

Multi-objective optimization in systematic conservation planning and the representation of genetic variability among populations

S. Schlottfeldt¹, M.E.M.T. Walter¹, A.C.P. L.F. Carvalho², T.N. Soares³,
M.P.C. Telles³, R.D. Loyola⁴ and J.A.F. Diniz-Filho⁴

¹Departamento de Ciência da Computação,
Instituto de Ciências Exatas, Universidade de Brasília, Brasília, DF, Brasil

²Departamento de Ciência da Computação, Universidade de São Paulo,
São Carlos, SP, Brasil

³Departamento de Biologia Geral, Instituto de Ciências Biológicas,
Universidade Federal de Goiás, Goiânia, GO, Brasil

⁴Departamento de Ecologia, Instituto de Ciências Biológicas,
Universidade Federal de Goiás, Goiânia, GO, Brasil

Corresponding author: S. Schlottfeldt
E-mail: shanass@unb.br

Genet. Mol. Res. 14 (2): 6744-6761 (2015)

Received August 21, 2014

Accepted February 13, 2015

Published June 18, 2015

DOI <http://dx.doi.org/10.4238/2015.June.18.18>

ABSTRACT. Biodiversity crises have led scientists to develop strategies for achieving conservation goals. The underlying principle of these strategies lies in systematic conservation planning (SCP), in which there are at least 2 conflicting objectives, making it a good candidate for multi-objective optimization. Although SCP is typically applied at the species level (or hierarchically higher), it can be used at lower hierarchical levels, such as using alleles as basic units for analysis, for conservation genetics. Here, we propose a method of SCP using a multi-objective approach. We used non-dominated sorting genetic algorithm II in order to identify the smallest set of local populations of *Dipteryx alata* (baru) (a Brazilian Cerrado species) for conservation, representing

the known genetic diversity and using allele frequency information associated with heterozygosity and Hardy-Weinberg equilibrium. We worked in 3 variations for the problem. First, we reproduced a previous experiment, but using a multi-objective approach. We found that the smallest set of populations needed to represent all alleles under study was 7, corroborating the results of the previous study, but with more distinct solutions. In the 2nd and 3rd variations, we performed simultaneous optimization of 4 and 5 objectives, respectively. We found similar but refined results for 7 populations, and a larger portfolio considering intra-specific diversity and persistence with populations ranging from 8-22. This is the first study to apply multi-objective algorithms to an SCP problem using alleles at the population level as basic units for analysis.

Key words: Conservation planning; Multi-objective optimization; Metaheuristics; Genetic variability; Biodiversity

INTRODUCTION

The current biodiversity crisis is forcing scientists to develop systematic strategies for effectively achieving conservation goals. The overall underlying principle of this strategy lies in systematic conservation planning (SCP), which involves a series of decisions in order to determine the most cost-effective method of investing in conservation measures (Margules and Pressey, 2000). The overall reasoning is to develop a protocol in which a set of conservation targets and goals are defined and achieved in the most objective and rational manner possible.

Although SCP is typically applied to species or even hierarchically higher levels (Brooks et al., 2004; Pressey, 2004), it is possible to solve a series of problems at much lower hierarchical level. Alleles from molecular analysis at the population level can be used as basic units in the context of conservation genetics (Diniz-Filho and Telles, 2002; Diniz-Filho et al., 2012) based on the idea of intra-specific conservation prioritization. This idea has its roots in the early 1990's in the debate of evolutionary significant units or management units for conservation (Fraser and Bernatchez, 2001). Although the definition of evolutionary significant units or management units provides the concept of intra-specific units (e.g., subspecies or local varieties) that can be used as conservation targets, it is not able to deal with continuous genetic variation at the species level (Diniz-Filho and Telles, 2002) and, more importantly, does not provide a method for identifying intra-specific variation units or components that should be prioritized in the context of SCP.

The complementarity concept is in the foundation of SCP and is mathematically modeled by the set covering problem, a classical problem in algorithm complexity theory that was shown to be NP-complete (Cormen et al., 2001). Understanding that a problem is NP-complete provides an indication of the difficulty of the problem. In general, it is not difficult to verify whether an answer to an NP-complete problem is correct. However, whether the solution is efficient must be determined by testing all possible options until finding one that solves the problem correctly. NP-complete problems arise in several real-world applications and, in practice, knowing that a problem is NP-complete can prevent the spending of time attempting to determine a polynomial-time algorithm to solve it exactly when such an algorithm likely does not exist.

Independently of the hierarchical level, the SCP problem can be stated as follows: to select a set of sites (among several available sites) to minimize the cost of conservation and, at the same time, maximize the natural feature representation, which can be modeled by the minimum set covering problem. This approach presents at least 2 conflicting objectives, as one is usually interested in the conservation of some biological features (e.g., alleles, species, or any other biological unit) subjected to conflicting interests in land use (e.g., human population living in the region, price of land, agricultural potential of the area, or probability of habitat loss). In this context, the conflicting objectives make the problem a perfect candidate for multi-objective optimization. Furthermore, other parameters can be considered by adding socio-economical costs to the areas or by minimizing their spatial aggregation; this adds dimensions to SCP, which is already multi-objective in its origin. Indeed, several real world problems involve simultaneous optimization of multiple conflicting objectives, which should be analyzed as independent dimensions rather than combined into a single weighted function.

These optimization problems with more than one objective are referred to as vector optimization or multi-objective problems (Zitzler et al., 2002; Coello-Coello et al., 2007). In these cases, there is no single optimal solution, but rather a set of solutions that should be considered to be equivalent in the absence of information regarding the relevance of each objective relative to the others (Fonseca and Fleming, 1995). Such solutions, known as non-dominated, are optimal in the sense that none can be declared the best when all objectives are considered (Zitzler et al., 2002). These solutions are called Pareto optimal solutions, and the graph of the solutions form the Pareto front (Coello-Coello et al., 2007).

Multi-objective evolutionary algorithms (MOEA) have been successfully applied for multi-objective problems. The positive aspects of MOEA include efficient solution space exploration, parallelism, ability to escape of local optima, capacity to handle complex problems for which it is not possible (or at least it is difficult) to obtain a detailed description, and they are less susceptible to the shape or continuity of functions (Coello-Coello et al., 2007).

In SCP, the development of algorithms and tools for decision support began in the early 1980s (Sarkar, 2012) and became an important element in conservation biology research. Several approaches have been suggested over the past decades, ranging from a simple scoring system to more complex optimization techniques (Table 1).

In their origin, such algorithms sequentially select complementary sites until all species are represented in typical greedy algorithm behavior. However, greedy algorithms are not guaranteed to identify optimal solutions (Possingham et al., 2000). Nevertheless, this class of algorithms has distinguished importance in the development of algorithms and tools for SCP. In contrast, the exact approach (which ensures the production of optimal solutions) was initially discussed by Cocks and Baird in 1989 (Sarkar, 2012). However, as SCP is an NP-complete problem, even the available software packages computing exact algorithms cannot solve some large data sets (Pressey et al., 1996), which is a common limitation in the SCP context. Because of these characteristics, another approach used for SCP involves metaheuristics, a method for solving an optimization problem using a combination of random choices and historical knowledge of previous results computed by the method, such that the heuristic explores the solution space. However, it is worth noting that this technique does not guarantee optimal solutions. The mainly metaheuristics used for SCP include simulated annealing [SPEXAN (Ball, 2000), SITES (Possingham et al., 2000), and Marxan (Ardron et al., 2010)], as well as the tabu search [ConsNet (Ciarleglio, 2010)].

Table 1. Methods and associated strategies proposed to solve the SCP problem.

Methods	Solution strategies				
	Scoring	Greedy	LP/ILP*	Rules	Metaheuristics
					Simulated annealing Tabu search Evolutionary
Kirkpatrick's Algorithm (Kirkpatrick, 1983)					
Ackery and Vane-Wright's Algorithm (Ackery and Vane-Wright, 1984)	X	X			
Margules and Nicholls' Algorithm (Margules and Nicholls, 1987) ¹		X		X	
Margules, Nicholls and Pressey's Algorithms (Margules et al. 1988) ²		X		X	
Nicholls and Margules' Algorithm (Nicholls and Margules, 1993) ³		X		X	
Rebello and Siegfried's Algorithm (Rebello and Siegfried, 1990) ⁴		X		X	
Exact algorithms (software packages as LINDO, CPLEX, OSL) (Underhill, 1994; Arthur et al., 1997)			X		
C-Plan (Pressey et al., 1996) ⁵		X		X	
SPEXAN (Ball, 2000)					
SITES (Possingham et al., 2000)				X	
Marxan (Possingham et al., 2000; Ardron et al., 2010) ⁷				X	
ResNet (Garson et al., 2002) ⁶		X		X	
Target (Barton et al., 2004)		X		X	
Zonation (Moilanen, 2005) ⁸		X			
MultiSync (Moffett et al., 2005) ⁹				X	
ConsNet (Clarke et al., 2010) ¹⁰				X	
SCP-NSGA-II (Schlotfeldt et al., 2012)				X	X

Heuristics to improve results: ¹Tiebreak criterion (area). ²Tiebreak criterion (rarity, frequency). ³Tiebreak criterion (rarity, distance, area). ⁴Tiebreak criterion (richness). ⁵Tiebreak criterion. ⁶Tiebreak criterion (complementarity, proximity). ⁷Maximization of compactness, minimization of perimeter. ⁸Connectivity, BLQ (boundary quality penalty), BLP (boundary length penalty). ⁹Multicriteria analysis. ¹⁰MASTS (modular abstract self-learning tabu search), DNS (dynamic neighborhood selection), RBO (rule-based objectives). *LP/ILP: Linear programming and integer linear programming.

In this study, we propose a more sophisticated and general solution for the SCP problem based on multi-objective optimization, which allowed us to cope with more than one objective. This provided more flexibility by including additional objectives, adding more complexity, and increasing the power of decision. In particular, we used non-dominated sorting genetic algorithm II (NSGA-II, a state-of-the-art MOEA) to search for optimal solutions. Our hypothesis was that NSGA-II could identify the smallest set of local populations of *Dipteryx alata* (also known as baru) that should be conserved to represent the known genetic diversity of this Brazilian Cerrado species, thus focusing on the *in situ* strategy. However, rather than simply representing the known alleles, the proposed approach begins with allele frequency information and incorporates information about heterozygosity in the local population and Hardy-Weinberg equilibrium. By including these 2 characteristics, local populations can be better represented in terms of their genetic diversity, allowing identification of sets of populations with a higher probability of persistence overtime. This is the first study to apply multi-objective optimization algorithms to an SCP problem with more than 2 objectives using alleles from molecular analysis at the population level as basic units.

MATERIAL AND METHODS

Data

We used data from *D. alata* (a Fabaceae tree species widely distributed and endemic to Brazilian Cerrado) consisting of 55 alleles from 9 microsatellite loci (Table 2) coding for a total of 642 individual trees sampled in 25 local populations distributed throughout species' geographical range (Figure 1), with sample sizes within the local populations ranging from 12-32 (Tables 3 and 4) (Diniz-Filho et al., 2012; Soares et al., 2012).

Based on the sampled data, we produced 4 matrices used as input for our MOEA:

1) Matrix A: an allele-by-site presence-absence matrix. Here, the population can be understood as a site, as each sampled tree was GPS-georeferenced. In matrix $A_{p \times a}$, $p = 25$ (populations), $a = 55$ (alleles), and a_{ij} represents the occurrence of allele j in population i .

2) Matrix B: allele frequencies within local populations, where allele frequency over population was normalized in order to minimize possible distortions due to different numbers of individuals sampled among populations (i.e. differences in the sample sizes). In matrix $B_{p \times a}$, $p = 25$ (populations), $a = 55$ (alleles), and b_{ij} is the normalized frequency of allele j in population i .

3) Matrix C: heterozygosity per population per locus. In matrix $C_{p \times l}$, $p = 25$ (populations), $l = 9$ (loci), and c_{ij} represents the amount of heterozygosity of locus j in population i .

4) Matrix D: we used the Hardy-Weinberg equilibrium (HWE) concept which states that in the absence of evolutionary pressure, allele and genotype frequencies will remain the same along generations. The matrix D is composed of the probabilities of chi-square tests for the HWE of each locus in each local population, such that higher P values indicate a population closer to equilibrium. In matrix $D_{p \times l}$, $p = 25$ (populations), $l = 9$ (loci), and d_{ij} represents the expected HWE of locus j in population i .

Modeling

Our problem was to identify solutions with the smallest set of *D. alata* local populations representing the genetic diversity of the species for its conservation and persistence. We considered 3 variations for the problem as follows.

Table 2. Alleles identified from 9 sequenced microsatellite from *Dipteryx alata*.

	Bm164	DaE06	DaE12	DaE20	DaE34	DaE41	DaE63	DaE67	DaE46	
	156	212	216	146	118	208	170	104	244	
	158	216	218	154	120	210	176	106	247	
	165	220	219	156	122	214		108	250	
	168		220	158	124			110	253	
	170		222		126			112		
	174				128			114		
	176				130			116		
	178				132			118		
					134			120		
					136			122		
					138			124		
					142					
					146					
					148					
					150					
Total	8	3	5	4	15	3	2	11	4	Total 55

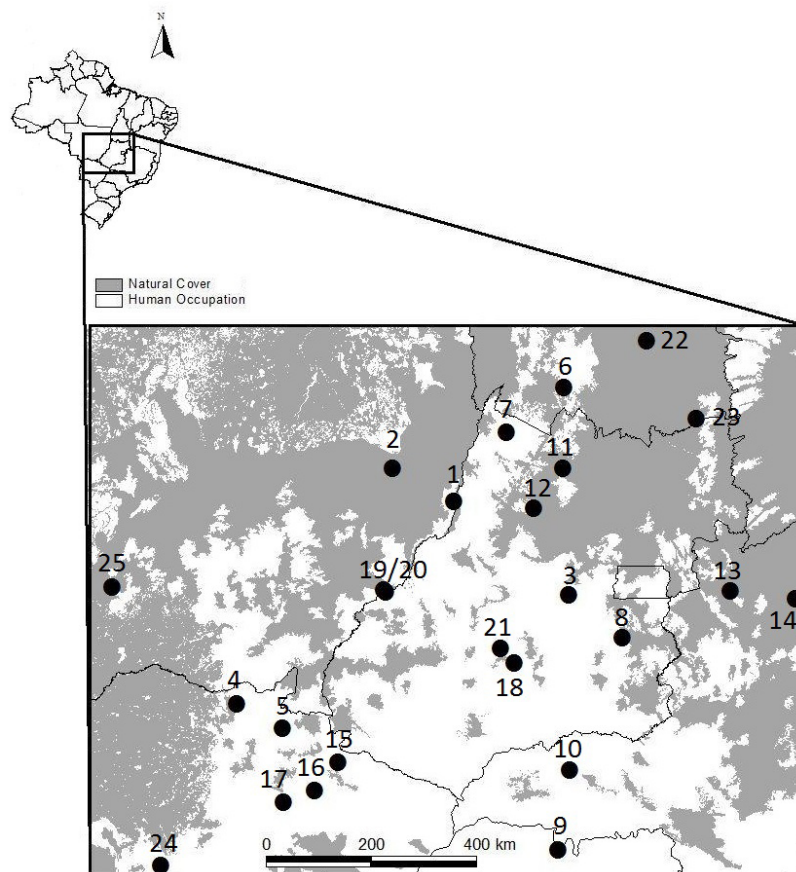
**Figure 1.** Geographic location of the 25 local populations of *Dipteryx alata* in Central Brazil analyzed using SCP methods based on microsatellite allelic variation. The region shown in dark tone remains covered by natural remnants of Cerrado vegetation.

Table 3. Populations and number of sampled individuals.

Population No.	Population name	No. of sampled individuals
1	CMT	32
2	ABMT	32
3	PGO	32
4	SMS	31
5	AMS	32
6	ATO	32
7	SMGO	32
8	LGO	32
9	ISP	31
10	MAMG	32
11	ENGO	12
12	STGO	12
13	AMG	32
14	PMG	32
15	PMS	13
16	PCMS	13
17	CMS	13
18	IGO	13
19	RAMT	27
20	RAGO	37
21	JGO	32
22	NTO	12
23	ARTO	15
24	AQMS	31
25	CAMT	30
Total	-	642

Table 4. Partial and schematic representation of collected data from the 642 individual trees sampled in 25 local populations throughout *Dipteryx alata*'s geographical range.

Sampled tree	Allele																	
	Bm164		DaE06		DaE12		DaE20		DaE34		DaE41		DaE63		DaE67		DaE46	
1CMT	158	158	216	216	220	220	154	154	110	110	126	126	208	208	176	170	253	253
2CMT	170	158	216	216	220	220	154	154	116	114	126	126	210	210	176	170	253	253
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
29CAMT	176	174	220	220	220	218	154	154	114	110	132	132	208	208	176	176	250	250
30CAMT	156	156	220	220	220	218	154	154	114	110	124	124	208	208	176	176	250	250

Variation 1: Two objective optimization (2-D optimization)

In this first variation, we reproduced the experiment of Diniz-Filho et al. (2012), but using a multi-objective optimization approach. We applied NSGA-II rather than simulated annealing. The latter was originally used by Diniz-Filho et al. (2012) and employs a monoobjective optimization approach while dealing with SCP.

The purpose of this first problem variation was to determine the smallest set of *D. alata* local populations (Equation 1) that should be preserved in order to represent the genetic diversity of the species, targeting its *in situ* conservation, i.e. each 1 of the 55 alleles should be represented at least once (Equation 2).

Using matrix A, described in Subsection Data, a candidate solution for the problem is a vector $x = \{x_1, x_2, \dots, x_{24}, x_{25}\}$, where $x_i \in \{0, 1\}$, such that $x_i = 1$, if population i is selected to compose the solution; or 0, otherwise.

The aim was to obtain:

$$\min \left(\sum_{i=1}^p x_i \right) \quad (\text{Equation 1})$$

Subject to:

$$\forall j \in \{1, 2, \dots, n\}, \sum_{i=1}^p a_{ij} x_i \geq 1 \quad (\text{Equation 2})$$

where $p = 25$ (populations) and $n = 55$ (alleles).

Regarding the multi-objective optimization approach, there are 2 objectives to be optimized:

- 1) Minimize the number of selected populations and
- 2) Maximize the number of alleles.

For simplicity, in the 2nd objective function, we used the number of missing alleles (those not present in the solution); therefore we worked with 2 minimization functions (Equations 3 and 4):

$$\min(f_1(\vec{x})) = \min(\text{number_of_populations}(\vec{x})) \quad (\text{Equation 3})$$

$$\min(f_2(\vec{x})) = \min(\text{missing_alleles}(\vec{x})) = \min(55 - \text{represented_alleles}(\vec{x})) \quad (\text{Equation 4})$$

Variation 2: 4 objective optimization (4-D optimization)

Variation 1 solutions ensure that all 55 alleles would be represented, but not their persistence over time. One attempt to cope with this limitation would be to maximize the allele's frequency, while simultaneously prioritizing the presence of heterozygosity or HWE in the set of selected populations for conservation.

Therefore, we obtained 3 more objectives (Equations 5-7) that were combined and added to Equations 3 and 4 in order to obtain a more consistent and refined solution to predict *D. alata* persistence:

- 1) Using matrix B to maximize the total allele frequency (Equation 5).

$$\max(f_3(\vec{x})) = \max(\text{allele_frequency}(\vec{x})) \quad (\text{Equation 5})$$

2) Using matrix C to maximize the heterozygosity of local populations:

$$\max(f_4(\vec{x})) = \max(\text{heterozygosity}(\vec{x})) \quad (\text{Equation 6})$$

2) Using matrix D to maximize the HWE level in populations (Equation 7):

$$\max(f_5(\vec{x})) = \max(\text{HWE}(\vec{x})) \quad (\text{Equation 7})$$

Each equation (each dimension) is an objective function that can be optimized. We executed 10,000 independently runs of NSGA-II (see Subsection Implementation) for each combination of the 4 objectives for optimization as follows:

- 1) Equations 3, 4, 5, and 6;
- 2) Equations 3, 4, 5, and 7;

Variation 3: 5 objective optimization (5-D optimization)

In this variation, we performed the optimization considering all the previous stated objectives. Five objectives were optimized simultaneously, including: number of populations, number of missing alleles, allele frequency, heterozygosity, and HWE (Equations 3-7).

NSGA-II

Evolutionary algorithms (EA) are inspired by biological evolution and use operators based on mutation, recombination, and natural selection (Bäck, 1996). Candidate solutions in EA play the role of individuals in a population. The results of previous studies suggest that EA are particularly appropriated for finding Pareto optimal solutions, particularly because they can efficiently process a set of solutions in parallel. Fonseca and Fleming (1995) as well as Valenzuela-Rendón (1998) suggested that multi-objective optimization is a research area in which EA can produce better results than traditional optimization techniques.

We used the NSGA-II (Deb et al., 2002), a state-of-the-art in MOEA (López-Jaimes and Coello-Coello, 2009). NSGA-II is a fast and elitist-based algorithm, in which individuals are classified based on a rank order, built on a dominance relationship, and a crowding operator. The best individuals are selected and evolutionary operators (crossover and mutation) are applied.

Briefly, the population is randomly initialized; prior to selection, the population is separated into categories (ranks) constructed based on dominance. For each non-dominated individual, a rank value of 1 is assigned, meaning that it belongs to the 1st Pareto front, which allows individuals to have the same potential to be selected. These individuals are removed from the population and the process of classifying the remaining individuals in their respective front ranks continues, e.g., individuals in the 2nd front receive a rank value of 2, and so on.

After assigning a rank to each individual, a value of crowding (an agglomeration comparison operator that allows prioritizing less crowded regions of solution space), is calculated. Individuals are ranked based in ascending order of their rank values and in descending order of their crowding values. The best individuals are selected and crossover and mutation operators are applied. The process continues until a stop condition is reached, e.g., a specified number of generations or a specified number of objective function evaluations.

Implementation

Candidate solutions were encoded as a binary vector of length L , where L is the number of populations (in this case, 25). Each element of the vector stores a 0 or 1 value, where 1 indicates that the corresponding population was selected to integrate a candidate solution; otherwise, a value of 0 is used.

At each execution, 500 initial solutions were randomly generated. These solutions were then evolved using NSGA-II, implemented in Matlab®.

Before running the experiments, we used a sample set to empirically estimate the most suitable parameter values, which were set to: population size = 500 individuals; crossover operator = single point crossover; crossover probability = 0.90; mutation probability = $1/L$ (where L is the number of populations); selection by binary tournament; mutation rate = 0.5; and number of objective functions evaluation = 100,000.

After obtaining the configuration parameters, 10,000 runs of NSGA-II for each problem variation (described in 2.2.1) were carried out. The tests were performed on 2 servers, a Hewlett-Packard ProLiant DL585 G7, 4xAMD 2.8Ghz 16-cores, 512 GB RAM and a Hewlett-Packard ProLiant DL385p Gen8, 2xAMD 2.8Ghz 16-cores, 256 GB RAM.

Null model

A null model is an attempt to generate value distributions for a given variable of interest in the absence of the process under study. In experimental sciences, this allows for a “controlled situation” (Paes and Blinder, 1995; Gotelli and Graves, 1996). The main goal of using a null model is to show that the experimental results would not have emerged from randomly generated data. In this study, 10,000 populations of 500 individuals (in a total of 5,000,000 individuals) were randomly generated in order to determine whether the same results would be found without the execution of NSGA-II.

RESULTS

Variation 1: 2-D optimization

We found that the smallest population set needed to represent all 55 alleles had size 7 (Figure 2), corroborating results of Diniz-Filho et al. (2012). However, it must be highlighted that while the previous study found only 2 distinct solutions, we found 6 different solutions for the investigated problem using multi-objective optimization. Four were new solutions (S3, S4, S5, S6) and 2 were the same as those determined by Diniz-Filho et al. (2012) (S1* and S2*) (Table 5).

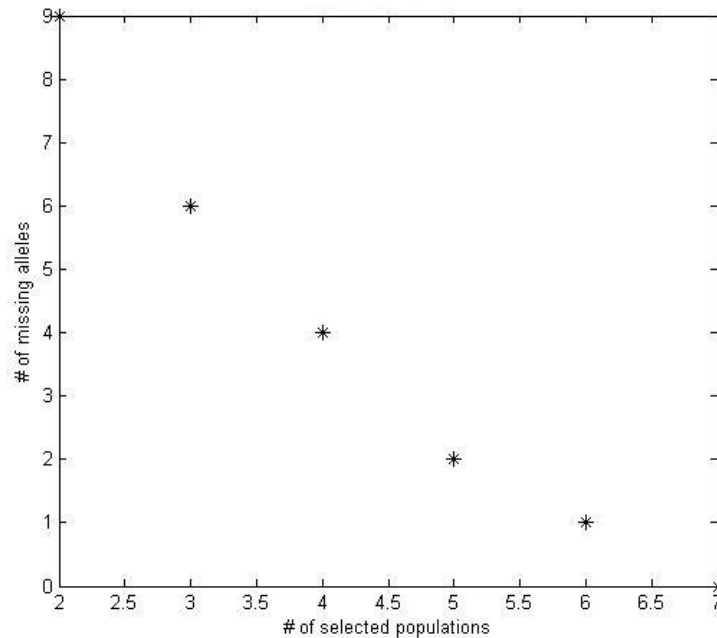


Figure 2. Pareto front obtained for problem Variation 1. Because there were 2 objectives (number of selected populations and of missing alleles), the results were plotted on a 2-D graph.

Table 5. Six solutions found for problem Variation 1. Solutions are S_i , where $i = 1-6$. In columns S_i , a value 1 indicates that the population j ($j = 1-25$) was included in the solution; 0, otherwise.

Population #	Population name	S_1^*	S_2^*	S_3	S_4	S_5	S_6	Irreplaceability*	Irreplaceability
1	CMT	1	1	1	1	1	1	1.00	1.00
2	ABMT	0	0	0	0	0	0	0.00	0.00
3	PGO	0	0	0	0	0	0	0.00	0.00
4	SMS	0	0	0	0	0	1	0.00	0.17
5	AMS	1	1	1	1	1	1	1.00	1.00
6	ATO	0	0	0	0	0	0	0.00	0.00
7	SMGO	0	0	0	0	1	0	0.05	0.17
8	LGO	0	0	0	0	0	0	0.00	0.00
9	ISP	0	0	0	0	0	0	0.00	0.00
10	MAMG	0	0	0	0	0	0	0.00	0.00
11	ENGO	0	0	0	0	0	0	0.00	0.00
12	STGO	0	0	0	0	0	0	0.00	0.00
13	AMG	0	0	0	0	0	0	0.00	0.00
14	PMG	0	0	0	0	0	0	0.00	0.00
15	PMS	0	1	0	1	0	0	0.50	0.33
16	PCMS	0	0	0	0	0	0	0.05	0.00
17	CMS	1	1	1	1	1	1	0.95	1.00
18	IGO	1	0	1	0	0	0	0.50	0.33
19	RAMT	1	1	1	1	1	1	1.00	1.00
20	RAGO	0	0	0	0	0	0	0.00	0.00
21	JGO	0	0	1	1	1	1	0.57	0.67
22	NTO	0	0	0	0	0	0	0.00	0.00
23	ARTO	0	0	0	0	0	0	0.00	0.00
24	AQMS	1	1	0	0	0	0	0.43	0.33
25	CAMT	1	1	1	1	1	1	1.00	1.00
-	Total	7	7	7	7	7	7	-	-

*Results found by Diniz-Filho et al. (2012).

Irreplaceability values, shown in the last 2 columns of Table 5, were obtained by considering the frequency by which a given population appeared in the solutions considering all determined solutions, including 6 in our experiment and 2 found by Diniz-Filho et al. (2012).

Local populations converging to 1 (1-CMT, 5-AMS, 17-CMS, 19-RAMT, 25-CAMT) were often irreplaceable, so that if they were lost, the conservation goal would not be achieved.

Variations 2 and 3: 4-D and 5-D optimizations

Results for 4 and 5 simultaneously optimized objectives were similar (Table 6), including the irreplaceability values shown in Figure 3.

Table 6. Partial results for minimum set of selected populations found by simultaneously optimizing 4 and 5 objectives.

No. of alleles (No. of missing alleles)	No. of selected populations	(Solution) selected populations	Variation 2		Variation 3
			4-D Heterozygosity	4-D HWE	5-D Heterozygosity & HWE
55(0)	7	(S1) 1-5-17-18-19-24-25	X	X	X
		(S2) 1-5-15-17-19-24-25	X	X	X
		(S3) 1-5-17-18-19-21-25	X	X	X
		(S4) 1-5-15-17-19-21-25	X	X	X
55(0)	Subtotal 8		4	4	4
		1-2-5-15-17-19-21-25	X	X	X
		1-2-5-15-17-19-24-25	X	X	X
		1-2-5-17-18-19-21-25	X	X	X
		1-2-5-17-18-19-24-25	X	X	X
		1-3-5-17-18-19-24-25	X	X	X
		1-4-5-15-17-19-21-25	X	X	X
		1-4-5-15-17-19-24-25	X	X	X
		1-4-5-17-18-19-21-25	X	X	X
		1-4-5-17-18-19-24-25	X	X	X
		1-5-9-17-18-19-24-25	X	X	X
		1-5-10-15-17-19-21-25	X	X	X
		1-5-10-15-17-19-24-25	X	X	X
		1-5-10-17-18-19-24-25	X	X	X
		1-5-13-15-17-19-21-25	X	X	X
		1-5-7-15-16-19-24-25		X	X
		1-5-7-16-18-19-24-25	X		X
		1-5-10-17-18-19-21-25	X	X	
		1-5-12-15-17-19-24-25	X		X
		1-5-12-17-18-19-21-25	X	X	
		1-5-13-15-17-19-24-25	X		X
		1-5-14-17-18-19-24-25		X	X
		1-5-15-17-18-19-21-25	X		X
		1-5-15-17-19-22-24-25	X		X
		1-5-16-17-18-19-24-25		X	X
		1-3-5-15-17-19-21-25	X		
		1-3-5-17-18-19-21-25		X	
		1-5-7-15-17-19-24-25			X
		1-5-8-17-18-19-24-25		X	
		1-5-9-15-17-19-24-25	X		
		1-5-9-17-18-19-21-25	X		
		1-5-11-15-17-19-24-25			X
		1-5-11-17-18-19-24-25			X
		1-5-12-17-18-19-24-25			X
		1-5-14-15-17-19-21-25		X	
		1-5-15-16-17-19-24-25	X		
		1-5-15-17-19-20-24-25	X		
		1-5-15-17-19-21-22-25	X		
		1-5-15-17-19-21-23-25			X
		1-5-17-18-19-20-21-25		X	
		1-5-17-18-19-21-22-25		X	
		1-5-17-18-19-21-24-25			X
	Subtotal		27	24	28
	Total		31	28	32

The optimization objectives are: 1) minimize number of selected population (set size); 2) minimize number of missing alleles; 3) maximize allele frequency; 4) maximize heterozygosity; 5) maximize HWE. Solutions Si, where i = 1-4, are the same corresponding to solutions on Table 5. X indicates that the solution was found by the corresponding approach.

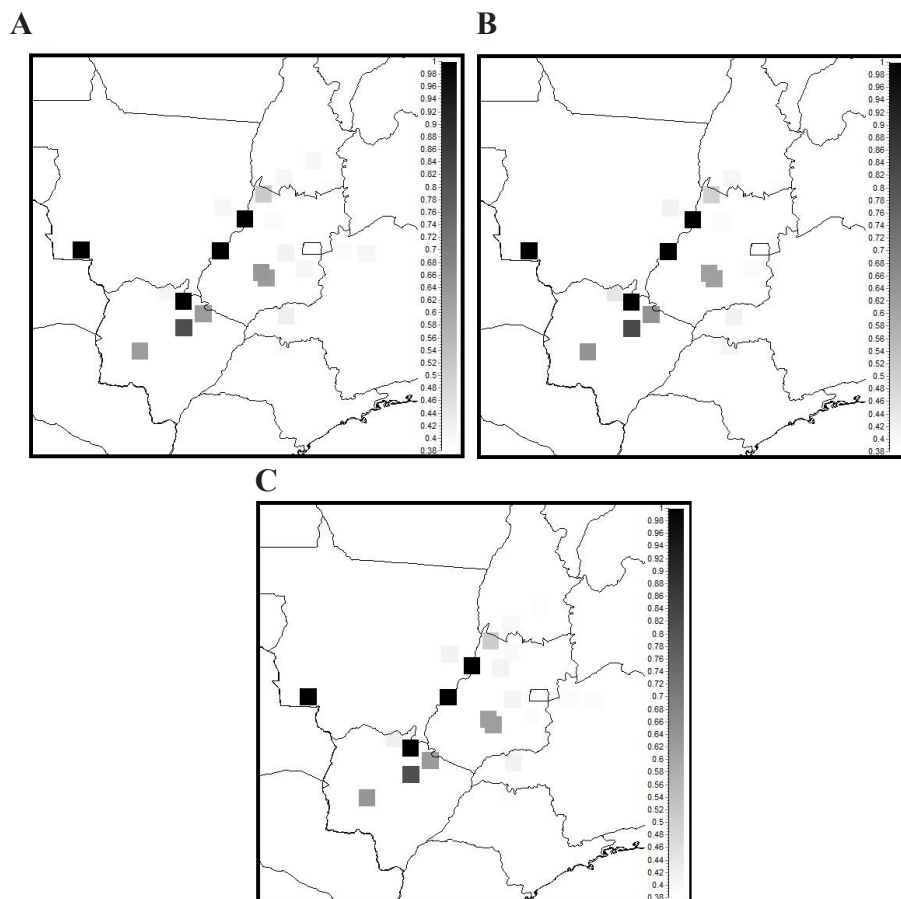


Figure 3. Irreplaceability values for (A) Variation 2 (4-D optimization) with heterozygosity as the 4th objective; (B) Variation 2 with HWE as the 4th objective and (C) Variation 3 (5-D optimization). Irreplaceability values were based on the frequency that a given population appeared in distinct solutions. A value converging to 1 means that the given local population was generally irreplaceable, in the sense that if it was not present in the solution, the conservation goal may not be achieved. Experimental irreplaceability values found for Variations 2 and 3 were very similar.

It is worth note that, in Variation 1, 7 was the minimum set size necessary to represent all 55 alleles, but because there were additional objectives, we used multi-objective optimization to obtain a much larger portfolio of solutions with populations ranging from 7-22 (only the solutions with a set of 7 and 8 populations are shown in Table 6), offering decision-makers a larger spectrum of options that fulfill the stated objectives.

We observed no hierarchy among the results shown in Table 6, indicating that all found solutions are optimal in the considered context, i.e., none can be declared the best when all optimized objectives are considered.

For 7 selected populations, both Variations 2 and 3 (4-D and 5-D optimizations) identified 4 from the 6 solutions identified in Variation 1, corresponding to S1, S2, S3, and S4 shown in Table 5. These 4 solutions are shown graphically in the right lower corner of the graph in Figure 4.

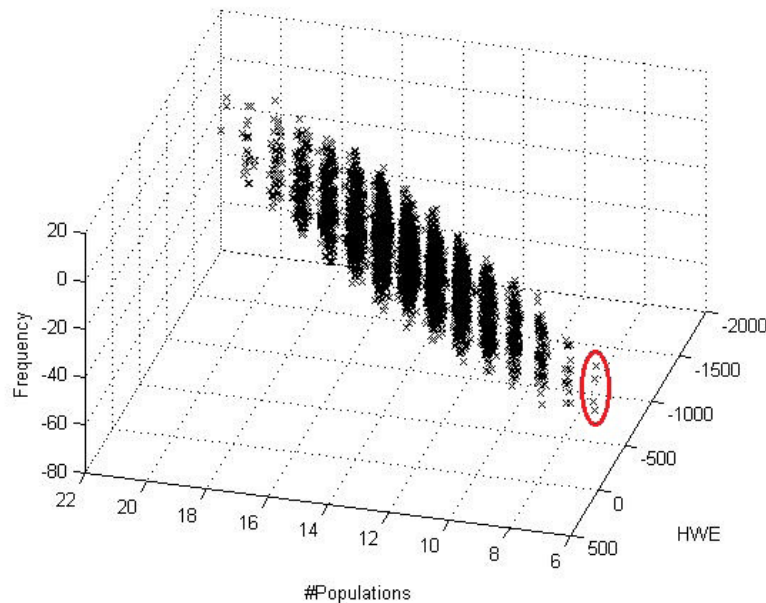


Figure 4. Pareto front obtained for problem Variation 3 (5-D optimization). There were 5 objectives (number of populations, allele frequency, HWE, heterozygosity, and number of missing alleles); the first 3 are plotted in this 3-D graph. Solutions with 7 populations are shown on the lower right corner (as indicated). It is worth note that the value for missing alleles of all solutions in this Pareto front was 0.

Table 7 shows the percentages for solutions with 7 populations found by optimization performed in Variations 2 and 3.

Table 7. Distribution (in percentage) of the solutions with 7 populations found by 4-D and 5-D optimizations. Solution Si, where i = 1-4, are the same corresponding to solutions on Table 5.

Solution	Variation 2		Variation 3
	4-D Heterozygosity	4-D HWE	5-D Heterozygosity & HWE
S1	22.78%	25.26%	1.58%
S2	20.06%	24.53%	98.34%
S3	31.44%	24.60%	0.02%
S4	25.72%	25.60%	0.05%

Null model

From 5,000,000 randomly generated individuals (10,000 populations of 500 individuals each), we obtained 351,547 with 7 populations (lower bound), approximately 7% of all randomly generated individuals. Furthermore, when compared to the results obtained for problem Variation 1, only 3 of the 351,547 results represented all 55 alleles (corresponding to 0.00006% of the randomly generated individuals). This indicates that NSGA-II experimental results were not generated by chance.

For each randomly generated individual, we determined heterozygosity and HWE values. Except for the 3 solutions with no missing alleles, the other 351,544 randomly generated solutions with 7 populations (number of populations presented in Table 6), heterozygosity and HWE values were worse (lesser) when compared to those found using NSGA-II.

DISCUSSION

The most popular methods, algorithms, and tools used for the SCP problem involve a monobjective approach; they aggregate the distinct objectives into a single function. Hence, despite the clear multi-objective aspect of SCP, optimization models often treat this as monobjective by assigning different weights to different objectives of the problem in order to aggregate the objectives into a single objective function, known as a fitness function (Zitzler et al., 2002). However, when 2 criteria represent different value systems, it may be impossible to combine such criteria in a meaningful way (e.g., comparing the preservation of endangered species and the promotion of economic development). A single objective function can result in the association of utterly disparate elements, requiring assumptions that decision-makers may consider inadequate and even leading to inaccurate results (Ciarleglio et al., 2010). This prompted us to use a multi-objective approach for solving SCP.

Moreover, it is not a trivial task to properly weight different conflicting objectives in order to combine them accordingly, and it is generally necessary to have expert knowledge that is not always available. In fact, depending on how the objectives are associated to compose the unique fitness function, one may obtain completely different results.

By applying multi-objective optimization, it is possible to overcome these issues because the objectives can simultaneously be optimized independently from one another.

Diniz-Filho et al. (2012) recently proposed an explicit complementary approach that could be used to optimize conservation of genetic variability, expressed as allelic variation (presence-absence of alleles) derived from microsatellite loci, and solved an SCP problem for the conservation of *D. alata* using a simulated annealing algorithm.

Nonetheless, these previous attempts were monobjective as well as used only the presence-absence of alleles in local populations, which is not as informative as using the allele frequencies directly. This approach more properly reflects the ecological and evolutionary processes driving genetic diversity in local populations and may be more related to population persistence. Using allele frequencies is equivalent to applying more complex characteristics of a species, such as abundance and environmental suitability in SCP higher hierarchical levels, potentially improving the long-term persistence of the conservation networks and providing the opportunity to test the multi-objective method in this new conservation problem at lower hierarchical level than the species level.

As stated above, in standard applications of SCP using software such as Marxan and based on simulated annealing (Table 1), the constraints are typically expressed as weights, obtained from a complex function combining several conflicting attributes. Hence, the approach proposed here, in addition to being the first to use an evolutionary algorithm to address the SCP problem, is also the first to use a legitimate multi-objective approach. This allows more flexibility by including additional decision objectives, as well as adds more complexity to results while increasing decision-maker options.

In Variation 1, developed to reproduce the experiment of Diniz-Filho et al. (2012), we identified a larger number of distinct solutions (6 rather than 2 as previously reported).

For Variations 2 and 3, we found similar results and significantly expanded the portfolio of solutions when considering options globally. Thus, in Variation 1, the method was not able to identify new options with more than 7 populations. By including a larger number of objectives, we introduced a degree of flexibility that allowed the identification of other solutions with more populations but with optimized allele frequency, heterozygosity, or HWE. Although it is desirable to have fewer populations, this is advantageous because a minimal set of populations representing all alleles is not necessarily the one with the best results when considering intra-specific diversity and persistence.

A key point is that Variations 2 and 3 were able to refine the Variation 1 solutions, indicating the optimization of persistence features (S1-S4) (Table 6). Additionally, the distribution (in percentage) of solutions with 7 populations found using 4-D optimizations (Table 7 2nd and 3rd columns) showed no significant differences between the results obtained by optimization performed in Variation 2, as they were similarly distributed from S1-S4. However, for 5-D optimization (Table 7 last column), although there was no hierarchy among these solutions (S1-S4) in the sense that they were optimal when all optimized objectives were considered, the results for Variation 3 clearly highlighted S2, as S2 corresponds to 98.34% of the solutions with 7 populations. This result suggests that optimization of 5 objectives can be used to further refine the results.

As the decision-makers' portfolio increases considerably, the method can indicate the most adequate options in this portfolio, which is advantageous.

In addition, it appears that the simultaneous optimization of heterozygosity and HWE in Variation 3 is advantageous, as applying it allowed the method to identify more distinct solutions for 8 populations. There were 28 solutions compared to 27 and 24 from Variation 2, which used heterozygosity and HWE, respectively, as the 4th objective (Table 6).

Considering that for 7 populations, the same results were found using 4-D and 5-D optimizations, it can be said that heterozygosity and HWE guided the set of solutions in the same direction. For 8 populations, of the 41 distinct solutions, 14 (34%) were determined using the 4-D and the 5-D optimizations, and 24 (58%) by at least 2 of the 3 optimizations performed.

The results from the null model assured that solutions obtained using NSGA-II did not emerge randomly. Only 3 results from the null model included 7 populations and all 55 alleles, while more than 65,000 solutions retrieved using NSGA-II presented these characteristics.

The possibility of dealing with these more complex situations clearly shows the advantage of our method as compared with the standard approach based on simulated annealing, as implemented by Diniz-Filho et al. (2012).

Considering that the most commonly used tools for SCP apply algorithms on a monobjective approach, our results show that dealing with various conflicting objectives in more than one dimension (i.e. using multi-objective optimization) allows for a more sophisticated and general solution to the SCP problem. This can be observed by the variety of different solutions generated using our method (instead of a single point generated using monobjective methods), thus improving the decision-maker support.

In conclusion, we used a multi-objective approach to solve variations of the SCP problem with more than two objectives, which added complexity and increased the decision-makers options. We showed that a multi-objective algorithm is more powerful and opens more possibilities than the methods previously used, such as simulated annealing.

We implemented a more refined search for optimal solutions to the problem of finding the smallest set of local populations of *D. alata* that should be conserved in order to represent genetic diversity based on allele frequency information associated with heterozygosity and

Hardy-Weinberg equilibrium. This was the first time these parameters (objectives) were used in the context of SCP.

We found that the smallest set of populations needed to represent all alleles under study was 7, corroborating the 2 solutions determined by Diniz-Filho et al. (2012), but we obtained more options of distinct solutions (the previous 2 solutions as well as 4 additional solutions, for a total of 6). By optimizing 4 and 5 objectives simultaneously (4-D and 5-D optimizations), we found 4 solutions for 7 populations, refining the 6 previously determined solutions. Additionally, we obtained a larger portfolio in terms of intra-specific diversity and persistence with populations ranging from 8-22. In particular, for 8 populations we found 41 solutions.

Although additional experiments should be conducted to improve the NSGA-II fitness function in a more precise and controlled manner, our results demonstrate the advantages of the new approach with respect to previous solutions. Additionally, this was the first attempt to apply multi-objective algorithms to an SCP problem with more than 2 dimensions based on molecular data at the population level as basic units. Our results can be used to propose a more specialized multi-objective algorithm for SCP problems, allowing researchers to deal with such problems in a more efficient and appropriate manner.

ACKNOWLEDGMENTS

S. Schlottfeldt wishes to thank the University of York and Prof. Jon Timmis for the PhD stay and the support from CNPq, through the Science without Borders Program. The research program in molecular ecology and conservation of Cerrado plants has been continuously supported by several grants and fellowships to the research network GENPAC (Geographical Genetics and Regional Planning for Natural Resources in Brazilian Cerrado) from CNPq/MCT/CAPES (projects #564717/2010-0 and 563624/2010-8) and by the “Núcleo de Excelência em Genética e Conservação de Espécies do Cerrado” - GECER (PRONEX/FAPEG/CNPq CP 07-2009). Field work was supported by Systema Natura e Consultoria Ambiental LTDA and work by M.E.M.T. Walter, A.C.P.L.F. Carvalho, M.P.C. Telles, R.D. Loyola, and J.A.F. Diniz-Filho have been continuously supported by productivity fellowships from CNPq.

REFERENCES

- Ackery PR and Vane-Wright RI (1984). Milkweed butterflies, their cladistics and biology: being an account of the natural history of the Danainae, a subfamily of the Lepidoptera, Nymphalidae. Natural History Museum, London.
- Ardron JA, Possingham HP and Klein CJ (2010). Marxan Good Practices Handbook, Pacific Marine Analysis and Research Association (PacMARA), Victoria.
- Arthur JF, Hachey M, Sahr K, Huso M, et al. (1997). Finding all optimal solutions to the reserve site selection problem: formulation and computational analysis. *Environ. Ecol. Stat.* 4: 153-165.
- Bäck T (1996). Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms. Oxford University Press, Oxford.
- Ball IR (2000). Mathematical Applications for Conservation Ecology: The Dynamics of Tree Hollows and the Design of Nature Reserves. Doctoral Thesis. Depts. of Applied Mathematics, Environmental Science and Management, University of Adelaide, Adelaide.
- Barton DN, Rusch G, Gjershaug JO, Faith DP, et al. (2004). TARGET as a tool for prioritising biodiversity conservation payments on private land - a sensitivity analysis. Technical Report SNR 4856-2004. Norwegian Institute for Water Research (NIVA), Oslo.
- Brooks TM, da Fonseca GAB and Rodrigues ASL (2004). Species, data, and conservation planning. *Conserv. Biol.* 18: 1682-1688.

- Ciarleglio M, Barnes JW and Sarkar S (2010). ConsNet: A tabu search approach to the spatially coherent conservation area network design problem. *J. Heuristics* 16: 537-557.
- Coello-Coello CA, Lamont GB and Veldhuizen DAV (2007). Evolutionary Algorithms for Solving Multi-Objective Problems. 2nd edn. Springer-Verlag, New York.
- Cormen TH, Stein C, Rivest RL and Leiserson CH (2001). Introduction to Algorithms. 2nd edn. MIT Press, Cambridge, MA.
- Deb K, Agarwal S, Pratap A and Meyarivan T (2002). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *IEEE T. Evol. Comput.* 6: 182-197.
- Diniz-Filho JAF and Telles MPC (2002). Spatial autocorrelation analysis and the identification of operational units for conservation in continuous populations. *Conserv. Biol.* 16: 924-935.
- Diniz-Filho JAF, Loyola RD, Melo DB, Oliveira G, et al. (2012). Planning for optimal conservation geographical genetic variability within species. *Conserv. Genet.* 13: 1085-1093.
- Fonseca CM and Fleming PJ (1995). An overview of evolutionary algorithms in multi-objective optimization. *Evol. Comput.* 3: 1-16.
- Fraser DJ and Bernatchez LB (2001). Adaptive evolutionary conservation: toward a unified concept for defining conservation units. *Mol. Ecol.* 10: 2741-2752.
- Garson J, Aggarwal A and Sarkar S (2002). ResNet Manual Version 1.2. User Manual. Biodiversity and Biocultural Conservation Laboratory, University of Texas at Austin, Austin.
- Gotelli NJ and Graves GR (1996). Null Models in Ecology. Smithsonian Institution Press, Washington, DC.
- Kirkpatrick JB (1983). An interactive method for establishing priorities for the selection of nature reserves - an example from Tasmania. *Biol. Conserv.* 25: 127-134.
- López-Jaimes A and Coello-Coello CA (2009). Multi-Objective Evolutionary Algorithms: A Review of the State-of-the-Art and some of their Applications in Chemical Engineering (Pandu RG, ed.). Multi-Objective Optimization Techniques and Applications in Chemical Engineering, World Scientific, Singapore, 61-90.
- Margules CR and Nicholls A (1987). Assessing the conservation value of remnant habitat "islands": Mallee patches on the Western Eyre peninsula (Saunders DA, Arnold GW, Burbridge AA and Hopkins AJM, eds.). Nature Conservation: The role of remnants of native vegetation. Surrey Beatty and Sons, Baulkham Hills, 89-102.
- Margules CR and Pressey RL (2000). Systematic conservation planning. *Nature* 405: 243-253.
- Margules CR, Nicholls A and Pressey R (1988). Selecting networks of reserves to maximize biological diversity. *Biol. Conserv.* 43: 63-76.
- Moffett A, Garson J and Sarkar S (2005). MultCSync: a software package for incorporating multiple criteria in conservation planning. *Environ. Modell. Softw.* 20: 1315-1322.
- Moilanen A (2005). Reserve selection using nonlinear species distribution models. *Am. Nat.* 165: 695-706.
- Nicholls A and Margules CR (1993). An updated reserve selection algorithm. *Biol. Conserv.* 64: 165-169.
- Paes E and Blinder PB (1995). Modelos nulos e processos de aleatorização: algumas aplicações em Ecologia de Comunidades (Peres-Neto PR, Valentin JL, Fernandez FAZ, eds.). Oecologia Brasiliensis Volume II: Tópicos em Tratamento de Dados Biológicos. Instituto de Biologia, UFRJ, Rio de Janeiro, 119-139.
- Possingham HP, Ball I and Andelman S (2000). Mathematical methods for identifying representative reserve networks (Ferson S and Burgman M, eds.). Quantitative methods for conservation biology. Springer-Verlag, New York, 291-305.
- Pressey RL (2004). Conservation planning and biodiversity: Assembling the best data for the job. *Conserv. Biol.* 18: 1677-1681.
- Pressey RL, Possingham HP and Margules CR (1996). Optimality in reserve selection algorithms: when does it matter and how much? *Biol. Conserv.* 76: 259-267.
- Rebello AG and Siegfried W (1990). Protection of Fynbos vegetation: ideal and real-world options. *Biol. Cons.* 54: 15-31.
- Sarkar S (2012). Complementarity and the selection of nature reserves: algorithms and the origins of conservation planning, 1980-1995. *Arch. Hist. Exact Sci.* 66: 397-426.
- Schlottfeldt S, Walter MEMT, Diniz-Filho JAF and Telles MPC (2012). Multi-objective Optimization in Systematic Conservation Planning to Represent Genetic Variability within Species. In: 8th International Conference on Ecological Informatics, ISEI2012, Brasília.
- Soares TN, Melo DB, Resende LV, Vianello RP, et al. (2012). Development of microsatellite markers for the neotropical tree species *Dipteryx alata* (Fabaceae). *Am. J. Bot.* 99: e72-e73.
- Underhill LG (1994). Optimal and suboptimal reserve selection algorithms. *Biol. Conserv.* 70: 85-87.
- Valenzuela-Rendón M (1998). Reinforcement learning in the fuzzy classifier system. *Expert Syst. Appl.* 14: 237-247.
- Zitzler E, Laumanns M and Thiele L (2002). SPEA2: Improving the Strength Pareto Evolutionary Algorithm for Multi-objective Optimization (Giannakoglou K, Tsahalis D, Periaux J, Papaliliou K, et al., eds.). Evolutionary Methods for Design, Optimisation and Control with Application to Industrial Problems. Proceedings of the EUROGEN2001 Conference, Athens, Greece, September 19-21, 2001, 95-100.