A probabilistic methodology to reconstruct biodiversity from implicit interactions with indicator species along latitudinal co-occurrence networks

Ilhem Bouderbala[1, 2, 3,*], Junior A. Tremblay [4,5], Daniel Fortin[3], Antoine Allard[2], Patrick Desrosiers[2,6],

October 28, 2023

[1] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton (Ab), Canada T6G 2R3 (current affiliation)

[2] Département de physique, de génie physique et d'optique, Université Laval, Québec (Qc), Canada G1V 0A6

[3] Département de biologie, Université Laval, Québec (Qc), Canada G1V 0A6

[4] Division de la recherche sur la faune, Environnement et Changement Climatique Canada, Québec (Qc), Canada, G1J 03

[5] Département des sciences du bois et de la forêt, Université Laval, Québec (Qc), Canada, G1V 0A6

[6] Centre de recherche CERVO, Québec (Qc), Canada G1J 2G3

Abstract

With the ongoing decline in biodiversity, there is a need for efficient field monitoring techniques. Indicator species (IS) emerged as a promising tool to monitor diversity because their presence indicates a maximum number of conditionally co-occurring species. Accordingly, there is empirical evidence that few IS could predict local richness. However, species richness is often insufficient to characterize biodiversity. We aim to assess the effectiveness of IS for biodiversity reconstruction based on their co-occurrence with other species. We develop probabilistic models that use IS to predict the occurrence probability of co-occurring species. We predict the occurrence of species based on (1) their conditional occurrence probability with IS and (2) the occurrence probability of IS. We test the approach with field observations of birds in the Côte-Nord region of Québec. First, we identify climate networks (based on temperature and precipitation) over which bird associations remain relatively stable to identify the most locally relevant

*bouderba@ualberta.ca

IS. We use four network-based methods to select the IS according to betweenness, betweenness-closeness centralities, and positive and negative co-occurrences. From co-occurrence networks, we conclude that the latitudinal climate gradient impacts the nature of both biotic interactions and IS composition. Indeed, the proportion of negative links, which reflect competition and avoidance liaisons, was about 25.2% in the southern network, whereas it was 13.2% in the northern network. Moreover, almost a complete turnover in the composition of the IS was observed between the northern and southern networks, with only two common IS among 17. Regarding the effectiveness in the reconstruction of assemblages occurrence, we observed a strong negative correlation ($r \leq -0.75$) between the percentage of sites occupied and the dissimilarity between the original and the estimated occurrences for a given species. More precisely, species must be present in more than 25% and 33% of northern and southern sites to recover well from its co-occurrence with IS. Depending on climate clusters, our approach enables us to recover the occurrence up to 100% and 70% of species present in more than 20% of sites of northern and southern groups, respectively. However, the accuracy decreases to 9% and 12% of northern and southern groups, respectively, when we include less-present species. The higher success at the north sites reflects the lower species richness observed at higher latitudes. In conclusion, our method demonstrates that it is possible to predict local species assemblages based on the occurrence and absence of IS. Nevertheless, the relatively low success of less present species illustrates the need for further theoretical development to reconstruct biodiversity, mainly to recover the occurrence of rare species.

Keywords: Biodiversity reconstruction, indicator species, co-occurrence networks, probabilistic species co-occurrence analysis, boreal forests, bird assemblage, climate gradient.

1 Introduction

Climate change is a significant threat to biodiversity and can potentially disrupt ecosystems' functioning (Thuiller et al., 2011; Pachauri et al., 2014; Zhang and Liang, 2014). Rising global temperatures and changing weather patterns are causing shifts in habitats and altering the availability of resources necessary for species survival (Duveneck et al., 2014). In particular, in Canada, the rise in the temperature led to the northern migration of thermophilous hard-wood trees, which was replacing boreal conifers (Boulanger and Puigdevall, 2021). In addition to climate change, ecosystems are also impacted by other anthropogenic changes such as land use, pollution, invasive species and the occurrence of extreme climatic events (Masson-Delmotte et al., 2018; Rechkemmer and von Falkenhayn, 2009; Labadie et al., 2021). While climate change plays a significant role in biodiversity loss, land-use changes may be a more vital driver of biodiversity loss in specific ecosystems, particularly in the short term. The cumulative effect of habitat change, occurring through anthropogenic disturbances, with the influence of climate and meteorological conditions, impacts heavily ecosystem functioning (Boulanger et al., 2017; Micheletti et al., 2021). Indeed, the coupled effect of climate change and land use leads to biodiversity alternation mainly caused by species turnover (Pachauri

et al., 2014; Cadieux et al., 2020; Bouderbala et al., 2023a). These changes in biodiversity could also exert pressure on agroecosystems and result in a loss of ecosystem services (Kelly and Goulden, 2008; Hillebrand et al., 2010).

The significance of monitoring biodiversity and the development of reconstruction tools has become increasingly crucial (Azeria et al., 2009; Dommain et al., 2020). Monitoring biodiversity allows us to understand the causes and current state of biodiversity loss to identify the main drivers of biodiversity decline or increase (Bouderbala et al., 2023a), such as habitat destruction, pollution, climate change, and invasive species. Moreover, monitoring biodiversity helps plan mitigation strategies and implement effective conservation actions (Stralberg et al., 2015). Furthermore, monitoring biodiversity is necessary to evaluate the effectiveness of conservation measures and adjust them accordingly (Robichaud and Knopff, 2015; Labadie et al., 2023). However, identifying species of many taxa to reconstruct biodiversity needs expertise and could be costly (Kim and Byrne, 2006). Therefore, methods based on using IS to represent species assemblages are well used to monitor wildlife conservation and ecological restoration (Dufrêne and Legendre, 1997; Chu et al., 2021). Usually, few IS could be informative to monitor environmental changes and provide signals for plausible ecological shift (Siddig et al., 2016; Chu et al., 2021). To use a fixed set of IS over a large spatiotemporal scale, those species must effectively predict some indicator, such as species richness, despite environmental changes. However, the ongoing changes in climate conditions permanently shape biodiversity patterns and community composition (Santillán et al., 2018; Bouderbala et al., 2023b). Therefore, the predictive potential of IS over a large spatial extent must be studied cautiously.

The statistical methods used to determine IS are developing rapidly and was used as an informing management decision tool (Fleming et al., 2020; Chu et al., 2021). Thus, a method based on the null models of species co-occurrence classification to identify sets of species that may have similar and distinct response patterns to their environment has been developed by Azeria et al. (2009) and Terrigeol et al. (2022). Fleming et al. (2020) developed conceptual models of IS to include a direct relationship between an IS, ecosystem change drivers and latent processes and variables. Furthermore, Chu et al. (2021) developed an approach for incorporating species-conditional co-occurrence into models used to select marine IS.

In this work, we suggested a probabilistic model to assess the reconstruction of the biodiversity from the interaction with the indicator spaces along the latitudinal gradient. We quantified the effectiveness of the reconstruction based on the difference between the estimated assemblage occurrence using our approach and the original occurrence. We analyzed bird species assemblages in the Côte-Nord region of Québec, Canada, within an area of 114118 km^2. We develop a model to identify the IS based on a probabilistic species co-occurrence analysis (Veech, 2013) along a latitudinal gradient (Lurgi et al., 2020). The latitudinal co-occurrence networks conception is a unique possibility to identify how much the composition of the IS under latitudinal gradient is changing. The identified latitudinal climate clusters represented a spatial variation in short-term climate conditions. We built latitudinal co-occurrence networks among each climate cluster

and identified the indicators species among each network based on network metrics evaluated based on co-occurrence significance. The latitudinal analysis helps us to analyze how the biotic interactions are changing. This will allow us to understand how the change in climate conditions will alter community composition (Graham and Grimm, 1990) and change the strength of interspecific interactions (Parmesan and Yohe, 2003; Harley, 2011). Moreover, we quantified the biodiversity reconstruction's effectiveness by comparing the original occurrences and the estimated through our probabilistic approach. Therefore, we developed probabilistic models to estimate the occurrence of species in assemblages based on (1) their co-occurrence with the local IS of the assemblage and (2) the modelling of the occurrence of the IS. Afterward, we explicitly emphasized the conditions under which the biodiversity could be reconstructed from the interaction with the local IS.

2 Network-based community occurrence model

2.1 Framework

We aimed to predict species occurrence based on their interaction with IS and quantify the accuracy of this approach to predict species occurrence following the latitudinal gradient. We first modelled the occurrence of the selected IS based on a single-species model as a function of the climate variables and the occurrence of the other IS in the same climate region. The second step consisted of calculating the occurrence probability of species in the pool conditional to the presence and the absence of the IS. Finally, we combined the two precedent probabilities to estimate the community occurrence. We divided our approach into six steps (see Fig. 1).

2.2 Identification of climate spatial clusters

Static IS are most likely ineffective in monitoring biodiversity over large climate extent (Morelli, 2015; Terrigeol et al., 2022). Therefore, we specified spatial clusters according to a climatic gradient to consider the impact of climate variation on identifying IS and assemblages. We identified the different climate clusters using hierarchical clustering on principal components (package 'FactoMineR', Husson et al. (2013)) based on the two first principal components. We used the climate variables as input variables for the principal component analysis (PCA).

2.3 Co-occurrence networks

We built a co-occurrence network under each climate cluster to extract the IS. We used occurrence community data (species by site matrix) where 1 indicated a presence and 0 an absence. Species were considered

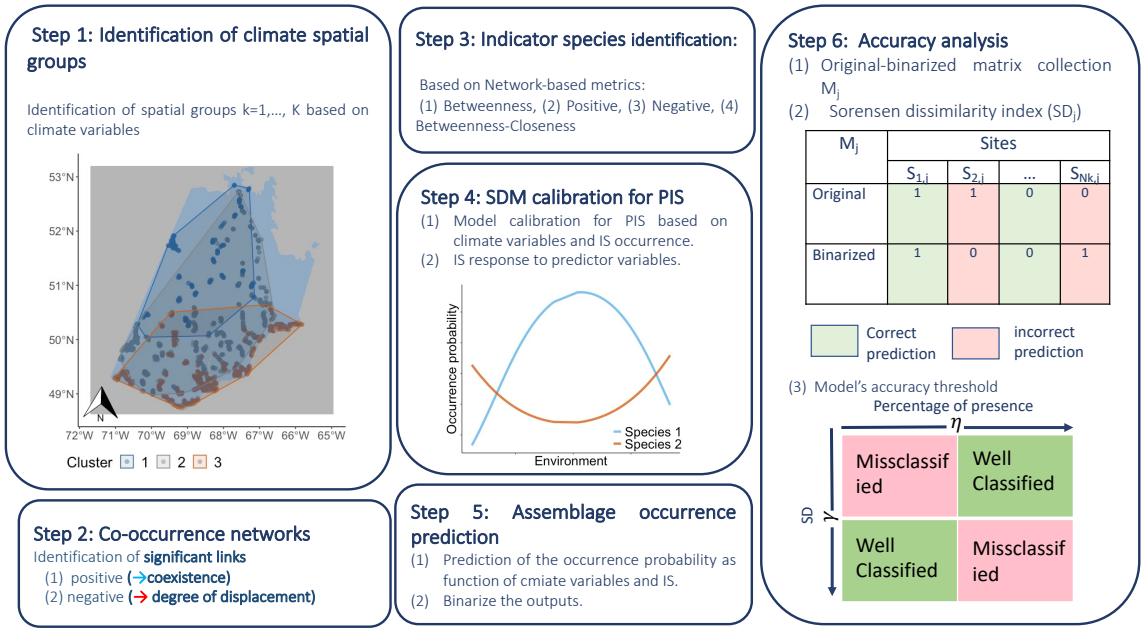


Figure 1: The prediction of species occurrence from IS occurrence. In Step 1, we identified spatial clusters based on the climate variables (based on the following four variables: precipitation, mean, minimum and maximum temperature). In Step 2, we built the co-occurrence networks based on Veech's probabilistic model (Veech, 2013). In step 3, we identified the IS for each climate region based on network-based metrics. In step 4, we calibrated the SDMs of the IS for each climate group. In step 5, we estimated the occurrence probability of the non-IS in the pool based on the conditional occurrence to the corresponding IS and the occurrence of the IS (Eq. 2). We estimated the accuracy of our approach in the last step through dissimilarity analysis between the observed occurrences and the estimated ones (step 5).

nodes, whereas links reflected positive (coexistence) or negative (displacement) liaisons. To evaluate the co-occurrence significance, we used probabilistic species co-occurrence analysis (Veech, 2013) ('coccur' package, Griffith et al. (2016)). Traditionally, building spatial networks of co-occurrence involved principally using null models based on data randomization (Araújo et al., 2011), which is considered a potential source of Type I and Type II errors (Veech, 2013). The probabilistic approach given in Veech (2013) is *distribution-free* and demonstrated to perform better than methods based on data randomization by having a lower Type I error rate and a low Type II error rate. The probability that two species co-occur at j sites with $\max\{0, N_1 + N_2 - N\} \leq j \leq \min\{N_1, N_2\}$ followed a hypergeometric distribution $H\left(N_2, \frac{N_1}{N}, N\right)$ (Veech, 2013):

$$P_j = \frac{\binom{N_1}{j} \times \binom{N-N_1}{N_2-j}}{\binom{N}{N_2}}$$

where N is the total of sites, N_1 and N_2 are the number of sites occupied by species one and two, respectively. $P_{lt} = \sum_{j=0}^{Q_{obs}} P_j$, $P_{gt} = \sum_{j=Q_{obs}}^{N} P_j$ are used as p_values for testing whether species one and species two co-occur significantly less often or significantly more often, respectively, than expected by chance. Q_{obs} is

the observed number of sites having both species. We included only species pairs with significant positive or negative co-occurrence at $\alpha = 0.05$.

2.4 Network-based IS identification

To simplify notations, we remove hereafter the variations between clusters and the method used to extract the IS; however, the procedure remains the same. Our methodology represented each climate cluster by its own set of IS. We considered the number (q) of the potential IS (PIS) as well as the number of IS ($1 \leq p < q$). We used four network-based metrics to select the IS belonging to each climate cluster. The first method was based on selecting the first p species with maximum significant positive co-occurrences (**Positive**). The second method chose the first p species with maximum significant negative co-occurrences (**Negative**). The third method was based on choosing the first p species with a maximum normalized betweenness centrality (**Betweenness**):

$$NB_i = \frac{2B_i}{(N_{Spe}-1)(N_{Spe}-2)}, B_i = \sum_{j \neq i, l \neq i}^{N_{Spe}} \frac{\sigma_{jl}(i)}{\sigma_{jl}},\ i \in \{1, ..., N_{Spe}\}$$

where $\sigma_{jl}(i)$ is the number of the shortest paths relating species j to species l passing by species i and σ_{jl} is the number of the shortest paths joining species j to species l (we used package 'igraph', Csardi and Nepusz (2006) to compute this metric). In the last method, PIS ($Spe_{B,q}$) were considered as the q species having the maximum of betweenness and the IS will be the most p heterogeneous species (maximum normalized closeness centrality) among $Spe_{B,q} = \cup_{i=1}^{q} i_{\bar{B}}$, where $i_{\bar{B}}$ is species order corresponding to the decreasing betweenness such that if $i_{\bar{B}} < j_{\bar{B}}$ then $B_i > B_j$. The betweenness-closeness (**BC**) centrality is given by:

$$BC_i = \frac{N_{Spe}-1}{\sum_{j=1, j \neq i}^{N_{Spe}} d_{ij}},\ i \in Spe_{B,q} \tag{1}$$

where d_{ij} is the length of the shortest path between species i and j. The selection of the IS according to the BC method consists of choosing the first p species with minimum BC (Eq. 1). This procedure prevented selecting highly associated species, which most likely represent the same assemblages. (See Fig. 2 for an illustration between IS selected from the four methods).

2.5 Species distribution models calibration for IS

We calibrated the species distribution models (SDM) of the jth IS in two steps: (1) Calibrating the SDM based only on climate variables (Y, by including all the potential quadratic and logarithmic transformations)

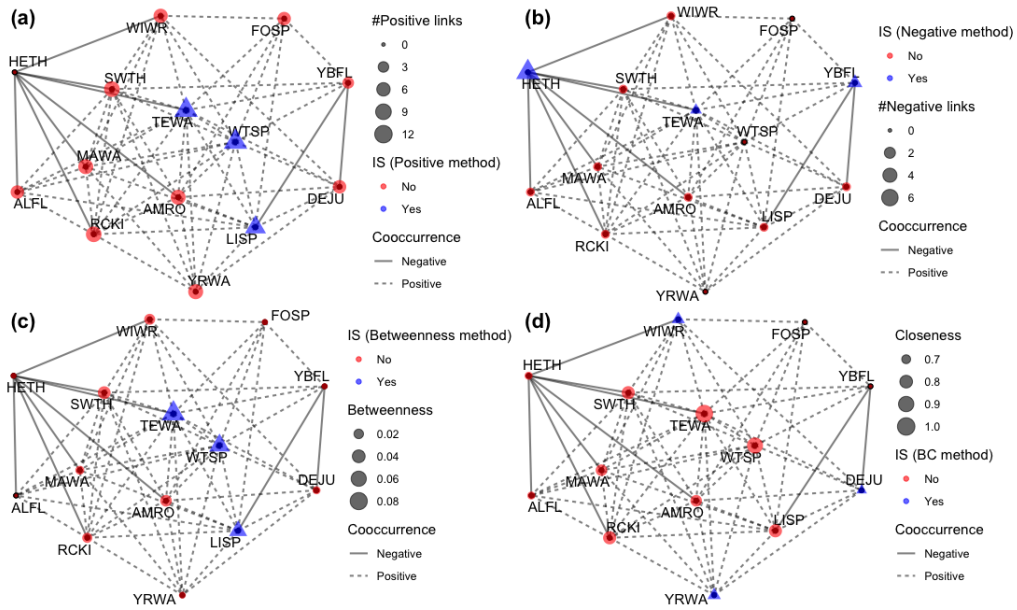


Figure 2: Example of the IS selected from the four metrics. The illustrated co-occurrence network corresponds to the filtered data (at least 20% of presence) for the northern Cluster. The red and blue links represent the negative and the positive significant co-occurrences. The four graphs differ only on the metric for selecting the three IS (triangles). (a) shows the method based on the number of positive c-occurrences, (b) indicates the method based on the number of negative c-occurrences, (c) expresses the method based on the maximum of the betweenness, (d) shows the method based on the minimum of closeness among the PIS having the maximum betweenness (**BC**).

and selecting variables through a step-wise procedure:

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

(2) Calibrating the final regression by adding the occurrence of the previous IS as independent variables:

$$log\left(\frac{P\left(\tilde{X}_j = 1|\tilde{X}_1, , ..., \tilde{X}_{j-1}, Y\right)}{1 - P\left(\tilde{X}_j = 1|\tilde{X}_1, , ..., \tilde{X}_{j-1}, Y\right)}\right) = \alpha_{0,j} + \alpha_{1,j}.Y_j^* + \beta_1\tilde{X}_1 + ... + \beta_{j-1}\tilde{X}_{j-1}, \quad (2)$$

$$j = 1, ..., p,$$

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

2.6 Community occurrence prediction

In our methodology, the occurrence probability of the jth non-IS depends implicitly on its interactions with the selected IS set. This probability is function of (1) the probability of occurrence of IS (Eq. 2) and (2) the probability of presence of the species conditional on the presence or the absence of IS (Eq. 5), which reflects

the interaction between the IS and the species in question. The occurrence probability of the jth non-IS is given by:

$$P\left(X_j\left(S_t\right)=1 \mid Y\left(S_t\right)\right)=\sum_{j^1=0}^{1}\sum_{j^2=0}^{1}...\sum_{j^p=0}^{1}P\Bigg(X_j\left(S_t\right)=1,\tilde{X}_1\left(S_t\right)=j^1,$$
$$\tilde{X}_2\left(S_t\right)=j^2,...,\tilde{X}_p\left(S_t\right)=j^p \mid Y\left(S_t\right)\Bigg),\, t=1,....,N,\, j\in\{1,N_{Spe}\},\, j\notin IS, \tag{3}$$

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

$$P\left(X_j\left(S_t\right)=1 \mid Y\left(S_t\right)\right)=\sum_{j^1=0}^{1}\sum_{j^2=0}^{1}...\sum_{j^p=0}^{1}$$
$$P\Bigg(X_j\left(S_t\right)=1 \mid \tilde{X}_1\left(S_t\right)=j^1,\tilde{X}_2\left(S_t\right)=j^2,...,\tilde{X}_p\left(S_t\right)=j^p,Y\left(S_t\right)\Bigg)\times \tag{4}$$
$$P\Bigg(\tilde{X}_1\left(S_t\right)=j^1,\tilde{X}_2\left(S_t\right)=j^2,...,\tilde{X}_p\left(S_t\right)=j^p \mid Y\left(S_t\right)\Bigg)$$

We assume independence between the occurrence of the jth species and the environmental conditions conditional to the IS. The first component in the right hand of the Eq. 4 is estimated as follows:

$$\hat{P}\Bigg(X_j\left(S_t\right)=1 \mid \tilde{X}_1\left(S_t\right)=j^1,\tilde{X}_2\left(S_t\right)=j^2,...,\tilde{X}_p\left(S_{t,k}\right)=j^p,Y\left(S_t\right)\Bigg)=$$
$$\begin{cases}\dfrac{\sum_{t=1}^{N}\mathbb{1}\left(X_j\left(S_t\right)=1,\tilde{X}_1\left(S_t\right)=j^1,...,\tilde{X}_p\left(S_t\right)=j^p\right)}{\sum_{t=1}^{N}\mathbb{1}\left(\tilde{X}_1\left(S_t\right)=j^1,...,\tilde{X}_p\left(S_t\right)=j^p\right)}, \\ \text{if } \sum_{t=1}^{N}\mathbb{1}\left(\tilde{X}_1\left(S_t\right)=j^1,...,\tilde{X}_p\left(S_t\right)=j^p\right)>0 \\ 0, \quad \text{otherwise}\end{cases} \tag{5}$$

Eq. 2 and Eq. 5 are then used to estimate the occurrence probability of all the other species in the pool:

$$\hat{P}\left(X_j\left(S_t\right)=1 \mid Y\left(S_t\right)\right)=\sum_{j^1=0}^{1}\sum_{j^2=0}^{1}...\sum_{j^p=0}^{1}$$
$$\frac{\sum_{t=1}^{N}\mathbb{1}\left(X_j\left(S_t\right)=1,\tilde{X}_1\left(S_t\right)=j^1,...,\tilde{X}_p\left(S_t\right)=j^p\right)}{\sum_{t=1}^{N}\mathbb{1}\left(\tilde{X}_1\left(S_t\right)=j^1,...,\tilde{X}_p\left(S_t\right)=j^p\right)}\times$$
$$\hat{P}\left(\tilde{X}_1\left(S_t\right)=j^1 \mid \tilde{X}_2\left(S_t\right)=j^2,...,\tilde{X}_p\left(S_t\right)=j^p,Y\left(S_t\right)\right)\times\cdots \tag{6}$$
$$\hat{P}\left(\tilde{X}_{p-1}\left(S_t\right)=j^{p-1} \mid \tilde{X}_p\left(S_t\right)=j^p,Y\left(S_t\right)\right)\times$$
$$\hat{P}\left(\tilde{X}_p\left(S_t\right)=j^p \mid Y\left(S_t\right)\right)$$

3 Accuracy analysis

3.1 Reconstruction dissimilarity

To quantify the performance of our approach and estimate the number of species well recovered based on their interaction with IS, we quantified the Sorensen dissimilarity (SD) between the original and the predicted binarized occurrence:

$$SD_j = \frac{PA_j + AP_j}{2PP_j + PA_j + AP_j}, j \in \{1, N_{Spe}\} \tag{7}$$

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 the converse. We combined all the methods that satisfy $SD \leq \gamma$ (we choose $\gamma = 0.5$) to compute the number of recovered species. The choice of the binarization threshold was based on the method that maximizes the accuracy between the real presence-absence data and the predictions from the 12 optimization methods, such as a maximum of true skill and Kappa statistics (Araújo et al., 2005; Allouche et al., 2006), developed in the package 'PresenceAbsence' (Freeman and Moisen, 2008).

3.2 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 $OP \geq \eta$ (**WellClass (Condi1)**) or $SD > \gamma$ and $OP < \eta$ (**WellClass (Condi2)**) otherwize it is misclassified (**Misclassification**). Our approach is based on selecting η as the occurrence percentage of the species which maximizes the percentage of species well classified (PWC):

$$\hat{\eta} = OP_{j*},\ j* = \underset{j=1,\ldots,N_{Spe}}{\operatorname{argmax}}\ PWC(j) \tag{8}$$

$$PWC(j) = 100 \times \frac{\sum_{l=1}^{\bar{N}_j} \mathbb{1}\left((SD_l \leq \gamma, OP_l \geq \eta) \bigvee (SD_l > \gamma, OP_l < \eta)\right)}{\bar{N}_j} \tag{9}$$

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

4 Model's application

4.1 Study area and occurrence data

The study area is located in the Côte-Nord region of Québec, Canada ($48°N$ to $53°N$, $65°W$ to $71°W$), within an area of 114118 km^2 (Fig. 1). The northern part of the study area belongs to the black spruce-

feather moss bioclimatic domain. The southern part of the study area belongs to the eastern balsam fir–white birch subdomain, mostly dominated by balsam fir and white spruce mixed with white (paper) birch.

Presence-absence data were collected between 2010 and 2018 to model species distributions. We used the data from the Atlas of Breeding Birds (Atlas des oiseaux nicheurs du Québec, 2018), which were based on species occurrences that were detected using unlimited distance 5-minute point counts (Bibby et al., 2000), which were collected during the breeding season (late May to mid-July) between 2010 and 2018. To quantify the importance of species occurrence on the performance, we compared the outcomes based on the original presence-absence matrix (**Full data**) and a subset including only species present in more than 20% of sites under each Cluster (**Filtered data**).

4.2 Climate predictor variables

To predict species occurrence, we used climate variables. Initially, we generated 17 potential climate variables at a 250-m resolution using the BioSim platform (Régnière et al., 2017), including the annual average of temperature and precipitation between 2004 and 2018 (see Table. 1 in Supporting Information for the description of all potential predictor variables). BioSIM simulates daily maximum and minimum temperatures, precipitation, water deficit, mean daily relative humidity and wind speed by matching georeferenced sources of daily weather data to spatially georeferenced points. BioSIM uses spatial regression to adjust weather data for differences in latitude, longitude, and elevation between the sources of weather data and each field location (for more details, see (Boulanger et al., 2018)). In our case, the spatially referenced points were 15,000 randomly located across the Province of Québec. In contrast, weather data were daily data originating from discrete weather stations within the province. We generated climate variables at a 250 m scale by spatially interpolating data from the 15,000 random points using kriging and elevation as drifting variables.

5 Results

5.1 Identification of climate spatial clusters

We used the following four climate variables to identify the principal components: precipitation, mean temperature, maximum temperature and minimum temperature along the Côte-Nord region in Québec. Three climate clusters were identified using hierarchical clustering (see Fig. 3a), pointing to a latitudinal solid climatic gradient (see Fig. 3b). Indeed, this climatic gradient was principally explained by the latitudinal shift in the temperature (see Fig. 3d). For instance, Cluster 1, which the northern, registered the lowest temperature variation with a mean temperature between -4.68°C and -0.28°C followed by Cluster 2 with a

mean temperature between -1.46°C and 1.53°C. Cluster 3 was the southern one with maximum temperature with mean temperature between 1.31°C and 5.44°C. The precipitation did not contribute as much as the temperature in the latitudinal explanation of the three climate clusters (see, for instance, the difference between Fig. 3d and Fig. 3e). We compared only Cluster 1 and 3 in this work (see S2 File) for the data under the two climate clusters).

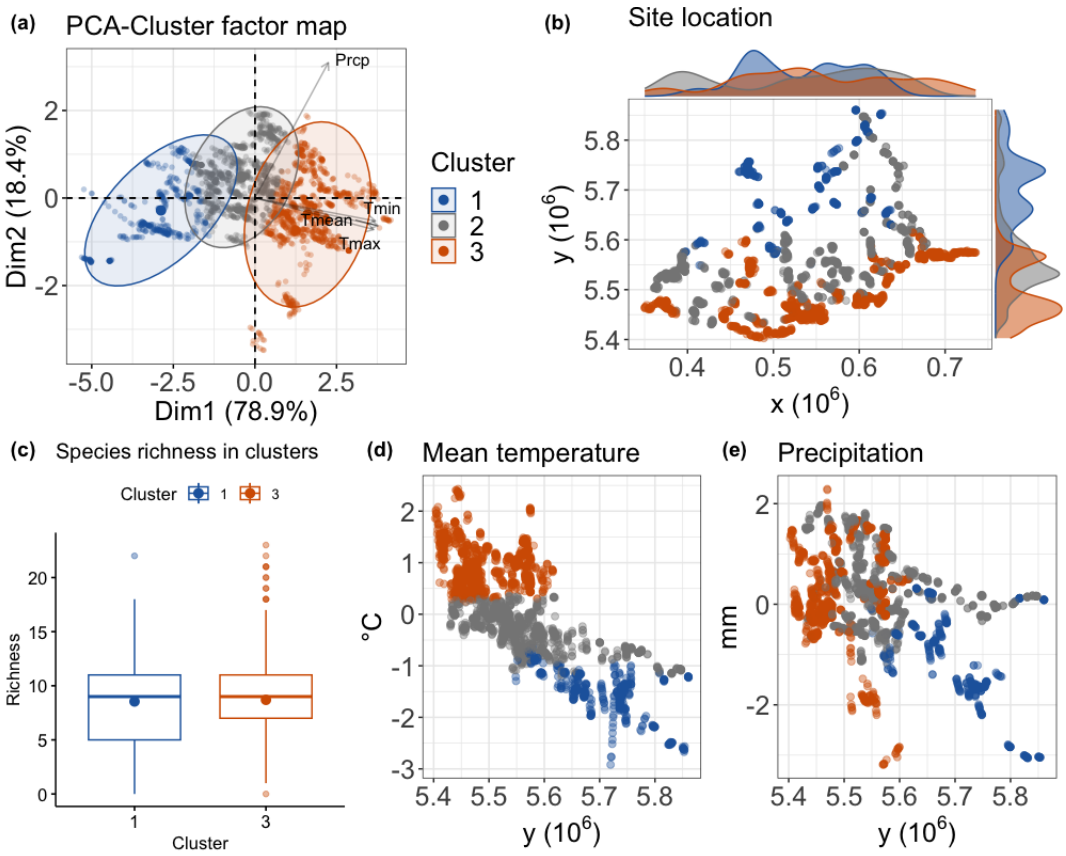


Figure 3: Identification of climate spatial clusters. (a) shows the factor map of Principal Component Analysis (PCA). The hierarchical clustering is applied to the PCA, and the results allowed three distinct homogeneous latitudinal climate clusters: Cluster 1 is northern, and Cluster 3 is southern. (b) The spatial distribution of the three climate clusters. (c) Species richness at the northern and southern clusters. (d)-(e) The distribution of temperature and precipitation within the three climate clusters.

5.2 Co-occurrence network analysis

We considered each climate cluster as a network. We analyzed the outputs of the northern, corresponding to Cluster 1, and the southern, corresponding to Cluster 3, co-occurrence networks. Cluster 3 included more species than Cluster 1 (88 versus 61 species respectively), with more than double the number of sites (979 versus 406 respectively). However, Cluster 3, which is the southern Cluster, registered almost double the percentage of negative co-occurrence links (see Fig. 4b) compared to the northern Cluster (25.2% versus

13.2% respectively). Also, more species have the highest co-occurrences (with a degree between 20 and 30) in the south than in the north. Inversely, there are more species with low co-occurrence (with a degree between 0 and 20) in the north than in the south (see Fig. 4c). Moreover, we observed a bimodal degree distribution in both networks, which means there are two groups of weak-interacting and strong-interacting species. The separation between weak-interacting and strong-interacting species is stronger in the northern as the separation between the two groups is sharper under Cluster 1 (see Fig. 4c).

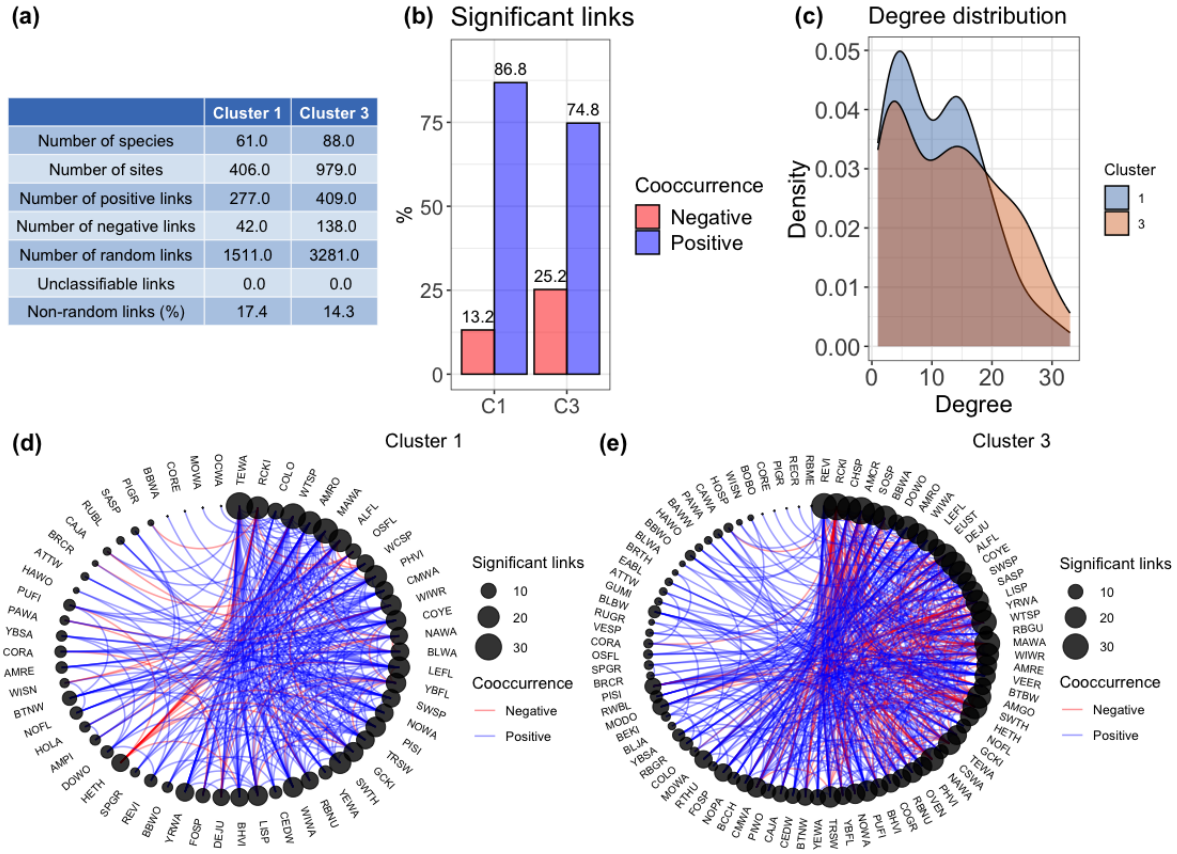


Figure 4: Networks characteristics. (a) The information related to the northern (Cluster 1) and southern (Cluster 3) networks. (b) The degree distribution of each network. (c) The percentage of positive and negative links (co-occurrences) within the northern (C1) and southern (C3) networks. (d)-(e) The northern and southern co-occurrence networks. The blue and red links represent the negative and positive significant co-occurrences. The size of the circle for each species represents the degree.

5.3 Network-based indicator species

We identified three IS based on the four metrics for north (Cluster 1) and south (Cluster 3) climate networks. We considered the most informative three species for each of the four network-based metrics. Only two species were common IS between the northern and southern networks. The other 15 species were identified in the northern but not in the southern networks and inversely (see Fig. 5a). For both networks, all metrics shared

at most one IS among the three (see Fig. 5b-c). We noticed that the selected IS based on betweenness centrality are the most similar to the three other sets of IS. Indeed, for Cluster 1 (northern network), the similarity between IS based on betweenness centrality and betweenness-closeness centrality and positive co-occurrences was 0.33. For Cluster 3 (southern network), the similarity between IS based on betweenness centrality with positive and negative co-occurrences was 0.33. Betweenness-closeness centrality was similar to betweenness centrality and negative co-occurrence but only under the northern network. Overall, the similarity in the selection of the IS was more substantial under the northern network than the southern network.

A significant difference in the habitat (see S1) of the IS between the northern and southern networks was observed. Although the majority of the IS were associated with early-to-mid succession forests, only one species was associated with late succession forests within only the northern network (*Yellow-bellied Flycatcher* (YBFL) species). All four generalist species were selected as IS only within the southern network (see S1 Table).

5.4 Performance of the indicator species in recovering assemblages occurrence

According to all four metrics, there is a negative relationship between species presence percentage and the dissimilarity between the true and the estimated occurrence according to our probabilistic method (see Fig. 6). Indeed, Pearson correlation between SD, calculated between the true occurrences and the estimated ones, and the percentage of presence for Cluster 1 was between -0.75 (based on the betweenness-closeness method) and -0.94 (based on the negative method). For Cluster 3, the correlation between SD and the percentage of presence was between -0.89 (based on the betweenness method) and -0.96 (based on the betweenness-closeness method). Therefore, the occurrence of highly present species will be more efficient to recover based on their co-occurrence with IS. To evaluate how the percentage of presence impacted the effectiveness in estimating species occurrence, we compared the percentage of well-recovered species between full and filtered data (see Fig. 7c). We considered a species as well-recovered if SD between the original and the estimated species occurrence is less than 0.5. Also, the filtered data included only species with a percentage of presence at least equal to 20%. For the fitted data, the percentage of well-recovered species was between 86% and 100% within the northern network (Cluster 1) and between 40% and 70% within the southern network (Cluster 3). However, the percentage of well-recovered species was between 16% and 28% within the northern network (Cluster 1) and between 9% and 12% within the southern network (Cluster 3). We estimated the threshold of the percentage of presence to ensure dissimilarity effectiveness less than 0.5 ($\gamma = 0.5$). We estimated a threshold of presence for the northern network $\hat{\eta}$ between 25% and 29% to ensure a percentage of well-classified species (PWC) between 95% and 100%. For the southern Cluster, the presence threshold equals 33%, ensuring a PWC between 98% and 100% (see Fig. 7d).

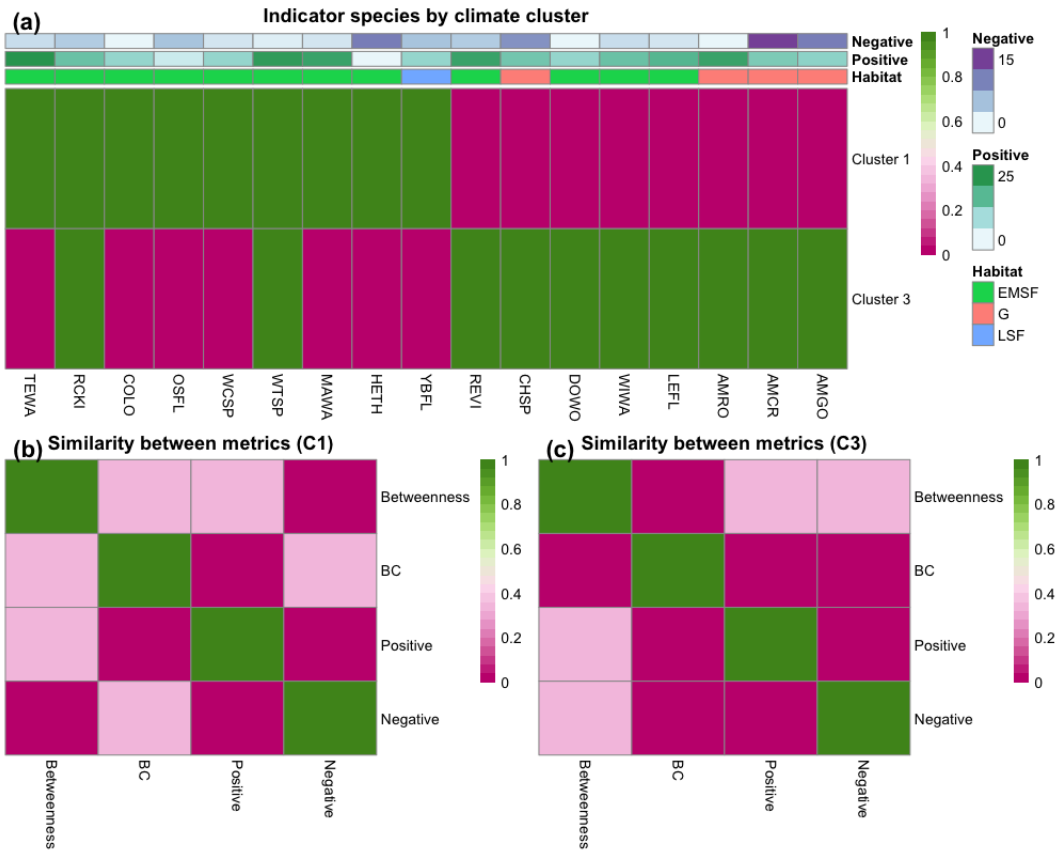


Figure 5: The similarity in the selection of IS is based on the four metrics: betweenness-closeness, betweenness, negative co-occurrences and positive co-occurrences within each of the northern and the southern climate clusters. (a) The identified IS according to all the metrics. Values of one and zero indicate, respectively, that the species is either an indicator or not an IS for the specified climate cluster. We also showed the species' habitat and the number of positive and negative significant co-occurrences for each species. Habitat shows the IS habitat: late succession forest (*LSF*), early-to-mid succession forest (*EMSF*) or generalist species (*G*) (see S1 and S1 File). (b)-(c) Similarity with a value of 0 indicates no common IS between a pair of metrics. A value of 1 indicates the same set of IS between the two metrics (see S1 Table).

6 Discussion

We aimed through this study to quantify how much the IS could be informative to reconstruct assemblages occurrence and under which conditions the reconstruction is effective. We combined a co-occurrence network analysis with species distribution models to calculate the probability of occurrence of bird species conditional to the presence and the absence of the IS among latitudinal networks. Gradual changes in climatic conditions in latitude indicate ongoing climate change. Using IS is a helpful and informative tool to monitor biodiversity (Elmendorf et al., 2015; Mimet et al., 2016). However, the limitations are mainly due to the change in the composition of the IS when the climate conditions change (Terrigeol et al., 2022). Accordingly, we noticed that the latitudinal gradient plays a crucial role in the change in the nature of the composition of the IS and the biotic interactions. Indeed, there is an almost complete turnover in the composition of the IS between

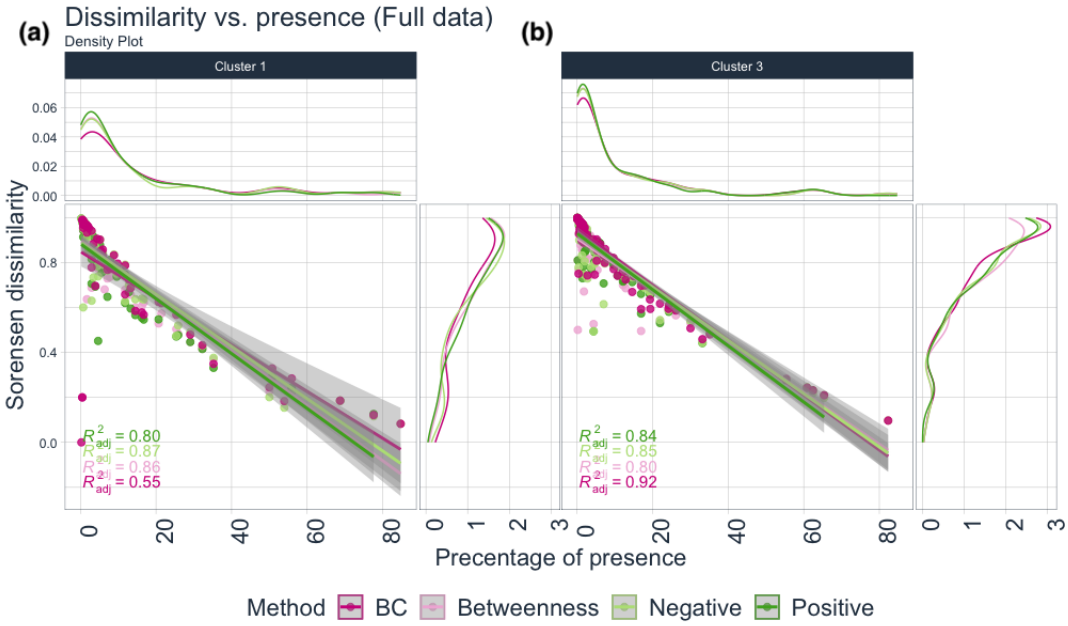


Figure 6: Cluster's assemblage dissimilarity vs. species' percentage of presence within each northern and the southern climate clusters. The Sorensen dissimilarity between the estimated occurrences and the original occurrences was computed and plotted with the percentage of presence of each species. (a) and (b) represent respectively the accuracy among the northern and southern networks respectively.

the northern and southern networks based on our network-based metrics. Furthermore, the percentage of negative co-occurrence, an indicator of competition and avoidance liaisons, is higher under the southern network and inversely for positive co-occurrences. Indeed, recent observations showed that climate change can change the abundance of species and alter biotic interaction with a loss of interactions through extinctions, range shifts, and changes in relative abundance (Blois et al., 2013). Accordingly, Milazzo et al. (2013) showed that species co-occurrence could switch to competitive displacement mode in rising temperatures. We showed that biodiversity reconstruction could be effective only when the species are abundant. The conditions about the minimum percentage of presence to effectively recover the assemblage occurrence becomes more severe going south. Therefore, species must be more present under the lowest latitudes to recover well by interacting with the identified IS.

The effectiveness in the reconstruction of assemblage occurrence

Based on our approach, we found that a reconstruction of the biodiversity from the co-occurrence with the IS could be achieved under some conditions related to the percentage of species' occupancy in the assemblage. More precisely, a reconstruction could be very efficient, with dissimilarity between the true occurrences and the estimated ones less than 0.5, if the presence percentage is more than 25% and 33% in the northern and the southern networks. Therefore, the percentage of presence toward the south is more severe. The lower

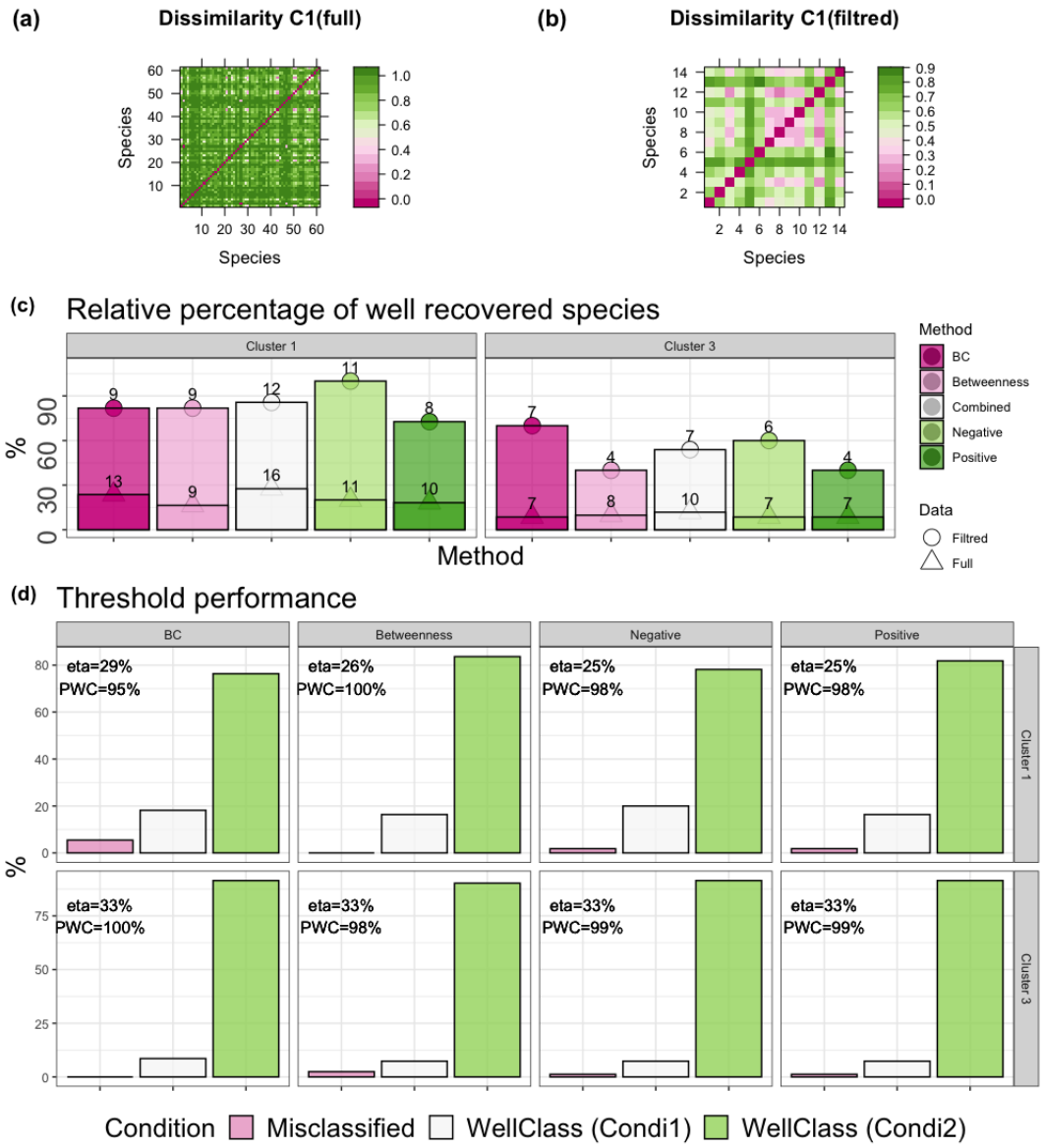


Figure 7: The difference in the relative percentage of well-recovered species between full and filtered occurrence data. (a)-(b) Illustration of dissimilarity between species occurrence using full and filtered data for Cluster 1. (c) The percentage of well-recovered species, with dissimilarity between the original occurrence and the predicted occurrences, is less than 0.5 between full and filtered data. (d) The threshold of performance η and PWC represent, respectively, the species occurrence probability threshold and percentage of species well classified (Eq. 8-9).

minimum request of the percentage of presence at the north sites reflects the lower species richness (see Fig. 3c) observed at higher latitudes. This pattern could be explained by the latitudinal diversity gradient where the richness incenses toward the Tropics (Pianka, 1966; Rohde, 1992; Hillebrand, 2004). Moreover, based on our approach, the recovery of the rare species is not possible based on the co-occurrence of IS representative for each climate cluster. This could be due to the spatial variation in species assemblages. Indeed, Blanchet et al. (2020) showed that often species co-occur without significant interaction, and consequently, it becomes

more complicated that they belong to a shared community under different climate conditions (Terrigeol et al., 2022).

Latitudinal gradient and the alternation of interspecific interactions

We compared two latitudinal networks to determine the ecological consequences of spatial gradient and thermal stress on the alternation of interspecific interaction. Latitudinal data offers the almost unique possibility of studying climate change's effect on the composition of assemblages. The latitudinal networks consist of arrays along the latitude axis representing co-occurring species (Pérez-García et al., 2016) within a homogeneous climate extent (Terrigeol et al., 2022). Anthropogenic climate change can alter interspecific interactions and produce unexpected changes in species distributions, community structure, and diversity (Harley, 2011). By comparing the northern and the southern networks, we predict a rise in the percentage of the negative co-occurrence liaisons toward the south. Indeed, the percentage of negative co-occurrences increased from 13.2% to 25.2% from the northern to the southern network. A clear shift in the temperature was noticed with a difference that reaches more than 6°C between the northern and the southern networks (see Fig. 3), which could be a fundamental reason for this change (Milazzo et al., 2013). Accordingly, the rise in the temperature could increase the competition and displacement could also lead to the switch to a competitive displacement mode (Milazzo et al., 2013) or to an increase in the predation intensity (Harley, 2011). A study by DeGregorio et al. (2015) demonstrated that climate change is expected to increase predation of bird nests by ratsnakes by 7%.

The composition of the indicator species along the latitudinal gradient

The latitudinal gradient also significantly impacted the IS composition. We observed an almost complete turnover of the IS between the northern and southern networks. This observed change in the composition of the IS is in line with other methods based on species richness (Terrigeol et al., 2022). Consequently, short-term climate change could lead to different IS for species assemblage. In fact, only *Ruby-crowned Kinglet* (RCKI) and *White-throated Sparrow* (WTSP) remained IS in both climate clusters among the 17 IS. RCKI and WTSP species are associated with early-to-mid succession forests (*EMSF*).

Conclusion

IS could be informative to recover assemblage occurrence, but only for highly present species. The low success, especially regarding the recovery of rare species, illustrates the need for more theoretical investigation of the modelling framework. Subsequently, through our analysis, we demonstrated that the latitudinal gradient plays an essential role in changing biotic interactions and the composition of the IS. Therefore, a latitudinal

network could describe the interactions between the layers; for example, the dispersal and/or changes in the composition of IS along the latitudinal axis through multiplex complex network analysis (Lurgi et al., 2020) could be informative to strengthen the model through the inclusion the interactions between networks.

Acknowledgements

This work was supported by the Sentinel North programme of Laval University, funded by the Canada First Research Excellence Fund. AA, DF and PD were also supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). We acknowledge Calcul Québec and Compute Canada for their technical support and computing infrastructures. We thank also the Québec Breeding Bird Atlas for supplying data. We would also like to thank the following partners: Regroupement QuébecOiseaux, Environment and Climate Change Canada and Birds Canada, together with all the volunteer participants who gathered data for the project. We are grateful to Louis-Paul Rivest for his valuable statistical suggestions and advice.

Supporting information

S1 Habitat types of bird species.

S1 Table Network-based indicator species.

S1 File The associated habitat of the potential bird species.

S2 File The occurrence data for the northern and southern climate clusters.

References

Allouche, O., Tsoar, A., and Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6):1223–1232.

Araújo, M. B., Rozenfeld, A., Rahbek, C., and Marquet, P. A. (2011). Using species co-occurrence networks to assess the impacts of climate change. *Ecography*, 34(6):897–908.

Araújo, M. B., Pearson, R. G., Thuiller, W., and Erhard, M. (2005). Validation of species–climate impact models under climate change. *Global Change Biology*, 11(9):1504–1513.

Atlas des oiseaux nicheurs du Québec (2018). Données obtenues en réponse à une demande présentée aux bureaux de l'Atlas (www.atlas-oiseaux.qc.ca). Regroupement QuébecOiseaux, Service canadien de la faune d'Environnement Canada et Études d'Oiseaux Canada. Québec, QC, Canada.

Azeria, E. T., Fortin, D., Hébert, C., Peres-Neto, P., Pothier, D., and Ruel, J.-C. (2009). Using null model analysis of species co-occurrences to deconstruct biodiversity patterns and select indicator species. *Diversity and Distributions*, 15(6):958–971.

Bibby, C. J., Burgess, N. D., Hill, D. A., Hillis, D. M., and Mustoe, S. (2000). *Bird census techniques*. Elsevier.

Blanchet, F. G., Cazelles, K., and Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, 23(7):1050–1063.

Blois, J. L., Zarnetske, P. L., Fitzpatrick, M. C., and Finnegan, S. (2013). Climate change and the past, present, and future of biotic interactions. *Science*, 341(6145):499–504.

Bouderbala, I., Labadie, G., Béland, J.-M., Boulanger, Y., Hébert, C., Desrosiers, P., Allard, A., and Fortin, D. (2023a). Effects of global change on bird and beetle populations in boreal forest landscape: An assemblage dissimilarity analysis. *Diversity and Distributions*, 29(6):757–773.

Bouderbala, I., Labadie, G., Béland, J.-M., Tremblay, J. A., Boulanger, Y., Hébert, C., Desrosiers, P., Allard, A., and Fortin, D. (2023b). Long-term effect of forest harvesting on boreal species assemblages under climate change. *PLOS Climate*, 2(3):1–20.

Boulanger, Y., Parisien, M.-A., and Wang, X. (2018). Model-specification uncertainty in future area burned by wildfires in Canada. *International Journal of Wildland Fire*, 27(3):164–175.

Boulanger, Y. and Puigdevall, J. P. (2021). Boreal forests will be more severely affected by projected anthropogenic climate forcing than mixedwood and northern hardwood forests in eastern Canada. *Landscape Ecology*, 36(6):1725–1740.

Boulanger, Y., Taylor, A. R., Price, D. T., Cyr, D., McGarrigle, E., Rammer, W., Sainte-Marie, G., Beaudoin, A., Guindon, L., and Mansuy, N. (2017). Climate change impacts on forest landscapes along the Canadian southern boreal forest transition zone. *Landscape Ecology*, 32(7):1415–1431.

Cadieux, P., Boulanger, Y., Cyr, D., Taylor, A. R., Price, D. T., Sólymos, P., Stralberg, D., Chen, H. Y., Brecka, A., and Tremblay, J. A. (2020). Projected effects of climate change on boreal bird community accentuated by anthropogenic disturbances in western boreal forest, Canada. *Diversity and Distributions*, 26(6):668–682.

Chu, T.-J., Shih, C.-H., Lu, Y.-M., Shih, Y.-J., Wang, J.-Q., and Huang, L.-M. (2021). Incorporating species-conditional co-occurrence when selecting indicator species to monitor restoration after mangrove removal from the siangshan wetland, taiwan. *Journal of Marine Science and Engineering*, 9(10).

Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems:1695.

DeGregorio, B. A., Westervelt, J. D., Weatherhead, P. J., and Sperry, J. H. (2015). Indirect effect of climate change: Shifts in ratsnake behavior alter intensity and timing of avian nest predation. *Ecological Modelling*, 312:239–246.

Dommain, R., Andama, M., McDonough, M. M., Prado, N. A., Goldhammer, T., Potts, R., Maldonado, J. E., Nkurunungi, J. B., and Campana, M. G. (2020). The challenges of reconstructing tropical biodiversity with sedimentary ancient dna: A 2200-year-long metagenomic record from bwindi impenetrable forest, uganda. *Frontiers in Ecology and Evolution*, 8.

Dufrêne, M. and Legendre, P. (1997). Species assemblages and indicator species:the need for a flexible asymmetrical approach. *Ecological Monographs*, 67(3):345–366.

Duveneck, M. J., Scheller, R. M., White, M. A., Handler, S. D., and Ravenscroft, C. (2014). Climate change effects on northern Great Lake (USA) forests: A case for preserving diversity. *Ecosphere*, 5(2):art23.

Elmendorf, S. C., Henry, G. H. R., Hollister, R. D., Fosaa, A. M., Gould, W. A., Hermanutz, L., Hofgaard, A., Jónsdóttir, I. S., Jorgenson, J. C., Lévesque, E., Magnusson, B., Molau, U., Myers-Smith, I. H., Oberbauer, S. F., Rixen, C., Tweedie, C. E., and Walker, M. D. (2015). Experiment, monitoring, and gradient methods used to infer climate change effects on plant communities yield consistent patterns. *Proceedings of the National Academy of Sciences*, 112(2):448–452.

Fleming, J., Sutherland, C., Sterrett, S. C., and Campbell Grant, E. H. (2020). A latent process model approach to improve the utility of indicator species. *Oikos*, 129(12):1753–1762.

Freeman, E. A. and Moisen, G. (2008). PresenceAbsence: An r package for presence absence analysis. *Journal of Statistical Software*, 23(11):1–31. eafreeman@fs.fed.us.

Graham, R. W. and Grimm, E. C. (1990). Effects of global climate change on the patterns of terrestrial biological communities. *Trends in Ecology and Evolution*, 5(9):289–292.

Griffith, D. M., Veech, J. A., and Marsh, C. J. (2016). Cooccur: probabilistic species co-occurrence analysis in r. *Journal of Statistical Software*, 69:1–17.

Harley, C. D. G. (2011). Climate change, keystone predation, and biodiversity loss. *Science*, 334(6059):1124–1127.

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *The American Naturalist*, 163(2):192–211. PMID: 14970922.

Hillebrand, H., Soininen, J., and Snoijs, P. (2010). Warming leads to higher species turnover in a coastal ecosystem. *Global Change Biology*, 16(4):1181–1193.

Husson, F., Josse, J., Le, S., and Mazet, J. (2013). Factominer: multivariate exploratory data analysis and data mining with r. *R package version*, 1(1.29).

Kelly, A. E. and Goulden, M. L. (2008). Rapid shifts in plant distribution with recent climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33):11823–11826.

Kim, K. C. and Byrne, L. B. (2006). Biodiversity loss and the taxonomic bottleneck: emerging biodiversity science. *Ecological Research*, 21(6):794.

Labadie, G., Hardy, C., Boulanger, Y., Vanlandeghem, V., Hebblewhite, M., and Fortin, D. (2023). Global change risks a threatened species due to alteration of predator–prey dynamics. *Ecosphere*, 14(3):e4485.

Labadie, G., McLoughlin, P. D., Hebblewhite, M., and Fortin, D. (2021). Insect-mediated apparent competition between mammals in a boreal food web. *Proceedings of the National Academy of Sciences of the United States of America*, 118(30).

Lurgi, M., Galiana, N., Broitman, B. R., Kéfi, S., Wieters, E. A., and Navarrete, S. A. (2020). Geographical variation of multiplex ecological networks in marine intertidal communities. *Ecology*, 101(11):e03165.

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., et al. (2018). Global Warming of 1.5 °C. An IPCC Special Report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Technical report, IPCC.

Micheletti, T., Stewart, F. E. C., Cumming, S. G., Haché, S., Stralberg, D., Tremblay, J. A., Barros, C., Eddy, I. M. S., Chubaty, A. M., Leblond, M., Pankratz, R. F., Mahon, C. L., Van Wilgenburg, S. L., Bayne, E. M., Schmiegelow, F., and McIntire, E. J. B. (2021). Assessing pathways of climate change effects in SpaDES: An application to bBoreal Landbirds of Northwest Territories Canada. *Frontiers in Ecology and Evolution*, 9(October).

Milazzo, M., Mirto, S., Domenici, P., and Gristina, M. (2013). Climate change exacerbates interspecific interactions in sympatric coastal fishes. *Journal of Animal Ecology*, 82(2):468–477.

Mimet, A., Pellissier, V., Houet, T., Julliard, R., and Simon, L. (2016). A holistic landscape description reveals that landscape configuration changes more over time than composition: Implications for landscape ecology studies. *PLOS ONE*, 11(3):1–16.

Morelli, F. (2015). Indicator species for avian biodiversity hotspots: Combination of specialists and generalists is necessary in less natural environments. *Journal for Nature Conservation*, 27:54–62.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P., et al. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC, Gland, Switzerland.

Parmesan, C. and Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421(6918):37–42.

Pianka, E. R. (1966). Latitudinal gradients in species diversity: A review of concepts. *The American Naturalist*, 100(910):33–46.

Pérez-García, J. M., Sebastián-González, E., Botella, F., and Sánchez-Zapata, J. A. (2016). Selecting indicator species of infrastructure impacts using network analysis and biological traits: Bird electrocution and power lines. *Ecological Indicators*, 60:428–433.

Rechkemmer, A. and von Falkenhayn, L. (2009). The human dimensions of global environmental change: Ecosystem services, resilience, and governance.

Régnière, J., Saint-Amant, R., Béchard, A., and Moutaoufik, A. (2017). BioSIM 11: User's manual. Technical report, Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Québec, QC Canada.

Robichaud, C. B. and Knopff, K. H. (2015). Biodiversity offsets and caribou conservation in alberta: opportunities and challenges. *Rangifer*, 35(2):99–122.

Rohde, K. (1992). Latitudinal gradients in species diversity: The search for the primary cause. *Oikos*, 65(3):514–527.

Santillán, V., Quitián, M., Tinoco, B. A., Zárate, E., Schleuning, M., Böhning-Gaese, K., and Neuschulz, E. L. (2018). Spatio-temporal variation in bird assemblages is associated with fluctuations in temperature and precipitation along a tropical elevational gradient. *PLOS ONE*, 13(5):1–15.

Siddig, A. A., Ellison, A. M., Ochs, A., Villar-Leeman, C., and Lau, M. K. (2016). How do ecologists select and use indicator species to monitor ecological change? insights from 14 years of publication in ecological indicators. *Ecological Indicators*, 60:223–230.

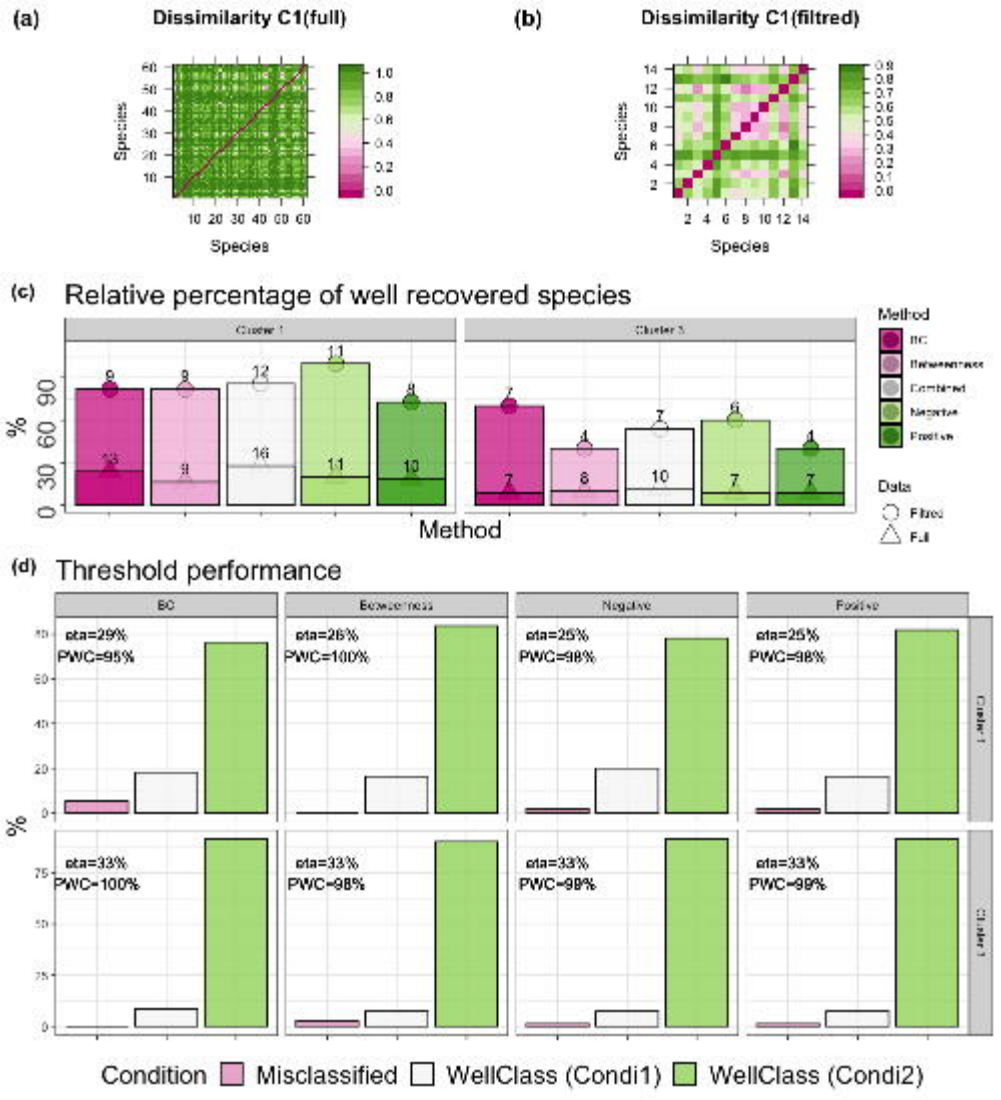
Stralberg, D., Bayne, E. M., Cumming, S. G., Sólymos, P., Song, S. J., and Schmiegelow, F. K. (2015). Conservation of future boreal forest bird communities considering lags in vegetation response to climate change: A modified refugia approach. *Diversity and Distributions*, 21(9):1112–1128.

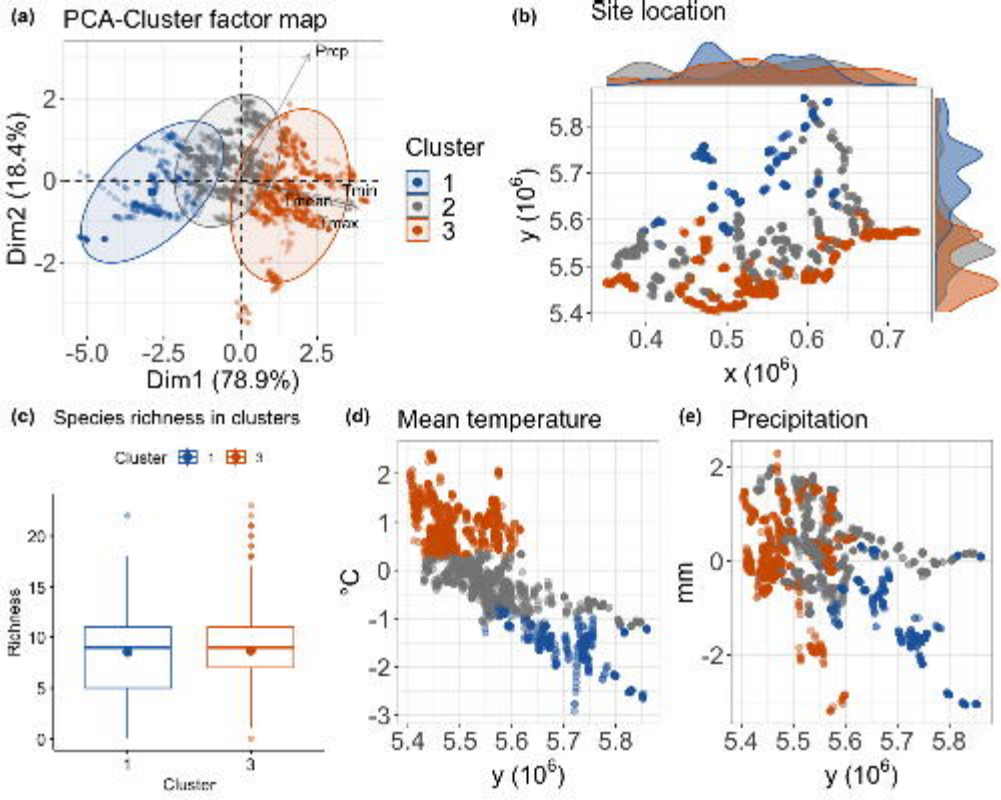
Terrigeol, A., Ewane Ebouele, S., Darveau, M., Hébert, C., Rivest, L.-P., and Fortin, D. (2022). On the efficiency of indicator species for broad-scale monitoring of bird diversity across climate conditions. *Ecological Indicators*, 137:108773.

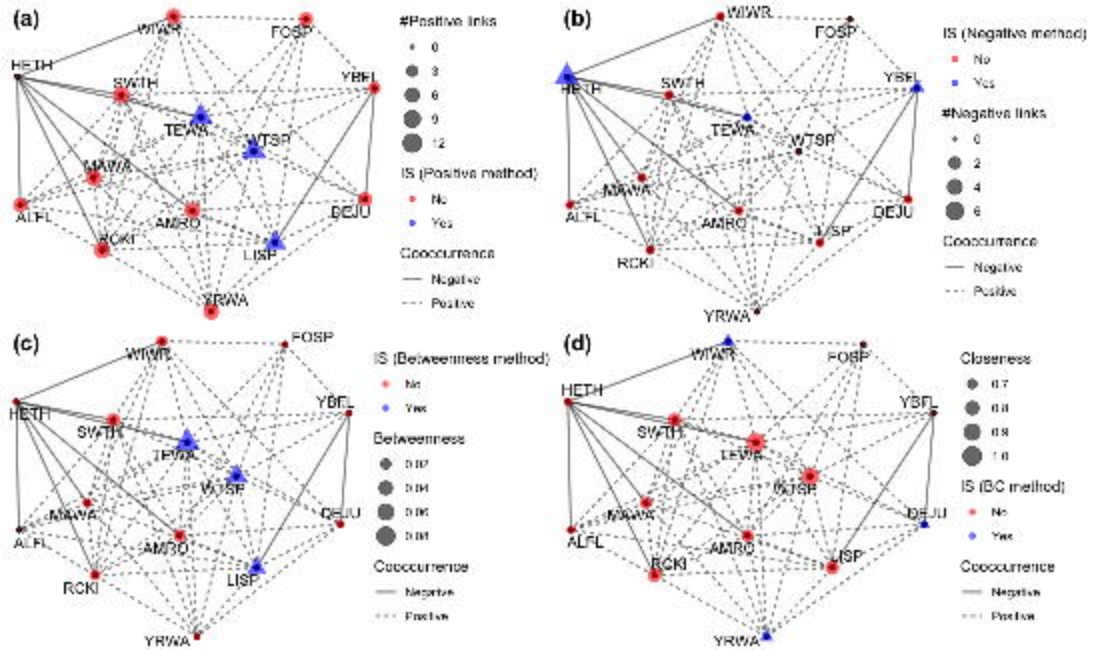
Thuiller, W., Lavergne, S., Roquet, C., Boulangeat, I., Lafourcade, B., and Araujo, M. B. (2011). Consequences of climate change on the tree of life in Europe. *Nature*, 470(7335):531–534.

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

Zhang, Y. and Liang, S. (2014). Changes in forest biomass and linkage to climate and forest disturbances over northeastern china. *Global Change Biology*, 20(8):2596–2606.


(a)
Dissimilarity C1(full)
Species
Species
(b)
Dissimilarity C1(filtred)
Species
Species
(c)
Relative percentage of well recovered species
Cluster 1
Cluster 3
%
Method
Method
BC
Betweenness
Combined
Negative
Positive
Data
Filtred
Full
(d)
Threshold performance
BC
Betweenness
Negative
Positive
Cluster 1
Cluster 3
%
eta=29%
PWC=95%
eta=26%
PWC=100%
eta=25%
PWC=98%
eta=25%
PWC=98%
eta=33%
PWC=100%
eta=33%
PWC=98%
eta=33%
PWC=98%
eta=33%
PWC=99%
Condition
Misclassified
WellClass (Condi1)
WellClass (Condi2)

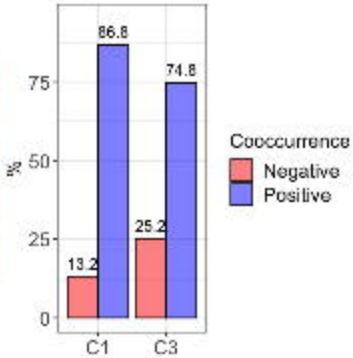

(a)
PCA-Cluster factor map
Prcp
Tmin
Tmean
Tmax
Cluster
1
2
3
Dim2 (18.4%)
Dim1 (78.9%)
(b)
Site location
y (10^6)
x (10^6)
(c)
Species richness in clusters
Cluster
Richness
(d)
Mean temperature
°C
y (10^6)
(e)
Precipitation
mm
y (10^6)


(a)
WIWR
FOSP
HETH
SWTH
YBFL
TEWA
WTSP
MAWA
ALFL
AMRO
DEJU
RCKI
LISP
YRWA
#Positive links
0
3
6
9
12
IS (Positive method)
No
Yes
Cooccurrence
Negative
Positive
(b)
WIWR
FOSP
HETH
SWTH
YBFL
TEWA
WTSP
MAWA
ALFL
AMRO
DEJU
LISP
RCKI
YRWA
IS (Negative method)
No
Yes
#Negative links
0
2
4
6
Cooccurrence
Negative
Positive
(c)
WIWR
FOSP
HETH
SWTH
YBFL
TEWA
WTSP
MAWA
DEJU
ALFL
AMRO
RCKI
LISP
YRWA
IS (Betweenness method)
No
Yes
Betweenness
0.07
0.04
0.06
0.08
Cooccurrence
Negative
Positive
(d)
WIWR
FOSP
HETH
YBFL
SWTH
TEWA
WTSP
MAWA
DEJU
ALFL
AMRO
RCKI
LISP
YRWA
Closeness
0.7
0.8
0.9
1.0
IS (BC method)
No
Yes
Cooccurrence
Negative
Positive

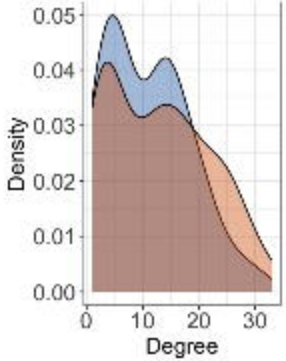
(a)

	Cluster 1	Cluster 3
Number of species	61.0	88.0
Number of sites	406.0	979.0
Number of positive links	277.0	409.0
Number of negative links	42.0	138.0
Number of random links	1511.0	3281.0
Unclassifiable links	0.0	0.0
Non-random links (%)	17.4	14.3

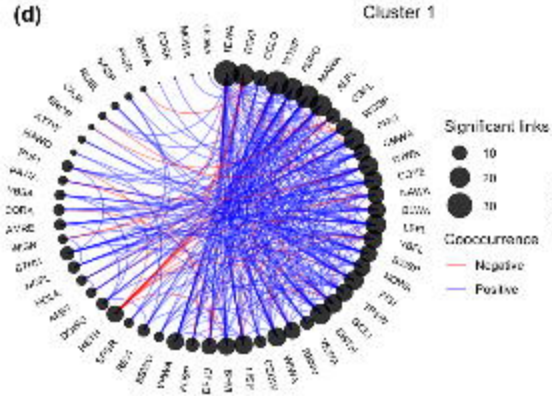
(b) Significant links



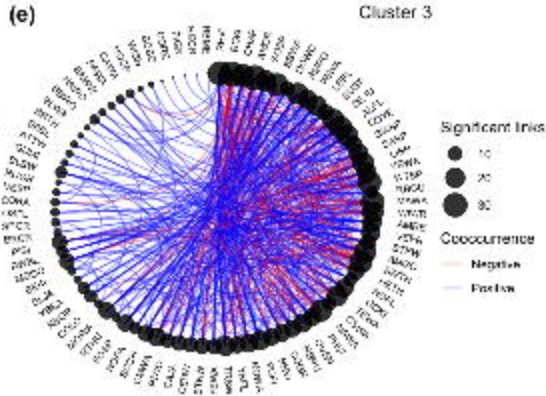
(c) Degree distribution

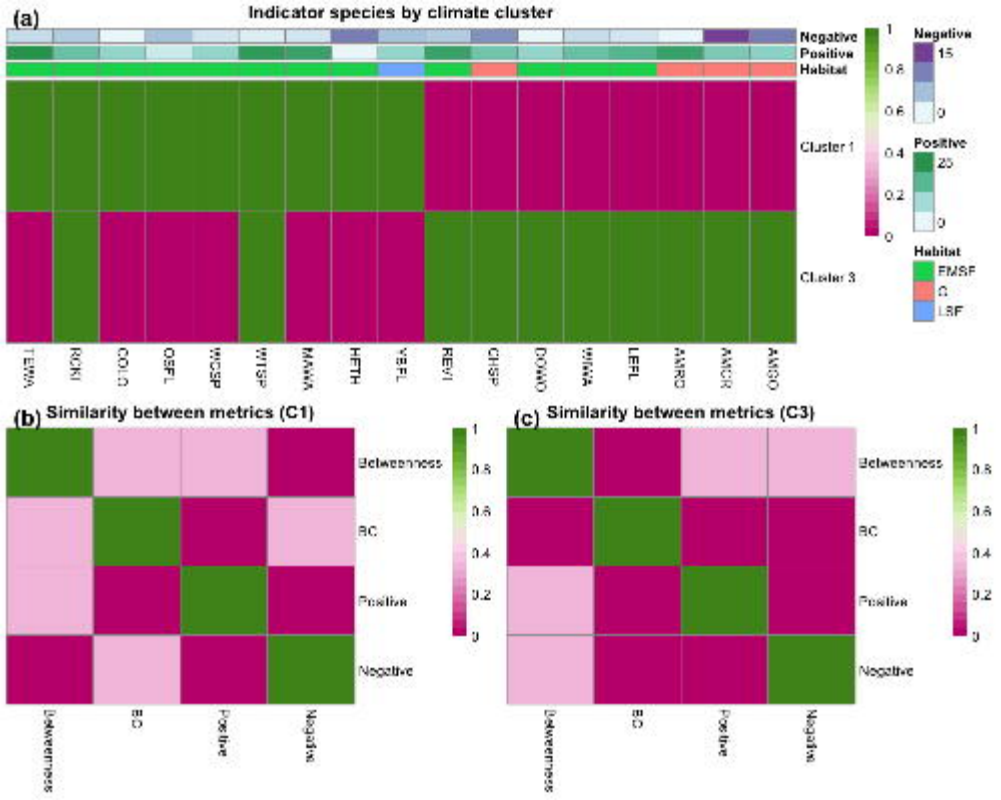


(d) Cluster 1



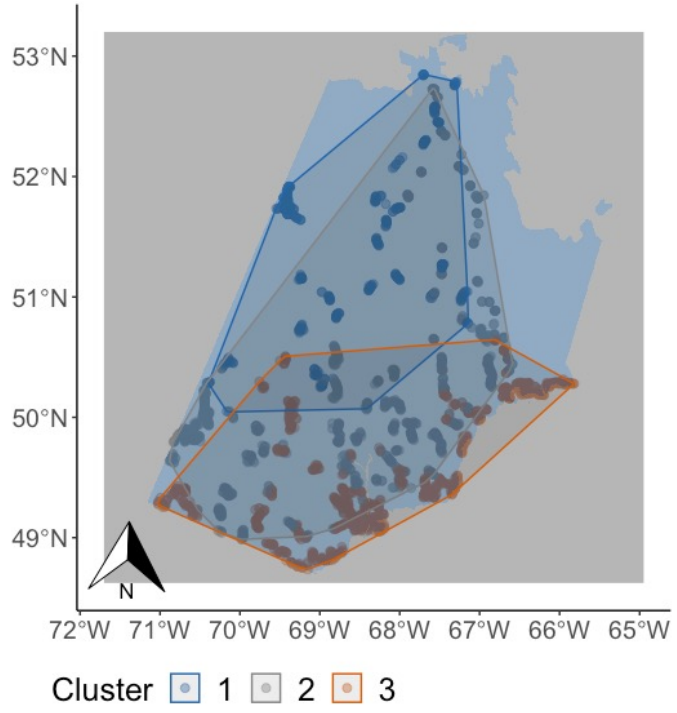
(e) Cluster 3




(a)
Indicator species by climate cluster
Negative
Positive
Habitat
Cluster 1
Cluster 3
Negative
15
0
Positive
25
0
Habitat
FMSF
C
LSF
1
0.8
0.6
0.4
0.2
0
TEMA
RCKI
COLO
OSFL
WCSP
WTSP
MAWA
HETH
YBFL
REVI
CHSP
DOWO
WIWA
LEFL
AMRO
AMGR
AMGO
(b)
Similarity between metrics (C1)
Betweenness
BC
Positive
Negative
(c)
Similarity between metrics (C3)
Betweenness
BC
Positive
Negative

Step 1: Identification of climate spatial groups

Identification of spatial groups k=1,…, K based on climate variables



Step 2: Co-occurrence networks

Identification of significant links

(1) positive (→coexistence)

(2) negative (→ degree of displacement)

Step 3: Indicator species identification:

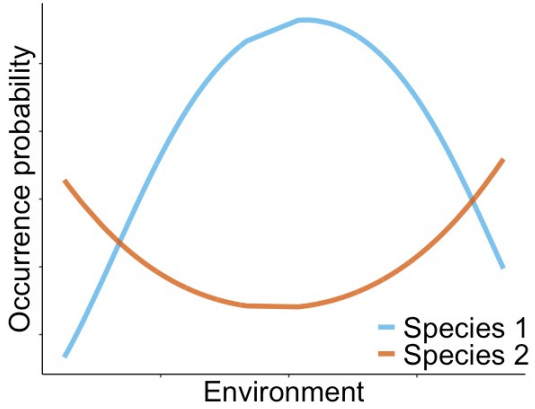
Based on Network-based metrics:

(1) Betweenness, (2) Positive, (3) Negative, (4) Betweenness-Closeness

Step 4: SDM calibration for PIS

(1) Model calibration for PIS based on climate variables and IS occurrence.

(2) IS response to predictor variables.



Step 5: Assemblage occurrence prediction

(1) Prediction of the occurrence probability as function of cmiate variables and IS.

(2) Binarize the outputs.

Step 6: Accuracy analysis

(1) Original-binarized matrix collection M_j

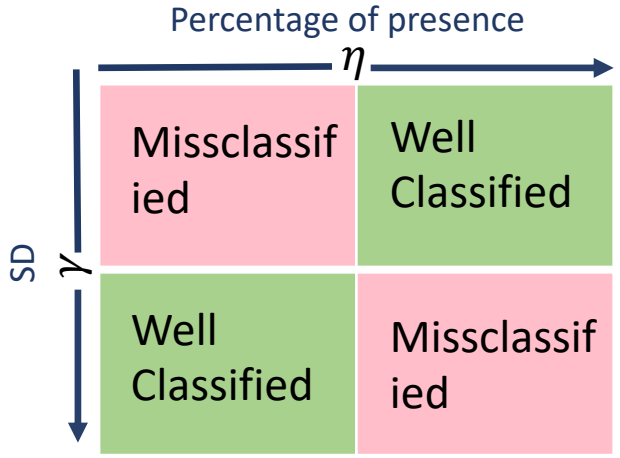
(2) Sorensen dissimilarity index (SD_j)

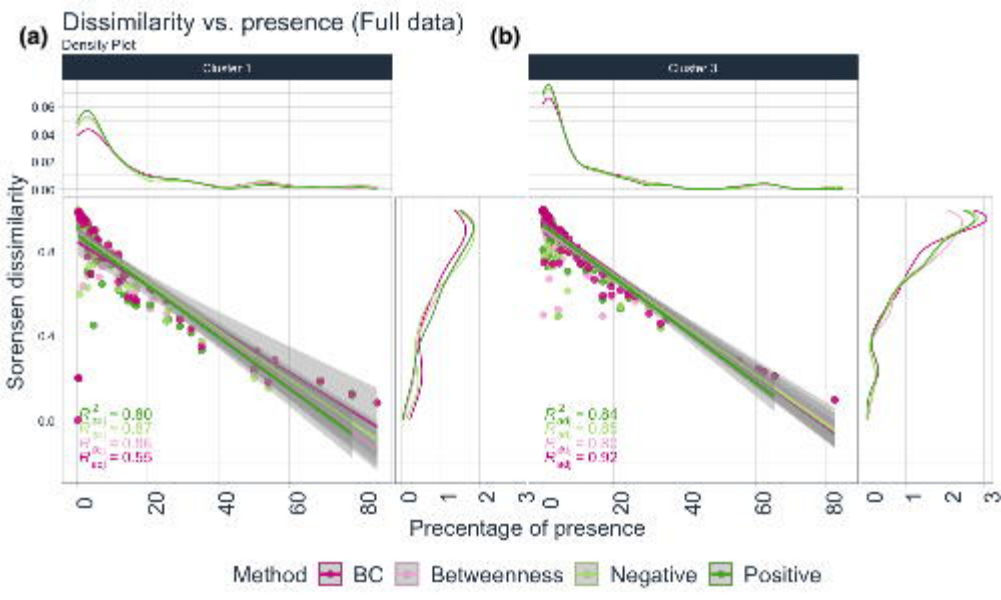
M_j	Sites			
	$S_{1,j}$	$S_{2,j}$	…	$S_{Nk,j}$
Original	1	1	0	0
Binarized	1	0	0	1

Correct prediction

incorrect prediction

(3) Model's accuracy threshold




(a)
Dissimilarity vs. presence (Full data)
Density Plot
Cluster 1
(b)
Cluster 3
Sorensen dissimilarity
Precentage of presence
Method
BC
Betweenness
Negative
Positive