Systematic Conservation Prioritization

With Prioritizr

May 2018

Prepared by Elizabeth Law workingconservation@gmail.com

Prioritizr credits to the Prioritizr developers, especially Jeffrey Hanson and Richard Schuster

NADC data credits to Leandro Macchi, Riccardo Torres, Mattias Bauman, Ignacio Gasparri, Tobias Kümmerle

---

Practical 1:

15 min	Introduction: what is systematic conservation planning	Systematic Conservation Prioritization.pdf
15 min	A first example of systematic conservation planning	NADC_simple.xlsx
20 min	Prioritizr for the simple example	Exercise 1.html
30 min	The NADC study area – data, gap analysis, basic prioritization, adding lock-in, or lock-outs, and connectivity (Exercise 2)	Exercise 2.html
5 min	Allocation of scenarios for practical 2	-

Homework: Finish off Exercise 2. Think about your scenarios! Prepare some code to try, and research some of the discussion points.

Practical 2:

60 min	Scenarios	Scenarios.html
30 min	Class presentations (5 min / group)	-

What is Systematic Conservation Prioritization?

Systematic conservation prioritization is a rigorous, repeatable, and structured approach to designing new protected areas that efficiently meet conservation objectives, while minimizing socioeconomic cost (Margules and Pressey 2000), when used as part of a decision process (which should also involve meaningful stakeholder consultations and reviews).

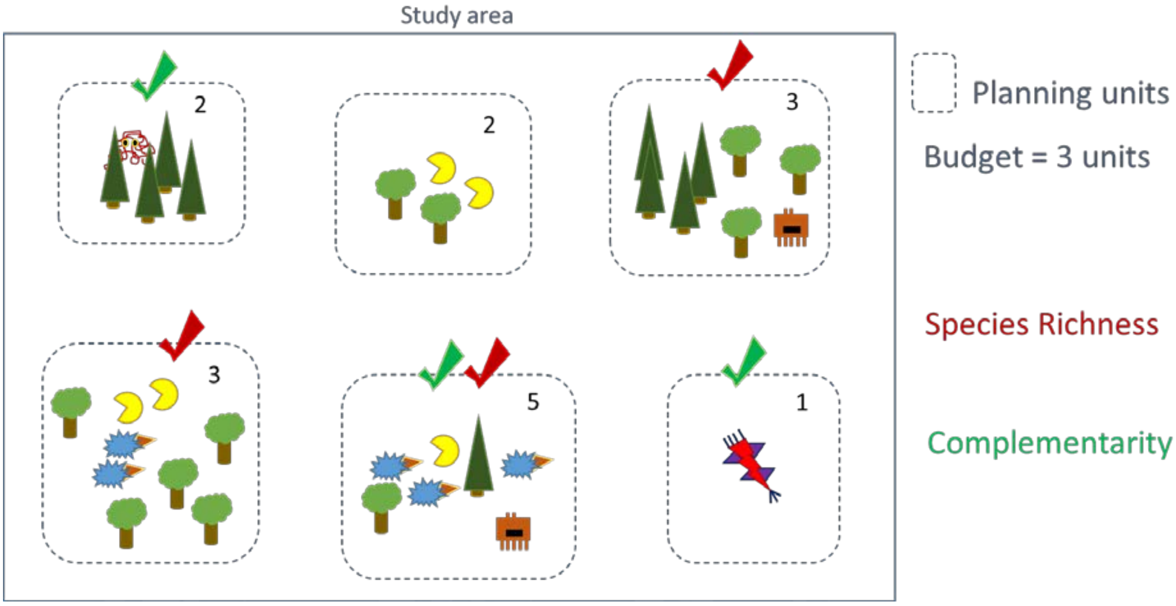
> **Rigorous:** based on evidence, defensible.
>
> **Repeatable:** if conditions change, then inputs can be updated, and sensitivity to input data variation tested.
>
> **Structured/Transparent**: not a 'black box'; objective is clear, subjective decisions clarified.
>
> **Efficient:** minimizes costs and/or maximizes benefits.

For example, when prioritizing places for new protected areas, you might want the resulting reserve system to be **representative** of the most species possible, and incur the least costs or conflict from other uses. This process takes account of the **complementarity** of different selected areas.

Representative: in terms of a protected areas, that these represent all species of concern; all species have enough of their population contained within the protected reserve system to persist.

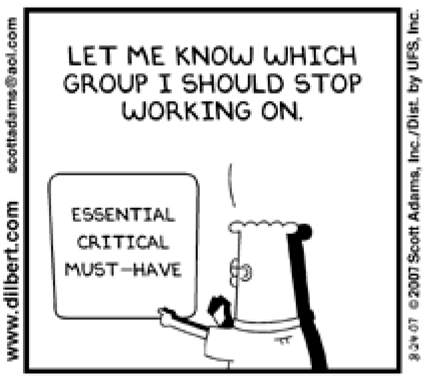
Complementarity: selection of a *set* that best satisfies objectives, not just selecting based on a single *a priori* ranking.

For example, we have six areas available to protect below. What is the species richness of each of them? If based on species richness only, which areas would you protect? How can we best represent all the species in a system? What about if we consider complementarity?



Why do this systematically?

Historically, conservation decision-making has often evaluated parcels opportunistically as they became available for purchase, donation, or under threat. Such decisions *may or may not* maximize the long-term persistence of target species or communities, or the biodiversity returns on dollars invested, in the absence of a landscape-level understanding of the distribution of target species and communities. Remember the 'high and far' bias in the location of protected areas!



Motivations for systematic prioritization include:

- Limited resources [funding, equipment, land, workers]
- Limited time
- (Too) many species requiring attention
- A problem with no clear, obvious solution, or involving trade-offs

Some of these limitations, but not all, are able to be modified. For example, there are many that advocate for increasing conservation funding, and not to accept conservation triage.

Triage: a system of prioritization, commonly conceptualised in conservation as a consequentialist, utilitarian prioritization, with the implication that some species are too ‘costly’ to save. Recommended reading: Wilson, K.A., Law, E.A. (2016) Ethics of conservation triage. Frontiers in Ecology and Evolution 4:112.

Remember: Conservation biology is a crisis discipline: “… one must act before knowing all the facts; crisis disciplines are … a mixture of science and art, and their pursuit requires intuition as well as information” Soulé (1985).

Elements of a prioritization are:

1. An objective(s): what do you want to minimize or maximize?
2. A list of available options: e.g. conservation actions in space and time.
3. Information on benefits: what each of the actions will achieve for each feature.
4. Information on costs: how much each action will cost.
5. (optional) Information on other constraints: e.g. threshold targets to achieve, budgets not to go over, areas that need to be ‘locked in’ to certain actions, or ‘locked out’ from other options.

Solving a problem by hand

Let’s have a go at solving a simple problem by ‘hand’. The data in the **NADC_simple.xlsx** file represent the costs and benefits from a study region. This region has approximately 100 cells. To keep things simple, and to focus on the process, let’s just assume for now that the costs are accurate, and the benefits (for 6 features) are also.

Your mission is to:

- Problem: select a reserve network
- Objective: minimize the cost
- Constraints: want to have the reserve network represent a threshold target of 17% for each feature.

The figures below show – by way of colour – the cost and benefit values, and can serve you as a visual guide.

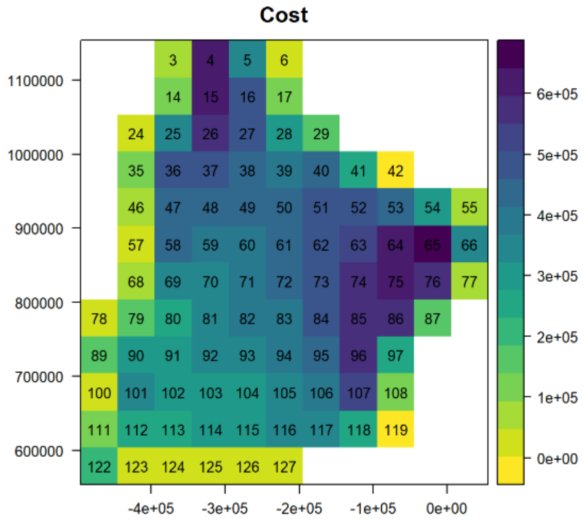
You can use the excel spreadsheet to ‘select’ cells (by giving them a 1 in the ‘select’ column). This will then automatically calculate the benefits that cell gives for the features. There is a summary table above which calculates

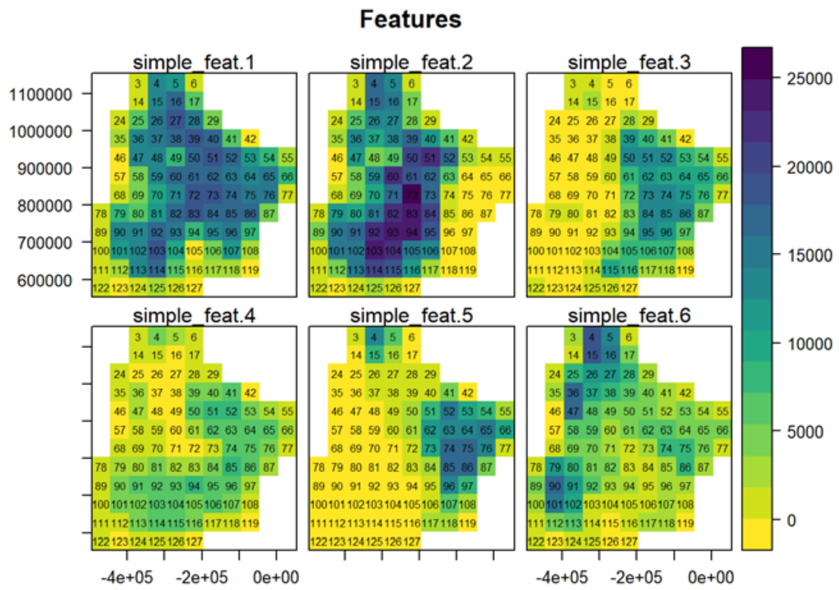
a) The total possible for all the features and the cost, and number of cells
b) The 17% target required
c) The amount we have already achieved in the selected cells
d) The amount we are missing from our target, for each of the features.

Remember, this is a trade off – we are trying to minimize the cost for achieving the threshold targets for the species.

Hint: Think about costs versus benefits. Is it possible to combine this into a single metric?

Hint 2: Excel is useful for sorting and calculating stuff!





	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	**Instructions**								# pu	cost	feat1	feat2	feat3	feat4	feat5	feat6
2	Only change things in the green 'select' column!						SUM all		95	30265184	902625	838826	342960	269285	293904	437880
3	Enter 1 into select column to select a cell						17% target				153446	142600	58303	45778	49964	74440
4	Change it back to 0 to deselect it						Sum selected		0	0	0	0	0	0	0	0
5	But you can add additional columns if you like						Missing target				153446	142600	58303	45778	49964	74440
6	The red cell is the total cost - you want to minimize this															
7	While achieving all the targets for the features															
8																
9	Data values								Selected values							
10																
11	puid	cost	feat1	feat2	feat3	feat4	feat5	feat6	select	cost	feat1	feat2	feat3	feat4	feat5	feat6
12	3	71342.69	1800	1720	738	720	208	2400	0	0	0	0	0	0	0	0
13	4	618140.5	13387.5	16980	1728	5295	13384	17850	0	0	0	0	0	0	0	0
14	5	365698.1	11707.5	13332	0	2332	7044	13090	0	0	0	0	0	0	0	0
15	6	22362.58	562.5	864	0	179	0	570	0	0	0	0	0	0	0	0
16	14	126465.6	637.5	380	156	81	0	850	0	0	0	0	0	0	0	0
17	15	622695.3	13875	16502	1836	1222	7180	17880	0	0	0	0	0	0	0	0
18	16	465912.8	17385	15234	0	0	2352	15550	0	0	0	0	0	0	0	0
19	17	100888.7	5850	4022	0	148	20	4100	0	0	0	0	0	0	0	0
20	24	9056.529	97.5	132	0	100	0	130	0	0	0	0	0	0	0	0

Problems, algorithms, and solutions

Prioritizing by hand is hard, no? And this was only with 95 cells, one cost, and 6 species. Just imagine if we had many, many more. This speaks to the need for automated processes – software – to do this for us. But before we talk about software, I just want to clarify a little more terminology.

Once we have a motivation for conservation, we need to define our objectives, and then define the problem. We can solve these problems with different algorithms. Different methods are suited to different types of problems, different data, etc.

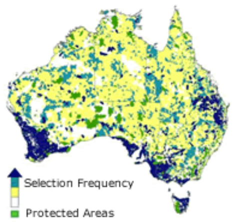
> **Problem:** The mathematical specification of the issue
>
> **Algorithm:** the methods used to solve it
>
> **Exact / mathematical methods:** mathematical methods, such as *linear* or *integer programming*, that find solutions with a known optimality, providing the problem is specified in a certain way (e.g. linear).
>
> **Heuristic / approximate methods:** rule-based methods, such as *simulated annealing* and *genetic algorithms*, that produce (good, but not necessarily optimal) answers. While we never can know how optimal our solutions are, these methods can handle really complex non-linear problems.

Typically conservation problems are so complex, involving many species and many potential actions, that approaching them by hand or intuition is not feasible. Luckily, a few conservation planning software tools exist. Here are just a few:

MARXAN
conservation solutions

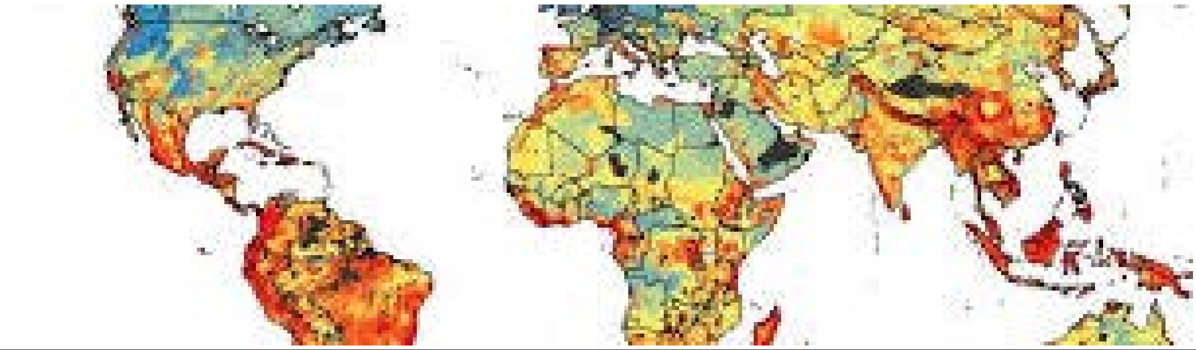
http://marxan.org/

Marxan is a suite of tools designed to help decision makers find good solutions to conservation planning problems. It is based on a minimum set or maximal coverage formulation, and includes options for connectivity, probability, and multiple zones. Marxan is the most frequently used conservation planning software and has been applied to hundreds of spatial conservation planning problems around the world. It is solved using an approximate method: a *simulated annealing* algorithm.



ZONATION https://www.helsinki.fi/en/researchgroups/metapopulation-research-centre/software#section-14300

Zonation produces a hierarchical prioritization of the landscape by iteratively removing the least valuable remaining cell while accounting for connectivity and generalized complementarity. That is, it defines and solves a problem using a set of *greedy heuristics* (which assume that locally optimal choices can lead to a global optimum). Its focus is on habitat quality and connectivity.



Prioritizr https://CRAN.R-project.org/package=prioritizr

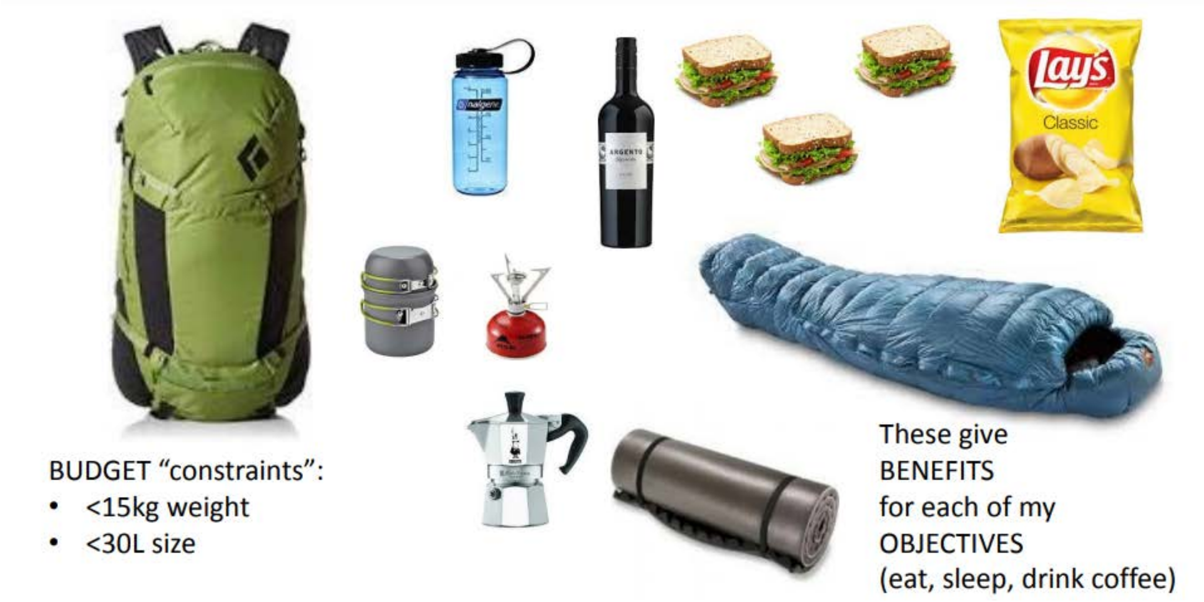
Prioritizr is an R package for solving systematic conservation prioritization problems using linear or *integer linear programming (ILP)* techniques. The package offers a flexible interface for creating conservation problems using a range of different objectives and constraints that can be tailored to the specific needs of the conservation planner. Conservation problems can be solved using a variety of commercial and open-source exact algorithm solvers. In contrast to the algorithms conventionally used to solve conservation problems, such as greedy heuristics or simulated annealing, the ILP algorithms used by prioritizr (implemented via Gurobi, lpsymphony, or Rsymphony) are guaranteed to find optimal solutions.

The basic problem is inherited from Marxan, but a linearization of the otherwise non-linear connectivity specifications can allow LP/ILP methods which can find much cheaper solutions of known optimality, often in a much shorter period of time than Marxan (Beyer et al. 2016). Currently developed to include the basic Marxan options, but also has options for including phylogenetic representation, among a number of alternate problem formulations.

We will be using this, with raster-based input, and a minimum set objective. It also handles some input in shapefiles, or as text files demanded by Marxan.

The Minimum Set Objective

Conservation prioritization is often set as a ‘minimum set objective’, aka the ‘knapsack’ problem:



In the context of systematic reserve design, the minimum set objective function seeks to find the set of planning units that minimizes the overall cost of a reserve network, while meeting a set of representation targets for the conservation features.

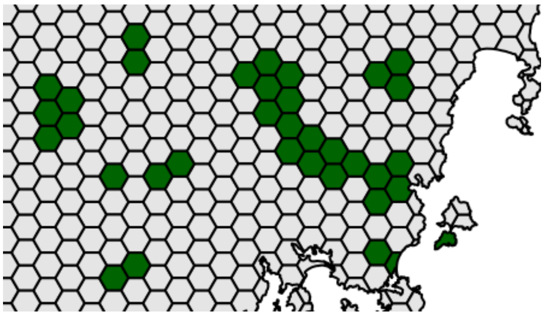
The basic reserve design problem for the minimum set objective function can be expressed mathematically for n planning units as:

$$Minimize \sum_{i=1}^{n} x_i c_i$$

$$subject\ to \sum_{j=1}^{n} x_j r_{ij} \geq T_i \forall\ i$$

Where x_i is a binary decision variable specifying whether planning unit i has been selected (1) or not (0), c_i is the cost of planning unit i, r_{ij} is the representation level of feature i in planning unit j, and T_i is the target for feature i. The first term is the objective function and the second is the set of constraints. In words this says ‘find the set of planning units that meets all the representation targets while minimizing the overall cost’. This objective is *equivalent to* a simplified Marxan reserve design problem, with the Boundary Length Modifier (BLM) set to zero. The BLM option adds to the minimization objective a sum of the exposed boundary length of the selected cells, in shorthand:

$$Minimize \sum_{i=1}^{n} x_i c_i + blm \sum_{i=1}^{n} boundary\ length$$



Where *blm* is a multiplier, so we can scale how much effect the boundary costs have on the final solution, and the *boundary length* is the length of the exposed boundary of selected PU's. We'll explore this a bit more in the examples.

A note on data preparation

Often the collation and preparation of the data are the most time consuming aspects of systematic conservation prioritization exercises. This includes defining the study area and the problem formulation, developing models, error checking collated data and models, and getting everything aligned and into the right formats. There are many, many decisions made along the way, and it is not unusual for this task to take up over a year.

How does Prioritizr do on our simple problem?

Let's see, and in the process become a little more familiar with R. We will:

1. Open R via RStudio
2. Install and load all the required packages
3. Load and explore the example data
4. Use Prioritizr to solve the problem

Open the Example_1.html file by double clicking on it, and follow the instructions therein.

NADC case study

The example above was pretty simplified. Let's have a look at a slightly more realistic problem.

Study area

Our study region is located in the semi-arid Northern Argentinian Dry Chaco, and covers ~172,800 km^2 across Salta, Santiago del Estero, Chaco and Formosa provinces. The landscape is a mosaic dominated by medium-tall (16-18m) semi-deciduous dry forests with shrubby understory, some natural grassland and flooded savannahs. Biodiversity includes 46 tree species, over 400 bird species, over 30 amphibian species and around 145 mammal species.

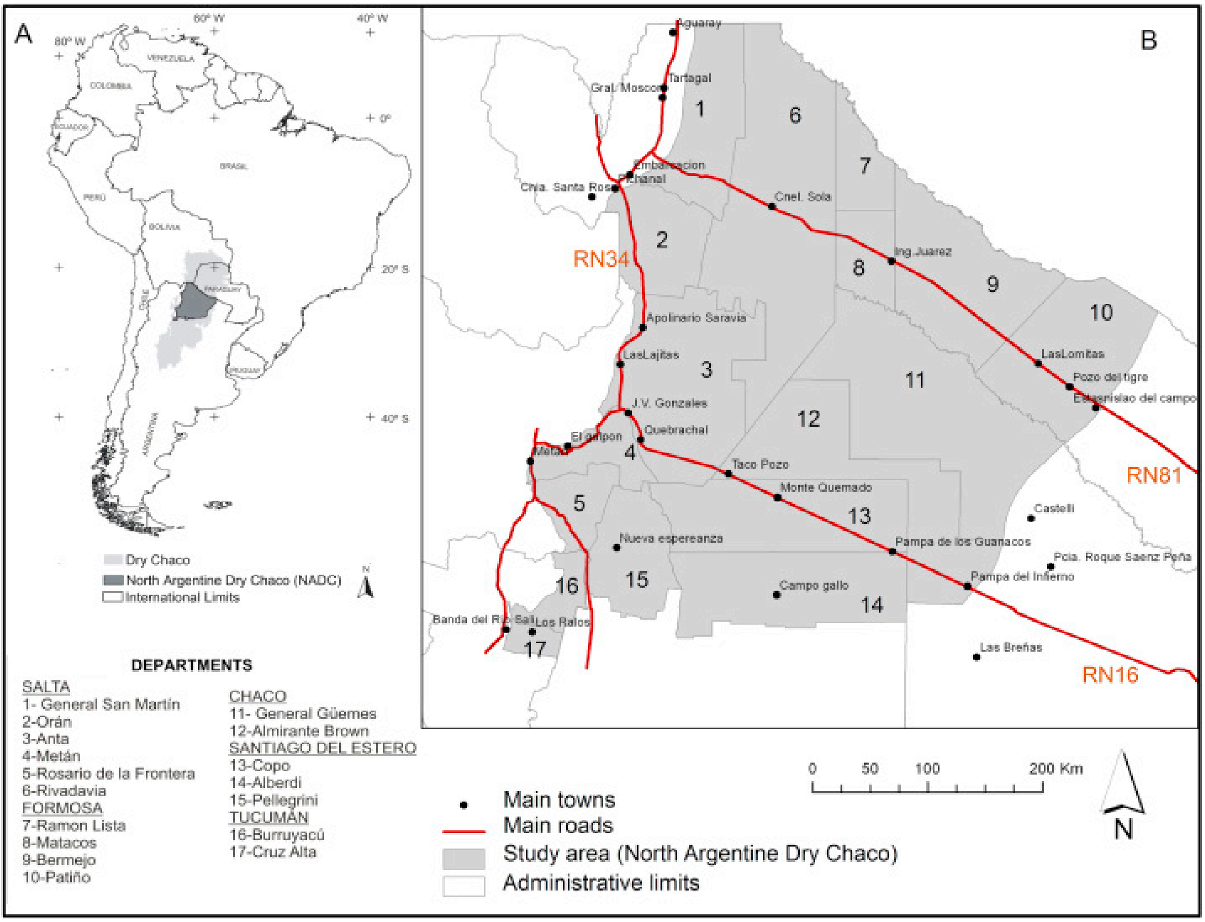


Figure from Gasparri et al 2015 https://doi.org/10.1016/j.jaridenv.2015.05.005

Camera trap photos from Asun.

The study region has a long history of human use, characterized by extensive livestock ranching, mainly of cattle, and selective logging. For the last few decades, there has been a hotspot of deforestation associated with expansion of cattle and cropping. The current actors include:

- Smallholder forest livestock systems: mainly subsistence raising of cattle or goats, under a forest cover. Grazing removes some undergrowth, e.g. around watering points, but rarely full deforestation. Limited cropping for fodder.
- Several indigenous communities: mainly in Formosa province, where the communities of the Toba qom, Pilagá and Wichí manage their land. These communities typically practice

subsistence hunting, fishing and fruit gathering, charcoal and firewood production, handicrafts, and limited cropping and livestock.

- Large agribusiness farms: grazing or cropping. Grazing is dominated by large agribusiness farms, including sown and natural pastures and silvopastures. Cropping is mainly soy, but also wheat, maize, and cotton. Some properties, often larger ones, are managed with tree-breaks.
- Protected areas: some more degraded than others.

Species

We selected 5 forest-associated birds, 5 other birds, and 5 mammals to focus on, to be representative of a range of different species distributions. Note there are many different ways to select species to focus on, including conservation interests, economic interests, and not least, data availability. Our species include:

SPID	Type	Group	SpeciesName	Common name
bf_1	Bird	forest	*Cacicus chrysopterus*	Golden-winged Cacique
bf_2	Bird	forest	*Eudromia formosa*	Quebracho crested tinamou
bf_3	Bird	forest	*Melanerpes candidus*	White woodpecker
bf_4	Bird	forest	*Spiziapteryx circumcincta*	Spot-winged falconet
bf_5	Bird	forest	*Sublegatus modestus*	Southern scrub flycatcher
bo_1	Bird	other	*Campylorhamphus trochilirostris*	Red-billed Scythebill
bo_2	Bird	other	*Crypturellus tataupa*	Tataupa tinamou
bo_3	Bird	other	*Dryocopus schulzi*	Black-bodied woodpecker
bo_4	Bird	other	*Guira guira*	Guira cuckoo
bo_5	Bird	other	*Poospiza torquata*	Ringed warbling finch
mm_1	Mammal	forest	*Catagonus wagneri*	Chacoan Peccary
mm_2	Mammal	forest	*Chaetophractus vellerosus*	Screaming hairy armadillo
mm_3	Mammal	forest	*Galictis cuja*	Lesser grison
mm_4	Mammal	forest	*Leopardus pardalis*	Ocelot
mm_5	Mammal	forest	*Procyon cancrivorus*	Crab-eating Raccoon

Prioritization options

There's many options we could do… over the next session we will have a look at the following (*optional scenarios in italics*):

	NADC example options
Objective	Minimise agricultural opportunity cost
Options	Selection as a reserve or not
Benefits	• Species metric (Area*quality) for 15 species of different groups • *Carbon stock*
Costs	Agricultural opportunity (potential soy profit metric)
Constraints	• Existing protected areas 'locked in' • *Existing intensive production areas 'locked out'* • *High density forest smallholders 'locked out'*
Additions (can be incorporated in different ways)	• Clustering specifications using the Boundary Length Modifier, added to the cost (minimization objective).

Data

Planning units: 1km x 1km raster cells

Cost data: Potential soy cropping production (model based on precipitation and distance to ports)

Species data: Species presence (Maxent models), weighted by relative abundance in forest & natural grassland (field values for birds, and expert opinion for mammals)

Protected areas: include both national and provincial protected areas

Intensive agricultural area: includes current area that is identified as crop in remote sensing analysis

Forest smallholder density: small-scale extensive cattle ranching under forest cover (e.g. 'puesto')

Carbon: Potential carbon stock in forest and natural grassland

NADC basic Prioritizr exercises

In this exercise, we will run a basic Prioritizr problem:

1. Open R via RStudio (you can get this at home from https://www.rstudio.com/)
2. Install and load the required packages
3. Load and explore the 10km resolution data
4. Construct a Prioritizr problem, with lock in and lock out options
5. Solve this, and explore the solutions
6. Add connectivity constraints to the problem
7. Solve, and explore the solutions

Open the Example_2.html file by double clicking on it, and follow the instructions therein.

NADC scenarios exercises

We will split into groups to explore several different scenarios, examples of analyses that we might want to do to understand our system better, and look at the sensitivity of our reserve selections to different options. At the end, we will summarize within each scenario, and present the results to the class (5 slides, 5 minutes).

This way, as a class we will be able to explore 5 different scenarios:

S1. Target – Cost trade-off

S2. Different species groups

S3. Connectivity

S4. Lock in and lock out

S5. Portfolios of solutions

Further descriptions of these, and suggested discussion points are available in the **Scenarios.html** file. We highly recommend that you have a look at this, and start thinking/researching/developing your code before the practical next week!!

If you are a rock star and finish early, you might like to have a go at the bonus scenario:

Bonus scenario: carbon trade-offs

Conservation of biodiversity is of course only one reason we might want to protect natural landscapes from (intensive) development. Other reasons include ecosystem services. One of the most common ecosystem services discussed is Carbon - or more precisely, mitigation of climate change through managing carbon stocks and flow in the landscape. In general, carbon is thought to be 'synergetic' with biodiversity: trees and forests typically house the species of conservation concern, and they also store carbon. But if we maximise or target carbon in the landscape, what does that mean for species conservation? And if we only consider species, what will that give us for carbon? Do these objectives trade off?

Note: The 'carbon' data in the input folder show predicted maximum carbon stock in forest and natural grassland.

Further concepts and reading

Objectives: *Remember* Mace 2014 Science:… that the mission of conservation has experienced a substantial shift over last decades:

- Focus on 'wild' places and 'intact nature' (intrinsic value, moral argument)
- Focus on protecting species and habitats from people (intrinsic value, moral argument)
- Focus on protecting nature for people (instrumental value, economic argument)
- Focus on protecting socio-ecological systems (intrinsic & instrumental)

What might be the objectives for a conservation prioritization in these cases? What might be the available options, the benefits, and the costs considered? What data would you need? How easy might this data be to get?

Data quality and availability: Often input data are assumed to be good/accurate, especially cost and benefit… but often they are not. Species Distribution Models are not always great, particularly when downloaded from sites that pre-prepare them for multiple uses (e.g. IUCN Red List)(Hint: know your data!). Data are often poorly transferrable across space and time (think climate change). Data availability is often a severe constraint, and may bias the features and costs considered. For example, many rare species of conservation concern are poorly researched, and cultural values are difficult to quantify. We often assume that the species we plan for can be adequate proxies for the ones we don't. But we might also like to include planning for features such as ecosystems/habitats, ecosystem processes e.g. that happen on a catchment scale, or special interactions that you might see on ecotones, or in climate refugia of different types.

Feasibility and effectiveness of actions are often assumed to be homogeneous… but are usually not. For example protected areas may prevent (much) development, but do little against hunting, invasive species and disease, and other habitat changes such as climate change or succession. And sometimes they might not even be effective at protecting from deforestation/development either.

Targets, baselines, and other reference points are a very non-trivial decision. Weighting different species – of conservation concern, phylogenetic distinctiveness – might be desirable. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1755-263X.2008.00042.x

Sinclair et al (In press) The use, and usefulness, of spatial conservation prioritizations: https://onlinelibrary.wiley.com/doi/epdf/10.1111/conl.12459

Armsworth et al. 2017 Factoring economic costs into conservation planning may not improve agreement over priorities for protection. https://www.nature.com/articles/s41467-017-02399-y

Armsworth 2014 Inclusion of costs in conservation planning depends on limited datasets and hopeful assumptions https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/nyas.12455

Kukkala and Moilanen 2012 Core concepts of spatial prioritisation in systematic conservation planning https://onlinelibrary.wiley.com/doi/abs/10.1111/brv.12008

Moilanen, Wilson, Possingham 2009. Spatial Conservation Prioritization: Quantitative Methods and Computational Tools. https://global.oup.com/academic/product/spatial-conservation-prioritization-9780199547777?cc=de&lang=en& (available in HU library)