Template to prepare preprints and manuscripts using markdown and github actions

Michael D. Catchen [1,2] Laura Pollock [1,2] Timothée Poisot [3,2] Andrew Gonzalez [1,2]

[1] McGill University [2] Québec Centre for Biodiversity Sciences [3] Université de Montréal

Correspondance to:

Michael D. Catchen — michael.catchen@mcgill.ca

Last revision: *November 28, 2022*

Abstract: TBD

Introduction

Earth's ecosystems are changing due to human activity.

However, we currently lack the data collected in a systematic way to adaquetly attribute change in biodiversity to particular drivers, and to filter out inherent temporal variation in ecological processes from deviations toward non-stationarity (TODO: wording) (**AndyDNAPaper?**).

Sampling is expensive.

What is a BON? Talk about other important aspects of this framework. We want the data to be open. We want to encourage colloboration across scales. Not top-down.

Methods

Combining many geospatial layers into a priority map

A set of geospatial layers, which potentially represent many different types of information. Here we consider each layer to fall into one of three categories: (1) ecological layers, which represent information about biological processes, e.g. species richness, trend in abundance, uncertainty in species occurrence, etc. (2) environmental layers, which represent information about the abiotic environment, e.g. climatic information, land-cover, elevation, climate velocity and uniqueness, and son. Finally, (3) cost layers: e.g. the physical accessability of certain locations, the amount of money per unit of time spend sampling a given location, etc. We denote the value of an arbitrary layer L_i at the coordinate (x, y) as $L_i^{(x,y)}$. We also denoted a generic function $g(L_i)$ where returns the group (ecological, environmental, cost) that the particular L_i belongs to. We also define a set of targets, which are the objective information to be obtained via sampling. (examples here).

The fundemental objects here are the weights matrix $\mathbf{W}$, and mixing vectors γ and ϕ. The weights matrix $\mathbf{W}$ is an $\mathcal{L} \times \mathcal{T}$ matrix, where $\mathcal{L}$ and $\mathcal{T}$ are the number of layers and targets respectively. The value of $\mathbf{w}_{i,t}$ is the relative importance of the i-th layer to the t-th target compared to all of the other layers in the group the i-th layer is in, $g(L_i)$. From this, and the layer group mixing vector ϕ, we can compute single layers for each target, where we similarly denote the value i-th target at location (x, y) as $T_i^{(x,y)}$. This can be

computed as

$$T_i^{(x,y)} = \sum_j \gamma_{g(L_j)} \cdot \mathbf{w}_{j,i} \cdot L_j^{(x,y)}$$

or in slightly more compact matrix form

$$\vec{\mathbf{T}}^{(x,y)} = \mathbf{W}^T(\vec{\gamma}^* \circ \vec{\mathbf{L}}^{(x,y)})$$

where $\vec{\gamma}^*$ is a vector of length $\mathcal{L}$ where $\gamma_i^* = g(L_i)$.

Finally, these target layers can then be combined into a single map as a weighted average based on the target mixing weights ϕ, i.e. the priorty map $\mathbf{P}$ at (x, y) can be computed as

$$\mathbf{P}^{(x,y)} = \sum_i \phi_i T_i^{(x,y)}$$

or in its complete matrix form

$$\mathbf{P}^{(x,y)} = \mathbf{W}^T(\vec{\gamma}^* \circ \mathbf{L}^{(x,y)}) \cdot \phi$$

where $\circ$ is the element-wise product and $\cdot$ is the inner product

[Figure 1 about here.]

Point selection algorithms

There is a long history of discourse in the literature about choosing a set of coordinates within a spatial domain that will result in a "best" sample.

- Environmental uniqueness
- Spatial balance
- Whatever the fuck cube sampling is doing
- Particular scale-dependence weirdness in ecology

Choosing and optimizing weights

There are a lot of caveats here.

- Rarely is there relevant available data from which to derive an evidence-based choice of the relative importance of each layer toward sampling targets.
- Often knowledge of species life-history matters for this, requires experts on particular species
- There is no a priori reason to believe there will be a positive trade-off possible between two targets.
- A particular choice of weight matrix may produce a priority map that is very sensitive to minor changes in the weights
- We should try to validate that our sampling sites work well, but this is a chicken and egg problem with our current data. So we want a max-entropy approach to real distributions from which we sample possible realized sampling outcomes and compare as we tweak weights.
- Need to assert the constraint that

$$\frac{\partial \mathbb{L}_i}{\partial T_i} > 0$$

  or otherwise that means the weights are not actually suited toward targets

Target tradeoffs

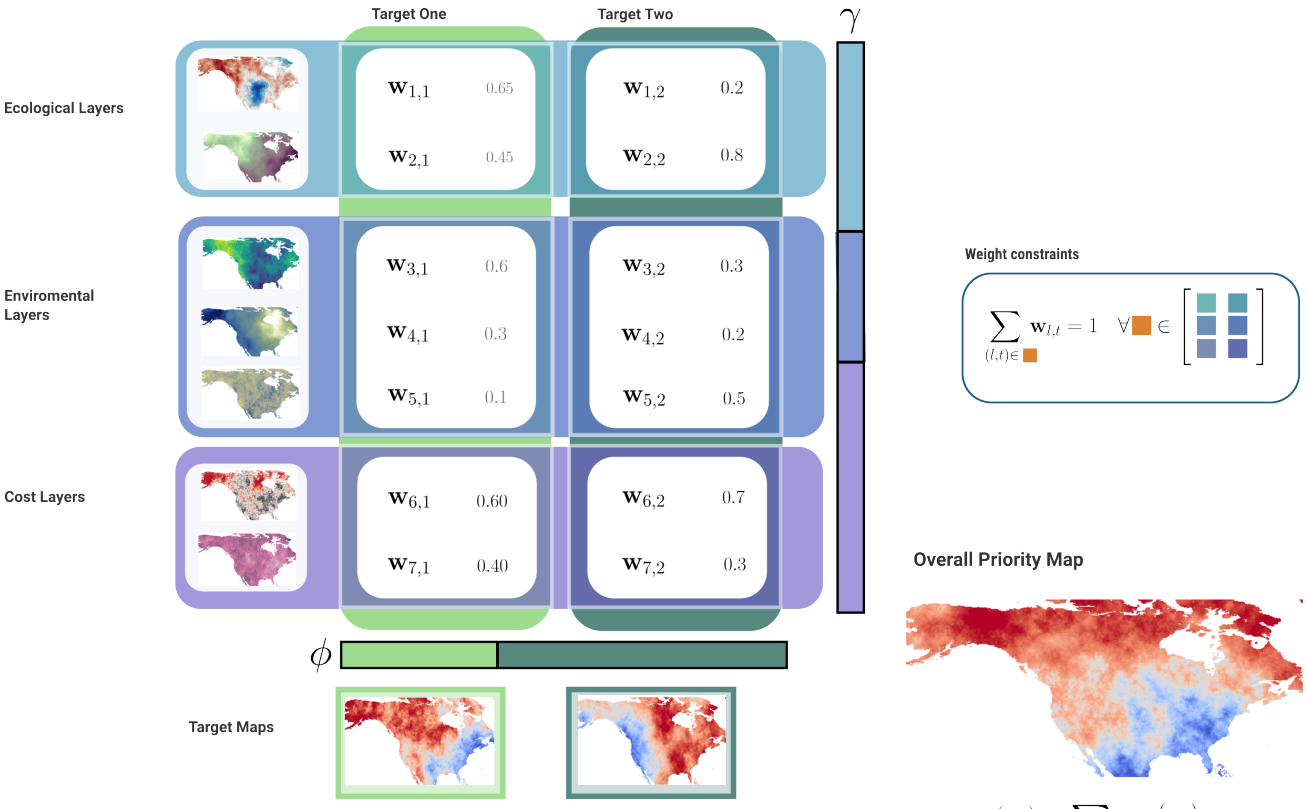
[Figure 2 about here.]

Sensitivity analysis

Validation through simulation of the sampling process

Case study: bird (or perhaps non migratory?) species TBD in Quebec

Discussion



$$T_i^{(x,y)} = \sum_j \gamma_{g(L_j)} \cdot \mathbf{w}_{j,i} \cdot L_j^{(x,y)}$$

$$\vec{\mathbf{T}}^{(x,y)} = \mathbf{W}^T(\gamma^* \circ \vec{\mathbf{L}}^{(x,y)})$$

$$P^{(x,y)} = \sum_i \phi_i T_i^{(x,y)}$$

$$\mathbf{P}^{(x,y)} = \mathbf{W}^T(\gamma^* \circ \vec{\mathbf{L}}^{(x,y)}) \cdot \phi$$

Figure 1: todo

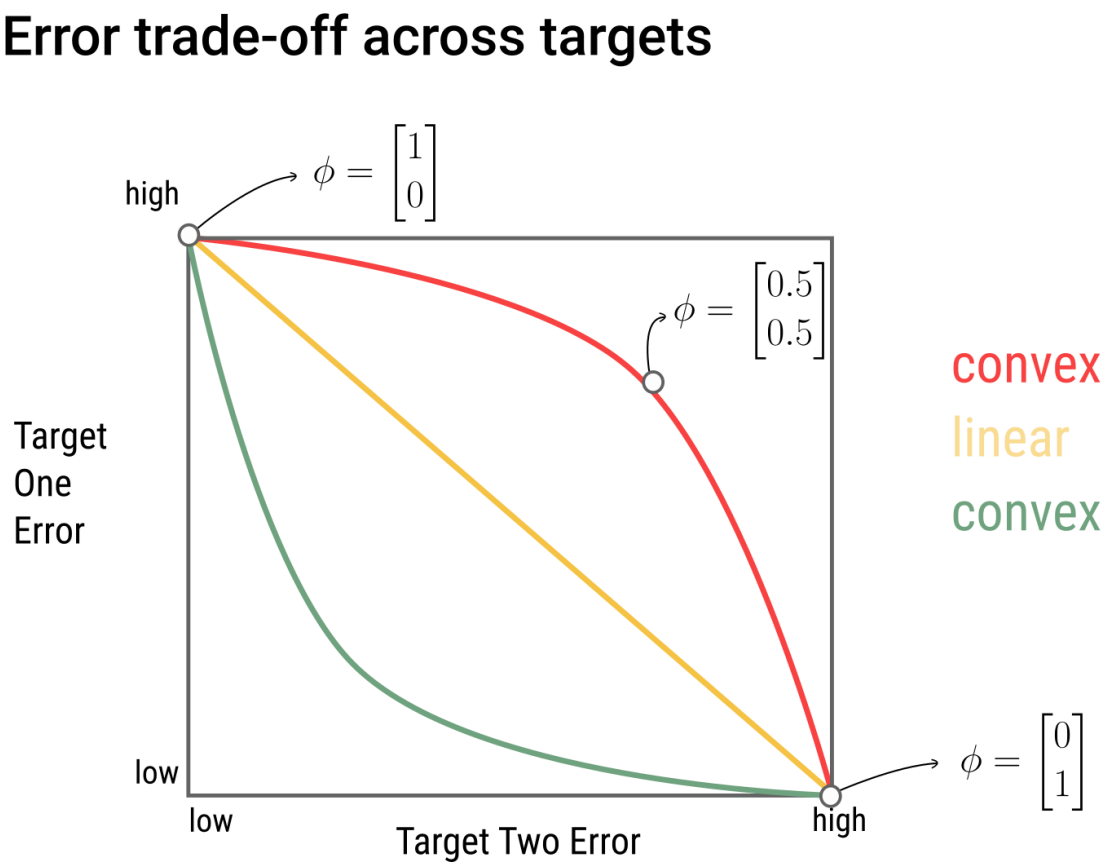


Figure 2: todo