On the effectiveness of reconstructing biodiversity from indicator species along a climate gradient

Ilhem Bouderbala, Daniel Fortin, Junior A. Tremblay, Antoine Allard, Louis-Paul Rivest, Patrick Desrosiers

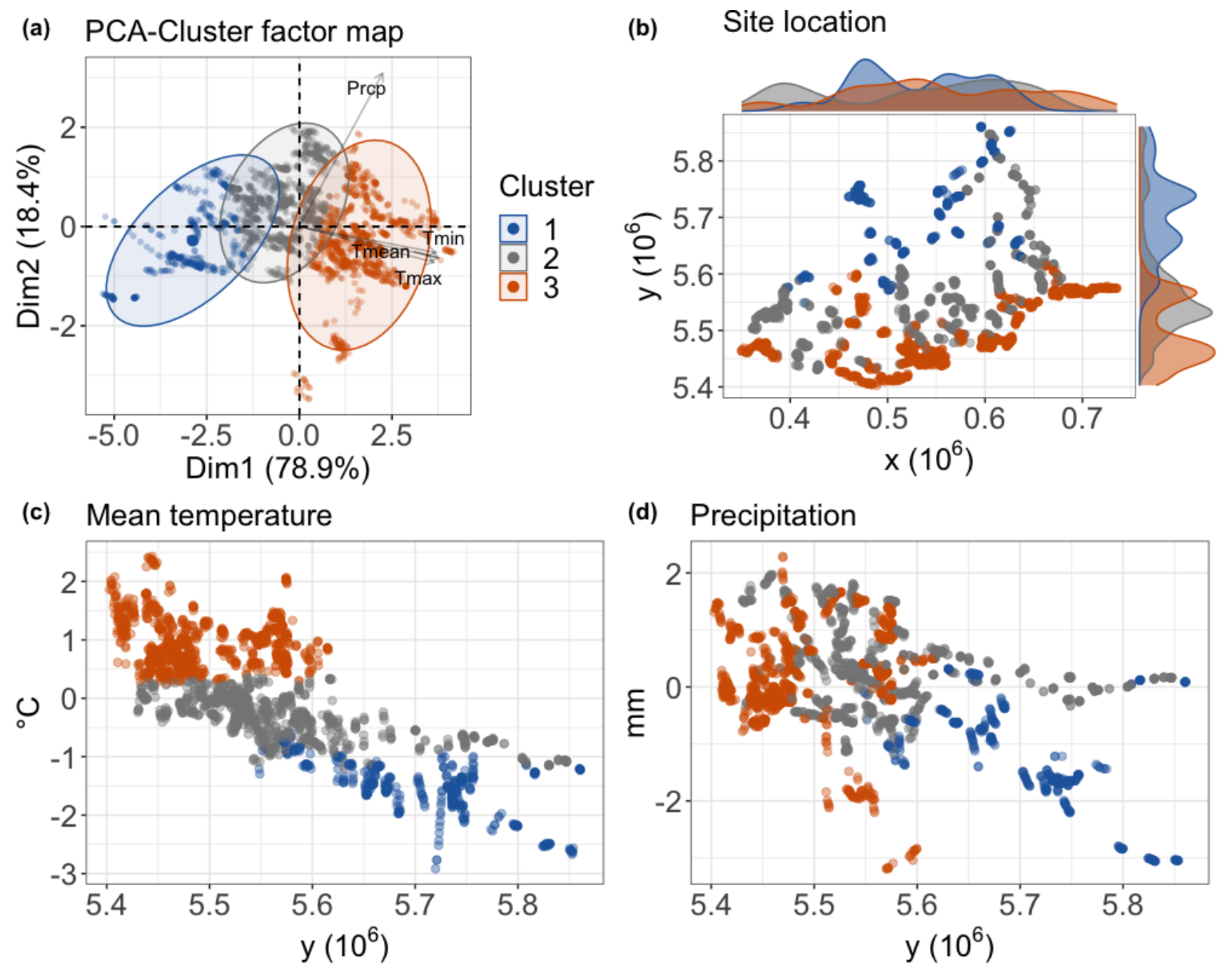
Laval University

Reconstructing biodiversity from indicator species

Indicator species emerged as a promising tool to monitor diversity because their presence indicates a maximum number of conditionally co-occurring species [1]. However, species richness is often insufficient to characterize biodiversity [2]. We aim to assess the effectiveness of indicator species for biodiversity reconstruction based on their co-occurrence with other species.

Identification of climate spatial clusters

We specified spatial clusters to consider the impact of climate variation on the identification of indicator species and assemblages [2]



Co-occurrence networks

To analyze community co-occurrence, we where based on a probabilistic species co-occurrence analysis [3]. The probability that two species co-occur at j sites, $\max\{0, N_1+N_2-N\} \leq j \leq \min\{N_1, N_2\}$ given in [3] simplify to a hypergeometric distribution $H\left(N_2, \frac{N_1}{N}, N\right)$ where N is the total of sites, N_1 and N_2 are the number of sites occupied by species 1 and species 2 respectively.

Network-based metrics

We denote by q the number of the potential indicator species (PIS) and by p the number of indicator species ($1 \leq p < q$). We used four network-based metrics to select the PIS:

1. maximum significant positive co-occurrences (**Positive**).
2. maximum significant negative co-occurrences (**Negative**).
3. maximum of the normalized betweenness centrality (**Betweenness**):
4. maximum closeness centrality among species having the maximum of betweenness (**BC**).

SDM calibration for indicator species

We calibrate the SDM of the jth indicator species of the hth combination in two steps:

$$log\left(\frac{P^h\left(\tilde{X}_j=1|Y\right)}{1-P^h\left(\tilde{X}_j=1|Y\right)}\right)=\alpha_{0,j}+\alpha_{1,j}.Y_j^*,\ j=1,...,p,$$

if $p>1$:

$$log\left(\frac{P^h\left(\tilde{X}_j=1|\tilde{X}_1,,...,\tilde{X}_{j-1},Y\right)}{1-P^h\left(\tilde{X}_j=1|\tilde{X}_1,,...,\tilde{X}_{j-1},Y\right)}\right)=\alpha_{0,j}+\alpha_{1,j}.Y_j^*+\beta_1\tilde{X}_1+...+\beta_{j-1}\tilde{X}_{j-1}, \quad (1)$$
$$j=1,...,p,$$

where Y_j^* and $\tilde{X}_j$ represent respectively the climate predictors and the occurrence of the jth IS.

Community occurrence prediction

The occurrence probability of the jth non-indicator species is given by:

$$P^h\left(X_j\left(S_t\right)=1\mid Y\left(S_t\right)\right)=\sum_{j^1=0}^{1}\sum_{j^2=0}^{1}...\sum_{j^p=0}^{1}$$
$$P^h\left(X_j\left(S_t\right)=1\mid \tilde{X}_1\left(S_t\right)=j^1,\tilde{X}_2\left(S_t\right)=j^2,...,\tilde{X}_p\left(S_t\right)=j^p,Y\left(S_t\right)\right)\times \quad (2)$$
$$P^h\left(\tilde{X}_1\left(S_t\right)=j^1,\tilde{X}_2\left(S_t\right)=j^2,...,\tilde{X}_p\left(S_t\right)=j^p\mid Y\left(S_t\right)\right)$$

where X_j represents the occurrence of the jth non-indicator species, S_t is the tth sampling site and N is the number of sites.

Accuracy analysis

We quantify the dissimilarity between the original and the predicted binarized vectors using Sorensen dissimilarity

$$SD_j^h=\frac{PA_j^h+AP_j^h}{2PP_j^h+PA_j^h+AP_j^h}, j\in\{1,N_{Spe}\} \quad (3)$$

PP_j represents the number of sites where species j is present in both vectors. PA_j represents the number of sites where species j is present in the predicted vector and absent in the observed vector, AP_j is the is the converse.

Model's accuracy threshold

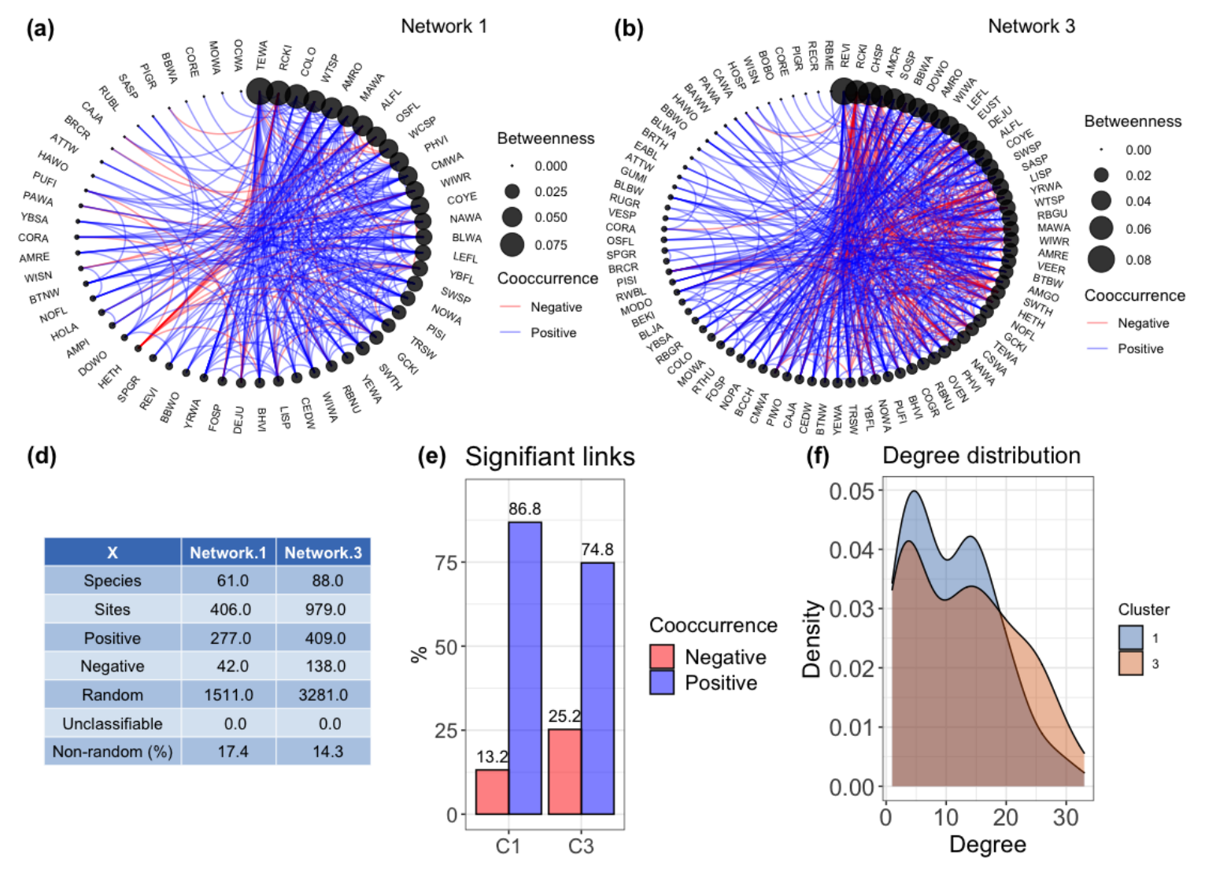
We estimate the species occurrence probability threshold (η) and considered a species as a well classified if $SD \leq \gamma$ and $OP \geq \eta$ or $SD < \gamma$ and $OP > \eta$ otherwise it is misclassified. Our approach is based on selecting η as the occurrence percentage of the species that maximizes the order percentage of species well classified (PWC):

$$\hat{\eta}=OP_{j*},\ j*=\underset{j=1,...,N_{Spe}}{\text{argmax}}\ PWC(j) \quad (4)$$

$$PWC(j)=100\times\frac{\sum_{l=1}^{\bar{N}_j}1\left(\left(SD_l^{h^*}\leq\gamma,OP_l\geq\eta\right)\bigvee\left(SD_l^{h^*}>\gamma,OP_l<\eta\right)\right)}{\bar{N}_j} \quad (5)$$

where $\bar{N}_j$ represents the number of species having $OP_l \leq OP_j, l=1,...,\bar{N}_j$.

Results



X	Network.1	Network.3
Species	61.0	88.0
Sites	406.0	979.0
Positive	277.0	409.0
Negative	42.0	138.0
Random	1511.0	3281.0
Unclassifiable	0.0	0.0
Non-random (%)	17.4	14.3

Figure 1. Co-occurrence networks

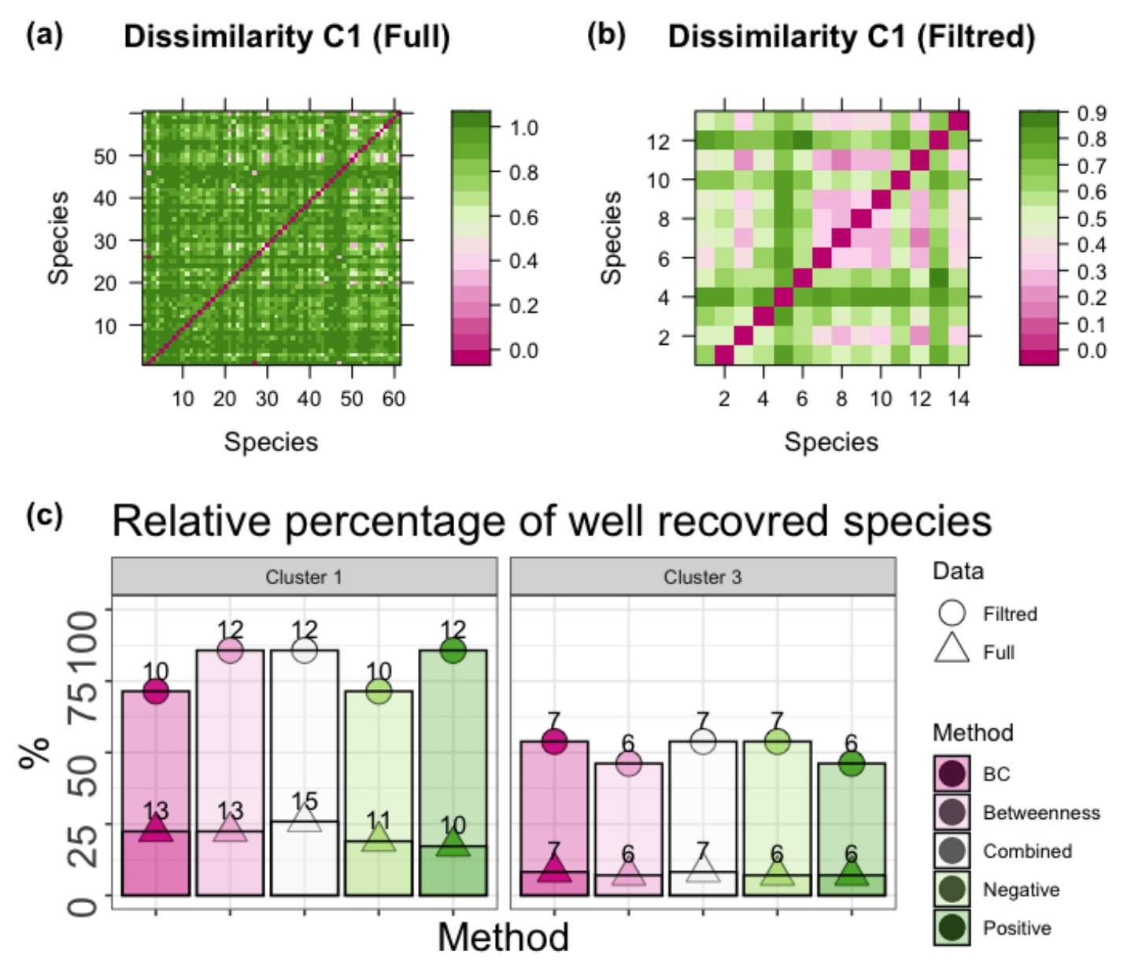
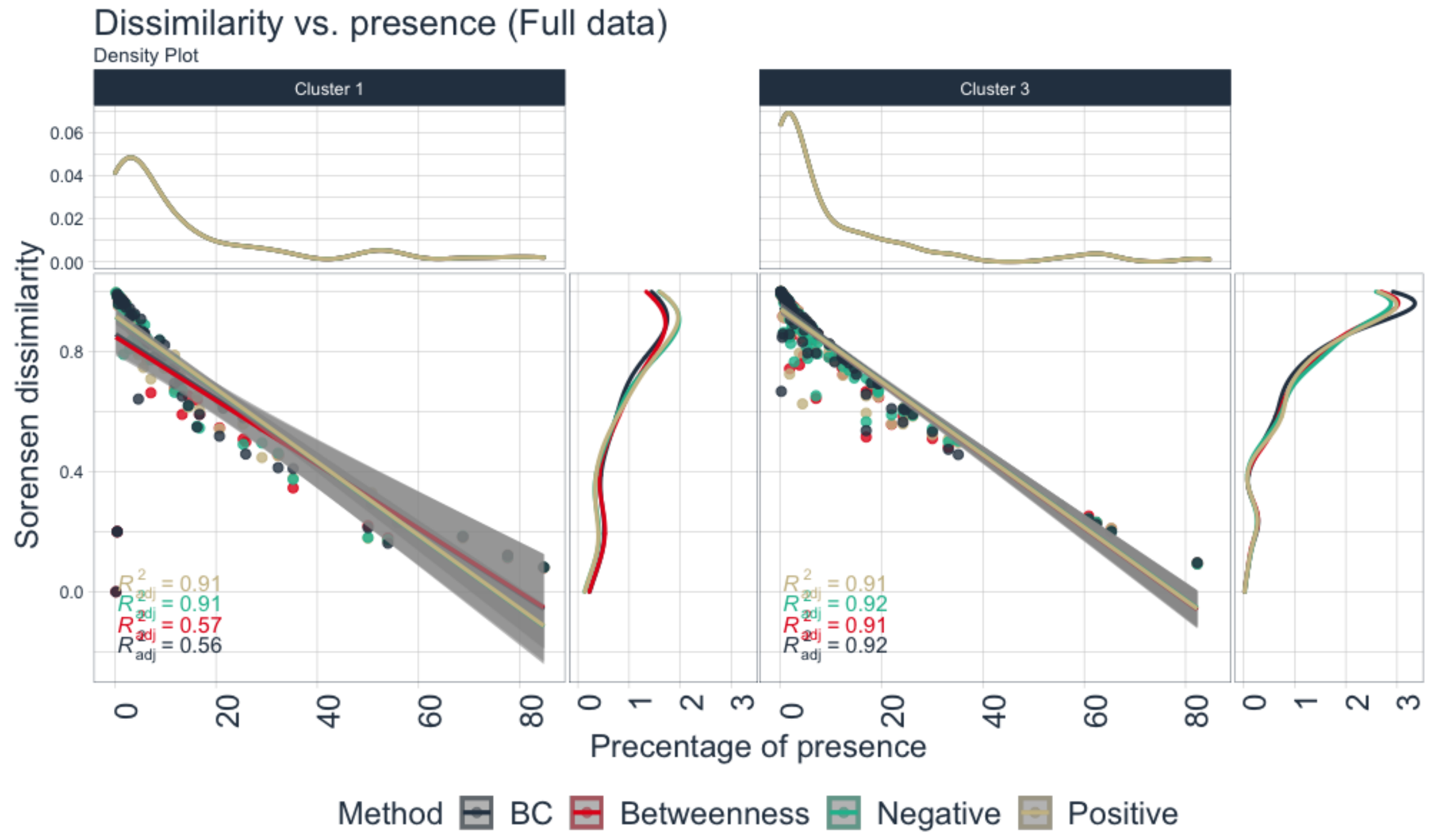
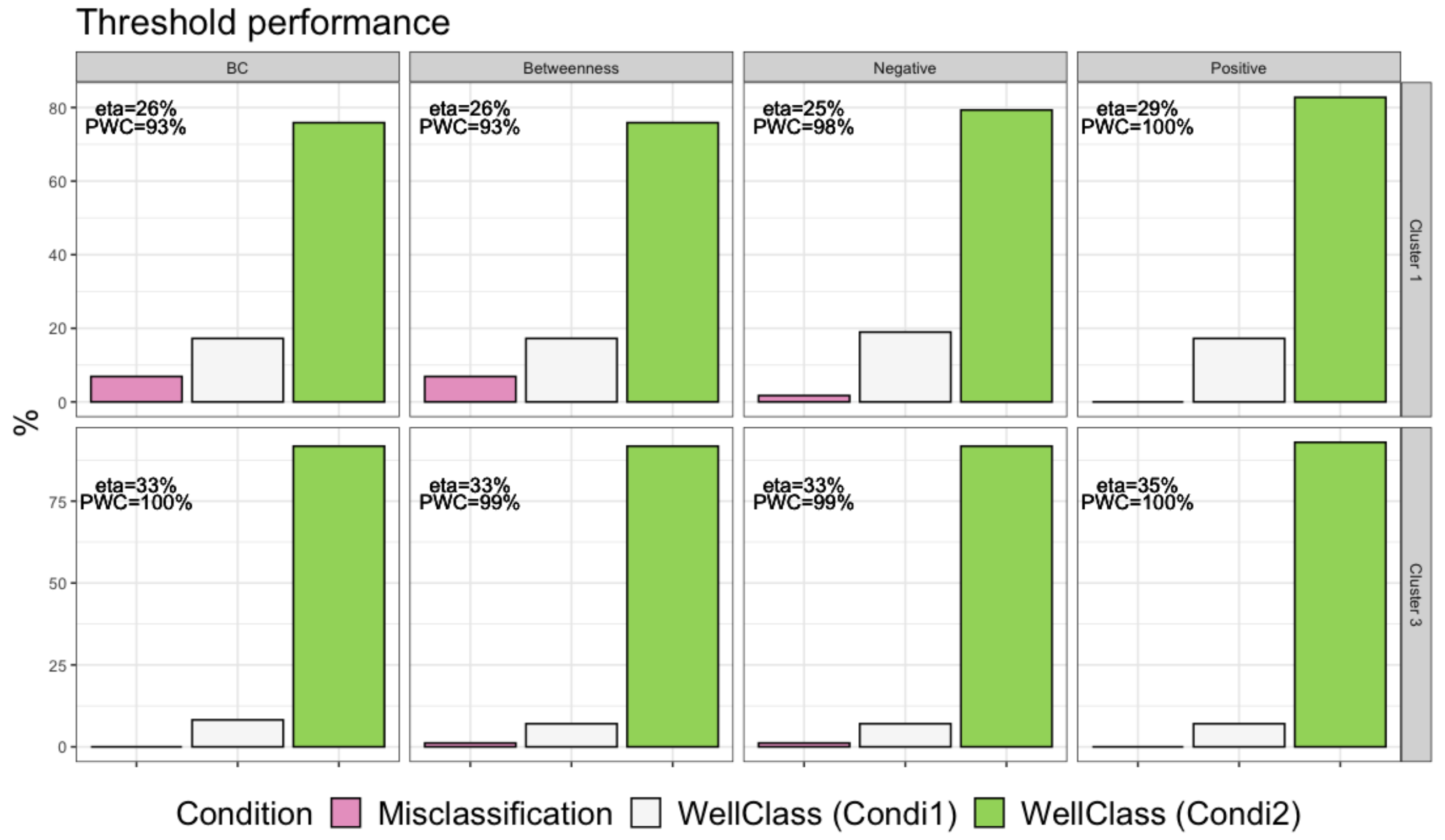


Figure 2. Accuracy analysis





Conclusion

- Depending on climate clusters, our approach enabled us to recover the occurrence of 86% and 54% of species present in more than 20% of sites of northern and southern groups, respectively.
- To exceed 0.5 of Sorensen similarity, the species must be present in more than 35% of sites.
- The higher success at the north sites could be due either to the lower species richness observed at higher latitudes or to the higher competition in the southern sites (the proportion of negative links was about 25% in the southern network whereas it was 13%).

References

[1] Ermias T. Azeria, Daniel Fortin, Christian Hébert, Pedro Peres-Neto, David Pothier, and Jean-Claude Ruel. Using null model analysis of species co-occurrences to deconstruct biodiversity patterns and select indicator species. *Diversity and Distributions*, 15(6):958–971, 2009.

[2] Alexandre Terrigeol, Sergio Ewane Ebouele, Marcel Darveau, Christian Hébert, Louis-Paul Rivest, and Daniel Fortin. On the efficiency of indicator species for broad-scale monitoring of bird diversity across climate conditions. *Ecological Indicators*, 137:108773, 2022.

[3] Joseph A. Veech. A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, 22(2):252–260, 2013.