely used to guide conservation planning (Hendry *et al.* 2010). This is perhaps in large part due to that fact that existing reserve selection tools focus on capturing either biodiversity patterns or processes—but not both.

Existing conservation planning tools are generally not well-suited for representing biodiversity processes (but see Faith 2003). To preserve biodiversity processes, a prioritisation must capture a representative sample of each species. For instance, to preserve predator-prey interactions, a prioritisation must preserve individuals from each predator and prey species in the same area. To preserve adaptive evolutionary processes, a prioritisation must preserve the adaptive landscape of each species—populations experiencing different selection pressures (Moritz 2002). To preserve neutral evolutionary processes, a prioritisation must secure individuals descended from each of the genetic lineages that comprise each species (Moritz 1994). Previous attempts to accommodate such into multi-species conservation planning have partitioned species’ ranges into different groups, for example at habitat discontinuities (e.g. Carvalho *et al.* 2011) as a pre-processing step to render a new set of “pseudo-species”. However, this approach assumes that the biodiversity processes that operate across a species’ range can be readily split into discrete units. Yet data on biodiversity processes is often continuous and hyper-dimensional, and often cannot be reduced to a few categories without significant information loss (Faith & Walker 1996).

One of the key issues in reserve selection is the lack of a unifying decision support tool that can accommodate data on biodiversity patterns and processes in a multi-species context. To begin to fill this gap, we present the **rapr** R package. This R package uses novel formulations of the reserve selection problem to provide decision makers with the tools to generate prioritisations that preserve both biodiversity patterns and processes. We aim to define the concepts behind the problem formulations. Furthermore, we aim to explore the functionality of the **rapr** R package by applying one of the formulations to a set of simulated species and two case studies.

Methods

PROBLEM FORMULATION

Biodiversity features are defined as the entities that the prioritisation is required to preserve (eg. species, ecosystems). Spatial attributes are defined as the variation across the species' range that the prioritisation is required to sample. They can be intrinsic (eg. genetic or trait variation) or extrinsic (eg. environmental variation) to the feature. These attributes should be related to the biodiversity processes that the decision maker aims to preserve.

Each attribute is conceptualised as a space. This space is termed an "attribute space". Each planning unit is thought to occupy a single point inside each space. For example, a decision maker may require a prioritisation that represents populations along climatic gradients. To achieve this, the decision maker might use an "climatic" attribute space with dimensions relating to mean annual temperature ($^{\circ}\text{C}$) and precipitation (mm). Any given combination of temperature and precipitation may be conceived as a point in this environmental space. By associating planning units with climatic data, they can be mapped from geographic space to this environmental attribute space.

Demand points are points that also exist in an attribute space. They are designated by the decision maker to indicate regions of the attribute space that the decision maker wishes to represent in the prioritisation (see below for discussion on how demand points can be generated for real-world datasets). The amount of variation in the attribute space that a prioritisation secures is a function of the distance between each demand point and each selected planning unit in the attribute space. The shorter the distances between the demand points and the planning units; the better the prioritisation is at securing the variation in the spatial attribute. To convert these amounts to a proportion—a meaningful unit for a decision maker—the distances between the selected planning units and the demand points are scaled by the distances between the demand points to the centroid of the demand points. In any attribute space there may exist points that are impossible (eg. mean annual rainfall -5 mm), or do not occur in the study area (eg. mean annual temperature 30°C in Antarctica). Additionally, there may be some regions that are desirable for some features and undesirable for others (eg. conditions known to be outside the physiological tolerance of certain species). Thus a

different set of demand points and weights are used for each attribute space and each feature. By placing demand points in desirable regions of an attribute space for a given feature, the decision maker can ensure that prioritisations secure the feature in planning units with spatial attributes that are desirable for that feature.

To illustrate these concepts, consider an example conservation planning scenario example involving an attribute space and demand points. We wish to develop a prioritisation for a single species that has four populations in the study area. However, we can only afford to preserve three populations. We aim preserve the adaptive landscape of the species, and to achieve this, we will preserve populations inhabiting different environmental conditions. To describe environmental variation, we obtain data on the environmental conditions (rainfall (mm) and temperature ($^{\circ}\text{C}$)) where each population is found. These environmental data are then used to construct a two-dimensional environmental attribute space. Next, we generated demand points as equi-distant points between the range of values where the populations were found. By comparing the distribution of the demand points to the distribution of the populations in the attribute space, we can identify a suitable prioritisation (Fig. 1). We can see that preserving both populations *A* and *C*, effectively “doubles-up” on the same environmental characteristics, resulting in considerable redundancy in the solution. Instead, a more representative sample of the intra-specific variation could be captured by securing populations *A*, *B*, and *D*. This example demonstrates how the inclusion of biodiversity processes can guide the reserve selection process.

The formulations used to express the reserve selection problem in the **rapr** R package are based on a combination of the **Marxan** reserve selection problem and the uncapacitated facility location problems (Cornuéjols *et al.* 1990). Although the **rapr** R package provides two novel formulations, for brevity, we will define the simpler formulation—referred to as the unreliable formulation—below and define the more complex version—the reliable formulation—in the Supporting Information S1. These formulations are named after the unreliable and reliable facility location formulations from which they are based upon (Cui *et al.* 2010). The key difference between these two formulations is that the reliable formulation explicitly considers the probability that planning units are occupied when calculating the proportion of an attribute space sampled in a solution. All mathematical terms defined hereafter are described in Table S1. For convenience, the cardinality of sets will be

denoted using the same symbol used to denote the variable.

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). Also, let A_j denote the area, and C_j denote the cost of preserving planning unit $j \in J$. To assess the extent to which each feature is secured in a given prioritisation, let q_{fj} denote the probability of feature f occupying planning unit j . The level of fragmentation associated with a prioritisation is parametrised as the net exposed boundary length. 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 represent coordinates for each attribute space $s \in S$. Let $I_{f_{si}}$ denote a set of demand points (indexed by i) for each feature $f \in F$ and each attribute space $s \in S$. Let $\lambda_{f_{si}}$ denote the weighting for each demand point $i \in I$, $f \in F$ and $s \in S$. Let $d_{f_{sij}}$ 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 $\delta_{f_{si}}$ denote the distance between each demand point $i \in I$ and the centroid of the demand points. Demand points with greater weight $\lambda_{f_{si}}$ are more important, and the optimal solution will be likely to select planning units close to highly weighted demand points. As a consequence, the decision maker will need to choose an appropriate weighting for each demand point.

Targets are used to ensure that prioritisations adequately preserve each species. Amount-based targets are used to ensure that the total amount of habitat preserved is sufficient. Let T_f denote the expected amount of area that needs to be preserved for each feature $f \in F$. Space-based targets ensure that a sufficient proportion of the intra-specific variation is secured. Let τ_{fs} denote the space-based targets for feature $f \in F$ and attribute space $a \in A$. For convenience, these both types of targets are expressed as proportions in the R package.

The control variables for the unreliable formulation are the B , T_s , and τ_{sa} variables.

T_s = amount target for feature f eqn 1a

τ_{sa} = representation target for feature f in attribute space a eqn 1b

B = boundary length modifier (BLM): penalise fragmented solutions eqn 1c

185 The decision variables are the X_j and Y_{fsij} variables.

$$X_j = \begin{cases} 1, & \text{if planning unit } j \text{ is selected for conservation action} \\ 0, & \text{otherwise} \end{cases} \quad \text{eqn 2a}$$

$$Y_{fsij} = \begin{cases} 1, & \text{if demand point } i \text{ is assigned to planning unit } j \text{ for feature } f \text{ in space } s. \\ 0, & \text{otherwise} \end{cases} \quad \text{eqn 2b}$$

186 Each demand point $i \in I$ for feature $f \in F$ and space $s \in S$ is assigned to a selected planning unit
 187 J where $X_j = 1$. The weighted distance between the demand point and its assigned planning unit
 188 $\lambda_{fsi}d_{fsij}$ is used to assess how well the demand point is represented in a given solution. Generally,
 189 demand points are assigned to the closest selected planning units (unless particularly low space-based
 190 targets are used).

191 The unreliable formulation (URAP) is defined as a multi-objective optimisation problem.

$$\text{(URAP)} \quad \text{Min} \quad \sum_{j=0}^{J-1} (X_j C_j) + \sum_{j=0}^{J-1} \sum_{k=j}^{J-1} X_j (1 - X_k) (Be_{jk}) + \quad \text{eqn 3a}$$

$$\text{s.t.} \quad \sum_{j=0}^{J-1} A_j q_{fj} \geq T_f \quad \forall 0 \leq f \leq F-1 \quad \text{eqn 3b}$$

$$1 - \frac{\sum_{i=0}^{I-1} \sum_{j=0}^{J-1} \lambda_{fsi} d_{fsij}^2 Y_{fsij}}{\sum_{i=0}^{I-1} \lambda_{fsi} \delta_{fsi}^2} \geq \tau_{fs} \quad \forall 0 \leq f \leq F-1, \quad \text{eqn 3c}$$

$$0 \leq s \leq S-1$$

$$\sum_{j=0}^{J-1} Y_{fsij} = 1 \quad \forall 0 \leq f \leq F-1, \quad \text{eqn 3d}$$

$$0 \leq s \leq S-1,$$

$$0 \leq i \leq I-1$$

$$Y_{fsij} \leq X_j \quad \forall 0 \leq f \leq F-1, \quad \text{eqn 3e}$$

$$0 \leq s \leq S-1,$$

$$0 \leq i \leq I-1,$$

$$0 \leq j \leq J-1$$

$$X_j, Y_{fsij} \in 0, 1 \quad \forall 0 \leq f \leq F-1, \quad \text{eqn 3f}$$

$$0 \leq s \leq S-1,$$

$$0 \leq i \leq I-1$$

192 The objective function (eqn 3a) determines the utility of a given prioritisation: a combination of
 193 the total cost of a prioritisation and how fragmented it is. Constraints (eqn 3b) ensure that all
 194 the amount-based targets are met. Constraints (eqn 3c) ensure that all the space-based targets
 195 are met for each feature and each attribute space. For each feature and attribute space, the total
 196 weighted distance between the demand points and their closest selected planning units is calculated
 197 $(\lambda_{fsi} d_{fsij} Y_{fsij})$. This total weighted distance is then scaled by the inherent variation in the demand
 198 points $(\lambda_{fsi} \delta_{fsi})$. The resulting fraction yields a proportion conceptually similar to the R^2 statistic
 199 used in k -means clustering analysis. The constraints ensure that this proportion must be equal to

or greater than the space-based target. Constraints (eqn 3d) ensure that only one planning unit is assigned to each demand point. Constraints (eqn 3e) ensure that demand points are only assigned to selected planning units. Constraints (eqn 3f) ensure that the X and Y variables are binary.

OPTIMISATION

The unreliable formulation is non-linear. However, the non-linear components can be linearised using existing techniques. The expression $X_j X_k$ in (eqn 3a) can be linearised using methods described by Beyer et al. (2016). Linearised versions of the problems can be solved using commercial exact algorithm solvers. The **rapr** R package provides functions to express conservation planning data as optimisation problems using linearised versions of the unreliable and reliable formulations. These optimisation problems can then be solved to generate prioritisations using the commercial **Gurobi** software suite (<http://www.gurobi.com>). Presently, academics can obtain a license at no cost from the Gurobi website. After installing the **Gurobi** software suite, users will need to install the **Gurobi** R package.

Examples

To understand the behaviour of the unreliable problem and showcase its value, we conducted a simulation study and two case studies. These studies involved generating solutions using only amount targets to represent prioritisations generated using conventional methods (eg. **Marxan**), and solutions using amount and space targets using the unreliable formulation. By comparing these solutions, we can gauge the benefits of explicitly including space targets in reserve selection. All analyses were performed in R (version 3.3.0; R Core Team 2016).

SIMULATION STUDY

Methods

We simulated a hypothetical study area with planning units arranged in 10×10 square grid (Fig. 2). We then simulated three species across this study area. The first species was simulated to represent

224 a hyper-generalist—occurring in all planning units with equal probability (Fig. 2a). This species’
 225 distribution was based on a uniform distribution (eqn 8a). The second species was simulated to
 226 represent a species with an idealised distribution (Fig. 2b). It has a core range area, and is less likely
 227 to be found in areas that are distant from the core area. This species’ distribution was simulated
 228 using the probability density function of a single multivariate normal distribution (eqn 8c). The
 229 third species represents a species with two distinct populations (Fig. 2c). This species’ distribution
 230 was simulated to depict a bimodal distribution, based on the combination of the probability density
 231 functions of two multivariate normal distributions (eqn 8d). The probability that each species
 232 inhabited a given planning unit was calculated using the X, Y coordinates of the planning unit and
 233 eqns 8a–8d. We used a geographic attribute space to showcase the behavior of the problem. For
 234 each species, demand points were generated by calculating the centroids of each planning units and
 235 their weights were set as the probability that the planning units were occupied.

$$P(\text{uniformly distributed species} | (x, y)) = 0.1 \quad \text{eqn 8a}$$

$$f(z, \mu, \Sigma) = (2\pi)^{-1} |\Sigma| e^{-\frac{1}{2}(z-\mu)'\Sigma^{-1}(z-\mu)} \quad \text{eqn 8b}$$

$$P(\text{normally distributed species} | (x, y)) = \frac{f\left(\begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} \begin{bmatrix} 12.58 & 0 \\ 0 & 12.5 \end{bmatrix}\right)}{2} \quad \text{eqn 8c}$$

$$P(\text{bimodally distributed species} | (x, y)) = \text{Max} \left\{ \begin{array}{l} f\left(\begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} -3.75 \\ -3.75 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}\right), \\ f\left(\begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} 3.75 \\ 3.75 \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}\right) \end{array} \right\} \quad \text{eqn 8d}$$

236 We generated four solutions for each species. First to represent solutions generated using conventional
 237 conservation planning methods, we generated solutions using only 20 % amount targets. Second to
 238 show how the addition of geographic targets can affect a prioritisation, we generated solutions using
 239 20 % amount targets and 75 % geographic targets. Third to represent solutions using conventional
 240 planning methods that penalise for fragmentation, we generated solutions using 20 % amount targets
 241 and a boundary length modifier of 1. Fourth to illustrate the combined affect using geographic

targets and penalising fragmentation, we generated solutions using using 20 % amount targets and 75 % geographic targets and a boundary length modifier of 1.

Results

The uniform species was simulated to occur with a constant probability of occupancy across the study area (Fig. 2a). The solution generated using 20 % amount-based targets (Fig. 3a) selected 20 planning units near the southern end of study area. This configuration is an artifact of the method used to solve this particular instance of the reserve selection problem. In fact, for this species, all prioritisations containing 20 planning units are optimal when considering only 20 % amount targets. In the absence of criteria to guarantee representativeness, reserve selection methods may or may not return solutions that secure a representative sample of the features. In this particular case, the solution does not secure a representative sample of the species' range (-23.64 % sampled).

The addition of a 75 % geographic target resulted in a solution that secured a representative sample of the uniform species' range (90.67 % sampled; Fig. 3b). Since all planning units have equal cost and an equal chance of being occupied, this solution has the same number of planning units as the solution generated using only amount targets (cf. Fig 3a). Although the use of amount and space based targets has addressed adequacy and representativeness objectives (respectively), they have resulted in a highly fragmented solution.

The addition of a positive boundary length modifier (BLM) parameter resulted in a well-connected solution that secured an adequate proportion (20 %) and representative sample (75.76 % sampled) of the species geographic distribution (Fig. 3d). This solution contains 20 planning units—a few more than the previously discussed solutions—to ensure that the solution secures a representative proportion of the species geographic distribution in a configuration that is not highly fragmented. This result suggests that the combination of space targets and boundary length modifiers may yield solutions that contain more planning units, since the solution generated using amount targets and a boundary length modifier (Fig. 3c) contained the sample number of planning units as the solution with just the amount targets (Fig. 3a). Overall, these results show that under the simplest of conditions, the reliance on just amount targets can cause an under-specified reserve selection problem that is unlikely to return a suitable solution for implementation.

270 The normally distributed species was simulated to contain a single core area where individuals are
271 most prevalent and marginal areas where individuals are less likely to occur (Fig. 2b). The solution
272 generated using 20 % amount targets contained 10 planning units and concentrated conservation
273 efforts in the core area (Fig. 3e). Since all planning units have equal costs and areas, this solution
274 contains the planning units with the highest probabilities of occupancy. Whilst this solution
275 satisfies the adequacy objective for a prioritisation in a cost-effective manner—it fails to fulfill the
276 representative objective for a prioritisation (59.09 % distribution sampled).

277 The solution generated using 20 % amount targets and 75 % geographic targets resulted in a solution
278 that is both adequate and representative in terms of the uniform species distribution (Fig. 3f).
279 This solution used 11 planning units to secure 20.02 % and sample 76.31 % of the normal species’
280 distribution. By using amount and geographic targets in the reserve selection problem, we have
281 obtained a solution that concentrates conservation effort in the range core and also the range margin.
282 However, similar to the corresponding solution for the uniform species (Fig. 3b), this solution is
283 highly fragmented.

284 By using 20 % amount targets and 75 % geographic targets and a boundary length modifier parameter
285 ($BLM = 1$), we obtained a solution that fulfills adequacy, representativeness and connectivity
286 objectives (Fig. 3h). This solution contains 12 planning units. Similar to the corresponding solution
287 for the uniform species (Fig. 3d), this solution contains more planning units than any other solutions
288 for this species (cf. Figs. 3e–3g). The results for the normally distributed species suggest that the
289 space targets can result in solutions that secure a more representative sample of the species’—even
290 for species without significant structure.

291 The bimodally distributed species was simulated to represent a species with a highly structured
292 distribution, and it contains two distinct populations (Fig. 2c). The solution generated using just
293 20 % amount targets assigned conservation effort to just one of the two populations (Fig. 3i). While
294 this solution secured an adequate proportion of the species’ distribution (20.18 % secured), it did
295 not sample a representative proportion of the range (12 % sampled). The addition of boundary
296 length modifiers to just amount targets resulted in a solution which sampled even less of the species
297 distribution (25.19 % sampled; Fig. 3k). However, the addition of geographic space targets resolved
298 these issues.

The solution generated using 20 % amount and 75 % geographic targets included planning units from both populations (Fig. 3j). Although this solution required more planning units to fulfill both targets ($n = 9$), this solution secured a representative sample of the bimodally distributed species (80.95 % range sampled). Similar to corresponding solutions generated for both the uniformly (Fig. 3b) and normally distributed species (Fig. 3f), this solution was also highly fragmented and we could obtain a more well-connected solution at the expense of selecting additional planning units ($n = 9$; Fig. 3l). The results for the bimodally distributed species suggest that species with highly structured populations or strong variation between individuals could benefit the most from prioritisations that are generated using space-based targets.

CASE STUDY 1

Methods

We investigated how space-based targets can be used in a multi-species planning context to generate a prioritisation that sufficiently preserves the realised niche for several species. By preserving the populations in suitable habitats with different environmental conditions, conservation planners can preserve the species' adaptive landscape and foster resilience against environmental change (Moritz 2002). We selected Queensland, Australia as the study area. This region is ideal for exploring the potential of niche-based targets because it contains a broad range of different habitats. We obtained data for 19 bioclimatic variables across the region (at 30'' resolution from www.worldclim.org; Hijmans *et al.* 2005) and subjected them to a principal components analysis (using ArcMap 10.3.1). We used the first two principal components (cumulatively explained 99.5 % of the total variation) to characterise the environmental variation across the study area (Fig. 4).

We selected four bird species that span a range of different evolutionary histories, distributions and ecologies: blue-winged kookaburra (*Dacelo leachii*), brown-backed honeyeater (*Ramsayornis modestus*), brown falcon (*Falco berigora*), and pale headed rosella (*Platycercus adscitus*). We then mapped the extent of occurrence for each species (Fig. 5). To do this, we obtained observation data from the Atlas of Living Australia across the whole of Australia (using the `ALA4R` R package; Raymond *et al.* 2015), spatially thinned the data to omit points within 10 km² of each other to

ameliorate the effects of sampling bias (using the developmental version of the `spThin` R package; www.github.com/mlammens/spThin; Aiello-Lammens *et al.* 2015), and fit 85 % minimum convex polygons (using the `adehabitatHR` R package; Calenge 2006). We used this method because it is trivially reproducible using freely available data.

We generated 500 demand points for each species (Fig. S2). To achieve this, for each species, we generated random points inside the species range and at each point extracted the principal component values at that location. We then fit hyperbox kernels to the distribution of these points to characterise the realised 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 `hypervolume` R package; Blonder *et al.* 2014). We then generated uniformly distributed points inside the species' kernels and estimated the density of the training points at the uniformly generated points. These uniformly distributed points and associated density estimated were used as demand point coordinates and weights (respectively).

We generated two solutions. The first solution was generated using 20 % amount-based targets for all species. The second solution was generated the same amount-based target with an additional 75 % niche-based for each species.

Results

Generally, the solution generated using amount targets preserved a representative sample of each the four bird species' niches—with one exception. Specifically, this solution sampled over 75 % of the realised niche of blue-winged kookaburra (89.54 %), brown falcon (92.31 %), and the pale-headed rosella (76.86 %). Yet it failed to achieve this for Brown-backed honeyeater (30.93 %). This result suggests that prioritisations generated using conventional methods may preserve a representative sample of most species realised niches (Figure S1). However, despite this, there may yet be species for which only a small fraction of their realised niche is preserved. By explicitly using space targets, conservation planners can generate prioritisations that are guaranteed to capture a representative sample of species' niches.

CASE STUDY 2

Methods

Here we used space-based targets to generate a prioritisation securing a representative sample of a species' intra-specific genetic variation. We used species occurrence and multilocus AFLP data collected by the international IntraBioDiv project in the European Alps (Meirmans *et al.* 2011; see Alvarez *et al.* 2009 for further explanation of data collection methods). Although this dataset contains multiple plant species, we used data for the betony-leaved rampion (*Phyteuma betonicifolium*), a wide-spread species with significant genetic structure (Aiello-Lammens *et al.* 2015). Members of the IntraBioDiv project collected data using a 20' longitude by 21' latitude grid (approx. 22.3 km \times 25 km; Fig. 7a). They visited every second grid cell, and if the betony-leaved rampion was detected in a cell, samples were collected from three individuals. Samples were genotyped using amplified fragment length polymorphisms (AFLP; Raymond *et al.* 2015), 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 the `cluster` R package; Maechler *et al.* 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 calculated the average of the values associated with individuals in each grid cell. These values were used to create a genetic attribute space (Figs. 7c and 7d). To assess spatial auto-correlation, we calculated Moran's I auto-correlation index for each NMDS axis using inverse great circle distances based on the grid cells' centroids (using the `ape` R package; Paradis *et al.* 2004).

The grid cells were used as planning units. The grid-cell averaged ordinations were used to describe the typical genetic characteristics of individuals in the planning unit. Since the number of planning units was relatively small, we used the same grid-cell averaged ordinations as demand points. To ensure that the solutions did not prioritise particularly costly areas, we included opportunity cost data (Fig. 7b). We obtained population density data (1 km resolution from the Global Rural-Urban

Mapping Project; GRUMP V1; CIESEN, Columbia University *et al.* 2011) and estimated the total population density inside each grid cell.

We generated two solutions. The first solution was generated using an 10 % amount-based target. The second solution was generated using the same amount-based target and an additional 95 % genetic-based target.

Results

The binary genetic data was ordinated into a two-dimensional space that described the main difference between individuals (stress = 0.17). These values were then averaged to the planning unit level to describe the typical genetic characteristics of individuals in each planning unit (Figs. 7c–7d). Planning units located near each other were found to contain individuals with similar genetic characteristics (Moran's I: NMDS axis 1, $I = 0.4$, $P < 0.001$; Moran's I: NMDS axis 2, $I = 0.32$, $P < 0.001$). 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 (Fig 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 (15.54 % sampled; Figs. 8a and 8c). This solution only sampled planning units that contained individuals belonging to one of the main two groups. On the other hand, the solution generated using additional genetic targets selected planning units belonging to both of the main groups. Note that since these solutions were generated using opportunity cost data, the solution generated using only the amount target is essentially the minimum number of least costly areas needed to fulfill the amount target. The only difference between the two solutions is a single planning unit. By swapping this single planning unit, the solution generated using amount and genetic targets was able to preserve a representative portion of the species' genetic variation 15.54 for only a minor increase in cost (99.41 total cost compared to 99.66 total cost).

Implications and future directions

The `rapr` R package provides a unified approach to reserve selection. One of the key advantages of this package is that it is general enough to incorporate any spatial variable as an attribute space. This package can accommodate intrinsic or extrinsic variation to the feature(s). For example, adaptation processes could be secured using environmental variation or genetic variation among loci under selection. Ecological processes, such as predator-prey interactions, could be secured by capturing intra-specific trophic variation. Different attribute spaces with different targets can be also specified for different species. As long as the variation can be described using Euclidean distances, the R package can be used to obtain a representative sample (note that data may require transformation to conform to this assumption). This `\texttt{R}` package provides decision makers with the tools to generate prioritisations that secure both biodiversity patterns and the processes that maintain biodiversity. Additionally, the package contains functionality to accommodate uncertainty in the distribution of features, and also identify suitably connected reserves. Both the simulated and case study species suggest that conservation planning exercises need to explicitly consider biodiversity processes during the reserve selection process to ensure they are captured.

The degree to which a prioritisation truly secures a representative sample of a feature depends on the attribute spaces and distribution of demand points chosen by the decision maker. Ultimately, the space-based targets are set as a proportion based on the distribution of the demand-points. As a consequence, if the decision maker uses an inappropriate set of spatial variables to construct an attribute space, or an inappropriate set of demand points, then the optimal solution will not be an effective prioritisation. We therefore stress that decision makers must carefully consider which biodiversity processes need to be reflected in the prioritisation, and which spatial data can be used to represent these processes. Note that maximising one process can be detrimental to another. For instance, to maximise the geographic spread of a prioritisation, reserves need to be further away from each other, yet to maximise the connectivity of a prioritisation, reserves need to be closer to each other. To assist in the selection of appropriate demand points, the R package provides several routines for generating demand points (see the `make.DemandPoints` function). These routines essentially use the distribution of a feature in the attribute space to define a polygon.

433 Demand points are then generated as random points within the polygon. A kernel is then fit to the
434 distribution of the feature in the space (using Blonder *et al.* 2014; Duong 2015), and the demand
435 points are weighted based on the estimated density of the feature at the demand points.

436 The formulation requires spatially comprehensive data to map attribute spaces to planning units.
437 For instance, in both case-studies, all planning units occupied by the species were associated with
438 values/coordinates in the attribute spaces. However, most real-world data sets are patchy—some
439 planning units will be occupied by species for which attribute space data is not available. To use
440 such patchy data with the `\texttt{rapr R}` package, the gaps in the data must first be filled in as a
441 pre-processing step. Spatially explicit models could be used to estimate values in locations that
442 are missing from a data set (eg. krigging or generalised dissimilarity models; Oliver & Webster
443 1990; Ferrier 2002). This approach has been successfully applied to a range of biological data sets
444 (Thomassen *et al.* 2010).

445 To maximise the long-term persistence of biodiversity, decision makers need to identify prioritisations
446 that preserve existing patterns of biodiversity and the processes that support them. To achieve this,
447 conservation planners need a decision support tool that can explicitly accommodate biodiversity
448 patterns and processes. Here, we developed the `rapr R` package to fill this void. By exploring the
449 functionality of this package using several simulated species, we found that including space-based
450 targets can radically change a prioritisation for the simplest of species.

451 **Acknowledgements**

452 JOH is funded by an Australian Postgraduate Award (APA) scholarship. RAF has an Australian
453 Research Council Future Fellowship. This work was supported by the Centre of Excellence for
454 Environmental Decisions (CEED) and the Landscape Ecology and Conservation Group (LEC) at
455 The University of Queensland.

Data accessibility

All data, code, and results are stored in an online repository (www.github.com/paleo13/rapr-manuscript) to permit replication and validation of this study.

References

- Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B. & Anderson, R.P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, **38**, 541–545.
- Alvarez, N., Thiel-Egenter, C., Tribsch, A., Holderegger, R., Manel, S., Schönswetter, P., Taberlet, P., Brodbeck, S., Gaudeul, M., Gielly, L., Küpfer, P., Mansion, G., Negrini, R., Paun, O., Pellecchia, M., Rioux, D., Schüpfer, F., Van Loo, M., Winkler, M., Gugerli, F. & IntraBioDiv Consortium. (2009). History or ecology? Substrate type as a major driver of patial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632–640.
- Ball, I., Possingham, H. & Watts, M.E. (2009). Marxan and relatives: software for spatial conservation prioritisation. *Spatial conservation prioritisation: Quantitative methods & computational tools* (eds A. Moilanen, K.A. Wilson & H. Possingham), pp. 185–189. Oxford University Press, Oxford, UK.
- Beyer, H.L., Dujardin, Y., Watts, M.E. & Possingham, H.P. (2016). Solving conservation planning problems with integer linear programming. *Ecological Modelling*, **328**, 14–22.
- Blonder, B., Lamanna, C., Violle, C. & Enquist, B.J. (2014). The n -dimensional hypervolume. *Global Ecology and Biogeography*, **23**, 595–609.
- CIESEN, Columbia University, International Food Policy Research Institute, The World Bank & Centro Internacional de Agricultura Tropical. (2011). Global rural-urban mapping project, Version 1 (GRUMP v1): Urban extents grid.
- Calenge, C. (2006). The package adehabitat for the R software: tool for the analysis of space and habitat use by animals. *Ecological Modelling*, **197**, 1035.

481 Carvalho, S.B., Brito, J.C., Crespo, E.J. & Possingham, H.P. (2011). Incorporating evolutionary
 482 processes into conservation planning using species distribution data: a case study with the western
 483 Mediterranean herpetofauna. *Diversity & Distributions*, **17**, 408–421.

484 Ciarleglio, M., Wesley Barnes, J. & Sarkar, S. (2009). ConsNet: new software for the selection of
 485 conservation area networks with spatial and multi-criteria analyses. *Ecography*, **32**, 205–209.

486 Cornuéjols, G., Nemhauser, G.L. & Wolsey, L.A. (1990). The uncapacitated facility location problem.
 487 *Discrete Location Theory* (eds P.B. Mirchandani & R.L. Francis), pp. 119–171. Wiley, New York.

488 Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M. & Wayne, R.K. (2000). Considering evolu-
 489 tionary processes in conservation biology. *Trends in Ecology & Evolution*, **15**, 290–295.

490 Cui, T.T., Ouyang, Y.F. & Shen, Z.J.M. (2010). Reliable facility location design under the risk of
 491 disruptions. *Operations Research*, **58**, 998–1011.

492 Duong, T. (2015). *ks: Kernel Smoothing, version 1.9.4*.

493 Faith, D.P. (2003). Environmental diversity (ED) as surrogate information for species-level biodi-
 494 versity. *Ecography*, **26**, 374–379.

495 Faith, D.P. & Walker, P.A. (1996). Environmental diversity: on the best-possible use of surrogate
 496 data for assessing the relative biodiversity of sets of areas. *Biodiversity & Conservation*, **5**, 399–415.

497 Ferrier, S. (2002). Mapping Spatial Pattern in Biodiversity for Regional Conservation Planning:
 498 Where to from Here? *Systematic Biology*, **51**, 331–363.

499 Gaston, K.J., Jackson, S.E., Cantu-Salazar, L. & Cruz-Pinon, G. ({2008}). The Ecological Perfor-
 500 mance of Protected Areas. *ANNUAL REVIEW OF ECOLOGY EVOLUTION AND SYSTEMAT-*
 501 *ICS, Annual Review of Ecology Evolution and Systematics*, **39**, 93–113.

502 Gower, J.C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, **27**,
 503 857–871.

504 Hendry, A.P., Lohmann, L.G., Conti, E., Cracraft, J., Crandall, K.A., Faith, D.P., Hauser, C., Joly,
 505 C.A., Kogure, K., Larigauderie, A., Magallon, S., Moritz, C., Tillier, S., Zardoya, R., Prieur-Richard,
 506 A.H., Walther, B.A., Yahara, T. & Donoghue, M.J. (2010). Evolutionary biology in biodiversity
 507 science, conservation, and policy: A call to action. *Evolution*, **64**, 1517–1528.

508 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution
 509 interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**,
 510 1965–1978.

511 Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L., Safi, K., Sechrest, W.,
 512 Boakes, E.H., Carbone, C., Connolly, C., Cutts, M.J., Foster, J.K., Grenyer, R., Habib, M., Plaster,
 513 C.A., Price, S.A., Rigby, E.A., Rist, J., Teacher, A., Bininda-Emonds, O.R.P., Gittleman, J.L.,
 514 Mace, G.M. & Purvis, A. (2009). PanTHERIA: a species-level database of life history, ecology, and
 515 geography of extant and recently extinct mammals. *Ecology*, **90**, 2648–2648.

516 Klein, C., Wilson, K., Watts, M., Stein, J., Berry, S., Carwardine, J., Smith, M.S., Mackey, B. &
 517 Possingham, H. (2009). Incorporating ecological and evolutionary processes into continental-scale
 518 conservation planning. *Ecological applications*, **19**, 206–217.

519 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2015). *cluster: Cluster Analysis*
 520 *Basics and Extensions*.

521 Margules, C.R. & Pressey, R.L. (2000). Systematic conservation planning. *Nature*, **405**, 243–253.

522 McNeely, J.A. (1994). Protected areas for the 21st-century: Working to provide benefits to society.
 523 *Biodiversity and Conservation*, **3**, 390–405.

524 Meirmans, P., Goudet, J., IntraBioDiv Consortium & Gaggiotti, O. (2011). Ecology and life history
 525 affect different aspects of the population structure of 27 high-alpine plants. *Molecular Ecology*, **20**,
 526 3144–3155.

527 Moilanen, A. (2007). Landscape Zonation, benefit functions and target-based planning: unifying
 528 reserve selection strategies. *Biological Conservation*, **134**, 571–579.

529 Moritz, C. (1999). Conservation units and translocations: strategies for conserving evolutionary
 530 processes. *Hereditas*, **130**, 217–228.

531 Moritz, C. (1994). Defining evolutionarily significant units for conservation. *Trends in Ecology &*
 532 *Evolution*, **9**, 373–375.

533 Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that
 534 sustain it. *Systematic Biology*, **51**, 238–254.

535 Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L.,
536 Solymos, P., Stevens, M.H.H. & Wagner, H. (2015). *vegan: community ecology package*.

537 Oliver, M.A. & Webster, R. (1990). Kriging: a method of interpolation for geographical information
538 systems. *International Journal of Geographical Information Systems*, **4**, 313–332.

539 Paradis, E., Claude, J. & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R
540 language. *Bioinformatics*, **20**, 289–290.

541 Pyke, C.R. & Fischer, D.T. (2005). Selection of bioclimatically representative biological reserve
542 systems under climate change. *Biological Conservation*, **121**, 429–441.

543 R Core Team. (2016). R: A language and environment for statistical computing.

544 Raymond, B., VanDerWal, J. & Belbin, L. (2015). *ALA4R: Atlas of Living Australia (ALA) Data*
545 *and Resources in R*.

546 Sanderson, E.W., Segan, D.B. & Watson, J.E.M. (2015). Global status of and prospects for
547 protection of terrestrial geophysical diversity. *Conservation Biology*, **29**, 649–656.

548 Thomassen, H.A., Buermann, W., Milá, B., Graham, C.H., Cameron, S.E., Schneider, C.J., Pollinger,
549 J.P., Saatchi, S., Wayne, R.K. & Smith, T.B. (2010). Modeling environmentally associated morpho-
550 logical and genetic variation in a rainforest bird, and its application to conservation prioritization.
551 *Evolutionary Applications*, **3**, 1–16.