

Species specific connectivity in reserve-network design using graphs

J. Orestes Cerdeira^{*1}, Leonor S. Pinto², Mar Cabeza³, Kevin J. Gaston⁴

Abstract

Systematic conservation planning applications based solely on the presence/absence of a large number of species are not sufficient to guarantee their persistence in highly fragmented landscapes. Recent developments have thus incorporated much desired spatial design considerations, and reserve-network connectivity has received increased attention. Nonetheless, connectivity is often determined without regard to species-specific responses to habitat fragmentation. But species differ in their dispersal ability and habitat requirements, making proximate priority areas necessary for some species, while undesirable for others. We present a novel approach that incorporates species-specific connectivity needs in reserve-network design. Importantly, our method differs from previous approaches in that connectivity is not part of the objective function, but part of the constraints, thus avoiding typical undesirable trade-offs that may result in high connectivity for some species but null connectivity for others. We use graphs to describe the dispersal pattern of each species and our goal is to iden-

^{*}This work was partially supported by FEDER/POCI 2010.

^{*}Corresponding author.

Email addresses: orestes@isa.utl.pt (J. Orestes Cerdeira), lspinto@iseg.utl.pt (Leonor S. Pinto), cabeza@cc.helsinki.fi (Mar Cabeza), k.j.gaston@sheffield.ac.uk (Kevin J. Gaston)

¹Centro de Estudos Florestais and Departamento de Matemática, Inst. Superior de Agronomia, Technical University of Lisbon (TULisbon), Tapada da Ajuda, 1349-017 Lisboa, PORTUGAL

²CEMAPRE and Inst. Sup. de Economia e Gestão, Technical University of Lisbon (TULisbon) Rua do Quelhas 6, 1200-781 Lisboa, PORTUGAL

³Metapopulation Research Group, Dept. of Biological and Environmental Sciences, P.O. Box 65, FIN-00014 University of Helsinki, FINLAND and Department of Biodiversity and Evolutionary Biology, National Museum of Natural Sciences, CSIC, Madrid, ES-28006, SPAIN

⁴Dept of Animal and Plant Sciences, Univ. of Sheffield, Sheffield, U.K. S10 2TN

tify minimum sets of reserves with connected sites for each of the species. This is not a trivial problem and we present three algorithms, one heuristic and two integer cutting algorithms that guarantee optimality, based on different 0-1 linear programming formulations. Applications to simulated data show that one of the algorithms that guarantee optimality is superior to the other, although both have limited application due to the number of sites and species they can manage. Remarkably, the heuristic can obtain very satisfactory solutions in short computational times, surpassing the limitations of the exact algorithms.

Key words: Reserve selection, systematic conservation planning, graphs, connectivity, algorithms, integer programming

1. Introduction

Systematic conservation planning is the process of using quantitative data to identify locations for conservation investment. It typically aims at maximising the number of species contained in selected units given a set of constraints. Such applications are often based solely on the presence/absence of a large number of species. However the ultimate goal of conservation is not to maximise current species occurrences but to maximise persistence of biodiversity (Cabeza and Moilanen, 2001; Cabeza and van Teeffelen, 2009; Pressey et al., 2007).

Aiming at maximising persistence is not a trivial task and consequently proxies for persistence are often used. These include maximising species abundances (Rodrigues et al., 2000), incorporating measures of site vulnerability (Wilson et al., 2005) and promoting spatial designs that minimise the impacts of habitat fragmentation (Cabeza et al., 2003; Possingham et al., 2000; van Teeffelen et al., 2006). Of all these proxies spatial design has received most attention, perhaps because habitat loss and fragmentation are the most important threat to biodiversity (Millennium Ecosystem Assessment, 2005) and rates of transformation of natural habitats continue to be high. Spatial design criteria include size, shape, replication, contiguity, connectivity, spacing and directional alignment. A proper spatial design thus buffers areas from external threats, provides

20 insurance against catastrophes, supports the regional persistence of fragmented
21 populations, and promotes adjustment of species ranges in response to climate
22 change.

23 In recent decades, reserve design tools have evolved to move from simple
24 spatial design guidelines derived from Island Biogeography or Metapopulation
25 theories to the inclusion of these principles quantitatively and objectively in
26 systematic conservation planning frameworks (see e.g. Williams et al. (2005)
27 for a review). While the first considerations of spatial attributes included simple
28 rules of adjacency when breaking ties in heuristic algorithms (Nicholls and Mar-
29 gules, 1993), a broad spectrum of approaches has developed since then. These
30 approaches deal with a number of alternative spatial design attributes such as
31 reserve compactness (e.g. minimising a linear combination of reserve size and
32 boundary length: Cabeza et al. (2003); McDonnell et al. (2002); Possingham et
33 al. (2000)), planning unit contiguity (Cerdeira and Pinto, 2005; Cerdeira et al.,
34 2005; Fuller et al., 2006; Önal and Briers, 2005; Önal and Wang, 2008; Shirabe,
35 2005), or cohesion and proximity (e.g. minimising the maximum distance be-
36 tween planning units or the sum of inter-planning unit distances: Fischer and
37 Church (2003); Önal and Briers (2002); Rothley (1999)). Optimisation model-
38 ing has also been used in corridor design (e.g. to minimise the amount or cost
39 of land needed to provide a corridor link between each reserve and every other
40 reserve: Sessions (1992); Williams (1998)).

41 With the methods outlined above, the desired level of connectivity is de-
42 termined subjectively, without regard to species-specific responses to habitat
43 fragmentation. However, as species differ in their dispersal ability and habitat
44 requirements, proximate priority areas may be necessary for some species, while
45 undesirable for others. But only a reduced number of studies have looked at
46 species-specific connectivity requirements in efficient reserve design (e.g. Cabeza
47 (2003); Cabeza et al. (2003); Moilanen and Cabeza (2002); Nicholson et al.
48 (2006)). Furthermore, this handful of studies has treated connectivity by incor-
49 porating it in the objective function. This means that the goal is maximising
50 occurrence and connectivity across species, which generates trade-offs that often

51 result in high connectivity for some species but null connectivity for others.

52 Here we present an alternative approach. We use graphs to describe the
53 dispersal pattern of each species, and consider species specific connectivity re-
54 quirements explicitly as part of the model constraints, and not as part of the
55 objective function. Our goal is to identify minimum sets of reserves with con-
56 nected sites for each of the species, connectivity here meaning sites occurring
57 within the dispersal range of the species. Note that we do not aim at a com-
58 pletely connected reserve network, but instead, each species is required to be
59 represented in a specified number of connected sites.

60 Spatial attributes increase the complexity of the reserve network design prob-
61 lem, as it often requires to be modeled as a non-linear expression. Linear expres-
62 sions are preferred because linear optimisation problems have the potential to
63 be solved exactly, while nonlinear problems are impossible or at least difficult
64 to solve to exact optimality. Consequently non-linear expressions for spatial
65 attributes are often solved with heuristic approaches. Nonetheless, some spa-
66 tial attributes, such as contiguity and compactness, have been solved with both
67 heuristic and exact methods (e.g. heuristic: Cabeza et al. (2003); McDonnell et
68 al. (2002); exact or both: Alagador and Cerdeira (2007); Cerdeira et al. (2005);
69 Önal and Wang (2008); Shirabe (2005)). Heuristics, such as greedy algorithms,
70 and metaheuristics such as simulated annealing or genetic algorithms can guar-
71 antee only approximate solutions. Contrastingly exact methods can in principle
72 find optimal solutions, at least if allowed to run to completion. However, very
73 large problems may not be soluble in reasonable amounts of time and thus
74 heuristics may be preferred.

75 To solve the problem we address here we present three algorithms, one heuris-
76 tic and two integer cutting algorithms, that guarantee optimality, based on
77 different 0-1 linear programming formulations. We apply the algorithms to sim-
78 ulated data sets of varying size, to evaluate and compare the practicability of
79 the approaches and the quality of the solutions produced by the heuristic.

80 Section 2 starts with some basic concepts of graphs and connectivity to give
81 a description of the problem. The 0-1 linear formulations and the integer cutting

82 algorithms are presented in Subsections 2.2 and 2.3. The heuristic is described
83 in Subsection 2.4. We conclude Section 2 explaining how the simulated species
84 distributions were generated, based on the neutral community theory. In Section
85 3 we report the main computational results, and in Section 4 we discuss and
86 compare the performance of the algorithms. We finish with conclusions and
87 final remarks in Section 5.

88 2. Material and methods

89 2.1. Graphs and connectivity

90 Graphs are mathematical objects suitable to describe the dispersal pattern
91 of species within a given region. We refer readers to Bondy and Murty (1976)
92 as a classical text book on graphs. A graph consists of a vertex set and an edge
93 set, where each edge is an unordered pair of vertices. To describe the distribu-
94 tion of a species s , vertices are used to represent their habitat sites, and edges
95 to specify the pairs of sites between which individuals from species s can move
96 directly. The graph of species s will be denoted by $G_s = (H_s, E_s)$, where H_s
97 is the set of vertices, and E_s the set of edges. A connected component of G_s
98 is a maximal subset of vertices C such that there is a path (i.e., a sequence
99 of edges with consecutive edges having a common vertex) linking any two ver-
100 tices of C . Note that the connected components of G_s distinguish the different
101 dispersal regions for the individuals of species s . The graph in Figure 1 has
102 vertex set equal to $\{1, 2, \dots, 16\}$ and two connected components: $\{1, 2, \dots, 9\}$
103 and $\{10, 11, \dots, 16\}$.

104 Identifying connected components is a basic problem in graphs, for which the
105 following efficient (linear time) algorithm can be used. Choose any vertex, mark
106 it and create a queue Q with the vertices which are adjacent to it. While Q
107 is not empty, remove an arbitrary vertex v from Q , mark it and add to Q all
108 unmarked vertices adjacent to v . When Q becomes empty a component has
109 been identified which consists of all the marked vertices. If an unmarked vertex

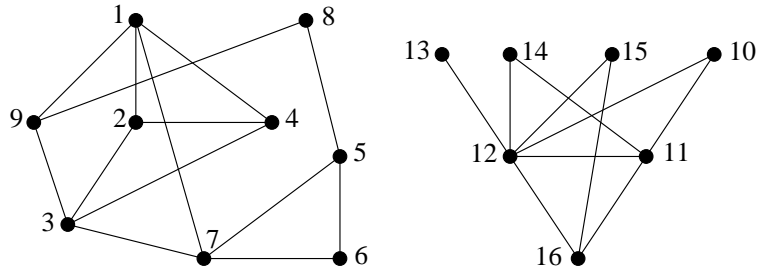


Figure 1: A graph with 16 vertices with two connected components

110 exists, remove from the graph the component already identified and repeat the
 111 procedure.

112 If S is the set of target species, $H = \cup_{s \in S} H_s$ denotes the union of the sites
 113 of all target species in the area under study. Given a subset H' from H , we
 114 use $H'_s = H' \cap H_s$ to represent the sites in H' where species s is represented.
 115 The graph with vertex set H'_s and whose edges are all the edges of G_s which
 116 connect pairs of vertices in H'_s is called the subgraph of G_s induced by H'_s and
 117 is denoted by $\langle H'_s \rangle$.

118 Suppose that a target t_s is assigned to each species s , indicating the minimum
 119 number of sites of H_s required for the protection of the species. Any subset H'
 120 of H which contains at least t_s vertices from H_s , i.e., $|H'_s| \geq t_s$, is called an
 121 s -cover. An S -cover is an s -cover for all s in S . We say that H' is an s -
 122 connected cover if $\langle H'_s \rangle$ (the subgraph of G_s induced by H'_s) has a connected
 123 component with t_s vertices. Assume that the graph represented in Figure 1
 124 is the graph G_s , with $H_s = \{1, 2, \dots, 16\}$, for some particular species s , and
 125 consider $V = \{1, 2, 3, 4, 5, 6, 10, 11, 12, 13\}$. The subgraph of G_s induced by V is the
 126 graph depicted in Figure 2, which has three connected components. If $t_s = 5$,
 127 V is an s -cover. However, since no component includes five or more sites, it is
 128 not an s -connected cover

129 We call H' an S -connected cover if it is an s -connected cover, for every s .
 130 Our goal is to find S -connected covers of minimum size.

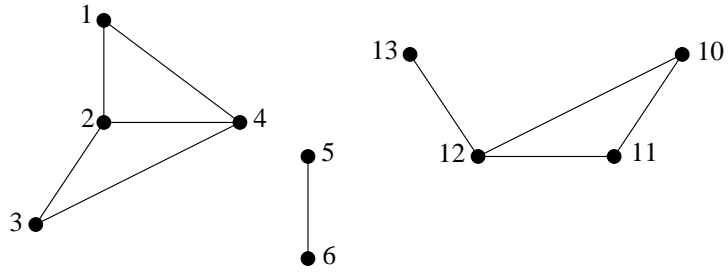


Figure 2: The subgraph of the graph of Figure 1 induced by $\{1, 2, 3, 4, 5, 6, 10, 11, 12, 13\}$

131 *2.2. A first model*

132 *2.2.1. Formulation*

133 Determining a reserve network with a minimum number of sites that has
 134 at least one site for every species is the set covering problem, a basic model in
 135 reserve network design (Possingham et al., 1993; Underhill, 1994; Reville and
 136 Williams, 2002).

137 A more general model, the multicovering problem (MCP), applies when the
 138 representation requirement of each species s in S is enlarged to a target number
 139 $t_s \geq 1$ sites, and a minimum size S -cover is desired.

140 The MCP is a well known problem in combinatorial optimisation (Hall and
 141 Hochbaum, 1986, 1992) and optimal solutions can be reached with integer linear
 142 programming techniques. To formulate the MCP as a 0-1 linear programming
 143 problem assign to each candidate site i of H a binary variable x_i indicating
 144 whether site i is included in the solution ($x_i = 1$) or not ($x_i = 0$). The MCP
 145 consists of

$$\min \sum_{i \in H} x_i \tag{1}$$

146 subject to:

$$\sum_{i \in H_s} x_i \geq t_s \quad s \in S, \tag{2}$$

147

$$x_i \in \{0, 1\} \quad i \in H. \tag{3}$$

148 If X is the set of sites in the solution, i.e., $X = \{i \in H : x_i = 1\}$, each
 149 inequality (2) ensures that $X_s = X \cap H_s$ is an s -cover. The objective function
 150 (1) seeks that the S -cover X has minimum size.

151 The MCP is a difficult problem (NP-hard, see Garey and Johnson (1979))
 152 but modern software can obtain optimal solutions for reasonably large instances
 153 (say a few thousand sites in H and some hundreds of species in S), in a few
 154 CPU seconds.

155 Typical MCP approaches result in a set of scattered sites. To ensure that the
 156 S -covers from the MCP are S -connected, constraints have to be added. These
 157 additional constraints can be derived from the following simple observation.
 158 Suppose K is a subset of sites such that no S -connected cover can be found in
 159 $H \setminus K$. Then, obviously every S -connected cover includes at least one site from
 160 K . Hence, and since a set of sites includes an S -connected cover if and only if
 161 it is also itself an S -connected cover, every inequality

$$\sum_{i \in K} x_i \geq 1 \quad K \subset H : H \setminus K \text{ is not an } S\text{-connected cover}, \quad (4)$$

162 is valid (i.e., satisfied by every S -connected cover), and together with (1),(2),(3)
 163 gives a 0-1 linear formulation for the minimum S -connected cover problem.

164 2.2.2. Algorithm

165 The large number of inequalities (4) heavily constrains the possibility of an
 166 integer linear programming (ILP) approach to deal with formulation (1),(2),(3),(4),
 167 even for problems of small dimensions.

168 To overcome this difficulty we developed an integer cutting algorithm. The
 169 algorithm starts by solving the MCP (1),(2),(3). If the S -cover $X = \{i \in H : x_i = 1\}$
 170 is S -connected, then X is an optimal S -connected cover and nothing
 171 more has to be done. Otherwise, consider the constraint (4) with $K = H \setminus X = \{i \in H : x_i = 0\}$,
 172 which is violated by the current solution x . If this constraint
 173 is added to the model, the next iteration either returns an optimal S -connected
 174 cover, or else identifies an inequality (4), which can be added to the model to
 175 eliminate the current solution from further consideration. This procedure can

176 be repeated until an S -connected cover is reached. Nevertheless, the number
 177 of constraints to be added until an S -connected cover is determined may make
 178 this procedure impractical. To accelerate the procedure we can work out the
 179 constraints according to the polyhedral combinatorics theory (see, for example,
 180 Pulleyblank (1983) or Schrijver (1995)). The same reasoning was used, in a
 181 similar context, by Cerdeira et al. (2005) to develop an integer cutting algorithm
 182 to identify the minimum number of sites satisfying the species representation
 183 targets in a unique connected component.

184 The key idea is to distinguish among the inequalities (4) those which are
 185 facet defining, from those which are implied by some other valid inequalities. If
 186 the algorithm determines an S -cover which is not S -connected, the inequality
 187 (4), with $K = \{i \in H : x_i = 0\}$, is violated by the current solution x . If the
 188 inequality is facet defining, it is a suitable cut to be added to the current model,
 189 and the algorithm proceeds to find a new S -cover. If the inequality is not facet
 190 defining, then some stronger inequalities exist that can cut more deeply the set
 191 of solutions of the current model, without violating any feasible solution. (We
 192 refer the reader to Cerdeira et al. (2005) for an explanation of facets and cuts
 193 in this context.) To illustrate this, consider a valid inequality (4) and suppose
 194 that for a given k in K , no S -connected cover exists in $(H \setminus K) \cup \{k\}$. Then,
 195 since every S -connected cover has to include at least one site from K different
 196 from k , $\sum_{i \in K \setminus \{k\}} x_i \geq 1$ is a valid inequality which clearly implies, and is not
 197 implied by $\sum_{i \in K} x_i \geq 1$, showing the latter is not facet defining.

198 The formulation (1), (2), (3), (4) is a particular case of a more general model
 199 for which Balas and Ng (1989) established the conditions for any valid inequality
 200 with coefficients 0 or 1 to be facet defining. (For technical reasons we assume
 201 that, for every site h in H , $H \setminus \{h\}$ is an S -connected cover.) Their result allow
 202 us to conclude that a valid inequality (4) is facet defining if and only if

- 203 **a)** for every k in K , $(H \setminus K) \cup \{k\}$ is an S -connected cover, and
- 204 **b)** for every h in $H \setminus K$ there is a site k in K such that $((H \setminus K) \setminus \{h\}) \cup \{k\}$
 205 is an S -connected cover.

206 Condition a) states that K is a minimal set (with respect to inclusion) for
 207 which the inequality (4) is valid, i.e., if K' is any proper subset of K , $\sum_{i \in K'} x_i \geq$
 208 1 is not valid. Above we showed that this condition is necessary for (4) to be
 209 facet defining.

210 To show that b) is also necessary, suppose that b) fails for some h in $H \setminus K$,
 211 i.e., whenever h is not selected, there is no S -connected cover that includes only
 212 one site of K . Hence, the inequality

$$x_h + \sum_{i \in K} x_i \geq 2 \quad (5)$$

213 is valid, as it states that if h is not selected (i.e., $x_h = 0$) at least two sites from
 214 K are needed to get an S -connected cover. Clearly, it is stronger than (4).

215 The constraint (5) belongs to the general type of inequalities

$$\sum_{i \in V} x_i \geq 2. \quad (6)$$

216 From the work of Balas and Ng (1989) it can also be derived when a valid
 217 inequality (6) is facet defining. To state the result we first introduce the 2-*cover*
 218 *graph* G_V associated to (6), which has vertex set V and an edge joining vertices
 219 v and u if and only if $(H \setminus V) \cup \{v, u\}$ is an S -connected cover. It follows from a
 220 result in Balas and Ng (1989) that a valid inequality (6) is facet defining if and
 221 only if

222 **c)** for every h in $H \setminus V$ there is a pair v, u in V such that $((H \setminus V) \setminus \{h\}) \cup \{v, u\}$
 223 is an S -connected cover, and

224 **d)** no connected component of graph G_V is bipartite.

225 A graph is bipartite if the vertices can be partitioned into two disjoint subsets
 226 A and B such that each edge connects a vertex from A to one from B .

227 When each of these conditions fails we can derive valid inequalities that
 228 imply (6).

229 If c) fails for some site h in $H \setminus V$, then

$$x_h + \sum_{i \in V} x_i \geq 3, \quad (7)$$

230 is valid. It expresses that any S -connected cover that does not includes h (i.e.,
 231 $x_h = 0$) has at least three sites from V .

232 Suppose there is a bipartite component of G_V with bipartition A and B .
 233 Then both inequalities

$$1.5 \sum_{i \in V \setminus (A \cup B)} x_i + \sum_{i \in A} x_i + 2 \sum_{i \in B} x_i \geq 3 \quad (8)$$

$$1.5 \sum_{i \in V \setminus (A \cup B)} x_i + 2 \sum_{i \in A} x_i + \sum_{i \in B} x_i \geq 3$$

234 are valid, and their sum equals three times (6). These inequalities state that no
 235 S -connected cover exists with two or less sites from V if these sites are either
 236 from A , or from B .

237 The integer cutting algorithm incorporates these polyhedral results.

238 In each step the current S -cover $X = \{i \in H : x_i = 1\}$ is tested to see if
 239 it is S -connected. If it fails, a set K , for which (4) is valid and that satisfies
 240 condition a), is identified. This is achieved with the following procedure. First,
 241 K is defined as $K = H \setminus X$. Next, for each species s , K_s is set to be $K_s = K \cap H_s$,
 242 and $H_s \setminus K_s$ is tested to see if it is an s -connected cover. If it is s -connected,
 243 set K is not modified and a new species s is considered. Otherwise, while there
 244 is a site k in K_s for which $(H_s \setminus K_s) \cup \{k\}$ is not an s -connected cover, K_s is
 245 updated accordingly to $K_s = K_s \setminus \{k\}$. When K_s is such that for every k in K_s ,
 246 $(H_s \setminus K_s) \cup \{k\}$ is an s -connected cover, set K is modified to become $K = K_s$,
 247 and the procedure continues with a new species s .

248 When all species have been considered, either X is an (minimum) S -connected
 249 cover, or else the resulting set K is such that the corresponding inequality (4)
 250 is valid, and verifies a).

251 If X is not an S -connected cover, the algorithm proceeds checking condition
 252 b) as follows. Take a site h of $H \setminus K$, and let K_h be initialized with $K_h = K$.
 253 For each species s such that $h \in H_s$, K_h is updated by removing those sites k for
 254 which $((H_s \setminus K) \setminus \{h\}) \cup \{k\}$ is not an s -connected cover. If $K_h = \emptyset$, condition
 255 b) fails, and inequality (5) holds. Otherwise, the procedure will continue with
 256 a new site h of $H \setminus K$. At the end either some h is found for which b) fails and

257 (5) holds, or else the inequality (4) is facet defining and is added to the current
 258 model, and the algorithm proceeds finding a new S -cover.

259 In the case condition b) fails, (5) is a valid inequality. To check whether it
 260 is facet defining the 2-cover graph associated to (5) is defined. Note that in this
 261 graph site h is adjacent to every site of K . Therefore, the graph is bipartite
 262 if and only if there are no edges linking pairs of sites in K . Let set E_K be
 263 initialized with $E_K := \{[u, v] : u \neq v \in K\}$, i.e., all the unordered pairs of
 264 different sites in K . For each species s such that $h \in H_s$, E_K is updated by
 265 removing those pairs of sites $[u, v]$ for which $((H_s \setminus K) \setminus \{h\}) \cup \{u, v\}_s$ is not an
 266 s -connected cover. If $E_K = \emptyset$, the 2-cover graph is bipartite, condition d) fails,
 267 and the inequality

$$2x_h + \sum_{i \in K} x_i \geq 3 \quad (9)$$

268 is valid. Actually, for this particular case (9) is the first inequality in (8) which
 269 implies the second one.

270 At the end of this procedure, either (9), which is stronger than (5), is added
 271 to the current model and the algorithm proceeds seeking a new S -cover, or else
 272 $V := K \cup \{h\}$ is defined and condition c) is examined.

273 To check c) we used the (non-bipartite) 2-cover graph G_V whose edge set
 274 (E_K together with $[h, u]$, for every u in K) has just been determined.

275 Take a site h of $H \setminus V$, and let E_h be initialized with every edge of the 2-cover
 276 graph G_V . For each species s such that $h \in H_s$, E_h is updated by removing
 277 those edges $[u, v]$ for which $((H_s \setminus V) \setminus \{h\}) \cup \{u, v\}_s$ is not an s -connected
 278 cover. If $E_h = \emptyset$, condition c) fails, and inequality (7) is valid. Otherwise, the
 279 procedure will continue with a new site h of $H \setminus V$. At the end either some
 280 h is found for which c) fails and (7) holds, or else the inequality (6) is facet
 281 defining. Accordingly, inequality (7) or (6) is added to the current model, and
 282 the algorithm proceeds to find a new S -cover.

283 This integer cutting algorithm, that will be referred to as IC, makes no use
 284 of the fact that, for every species s , there has to be a connected component of
 285 G_s with t_s sites selected. Next we present a formulation that takes this into

286 account.

287 2.3. A specialized model

288 2.3.1. Formulation

289 In this formulation additional variables Y_C^s associated with each connected
290 component C of the graph G_s of species s are considered.

291 Variables Y_C^s are used to ensure that the t_s sites required by the s -cover
292 inequalities (2) belong to the same component of G_s . This can be achieved with

$$\sum_{i \in C} x_i \geq t_s Y_C^s \quad C \in \mathcal{C}_s, s \in S, \quad (10)$$

$$\sum_{C \in \mathcal{C}_s} Y_C^s = 1 \quad s \in S, \quad (11)$$

$$Y_C^s \in \{0, 1\} \quad C \in \mathcal{C}_s, s \in S, \quad (12)$$

295 where \mathcal{C}_s denotes the set of connected components of G_s .

296 Conditions (10), (11), (12) constrain $X = \{i : x_i = 1\}$ to be an s -cover with
297 t_s sites in the component C for which $Y_C^s = 1$.

298 To ensure that the s -cover X is s -connected the following inequalities are
299 added

$$\sum_{i \in K} x_i \geq Y_C^s \quad K \subset C \in \mathcal{C}_s : \langle C \setminus K \rangle \text{ has no component with } t_s \text{ sites, } s \in S. \quad (13)$$

300 Constraints (13), together with (10), (11) and (12), state that for C such that
301 $Y_C^s = 1$, the subgraph induced by $X \cap C$ includes a connected component with
302 t_s sites.

303 Hence, finding a minimum size S -connected cover consists of (1) subject to
304 (3), (10), (11), (12), (13). We call this formulation the specialized model.

305 Compared with the formulation (1), (2), (3), (4), the specialized model ex-
306 hibits a feature which is very convenient for algorithmic proposes. The starting
307 point for the integer cutting procedure, which is the outcome of (1) subject
308 to (3), (10), (11), (12), is “closer” to an optimal S -connected cover than the
309 S -cover resulting from (1), (2), (3). More precisely, the lower bounds on the

310 minimum sizes of S -connected covers obtained from the specialized model with-
311 out the connectivity constraints (13) are likely to be significantly larger than
312 those resulting from dropping the connectivity constraints (4) in the first for-
313 mulation.

314 2.3.2. Algorithm

315 The algorithm, although being less involved, is similar to the algorithm IC
316 designed for the model described in Section 2.2.

317 First, the problem (1), (3), (10), (11), (12) is solved to produce an initial
318 S -cover $X = \{i \in H : x_i = 1\}$.

319 The algorithm proceeds checking, for each species s , if there is a connected
320 component C' of G_s such that the subgraph induced by $X \cap C'$ includes a
321 component with t_s sites. When this happens X is an s -connected cover, and
322 a new species s is considered. Otherwise, a set K is identified for which the
323 inequality (13) is valid, and minimal for inclusion. Set K is initialized with
324 $K = C \setminus X$, where C is the component marked by $Y_C^s = 1$, and it is sequentially
325 updated deleting from K some site k such that for $K := K \setminus \{k\}$ every connected
326 component of the subgraph $\langle C \setminus K \rangle$ has less than t_s sites. At the end of this
327 process K is such that (13) is valid, and K is minimal in the sense that, for
328 every k in K , some component of the subgraph induced by $(C \setminus K) \cup \{k\}$ has
329 at least t_s sites.

330 After considering all species, either X is an (minimum) S -connected cover
331 and nothing more has to be done, or else the valid connectivity inequalities (13)
332 previously determined are included in the current model and a new S -cover is
333 obtained.

334 We call this algorithm the sIC (specialized integer cutting).

335 2.4. Heuristic approach

336 We have also devised a heuristic for seeking minimum S -connected covers.

337 The heuristic makes use of the following procedure to turn an arbitrary S -
338 connected cover X into a minimal one. First, every site of X is marked as

339 non-considered. Then, some non-considered site i of X is selected. If $X \setminus \{i\}$ is
340 still an S -connected cover, X is updated with $X := X \setminus \{i\}$. Otherwise, site i is
341 marked as considered. When all sites in X have been considered, X is minimal.

342 In the implementation of this *minimal* procedure site i is selected with a
343 probability inversely proportional to its richness (number of species in i) among
344 the non-considered sites of the current X .

345 The heuristic, that we will refer to as GH, is a genetic type algorithm. Ge-
346 netic algorithms (Mühlenbein, 1997) start with an initial *population* of p feasible
347 solutions which are *mated* to produce children (i.e., other feasible solutions) that
348 inherit properties of their parents. The next *generation* will consist of elements
349 selected among those from the previous generation and their children.

350 To create each individual of the initial population, for each species s a set I_s
351 is initialized with a randomly selected site among the sites of the components
352 of G_s with at least t_s sites. While $|I_s| < t_s$, the set I_s is successively enlarged
353 adding a randomly selected site adjacent to some site in I_s . When all I_s are
354 defined, the minimal procedure turns $X = \cup_s I_s$ into a minimal S -connected
355 cover to become one of the p members of the initial population.

356 To form a child-bearing couple (F, M) , F is selected among the members
357 of the population with a probability inversely proportional to its size. Thus,
358 better S -connected covers are likely to be chosen. The mate M is selected in
359 the same way among the members of the population different from F .

360 The couple (F, M) generates an offspring O which is the outcome of the
361 minimal procedure on the input $X = F \cup M$. The algorithm counts the number
362 of existing replicates of O among the current population, and among the children
363 already created. If this number exceeds a given value r , the child is rejected
364 and is replaced by an S -connected cover obtained in the same way as each
365 individual of the initial population. The rationale for this is to avoid excessive
366 consanguinity that could lead the algorithm to get stuck in a population with
367 only a few different individuals.

368 Each new generation is formed by the p individuals, from among the previous
369 generation and their offspring, which have the lowest number of sites.

370 The number of generations is used as the stopping criterion.

371 For the computational tests reported in Section 3, the following specifications
372 of the algorithm were used. We defined $p = 100$ to be the size of the initial
373 population, which is also the number of individuals in each new generation. The
374 number of child-bearing couples is 50, which is half the size of the population.
375 Therefore, each new generation consists of the 100 elements with the lowest
376 number of sites, among the 100 individuals of the previous generation and the
377 50 offspring generated. We set $r = 10$ to be the upper bound on the number of
378 replicates allowed of each new child. We set the number of generations, which
379 is the stopping criterion, equal to 100. Thus, the algorithm stops at iteration
380 100, producing a list of 100, possibly not all distinct, S -connected covers.

381 2.5. *Simulating species distributions*

382 To compare the algorithms and test the limits of their computational prac-
383 ticability we generated virtual species distributions following the assumptions
384 of neutral community models.

385 Neutral community models are stochastic processes meant to explain, at least
386 in part, the patterns of distribution, abundance and diversity of ecological com-
387 munities. These models assume that all individuals have identical demographic
388 properties (Bell, 2001; Chave et al., 2002; Hubbell, 2001).

389 We implemented a neutral model in which individuals from a pool of 1000
390 species are distributed over 2500 sites, where each site is a cell from a 50×50
391 grid. At each iteration, and in each site:

- 392 (i) a single individual of each species is added to the community with proba-
393 bility $m = 0.001$;
- 394 (ii) each resident individual gives birth with probability $b = 0.5$ and dies with
395 probability $d = 0.5$;
- 396 (iii) with probability $u = 0.01$ each newborn moves to a random adjacent site
397 and continues to move until the criterion fails, and it settles in the site to
398 which it moved last;

399 (iv) if the number of individuals in the community exceeds $K = 100000$, excess
400 individuals are removed at random.

401 We started with 1000 individuals from each species in every site, and stopped
402 after 5000 iterations have been performed.

403 From the outcome of this procedure we produced several data sets for the
404 computational experiments we carried out. This was done by randomly select-
405 ing, for different values of n , an $n \times n$ square from the 50×50 grid, and subsets
406 S of 50, 100 and 150 species among the species represented on that square. We
407 then consider the species of S ordered by nondecreasing degrees of representa-
408 tion, and call the first 25% rare and the remaining 75% common. Representation
409 targets $t_s = 1$ are assigned to every common species s , and t_s equal to 3, 5 and
410 10, for each rare species s of S . Whenever G_s has no connected components
411 with t_s sites, then t_s is set to be equal to the maximum component size. Ten
412 instances were created for the same values of n , $|S|$ and t_s . This gives a total
413 of 90 instances with the same value of n .

414 We considered all species to have an equal level of dispersal in every instance,
415 and assume that species s can move directly between sites u and v from its
416 habitat sites H_s (i.e. $[u, v]$ is an edge of graph G_s) if and only if u and v have
417 a common edge or corner in the $n \times n$ grid. This allows the identification of
418 the connected components of each species directly in the grid, as the adjacency
419 relation describing the species s dispersal is the neighbourhood relation of sites
420 of H_s in the $n \times n$ grid.

421 3. Results

422 We performed computational tests to assess and compare the practicality
423 of the integer cutting algorithms IC and sIC, as well as the quality of the S -
424 connected covers determined by the heuristic GH.

425 The results reported here were obtained using an Intel Pentium IV, 2.8 GHz
426 with 504 MB RAM. The integer programming solver used was CPLEX 9.0.0.

427 With respect to the running times, both algorithms IC and sIC succeeded
428 in finding optimal S -connected covers in less than 60 CPU seconds for $n = 15$.
429 While, the integer cutting algorithm IC of Subsection 2.2 took an average of 7
430 CPU time seconds, the algorithm for the specialized model only once took more
431 than 1 second (1.3 sec).

432 Table 1 reports information on the running times of the algorithms IC and
433 sIC on instances where $n = 20$.

$ S $	t_s	< 15 sec.		15 – 60 sec.		1 – 30 min.		30 – 60 min.		not solved	
		IC	sIC	IC	sIC	IC	sIC	IC	sIC	IC	sIC
50	3	0	10	0	0	6	0	1	0	3	0
	5	0	8	1	2	4	0	0	0	5	0
	10	3	7	0	0	2	2	1	0	4	1
100	3	0	4	0	4	0	2	0	0	10	0
	5	0	5	0	2	0	3	0	0	10	0
	10	0	6	0	2	5	1	1	0	4	1
150	3	0	0	0	2	0	8	0	0	10	0
	5	0	3	0	1	0	5	0	0	10	1
	10	0	7	2	1	2	0	1	1	5	1

Table 1: Running times of the algorithms IC and sIC on 20×20 cells instances.

434 The first two columns refer to the number of species $|S|$ and the representa-
435 tions target t_s for each of the 10 instances considered in each row. The columns
436 *not solved* indicate, for each algorithm, the number of instances for which com-
437 putations were not finished at the end of 1 hour CPU time. The other columns
438 report the number of instances for which solutions were found within the time
439 indicated in the first row of the corresponding column. The four instances for
440 which the sIC algorithm did not succeed in finding minimum S -connected cov-
441 ers, were not solved either within 1 hour by the algorithm design for the first
442 model.

443 When we increased the size of the $n \times n$ grid to values of $n \geq 25$, the

444 algorithm IC, based on the initial model, was incapable of solving any instance
445 in less than 1 CPU hour.

446 For $n = 25$ the algorithm working on the specialized model also did not solve
447 34 of the 90 instances within 1 hour. However, 30 instances were solved within
448 no more than 1 CPU minute, and 22 more in less than 30 minutes.

449 With $n = 30$ no instance with 150 species was solved in less than 1 hour,
450 and only for those instances with 50 species minimum S -connected covers were
451 found within 1 minute.

452 Information about the running times of the algorithm for the specialized
453 model for $n = 25$ and $n = 30$ is given in Appendix A tables S1 and S2, respec-
454 tively.

455 For $n = 15$, except in one case, the list produced by the genetic heuristic on
456 each of the 90 instances included minimum size S -connected covers. The time
457 spent on the longest run was about 2.6 seconds. The heuristic did not find any
458 optimal solution in an instance with $|S| = 50$ species and representation target
459 $t_s = 3$. However, among the 100 solutions produced by the heuristic, 86 distinct
460 S -connected covers with 31 sites were found when 30 sites is known to be the
461 minimum size.

462 Two measures were used to assess the quality of the solutions obtained by
463 the GH algorithm on larger instances. Let O and H be the sizes of a minimum
464 S -connected cover and of the best solutions produced by the GH heuristic,
465 respectively. The measures are:

466 $\#opt$ - the number of instances for which H and O coincide, and

467 mre - the mean relative error, defined as the mean of the ratios $\frac{H-O}{O}$,

468 with respect to the 10 instances having the same values of n , $|S|$ and t_s .

469 In several instances with $n \geq 20$, the algorithms IC and sIC did not succeed
470 in identifying S -connected covers before the computations were interrupted, at
471 the end of the pre-established CPU time limit of 1 hour. Hence, we are not sure
472 about the true values of O . In those cases the value of O used to calculate $\#opt$
473 and mre is the size of the (unconnected) S -cover of the last iteration of the sIC

474 algorithm based on the specialized model. In such situations, the values of $\#opt$
 475 and mre may lead to a pessimist judgment of the accuracy of the heuristic in
 476 finding good solutions.

477 The values of $\#opt$ and mre for the instances with $n = 20$, $n = 25$ and $n = 30$
 478 are presented in Table 2. The CPU times varied between 3 and 34 seconds. The
 479 table does not include the values of the instances for which $n = 30$ and $|S| = 150$.
 480 For several of these instances, 1 hour was not sufficient for CPLEX to identify
 481 the initial S -cover that minimises (1) subject to (3), (10), (11), (12), and we
 482 have no alternative reliable lower bounds on the true values of O to evaluate,
 483 in a minimally credible way, the performance of the heuristic.

$ S $	t_s	mre	$\#opt$	mre	$\#opt$	mre	$\#opt$	
50	3	0.022	6	0.119	7	0.053	3	
	5	0.002	9	0.039	3	0.057	3	
	10	0.001	9	0.005	8	0.037	2	
100	3	0.023	5	0.054	2	0.090	1	
	5	0.012	5	0.039	4	0.079	0	
	10	0.003	6	0.018	0	0.086	0	
150	3	0.022	4	0.079	0	—	—	
	5	0.021	2	0.074	0	—	—	
	10	0.003	4	0.033	0	—	—	
			$n = 20$			$n = 25$	$n = 30$	

Table 2: Values of $\#opt$ and mre for the instances with 20×20 , 25×25 and 30×30 cells.

484 4. Discussion

485 The sIC algorithm based on the specialized model (1), (3), (10), (11), (12),
 486 (13), with more variables, is clearly better than the approach for the first for-
 487 mulation (1), (2), (3), (4). The ability to produce lower bounds on the sizes of
 488 the optimal S -connected covers significantly larger than those obtained using

489 the first formulation, is decisive for the superiority of the sIC algorithm. This
 490 is a consequence of the requirement incorporated in the specialized model forc-
 491 ing, for each species s , the existence of a connected component with t_s sites in
 492 every intermediate S -cover. As a result, the initial S -covers obtained with the
 493 specialized model are much “closer” to the optimal S -connected covers than the
 494 minimum S -cover solutions of (1), (2), (3). (See in Figure S2 of Appendix B the
 495 initial S -covers produced by algorithms IC and sIC in an instance with $n = 25$,
 496 $|S| = 100$ and $t_s = 5$.)

497 Whilst the use of the IC algorithm appears to be limited to instances with
 498 $400 = 20 \times 20$ sites, the bounds for the practicability of the sIC algorithm are
 499 instances consisting of $625 = 25 \times 25$ to $900 = 30 \times 30$ sites, and approximately
 500 50 species, sizes that can be considered quite reasonable for seeking guaranteed
 501 optimality for such an involved problem.

502 The genetic heuristic approach largely surpasses these limitations (it took
 503 139.9 seconds to run on the entire $2500 = 50 \times 50$ cells grid, with $|S| = 150$,
 504 $t_s = 10$), at the cost of guaranteed optimality. The computations of the GH
 505 algorithm are dominated by the *minimal* procedure designed to turn minimal
 506 (with respect to inclusion) any given S -connected cover X . Each execution
 507 entails, for each site $i \in X$ and each species s , with $i \in H_s$, defining the
 508 connected components of the graph $\langle (X \setminus \{i\})'_s \rangle$. To identify connected
 509 components we used the (linear time) approach described in Subsection 2.1,
 510 which is therefore called a number of times of order $|S||X|$ in each execution of
 511 the *minimal* procedure. The *minimal* procedure is used whenever a new child
 512 is generated, and when creating each member of the initial population. Thus,
 513 it is executed a number of times which is the order of the size of the initial
 514 population plus the number of offspring in each generation multiplied by the
 515 number of generations. This gives $100 + 50 \times 100 = 5100$ in our implementation.

516 Nevertheless, the GH heuristic is revealed to be capable of finding good solu-
 517 tions with no excessive computational weight. On average, for the 240 instances
 518 for which the *mre* values were calculated, the sizes of the S -connected covers
 519 produced by the GH algorithm do not exceed 4% of the optimal sizes. Actually,

520 the rate of 4% undervalues the quality of the solutions obtained, since the lower
521 bounds used to estimate O in expression mre are likely to be considerably lower
522 than the sizes of the corresponding minimum S -connected covers.

523 5. Conclusion and final remarks

524 We have presented three algorithms to solve a reserve design problem that
525 has been overlooked to date. Two integer cutting algorithms, that guarantee
526 optimality, and a genetic heuristic. Of the two integer cutting algorithms, the
527 specialized formulation is superior, being able to find solutions for most of the
528 instances assessed and in faster computational times. Nonetheless, both al-
529 gorithms have limitations regarding the magnitude of the instances for which
530 solutions can be found. It is unlikely that the most efficient of the two algo-
531 rithms is able to handle instances with more than 900 sites and more than 50
532 species. The heuristic algorithm surpasses these limitations. An instance with
533 2500 sites and 50 species was run in less than 2.5 CPU minutes. Despite run-
534 ning reasonably quickly, the solutions obtained are of good quality. Only in one
535 case, among 240 instances, was the size of the solution produced by the heuristic
536 larger 10% by of the minimum size.

537 We have considered that all sites are equally relevant, i.e., all variables x_i
538 have a coefficient equal to 1 in the objective function. However, in some situa-
539 tions it may be desirable to distinguish between different reserve networks with
540 the same number of sites (Rodrigues et al., 2000), for example those of differing
541 total area (where sites vary in size) or monetary value. In this case the objective
542 function (1) is replaced by

$$\min \sum_{i \in H} c_i x_i \quad (14)$$

543 where c_i is the cost (area, monetary value, or some other adequate measure) of
544 site i .

545 With respect to the integer cutting algorithms, this modification introduces
546 nothing more than giving the objective function (14) instead of (1) to the integer
547 programming solver.

548 The heuristic can also be easily be modified to handle (14) instead of (1).
549 In the *minimal* procedure, site i could be selected, among the sites in X , with
550 probability directly proportional to c_i . This will make sites with larger costs
551 more likely to be eliminated from the resulting minimal S -connected covers. In
552 addition, the selection of child-bearing couples, and the members of each new
553 generation should be made according to their costs instead of their sizes.

554 It may also be relevant, for conservation purposes, to require the representa-
555 tion of more than one population (or metapopulation) of certain species. This
556 means that instead of requiring for species s only one connected component with
557 t_s sites, it may be desirable that the S -connected covers have $n_s > 1$ different
558 components, each with at least t_s . It is not at all obvious how to incorporate this
559 generalization in the 0-1 formulations of Section 2. However, it is straightfor-
560 ward to have the genetic heuristic working on this generalization. It amounts to
561 generalizing the concept of the S -connected cover to incorporate the additional
562 requisites.

563 The heuristic can also accommodate other realistic assumptions, such as the
564 need for inclusion of a certain number of sites with specific functional purposes
565 (e.g. breeding, shelter and feeding grounds) for each species.

566 In every instance, we have assumed equal targets and equal dispersal dis-
567 tances for all species. These could be variable, without compromising the per-
568 formance of the algorithms. However, we realise that for practical conservation,
569 the difficulties are in setting meaningful targets and, in many cases, assessing
570 the dispersal capabilities for a large number of species. Nonetheless, reserve
571 designs that consider species specific connectivities should be encouraged, and
572 here we provide some tools to achieve such designs

573 *Acknowledgements*

574 We are grateful to Diogo Alagador for assistance and discussion.

575 **References**

- 576 Alagador, D., Cerdeira, J.O., 2007. Designing spatially-explicit reserve networks
577 in the presence of mandatory sites. *Biological Conservation* 137, 254-262.
- 578 Balas, E., Ng, S.M., 1989. On the set covering polytope: I. all facets with
579 coefficients in $\{0,1,2\}$. *Mathematical Programming* 43, 57-69.
- 580 Bell, G., 2001. Neutral macroecology. *Science* 293, 2413-2418.
- 581 Bondy, J.A., Murty, U.S.R., 1976. *Graph Theory with Applications*, North-
582 Holland, New York.
- 583 Cabeza, M., 2003. Habitat loss and connectivity of reserve networks in proba-
584 bility approaches to reserve design. *Ecology Letters* 6, 665-67.
- 585 Cabeza, M., Araújo, M.B., Wilson, R.J., Thomas, C.D., Cowley, M.J.R. Moilanen, A., 2003. Combining probabilities of occurrence with spatial reserve design. *Journal of Applied Ecology* 41, 252-262.
- 588 Cabeza, M., Moilanen, A., 2001. Design of reserve networks and the persistence
589 of biodiversity. *Trends in Ecology and Systematics* 16, 242-247.
- 590 Cabeza, M., van Teeffelen, A., 2009. Strategies of reserve selection. *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd., Chichester.
- 592 Cerdeira, J.O., Pinto, L.S., 2005. Requiring connectivity in the set covering
593 problem. *Journal of Combinatorial Optimization* 9, 35-47
- 594 Cerdeira, J.O., Gaston, K.J., Pinto, L.S., 2005. Connectivity in priority area
595 selection for conservation. *Environmental Modelling and Assessment* 10, 183-
596 192.
- 597 Chave, J., Muller-Landau, H.C., Levin, S. A., 2002. Comparing classical com-
598 munity models: theoretical consequences for patterns of diversity. *American Naturalist* 159, 1-23.
- 599

- 600 Fischer, D., Church, R.L., 2003. Clustering and compactness in reserve site
601 selection: An extension of the biodiversity management area selection model.
602 *Forest Science* 49, 555-65.
- 603 Fuller, T., Mungua, M., Mayfield, M., Snchez-Cordero, V., Sarkar, S., 2006.
604 Incorporating connectivity into conservation planning: A multi-criteria case
605 study from central Mexico. *Biological Conservation* 133, 131–142.
- 606 Garey, M.R., Johnson, D.S., 1979. *Computers and Intractability: A Guide to*
607 *the Theory of NP Completeness*, W.H. Freeman & Company, San Francisco.
- 608 Hall, N.G., Hochbaum, D.S., 1986. A fast approximation algorithm for the mul-
609 ticovering problem. *Discrete Applied Mathematics* 15, 35–40.
- 610 Hall, N.G., Hochbaum, D.S., 1992. The multicovering problem. *European Jour-*
611 *nal of Operational Research* 62, 323–339.
- 612 Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeog-*
613 *raphy*, Princeton University Press, Princeton.
- 614 McDonnell, M.D., Possingham, H.P., Ball, I.R., Cousins, E.A., 2002. Mathemat-
615 ical methods for spatially cohesive reserve design. *Environmental Modeling*
616 *and Assessment* 7, 107-14.
- 617 Millennium Ecosystem Assessment (MEA), 2005. *Ecosystems and Human Well-*
618 *Being: Biodiversity Synthesis*. World Resources Institute: Washington, DC.
- 619 Moilanen, A., 2005. Reserve selection using nonlinear species distribution mod-
620 els. *American Naturalist* 165, 695–706.
- 621 Moilanen, A., Cabeza, M., 2002. Single-species dynamic site selection. *Ecological*
622 *Applications* 12, 913–926.
- 623 Mühlenbein, H., 1997. Genetic algorithms, Aarts, E., Lenstrain, J.K. (Eds.),
624 *Local Search in Combinatorial Optimization*. John Wiley & Sons, pp. 137–
625 171.

- 626 Nicholls, A.O., Margules, C.R., 1993. An upgraded reserve selection algorithm.
627 *Biological Conservation* 64, 165-169.
- 628 Nicholson, E., Westphal, M.I., Frank, K., Rochester, W.A., Pressey, R.L., Lin-
629 denmayer, D., Possingham, H.P., 2006. A new method for conservation plan-
630 ning for the persistence of multiple species. *Ecology Letters* 9, 1049–1060.
- 631 Önal, H., Briers, R.A., 2002. Incorporating spatial criteria in optimum reserve
632 network selection. *Proceedings of the Royal Society of London B* 269, 1486-
633 1491.
- 634 Önal, H., Briers, R.A., 2005. Designing a conservation reserve network with min-
635 imal fragmentation: a linear integer programming approach. *Environmental*
636 *Modeling and Assessment* 10, 193–202, 2005.
- 637 Önal, H., Wang, Y., 2008. A graph theory approach for designing conservation
638 reserve networks with minimal fragmentation. *Networks* 52, 142–152.
- 639 Possingham, H., Day, J., Goldfinch, M., Salzborn, F., 1993. The mathematics
640 of designing a network of protected areas for conservation, in: Sutton, D.,
641 Cousins E., Pierce, C. (Eds.), *Decision Sciences, Tools for Today, Proceedings*
642 *of the 12th Australian Operations Research Conference, ASOR, Adelaide*, pp.
643 536-545, 1993.
- 644 Possingham, H., Ball, I., Andelman, S., 2000. Mathematical methods for iden-
645 tifying representative reserve networks, in: Ferson, S., Burgman, M. (Eds.)
646 *Quantitative Methods for Conservation Biology*. Springer, New York, pp. 291-
647 306.
- 648 Pressey, R.L., Cabeza, M., Watts, M.E., Cowling, R.M., Wilson, K.A., 2007.
649 Conservation planning in a changing world. *Trends in Ecology and Evolution*
650 22, 583-592.
- 651 Pulleyblank, W.R., 1983. Mathematical programming - The state of the art
652 in: Bachem, A., Grötschel, M., Korte, B. (Eds.), *Polyhedral Combinatorics*,
653 Springer, Berlin, pp. 312–345.

- 654 ReVelle, C., Williams, J.C., 2002. Reserve design and facility siting, in: Drezner,
655 Z., Hamacher, H.W. (Eds.), *Facility Location*, Springer, Berlin, pp. 307–328.
- 656 Rodrigues, A.S.L., Cerdeira, J.O., Gaston, K.J., 2000. Flexibility, efficiency,
657 and accountability: adapting reserve selection algorithms to more complex
658 conservation problems. *Ecography* 23, 565–574.
- 659 Rodrigues, A.S.L., Gregory, R., Gaston, K.J., 2000. Robustness of reserve se-
660 lection procedures under temporal species turnover. *Proceedings of the Royal
661 Society of London B* 267, 49-55.
- 662 Rothley, K.D., 1999. Designing bioserve networks to satisfy multiple, conflict-
663 ing demands. *Ecological Applications* 9, 741–750.
- 664 Schrijver, A., 1995. Polyhedral combinatorics, in: Graham, R.L., Grötschel, M.,
665 Lovász, L., (Eds.), *Handbook of Combinatorics*, Elsevier, Amsterdam, pp.
666 1649–1704.
- 667 Sessions, J., 1992. Solving for habitat connections as a Steiner network problem.
668 *Forest Science* 38, 203-207.
- 669 Shirabe, T., 2005. A model of contiguity for spatial unit allocation. *Geographical
670 Analysis* 37, 2-16.
- 671 Underhill, L., 1994. Optimal and suboptimal reserve selection algorithms. *Bio-
672 logical Conservation* 70, 85-87.
- 673 van Teeffelen, A., Cabeza, M., Moilanen, A., 2006. A comparison of reserve
674 selection algorithms with respect to the predicted persistence of species in
675 the reserve. *Biodiversity and Conservation* 15, 899–919.
- 676 Williams, J.C., 1998. Delineating protected wildlife corridors with multi-
677 objective programming. *Environmental Modeling and Assessment* 3, 77-86.
- 678 Williams, J.C., ReVelle, C. S., Levin, S.A., 2005. Spatial attributes and reserve
679 design models: A review. *Environmental Modeling and Assessment* 10, 163–
680 181.

681 Wilson, K.A., Pressey, R.L., Newton, A.N., Burgman, M.A., Possingham, H.P.,
682 Weston, C.J., 2005. Measuring and incorporating vulnerability into conserva-
683 tion planning. *Environmental Management* 35, 527-543.

684 **A. Running times of the sIC algorithms**

$ S $	t_s	< 15 sec.	15 – 60 sec.	1 – 30 min.	30 – 60 min.	not solved
50	3	5	5	0	0	0
	5	5	3	2	0	0
	10	4	2	3	0	1
100	3	0	3	5	2	0
	5	0	1	6	2	1
	10	0	1	2	0	7
150	3	0	0	0	0	10
	5	0	0	0	0	10
	10	0	1	4	0	5

Table S1: Running times of the sIC algorithm on 25×25 cells instances.

$ S $	t_s	< 15 sec.	15 – 60 sec.	1 – 30 min.	30 – 60 min.	not solved
50	3	6	1	3	0	0
	5	3	6	0	0	1
	10	3	0	5	0	2
100	3	0	0	3	1	6
	5	0	0	0	2	8
	10	0	0	0	0	10
150	3	0	0	0	0	10
	5	0	0	0	0	10
	10	0	0	0	0	10

Table S2: Running times of the sIC algorithm on 30×30 cells instances.

685 **B. Results of an instance with $n = 25$, $|S| = 100$ and $t_s = 5$**

686 Here we give some results obtained with an instance I with $625 = 25 \times 25$
687 sites, the number of species $|S| = 100$ and target representations $t_s = 5$ for rare
688 species.

689 A minimum size S -connected cover, consisting of 51 sites, is depicted in
690 Figure S1.

691 The initial S -covers produced by the algorithms IC and sIC are presented
692 in Figure S2a) and b), respectively.

693 The best S -connected covers obtained by the heuristic GH have 53 sites.
694 One of these solutions is depicted in Figure S3.

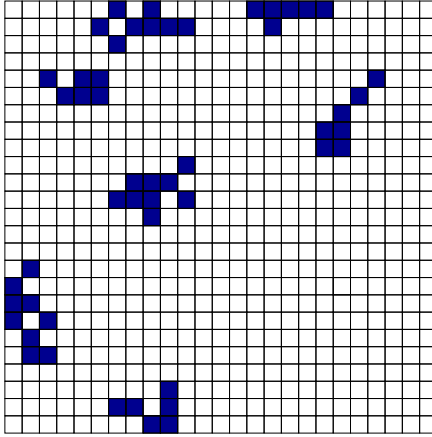


Figure S1: A 51 sites minimum size S -connected cover for instance I .

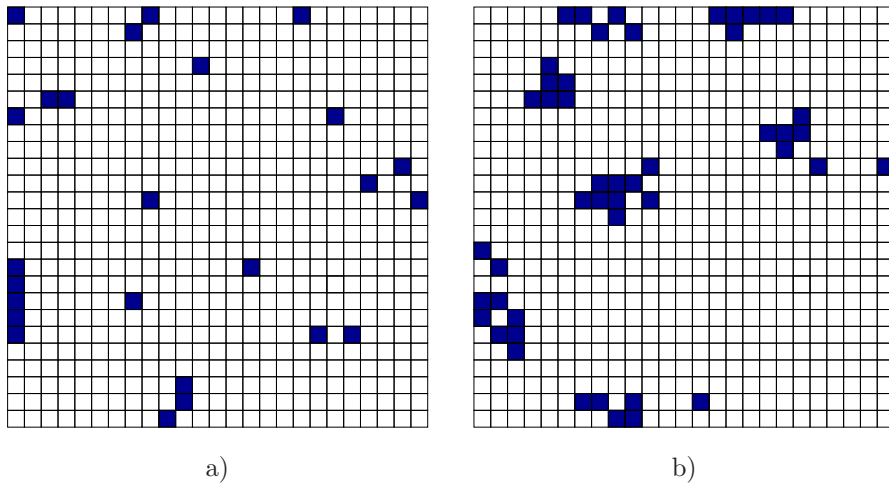


Figure S2: a) The initial 25 sites minimum size S -cover obtained with the IC algorithm, and b) The initial 48 sites S -cover obtained with the sIC algorithm, on instance I .

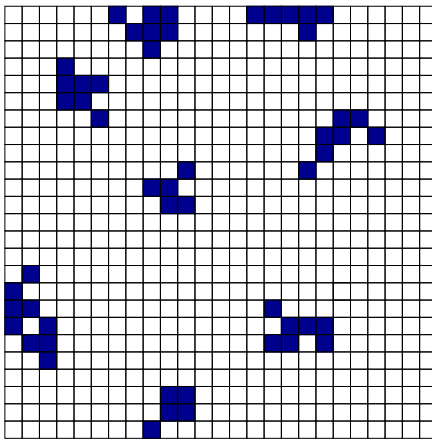


Figure S3: A 53 sites S -connected cover obtained by the heuristic GH on instance I .