

**Université de Montréal**

**Évaluation de l'unicité écologique sur des échelles spatiales  
étendues et continues**

par

**Gabriel Dansereau**

Département de sciences biologiques

Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)

en Sciences biologiques

26 avril 2021



# Université de Montréal

Faculté des arts et des sciences

---

Ce mémoire intitulé

## **Évaluation de l'unicité écologique sur des échelles spatiales étendues et continues**

présenté par

**Gabriel Dansereau**

a été évalué par un jury composé des personnes suivantes :

*Anne-Lise Routhier*

---

(président-rapporteur)

*Timothée Poisot*

---

(directeur de recherche)

*Pierre Legendre*

---

(codirecteur)

*Élise Filotas*

---

(membre du jury)



## Résumé

---

...sommaire et mots clés en français...



## **Abstract**

---

...summary and keywords in english...





# Table des matières

---

<b>Résumé .....</b>	<b>5</b>
<b>Abstract .....</b>	<b>7</b>
<b>Liste des tableaux .....</b>	<b>11</b>
<b>Table des figures.....</b>	<b>13</b>
<b>Liste des sigles et des abréviations .....</b>	<b>15</b>
<b>Remerciements.....</b>	<b>17</b>
<b>Introduction.....</b>	<b>19</b>
Avant-propos .....	19
Mise en contexte .....	20
Biodiversité .....	21
Développement de la diversité bêta .....	22
Modèles.....	25
Données.....	26
Enjeux spatiaux.....	27
<b>First Article. Evaluating ecological uniqueness on broad continuous scales using species                     distribution modelling .....</b>	<b>29</b>
1. Introduction .....	30

<b>Bibliographie.....</b>	<b>35</b>
---------------------------	-----------

## Liste des tableaux

---



## Table des figures

---



## Liste des sigles et des abréviations

---

KQ-Methode	Méthode des moindres carrés, de l'allemand <i>Methode der kleinsten Quadrate</i>
MCMC	Monte Carlo par chaînes de Markov, de l'anglais <i>Markov Chain Monte Carlo</i>
MSE	Erreur quadratique moyenne, de l'anglais <i>Mean Square Error</i>
NDR	Retract d'un voisinage, de l'anglais <i>Neighbourhood Deformation Retract</i>
OLS	Moindres carrés ordinaires, de l'anglais <i>Ordinary Least Square</i>
ZFC	Théorie des ensembles de Zermelo-Fraenkel avec l'axiome du choix





## Remerciements

---

...remerciements...



# Introduction

---

## Avant-propos

Ce mémoire est parti d'une idée assez simple. Celle de combiner les grandes bases de données en ligne sur biodiversité et les méthodes d'intelligence artificielle pour prédire la distribution géographique des sites de biodiversité exceptionnelle.

Il existe de nombreuses façons de mesurer la biodiversité et d'identifier les endroits où celle-ci est la plus exceptionnelle. Ces sites sont notamment importants à des fins de conservation et de gestion des aires protégées. Par contre, plusieurs de ces mesures ont été développées pour un petit nombre de sites ou d'espèces et nécessitent de connaître la composition en espèces exacte à un endroit donné. Or, la composition exacte est difficile, voire impossible, à connaître pour certains endroits, notamment en région éloignée, ce qui empêche l'utilisation de certaines mesures sur de très grandes échelles spatiales.

Des sources de données nouvelles se sont cependant développées récemment. Il s'agit notamment de bases de données en ligne massives comme GBIF, eBird et iNaturalist. Celles-ci sont imparfaites, mais contiennent tout de même beaucoup d'informations pouvant être mises à profit, notamment pour une évaluation de la biodiversité à grande étendue spatiale. Ces bases de données, largement basées sur des contributions citoyennes, sont cependant le reflet de biais déjà présents dans les données écologiques, notamment en faveur des villes, et ne règlent pas le manque de données en région éloignées.

D'un autre côté, nous assistons également au développement de méthodes prédictives en intelligence artificielle de plus en plus poussées et de plus en plus performantes. Celles-ci offrent

la possibilité d'obtenir des prédictions assez justes pour la distribution des espèces, ainsi que des inférences compensant entre autres pour le manque de données.

Par contre, regrouper ces différents éléments, les bases de données massives et les méthodes d'intelligence artificielle, pour la prédiction de la biodiversité soulève des questions d'applicabilité, puisque les mesures de biodiversité et le cadre théorique derrière celles-ci n'ont justement pas été développées pour une échelle aussi étendue et un si grand volume de données.

Mon projet est donc à la jonction de ces éléments. J'ai exploré les bases de données et les méthodes prédictives, puis j'ai cherché à vérifier comment elles peuvent s'intégrer avec les mesures de biodiversité actuelles, en particulier pour comprendre ce qu'elles font ressortir lorsque utilisées de cette façon. Je souhaite que ces travaux amènent une nouvelle vue des possibilités actuellement ouvertes, ainsi qu'une nouvelle compréhension des concepts soulevés. J'espère que ceux-ci pourront être utiles pour la conservation et la gestion des aires protégées.

## **Mise en contexte**

L'identification des zones clés de biodiversité est l'une des priorités pour la conservation et la gestion des aires protégées. En particulier, il y a actuellement un besoin pour développer des méthodes permettant d'identifier les sites les plus importants pour la biodiversité de façon efficace sur de grandes étendues spatiales. Or, identifier de tels endroits implique plusieurs questions complexes. En premier, il est nécessaire de définir ce que constituent des zones clés de biodiversité. Plusieurs définitions et plusieurs mesures ont été suggérées à ce sujet, mais elles varient généralement quand à l'étendue spatiale ou aux régions ciblées. Ensuite, au-delà de la définition de la biodiversité, il est nécessaire de trouver des données qui permettent d'évaluer avec justesse le caractère unique ou exceptionnel de la biodiversité à des sites donnés. La récolte de données en écologie est parfois difficile à réaliser à certains endroits, notamment en région éloignée. Nos connaissances des différents milieux ne sont pas égales non plus, alors que certains endroits plus proches des villes ou d'intérêt écologique particulier sont beaucoup mieux connus. Lorsque nécessaire, les observations directes peuvent parfois être remplacées par des prédictions réalisées à partir de données plus générales. Par contre, une panoplie de méthodes prédictives existent et la

plupart d'entre elles n'ont pas été évaluées spécifiquement avec certaines mesures de biodiversité. Finalement, il est également nécessaire d'adapter à la fois les mesures de biodiversité et les méthodes prédictives aux grandes étendues spatiales. La biodiversité varie parfois différemment en fonction des échelles. Il en est de même quand à la performance des mesures. Intégrer le tout peut donc s'avérer complexe et implique d'avoir une compréhension développée des définitions de la biodiversité, des données et des méthodes disponibles, ainsi que des facteurs pouvant influencer la biodiversité en fonction des échelles spatiales.

Dans mon mémoire, je me suis intéressé à cette question en cherchant à vérifier l'applicabilité d'une mesure donnée, celles des contributions locales à la diversité bêta, pour identifier les zones de biodiversité exceptionnelle à grande étendue spatiale. De plus, j'ai cherché à vérifier si cette méthode pouvait être appliquée à des prédictions de distribution d'espèces produites à partir de données provenant de grandes bases de données citoyennes. Mon mémoire est donc divisé en trois sections. La première comporte une mise en contexte, ainsi qu'une revue de littérature présentant les concepts pertinents. La seconde partie consiste en un article scientifique présentant les résultats de mes travaux et analyses. La dernière partie consiste en un retour sur les résultats, en lien avec la mise en contexte présentée dans la première section.

## **Biodiversité**

La biodiversité peut difficilement être séparée de sa dimension spatiale vu la diversité des espèces et la différence entre les espèces présentes en un endroit donné sur Terre. La diversité bêta, soit la variation dans la composition en espèces entre les sites d'une région géographique d'intérêt (Legendre et al., 2005), est donc une mesure essentielle de l'organisation de la biodiversité dans l'espace. En écologie des communautés, l'intérêt pour celle-ci est d'autant plus grand que la variation spatiale dans la composition en espèces permet de tester des hypothèses portant sur les processus qui génèrent et maintiennent la biodiversité dans les écosystèmes (Legendre et De Cáceres, 2013).

Dans le cadre du présent mémoire, trois phases importantes sont à retenir du développement du concept de diversité bêta, soit une première phase portant sur la définition de la diversité bêta

même, une deuxième sur son partitionnement et une troisième sur son utilisation comme mesure spatialement explicite pour évaluer l'unicité écologique de sites spécifiques. En effet, depuis les premières formulations des composantes de la diversité des espèces par Whittaker (1960), l'attention s'est progressivement tournée vers le partitionnement de ces composantes, menant entre autres à la formulation d'une mesure spatialement explicite par Legendre et De Cáceres (2013), puis à l'utilisation de celle-ci pour évaluer l'unicité écologique, notamment pour un très grand nombre de sites (Niskanen et al., 2017) ou même sur des distributions d'espèces prédites (Vasconcelos et al., 2018). Dans cette section, j'effectuerai donc une revue du développement de ces trois phases en lien avec l'évaluation de l'unicité écologique à grande échelle spatiale, ce qui constitue l'objectif de mon mémoire.

## Développement de la diversité bêta

Whittaker (1960) a détaillé trois composantes de la diversité des espèces au sein des communautés écologiques : 1) la diversité alpha, soit la richesse en espèce d'un site ou d'une communauté donnée, 2) la diversité beta, qui représente le degré de différenciation dans la composition des communautés au sein d'un environnement (ou d'un gradient), et 3) la diversité gamma, soit la richesse en espèces des communautés d'un environnement une fois regroupées (ou d'un ensemble de communautés), de sorte que la diversité gamma est le résultat (ou la conséquence) à la fois des diversités alpha et beta. Sous cette formulation initiale, la diversité bêta peut être mesurée comme  $\beta = \gamma / \bar{\alpha}$ , soit le ratio entre la diversité gamma et la diversité alpha moyenne des sites d'un échantillon, autrement dit le ratio entre le nombre d'espèces totales et le nombre d'espèces moyen (Whittaker, 1960, 1972). Elle peut également être mesurée à partir de mesures de similarités d'échantillons, comme le coefficient de communauté, le pourcentage de similarité ou une distance (Whittaker, 1972). Nous obtenons ainsi une seule valeur pour l'ensemble des sites visés, plutôt qu'une mesure pour chacun d'entre eux. Cette mesure est donc utile pour comparer des ensembles de sites, mais pas pour analyser les sites eux-mêmes.

La définition de la diversité bêta a légèrement été reformulée pour en arriver à celle-ci : la diversité bêta est la variation de la composition en espèces entre les sites d'une région géographique

d'intérêt (Legendre et al., 2005). Malgré cette définition, certains auteurs ont cependant distingué deux types de diversité bêta : le renouvellement des espèces, une mesure directionnelle utilisée, par exemple, pour montrer le changement de le long d'un gradient environnemental, et la variation de la composition, une mesure non directionnelle (Anderson et al., 2011 ; Vellend, 2001). L'intérêt s'est également porté sur la formulation d'une valeur seule permettant de quantifier la variation (Anderson et al., 2011 ; Koleff et al., 2003 ; Vellend, 2001 ; Whittaker, 1972) menant entre autres aux mesures additives et multiplicatives les plus populaires (Chao et al., 2012 ; Jost, 2007). Dans la plupart de ces formulations, la diversité bêta est cependant reliée aux diversités alpha et gamma, ce qui mena à Ellison (2010) à appeler à la formulation d'une mesure de diversité bêta indépendante de l'évaluation des deux autres formes de diversité. La diversité bêta mesurée comme la variance de la matrice de communautés utilisée par **Pelissier2003**, Legendre et al. (2005) et **Anderson2006** rempli cette condition. De plus, cette mesure offre l'avantage de pouvoir partitionner la variation pour tester des hypothèses sur l'origine et le maintien de la diversité bêta au sein des écosystèmes (Legendre et De Cáceres, 2013). Celle-ci peut être calculée de deux façons, soit par le calcul de la somme des carrés de la matrice des communautés ou par une matrice de dissimilarité (Legendre et De Cáceres, 2013).

La diversité bêta totale au sein d'une communauté peut, entre autres, être décomposée en contributions locales à la diversité bêta (*local contributions to beta diversity*, LCBD), ce qui permet d'identifier les sites possédant une composition en espèces exceptionnelle, donc une biodiversité unique (Legendre et De Cáceres, 2013). Cette mesure permet donc d'évaluer l'unicité écologique pour des sites précis, donc de façon spatialement explicite. La plupart des études ayant utilisé la mesure des LCBD l'ont utilisé à échelle locale, donc sur des étendues spatiales restreintes, et sur un petit nombre de sites (da Silva et Hernández, 2014 ; Heino et al., 2017 ; Heino et Grönroos, 2017 ; Legendre et De Cáceres, 2013). Par exemple, l'étude ayant présenté la mesure l'a utilisé sur des communautés de poissons échantillonnées à intervalles le long d'une rivière (Legendre et De Cáceres, 2013). Quelques études ont utilisé la mesure des LCBD sur de plus grandes étendues spatiales, donc comportant potentiellement une plus forte hétérogénéité spatiale, mais ces études comportaient un nombre de sites assez faible (Poisot et al., 2017 ; Taranu et al., 2020 ; Yang et al.,

2015). Quelques études récentes l'ont utilisée sur des données arrangées en grille, donc spatialement continues (D'Antracoli et al., 2020 ; Legendre et Condit, 2019 ; Tan et al., 2017 ; Tan et al., 2019). Cependant, celles-ci portaient également sur des échelles spatiales restreintes. Une étude récente l'a utilisée sur un grand nombre de sites (Niskanen et al., 2017). Cette dernière a également cherché à prédire la mesure directement en fonction de conditions environnementales. Une avenue intéressante a été explorée par Vasconcelos et al. (2018) en utilisant la mesure sur des prédictions réalisés par des modèles de distribution d'espèces en fonction des conditions climatiques actuelles et de scénarios de changements climatiques.

Un enjeu potentiel pouvant être soulevé pour l'utilisation de la mesure des LCBd est la nécessité est le besoin de données appropriées. Les exemples précédents montrent l'utilité de la mesure des LCBd pour évaluer l'unicité écologique dans différentes situations, y compris sur de grandes étendues spatiales, sur un grand nombre de sites et sur des prédictions de distributions d'espèces. Par contre, ces trois situations n'ont jamais été évaluées en même temps. Une raison potentielle, selon moi, est le manque de données appropriées à une telle situation. Le calcul des LCBd nécessite une matrice de communautés complète, donc que la composition en espèces de chaque site soit connue précisément. Or, il est difficile de connaître la composition précise des communautés en couvrant à la fois une grande étendue spatiale et un grand nombre de sites. Pour compenser, Niskanen et al. (2017) ont développé des modèles pour prédire la valeur des LCBd. Vasconcelos et al. (2018) ont de leur côté utilisé des modèles pour prédire la distribution des espèces, puis ont calculé les LCBd, ce qui s'apparente à l'utilisation originale. Par contre, leur utilisation s'est restreinte à la forêt Atlantique et au Cerrado au Brésil, ainsi qu'à environ 20 000 occurrences disponibles pour les anoues. Or, avec les développements récents des bases de données massives, il existe des espèces et des bases de données pour lesquelles nous possédons beaucoup plus d'occurrences sur des étendues spatiales encore plus grandes, dont il serait intéressant de tirer parti.

Je suggère donc de s'inspirer de leur démarche et de passer à un autre niveau, tout en cherchant à comprendre l'unicité écologique montrée par la mesure des LCBd à grande étendue spatiale. Organisme modèle plus mobile (plus grande étendue spatiale), populaire auprès des ornithologues amateurs, plus facile à repérer.



## Modèles

Le type de modèles utilisés par Vasconcelos et al. (2018) font partie de la grande famille des modèles de distribution d'espèces (*species distribution models*, ci-après SDM) (Guisan et Thuiller, 2005), qui servent notamment à prédire la distribution des espèces en fonction des conditions environnementales et d'observations déjà réalisées. Les assises théoriques derrière ces types de modèles remontent aux premières formulations de la niche écologique (Grinnell, 1917a, 1917b, 1924). Ils reposent également sur l'idée de l'hypervolume de Hutchinson (1957, 1959) selon laquelle les tolérances environnementales d'une espèce forment un hypervolume au sein duquel la présence de l'espèce est possible. L'un des premiers SDM, le modèle d'enveloppe climatique BIOCLIM (Nix, 1986), illustre particulièrement bien cette dépendance. Selon le modèle, la distribution potentielle des espèces devrait être contrainte au sein de l'étendue des conditions bioclimatiques où des observations ont été réalisées (Booth et al., 2014 ; Franklin, 2010). Le modèle classe les sites observés selon leur rang centile pour chaque variable bioclimatique fournie, puis attribue le score le plus élevé à la médiane, considérée comme l'endroit où les conditions conviennent le mieux à l'espèce (Hijmans et al., 2017). La valeur minimale parmi toutes les variables environnementales est ensuite interprétée comme la probabilité d'occurrence de l'espèce au site. Bien qu'il illustre assez simplement la relation entre les SDM et les concepts de niche et d'hypervolume, le modèle BIOCLIM est toutefois peu performant en comparaison avec des modèles plus récents (Elith et al., 2006). MAXENT (Phillips et al., 2017 ; Phillips et al., 2006 ; Phillips et Dudík, 2008), basé sur le principe d'entropie maximale, est l'un des modèles les plus utilisés dans le domaine des SDM (Booth et al., 2014). Plusieurs méthodes d'intelligence artificielle sont également performantes et très utilisées (Elith et al., 2006), notamment les forêts d'arbres décisionnels (*Random Forests*, RF) (Breiman, 2001) et les arbres de régressions fortifiés (*Boosted Regression Trees*, BRT) (Elith et al., 2008). Carlson (2020) ont récemment suggéré d'utiliser les *Bayesian Additive Regression Trees*

(BART)(Chipman et al., 2010) pour les SDM, une alternative prometteuse aux RF et BRT permettant d'obtenir de meilleurs résultats en réduisant le sur-ajustement, tout en permettant d'évaluer l'incertitude sous une formulation bayésienne.

Plusieurs méthodes ont été suggérées afin de réaliser des prédictions au niveau de la communauté à partir de SDM. La méthode la plus simple consiste à réaliser des prédictions séparées pour chaque espèce présente dans la communauté, puis à superposer les prédictions (*stacked SDM*, S-SDM), de façon à connaître la composition en espèces pour chaque site d'une région d'intérêt (Ferrier et al., 2002 ; Ferrier et Guisan, 2006). Des méthodes plus complexes ont également été suggérées, notamment par Guisan et Rahbek (2011), Pollock et al. (2014), Ovaskainen et al. (2017) et Staniczenko et al. (2017). Ces modèles ont l'avantage de prendre en compte plusieurs facteurs supplémentaires affectant la distribution des espèces, comme la co-occurrence entre les espèces, mais ils sont cependant plus complexes à réaliser. Malgré leur simplicité, les S-SDM offrent toutefois des résultats comparables aux autres modèles quant aux prédictions de valeurs concernant les communautés (Norberg et al., 2019 ; Zurell et al., 2020).

## Données

Nos connaissances sur la biodiversité souffrent de nombreuses lacunes. Hortal et al. (2015) a identifié 7 catégories, notamment sur la distribution des espèces (déficit Wallacéen), leur niche abiotique (déficit Grinellien) et leurs interactions biotiques (déficit Eltonien). Plusieurs mégaprojets de collecte et assemblage de données ont cours en ce moment et offriront de grandes opportunités d'avancement, mais ceux-ci devront toutefois s'accompagner d'une évaluation critique des déficits et de l'incertitude (Hortal et al., 2015).

De plus en plus de données sont disponibles en ligne, et ce, massivement. Il faut donc passer vers une approche dirigée par les données disponibles dans une optique de synthèse, ce qui permet généralement de révéler de nouvelles informations écologiques à partir de données existantes (Poisot et al., 2019).

De grandes bases de données sur la biodiversité fournissent des informations écologiques à exploiter, notamment eBird, GBIF et iNaturalist.

En même temps, nous disposons désormais de données de plus en plus précises sur les conditions environnementales partout sur le globe. Par exemple, WorldClim et CHELSA fournissent des données climatiques, alors que Copernicus et EarthEnv fournissent des informations sur l'utilisation du territoire. Dans les deux cas, ces informations sont parfois disponibles à des échelles spatiales très fines.

Plusieurs de ces méthodes SDM mentionnées plus tôt représentent toutefois des méthodes d'apprentissage supervisé, de sorte qu'elles ont besoin d'être entraînées sur des données déjà étiquetées. La principale conséquence au niveau des SDM est donc le besoin de disposer de données d'absence, en plus de données de présence, afin de pouvoir entraîner les algorithmes. Or, les données d'absence sont plus difficiles à obtenir, notamment en raison du problème du double-zéro (Legendre et Legendre, 2012).

La base de données *eBird* comporte toutefois un avantage à ce sujet, puisqu'il s'agit d'une base de données semi-structurée (Johnston et al., 2020). Les données (et donc les observations) y sont structurées par listes d'observations. En rapportant leurs observations, les utilisateurs doivent déclarer si celles-ci constituent une liste complète des espèces détectées lors de leur échantillonnage. Ainsi, cela permet un peu plus justement d'inférer la non-détection d'autres espèces

## Enjeux spatiaux

Plusieurs études ont cherché à comprendre les déterminants d'une forte mesure d'unicité écologique telle qu'évaluée par le mesure des LCBD.

Or, ceux-ci ont peu été étudiés à grande échelle spatiale et sur un très grand nombre de sites.

Selon la formulation initiale de Legendre et De Cáceres (2013), les LCBD devraient normalement identifier les sites les plus uniques, que ce soit en raison de leur nombre d'espèces élevé ou faible, d'une composition particulière dans une région ou en raison de la présence d'espèces rares.

Leur exemple initial a montré une relation négative entre la richesse spécifique et la valeur d'unicité (Legendre et De Cáceres, 2013). Ainsi, les sites les plus pauvres ressortent comme les plus uniques.

Cette relation négative également été observée dans la plupart des études ayant repris la mesure (Heino et al., 2017), mais d'autres études ont montré que la relation pouvait également être positive (Kong et al., 2017).

Certaines études ont avancé que cette relation serait influencée par la présence d'espèces rares dans la communauté (Yao et al., 2019). La relation serait positive lorsque la proportion d'espèces rares est plus élevée.

Par contre, la rareté est complexe à définir, en particulier sur de grandes étendues spatiales, où, par exemple, certaines espèces peuvent être très communes à échelle locale dans une région donnée, mais rares pour l'ensemble de la région, ce qui pourrait influencer la relation richesse-LCBD.

De plus, la diversité bêta totale augmente avec l'étendue spatiale (Barton et al., 2013) et dépend de l'échelle, notamment en raison de l'augmentation de l'hétérogénéité spatiale, ainsi qu'en raison du recoupement de bassins d'espèces locaux différents (Heino et al., 2015).

**First Article.**

# **Evaluating ecological uniqueness on broad continuous scales using species distribution modelling**

by

Gabriel Dansereau<sup>1</sup>, Pierre Legendre<sup>1</sup>, and Timothée Poisot<sup>1</sup>

(<sup>1</sup>) 1375 avenue Thérèse-Lavoie-Roux, Montréal, QC, Canada H2V 0B3  
Université de Montréal

This article was submitted in Global Ecology and Biogeography.

The main contributions of Gabriel Dansereau for this articles are presented.

- Developed and performed the analyses;
- Wrote the first version of the manuscript;

Timothée Poisot developed a preliminary version of the analyses.

Pierre Legendre and Timothée Poisot provided guidance on the analyses and interpretation of the results and revised the manuscript.

All authors read and approved the manuscript.

RÉSUMÉ. Le résumé en français.

**Mots clés :** Mots clés

ABSTRACT. The english abstract.

**Keywords:** Key words

## 1. Introduction

Beta diversity, defined as the variation in species composition among sites in a geographic region of interest (Legendre et al., 2005), is an essential measure to describe the organization of biodiversity through space. Total beta diversity within a community can be partitioned into local contributions to beta diversity (LCBD) (Legendre & De Cáceres, 2013), which allows the identification of sites with exceptional species composition, hence unique biodiversity. Such a method, focusing on specific sites, is useful for both community ecology and conservation biology, as it highlights areas that are most important for their research or conservation values. However, the use of LCBD indices is currently limited in two ways. First, LCBD indices are typically used on data collected over local or regional scales with relatively few sites, for example on fish communities at intervals along a river or stream (Legendre & De Cáceres, 2013). Second, LCBD calculation methods require complete information on community composition, such as a community composition matrix  $Y$ ; thus, they are inappropriate for partially sampled sites (e.g. where data for some species is missing), let alone for unsampled ones. Accordingly, the method is of limited use to identify areas with exceptional biodiversity in regions with sparse sampling. However, predictive approaches are increasingly common given the recent development of computational methods, which often uncover novel ecological insights from existing data (Poisot et al., 2019), including in unsampled or lesser-known locations, as well as larger spatial scales. Here, we examine whether the LCBD method can assess ecological uniqueness over broad and continuous scales based on predictions of species distributions and evaluate whether this reveals novel ecological insights regarding the identification of exceptional biodiversity areas.

Species distribution models (SDMs) (Guisan & Thuiller, 2005) can bring a new perspective to LCBBD studies by filling in gaps and performing analyses on much broader scales. In a community matrix  $Y$ , such as required for LCBBD calculation, ecological communities are abstracted as assemblages of species present at different sites. Viewing communities as such opens the perspective of predicting community composition from predictions of individual species, which is precisely the aim of SDMs. Community-level modelling from SDMs is not an especially novel idea (Ferrier et al., 2002; Ferrier & Guisan, 2006), but it is increasingly relevant with the advent of large-scale, massive, and open data sources on species occurrences, often contributed by citizens, such as eBird and GBIF. At their core, SDMs aim at predicting the distribution of a species based on information about where the species was previously reported, matched with environmental data at those locations, and then make predictions at other (unsampled) locations based on their respective environmental conditions. However, going from single-species SDMs to a whole community is not a trivial task, and many solutions have been suggested, such as stacked species distribution models (S-SDMs) (Ferrier & Guisan, 2006), spatially explicit species assemblage modelling (SESAM) (Guisan & Rahbek, 2011), joint species distribution models (JSDMs) (Pollock et al., 2014), and hierarchical modelling of species communities (HMSC) (Ovaskainen et al., 2017). These alternative methods all have different strengths, but even S-SDM, in a sense the most simple and less community-specific method, has been shown to provide reliable community predictions (Norberg et al., 2019; Zurell et al., 2020). This is important, as in the context of large-scale studies with a high number of sites and species, reducing the model complexity with a simpler yet efficient model such as an S-SDM can reduce the number of computations in an important way. Regardless of the method used, community-level analyses can be applied to the resulting community prediction, but this has been lacking for community measures other than species richness (Ferrier & Guisan, 2006). Notably, the LCBBD framework has, to our knowledge, never been applied to SDM results. The computation of local contributions to beta diversity (LCBD) on SDM predictions, however, raises the issue of calculating the uniqueness scores on much larger community matrices than on the typical scales on which it has been used.

The total number of sites will increase (1) because of the continuous scale of the predictions, as there will be more sites in the region of interest than the number of sampled sites, and (2) because of the larger spatial extent allowed for the SDM predictions. A high number of SDM-predicted sites with a large extent opens up the possibility of capturing a lot of variability of habitats and community composition, but also many very similar ones, which could change the way that exceptional sites contribute to the overall variance in the large-scale community. LCBD scores have typically been used at local or regional scales with relatively few sites (da Silva & Hernández, 2014; Heino et al., 2017; Heino & Grönroos, 2017; Legendre & De Cáceres, 2013). Some studies did use the measure over broader, near-continental extents (Poisot et al., 2017; Taranu et al., 2020; Yang et al., 2015), but the total number of sites in these studies was relatively small. Recent studies also investigated LCBD and beta diversity on sites distributed in grids or as pixels of environmental raster layers, hence continuous scales, but these did not cover large extents and a high number of sites (D’Antraccoli et al., 2020; Legendre & Condit, 2019; Tan et al., 2017; Tan et al., 2019). Niskanen et al. (2017) predicted LCBD values of plant communities (and three other diversity measures) on a continuous scale and a high number of sites (> 25 000) using Boosted Regression Trees (BRTs). However, they modelled the diversity measures directly instead of modelling species distributions first, as we are suggesting here. They obtained lower predictive accuracy for LCBD than for their other diversity measures, mentioning that it highlighted the challenge of predicting LCBD specifically. They also computed LCBD indices at a regional scale, not a continental one, while using a fine spatial resolution (1 km x 1 km). Therefore, the distribution of LCBD values at broad, continuous scales with a high number of sites and predicted species assemblages remains to be investigated.

Measuring ecological uniqueness from LCBD indices on extended continuous scales also raises the question of which sites will be identified as exceptional and for what reason. The method intends that sites should stand out and receive a high LCBD score whenever they display an exceptional community composition, be it a unique assemblage of species that may have a high conservation value or a richer or poorer community than most in the region (Legendre & De Cáceres,



2013). Both the original study and many of the later empirical ones have shown a negative relationship between LCBD scores and species richness (da Silva & Hernández, 2014; Heino et al., 2017; Heino & Grönroos, 2017; Legendre & De Cáceres, 2013), although other studies observed both negative and positive relationships at different sites (Kong et al., 2017) or quadrats (Yao et al., 2021). Therefore, this relationship should still be investigated, especially at broad continuous scales, where LCBD indices have not yet been used. Total beta diversity increases with spatial extent (Barton et al., 2013) and is strongly dependent on scale, notably because of higher environmental heterogeneity and sampling of different local species pool (Heino et al., 2015), which could potentially add some variation to the relationship. Neither the previous studies at broad spatial extents (Poisot et al., 2017; Taranu et al., 2020; Yang et al., 2015), on spatially continuous data (D’Antraccoli et al., 2020; Tan et al., 2019), or on a high number of sites (Niskanen et al., 2017) have specifically measured the variations of the richness-LCBD relationship according to different regions and spatial extents. These studies brought forward relevant elements which now need to be combined.

This study shows that species distribution modelling offers relevant LCBD and community-level predictions on broad spatial scales, similar to those obtained from occurrence data and providing uniqueness assessments in poorly sampled regions. Our results further highlight a changing relationship between site richness and LCBD values depending on (i) the region on which it is used, as species-poor and species-rich regions display different uniqueness profiles; and on (ii) the scale at which it is applied, as increasing the spatial extent can merge the uniqueness profiles of contrasting subregions to create a new, distinct one at a broader scale. Hence, our method could prove useful to identify beta diversity hotspots in unsampled locations on large spatial scales, which could be important targets for conservation purposes.



## Bibliographie

---

- Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., Sanders, N. J., Cornell, H. V., Comita, L. S., Davies, K. F., Harrison, S. P., Kraft, N. J. B., Stegen, J. C. et Swenson, N. G. (2011). Navigating the Multiple Meanings of  $\beta$  Diversity : A Roadmap for the Practicing Ecologist. *Ecology Letters*, 14(1), 19-28. <https://doi.org/10.1111/j.1461-0248.2010.01552.x>
- Barton, P. S., Cunningham, S. A., Manning, A. D., Gibb, H., Lindenmayer, D. B. et Didham, R. K. (2013). The Spatial Scaling of Beta Diversity. *Global Ecology and Biogeography*, 22(6), 639-647. <https://doi.org/10.1111/geb.12031>
- Booth, T. H., Nix, H. A., Busby, J. R. et Hutchinson, M. F. (2014). BIOCLIM : The First Species Distribution Modelling Package, Its Early Applications and Relevance to Most Current MaxEnt Studies. *Diversity and Distributions*, 20(1), 1-9. <https://doi.org/10.1111/ddi.12144>
- Breiman, L. (2001). Random Forests. *Machine learning*, 45(1), 5-32.
- Carlson, C. J. (2020). Embarcadero : Species Distribution Modelling with Bayesian Additive Regression Trees in R. *Methods in Ecology and Evolution*, 11(7), 850-858. <https://doi.org/10.1111/2041-210X.13389>
- Chao, A., Chiu, C.-H. et Hsieh, T. C. (2012). Proposing a Resolution to Debates on Diversity Partitioning. *Ecology*, 93(9), 2037-2051. <https://doi.org/10.1890/11-1817.1>
- Chipman, H. A., George, E. I. et McCulloch, R. E. (2010). BART : Bayesian Additive Regression Trees. *Annals of Applied Statistics*, 4(1), 266-298. <https://doi.org/10.1214/09-AOAS285>

- da Silva, P. G. et Hernández, M. I. M. (2014). Local and Regional Effects on Community Structure of Dung Beetles in a Mainland-Island Scenario. *PLOS ONE*, 9(10), e111883. <https://doi.org/10.1371/journal.pone.0111883>
- D'Antraccoli, M., Bacaro, G., Tordoni, E., Bedini, G. et Peruzzi, L. (2020). More Species, Less Effort : Designing and Comparing Sampling Strategies to Draft Optimised Floristic Inventories. *Perspectives in Plant Ecology, Evolution and Systematics*, 45, 125547. <https://doi.org/10.1016/j.ppees.2020.125547>
- Elith, J., Leathwick, J. R. et Hastie, T. (2008). A Working Guide to Boosted Regression Trees. *Journal of Animal Ecology*, 77(4), 802-813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M. M., Peterson, A. T., ... Et Zimmermann, N. E. (2006). Novel Methods Improve Prediction of Species' Distributions from Occurrence Data. *Ecography*, 29(2), 129-151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>
- Ellison, A. M. (2010). Partitioning Diversity. *Ecology*, 91(7), 1962-1963. <https://doi.org/10.1890/09-1692.1>
- Ferrier, S., Drielsma, M., Manion, G. et Watson, G. (2002). Extended Statistical Approaches to Modelling Spatial Pattern in Biodiversity in Northeast New South Wales. II. Community-Level Modelling. *Biodiversity & Conservation*, 11(12), 2309-2338. <https://doi.org/10.1023/A:1021374009951>
- Ferrier, S. et Guisan, A. (2006). Spatial Modelling of Biodiversity at the Community Level. *Journal of Applied Ecology*, 43(3), 393-404. <https://doi.org/10.1111/j.1365-2664.2006.01149.x>
- Franklin, J. (2010). *Mapping Species Distributions : Spatial Inference and Prediction*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511810602>
- Grinnell, J. (1917a). Field Tests of Theories Concerning Distributional Control. *The American Naturalist*, 51(602), 115-128. <https://doi.org/10.1086/279591>

- Grinnell, J. (1917b). The Niche-Relationships of the California Thrasher. *The Auk*, 34(4), 427-433.  
<https://doi.org/10.2307/4072271>
- Grinnell, J. (1924). Geography and Evolution. *Ecology*, 5(3), 225-229. <https://doi.org/10.2307/1929447>
- Guisan, A. et Rahbek, C. (2011). SESAM – a New Framework Integrating Macroecological and Species Distribution Models for Predicting Spatio-Temporal Patterns of Species Assemblages. *Journal of Biogeography*, 38(8), 1433-1444. <https://doi.org/10.1111/j.1365-2699.2011.02550.x>
- Guisan, A. et Thuiller, W. (2005). Predicting Species Distribution : Offering More than Simple Habitat Models. *Ecology Letters*, 8(9), 993-1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Heino, J., Bini, L. M., Andersson, J., Bergsten, J., Bjelke, U. et Johansson, F. (2017). Unravelling the Correlates of Species Richness and Ecological Uniqueness in a Metacommunity of Urban Pond Insects. *Ecological Indicators*, 73, 422-431. <https://doi.org/10.1016/j.ecolind.2016.10.006>
- Heino, J. et Grönroos, M. (2017). Exploring Species and Site Contributions to Beta Diversity in Stream Insect Assemblages. *Oecologia*, 183(1), 151-160. <https://doi.org/10.1007/s00442-016-3754-7>
- Heino, J., Melo, A. S., Bini, L. M., Altermatt, F., Al-Shami, S. A., Angeler, D. G., Bonada, N., Brand, C., Callisto, M., Cottenie, K., Dangles, O., Dudgeon, D., Encalada, A., Göthe, E., Grönroos, M., Hamada, N., Jacobsen, D., Landeiro, V. L., Ligeiro, R., ... Et Townsend, C. R. (2015). A Comparative Analysis Reveals Weak Relationships between Ecological Factors and Beta Diversity of Stream Insect Metacommunities at Two Spatial Levels. *Ecology and Evolution*, 5(6), 1235-1248. <https://doi.org/10.1002/ece3.1439>
- Hijmans, R. J., Phillips, S., Leathwick, J. et Elith, J. (2017). dismo : Species Distribution Modeling. <https://CRAN.R-project.org/package=dismo>
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M. et Ladle, R. J. (2015). Seven Shortfalls That Beset Large-Scale Knowledge of Biodiversity. *Annual Review of*

- Ecology, Evolution, and Systematics*, 46(1), 523-549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22(0), 415-427. <https://doi.org/10.1101/SQB.1957.022.01.039>
- Hutchinson, G. E. (1959). Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? *The American Naturalist*, 93(870), 145-159. <https://doi.org/10.1086/282070>
- Johnston, A., Hochachka, W. M., Strimas-Mackey, M. E., Gutierrez, V. R., Robinson, O. J., Miller, E. T., Auer, T., Kelling, S. T. et Fink, D. (2020). Analytical Guidelines to Increase the Value of Citizen Science Data : Using eBird Data to Estimate Species Occurrence. *bioRxiv*, 574392. <https://doi.org/10.1101/574392>
- Jost, L. (2007). Partitioning Diversity into Independent Alpha and Beta Components. *Ecology*, 88(10), 2427-2439. <https://doi.org/10.1890/06-1736.1>
- Koleff, P., Gaston, K. J. et Lennon, J. J. (2003). Measuring Beta Diversity for Presence–Absence Data. *Journal of Animal Ecology*, 367-382. [https://doi.org/10.1046/j.1365-2656.2003.00710.x@10.1111/\(ISSN\)1365-2656.BIODIV](https://doi.org/10.1046/j.1365-2656.2003.00710.x@10.1111/(ISSN)1365-2656.BIODIV)
- Kong, H., Chevalier, M., Laffaille, P. et Lek, S. (2017). Spatio-Temporal Variation of Fish Taxonomic Composition in a South-East Asian Flood-Pulse System. *PLOS ONE*, 12(3), e0174582. <https://doi.org/10.1371/journal.pone.0174582>
- Legendre, P., Borcard, D. et Peres-Neto, P. R. (2005). Analyzing Beta Diversity : Partitioning the Spatial Variation of Community Composition Data. *Ecological Monographs*, 75(4), 435-450. <https://doi.org/10.1890/05-0549>
- Legendre, P. et Condit, R. (2019). Spatial and Temporal Analysis of Beta Diversity in the Barro Colorado Island Forest Dynamics Plot, Panama. *Forest Ecosystems*, 6(1), 7. <https://doi.org/10.1186/s40663-019-0164-4>
- Legendre, P. et De Cáceres, M. (2013). Beta Diversity as the Variance of Community Data : Dissimilarity Coefficients and Partitioning. *Ecology Letters*, 16(8), 951-963. <https://doi.org/10.1111/ele.12141>

- Legendre, P. et Legendre, L. (2012). *Numerical Ecology* (Third English edition). Elsevier. <http://www.sciencedirect.com/science/bookseries/01678892/24>
- Niskanen, A. K. J., Heikkinen, R. K., Väre, H. et Luoto, M. (2017). Drivers of High-Latitude Plant Diversity Hotspots and Their Congruence. *Biological Conservation*, 212, 288-299. <https://doi.org/10.1016/j.biocon.2017.06.019>
- Nix, H. A. (1986). A Biogeographic Analysis of Australian Elapid Snakes. *Atlas of elapid snakes of Australia*, 7, 4-15.
- Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., ... Et Ovaskainen, O. (2019). A Comprehensive Evaluation of Predictive Performance of 33 Species Distribution Models at Species and Community Levels. *Ecological Monographs*, 89(3), e01370. <https://doi.org/10.1002/ecm.1370>
- Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F. G., Duan, L., Dunson, D., Roslin, T. et Abrego, N. (2017). How to Make More out of Community Data ? A Conceptual Framework and Its Implementation as Models and Software. *Ecology Letters*, 20(5), 561-576. <https://doi.org/10.1111/ele.12757>
- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. et Blair, M. E. (2017). Opening the Black Box : An Open-Source Release of Maxent. *Ecography*, 40(7), 887-893. <https://doi.org/10.1111/ecog.03049>
- Phillips, S. J., Anderson, R. P. et Schapire, R. E. (2006). Maximum Entropy Modeling of Species Geographic Distributions. *Ecological Modelling*, 190(3), 231-259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Phillips, S. J. et Dudík, M. (2008). Modeling of Species Distributions with Maxent : New Extensions and a Comprehensive Evaluation. *Ecography*, 31(2), 161-175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>

- Poisot, T., Guéveneux-Julien, C., Fortin, M.-J., Gravel, D. et Legendre, P. (2017). Hosts, Parasites and Their Interactions Respond to Different Climatic Variables. *Global Ecology and Biogeography*, 26(8), 942-951. <https://doi.org/10.1111/geb.12602>
- Poisot, T., LaBrie, R., Larson, E., Rahlin, A. et Simmons, B. I. (2019). Data-Based, Synthesis-Driven : Setting the Agenda for Computational Ecology. *Ideas in Ecology and Evolution*, 12. <https://doi.org/10.24908/iee.2019.12.2.e>
- Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A. et McCarthy, M. A. (2014). Understanding Co-Occurrence by Modelling Species Simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5(5), 397-406. <https://doi.org/10.1111/2041-210X.12180>
- Staniczenko, P. P. A., Sivasubramaniam, P., Suttle, K. B. et Pearson, R. G. (2017). Linking Macroecology and Community Ecology : Refining Predictions of Species Distributions Using Biotic Interaction Networks. *Ecology Letters*, 20(6), 693-707. <https://doi.org/10.1111/ele.12770>
- Tan, L., Fan, C., Zhang, C., von Gadow, K. et Fan, X. (2017). How Beta Diversity and the Underlying Causes Vary with Sampling Scales in the Changbai Mountain Forests. *Ecology and Evolution*, 7(23), 10116-10123. <https://doi.org/10.1002/ece3.3493>
- Tan, L., Fan, C., Zhang, C. et Zhao, X. (2019). Understanding and Protecting Forest Biodiversity in Relation to Species and Local Contributions to Beta Diversity. *European Journal of Forest Research*, 138(6), 1005-1013. <https://doi.org/10.1007/s10342-019-01220-3>
- Taranu, Z. E., Pinel-Alloul, B. et Legendre, P. (2020). Large-Scale Multi-Trophic Co-Response Models and Environmental Control of Pelagic Food Webs in Québec Lakes. *Oikos*, n/a(n/a). <https://doi.org/10.1111/oik.07685>
- Vasconcelos, T. S., do Nascimento, B. T. M. et Prado, V. H. M. (2018). Expected Impacts of Climate Change Threaten the Anuran Diversity in the Brazilian Hotspots. *Ecology and Evolution*, 8(16), 7894-7906. <https://doi.org/10.1002/ece3.4357>
- Vellend, M. (2001). Do Commonly Used Indices of  $\beta$ -Diversity Measure Species Turnover? *Journal of Vegetation Science*, 12(4), 545-552. <https://doi.org/10.2307/3237006>



- Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, 30(3), 279-338. <https://doi.org/10.2307/1943563>
- Whittaker, R. H. (1972). Evolution and Measurement of Species Diversity. *Taxon*, 21(2/3), 213-251. <https://doi.org/10.2307/1218190>
- Yang, J., Sorte, F. A. L., Pyšek, P., Yan, P., Nowak, D. et McBride, J. (2015). The Compositional Similarity of Urban Forests among the World's Cities Is Scale Dependent. *Global Ecology and Biogeography*, 24(12), 1413-1423. <https://doi.org/10.1111/geb.12376>
- Yao, J., Huang, J., Ding, Y., Xu, Y., Xu, H. et Zang, R. (2021). Ecological Uniqueness of Species Assemblages and Their Determinants in Forest Communities. *Diversity and Distributions*, 27(3), 454-462. <https://doi.org/10.1111/ddi.13205>
- Yao, J., Zhang, C., Cáceres, M. D., Legendre, P. et Zhao, X. (2019). Variation in Compositional and Structural Components of Community Assemblage and Its Determinants. *Journal of Vegetation Science*, 30(2), 257-268. <https://doi.org/10.1111/jvs.12708>
- Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T. et Wüest, R. O. (2020). Testing Species Assemblage Predictions from Stacked and Joint Species Distribution Models. *Journal of Biogeography*, 47(1), 101-113. <https://doi.org/10.1111/jbi.13608>