

Evaluating ecological uniqueness over broad spatial extents using species distribution modelling

Gabriel Dansereau^{1,2,*} Pierre Legendre^{1,2} Timothée Poisot^{1,2}

¹ Département de sciences biologiques, Université de Montréal, Montréal, Canada ² Québec

Centre for Biodiversity Science, Montréal, Canada

* gabriel.dansereau@umontreal.ca

This work is released by its authors under a CC-BY 4.0 license



Last revision: *February 23, 2022*

Abstract: Local contributions to beta diversity (LCBD) can be used to identify sites with high ecological uniqueness and exceptional species composition within a region of interest. Yet, these indices are typically used on local or regional scales with relatively few sites, as they require information on complete community compositions difficult to acquire on larger scales. Here, we investigated how LCBD indices can be predicted over broad spatial extents using species distribution modelling and examined the effect of scale changes on beta diversity quantification. We used Bayesian additive regression trees (BARTs) to predict warbler species distributions in North America based on observations recorded in the eBird database. We then calculated LCBD indices for observed and predicted data and compared the site-wise difference using direct comparison, a spatial association test, and generalized linear regression. We also examined the relationship between LCBD values and species richness in different regions and at various spatial extents. Our results showed that species distribution models provided uniqueness estimates highly correlated with observed data. The form and variance of the LCBD-richness relationship varied according to the region and the total extent size. The relationship was also affected by the proportion of rare species in the communities. Therefore, sites identified as unique over broad spatial extents may vary according to regional characteristics. These results show that species distribution modelling can be used to predict ecological uniqueness over broad spatial extents, which could help identify beta diversity hotspots and important targets for conservation purposes in unsampled locations.

1 **Introduction**

2 Beta diversity, defined as the variation in species composition among sites in a geographic re-
3 gion of interest (Legendre, Borcard, and Peres-Neto 2005), is an essential measure to describe
4 the organization of biodiversity through space. Total beta diversity within a community can be
5 partitioned into local contributions to beta diversity (LCBD) (Legendre and De Cáceres 2013),
6 which allow the identification of sites with exceptional species composition, hence unique biodi-
7 versity and potential conservation value. Sites with unique community composition often differ
8 from those with high species richness, possibly as they harbour rare species or help maintain
9 beta diversity (da Silva, Hernández, and Heino 2018; Heino et al. 2017; Landeiro et al. 2018).
10 Hence, focusing on uniqueness can prove helpful as a complementary approach to species rich-
11 ness (Heino and Grönroos 2017; da Silva, Hernández, and Heino 2018; Yao et al. 2021; Dubois,
12 Proulx, and Pellerin 2020). However, the use of LCBD indices is currently limited in two ways.
13 First, LBCD indices are typically used on data collected over local or regional scales with rela-
14 tively few sites, for example, on fish communities at intervals along a river or stream (Legendre
15 and De Cáceres 2013). Second, LCBD calculation methods require complete information on
16 community composition; thus, they are inappropriate for partially sampled sites (e.g., where
17 data for some species are missing or uncertain) and cannot directly provide assessments for un-
18 sampled ones. Accordingly, this method is of limited use to identify areas with exceptional
19 biodiversity in regions with sparse sampling. However, predictive approaches offer an oppor-
20 tunity to overcome such limitations, as computational methods often uncover novel ecological
21 insights from existing data (Poisot et al. 2019), including in lesser-known locations and on larger
22 spatial scales.

23 Species distribution models (SDMs) (Guisan and Thuiller 2005) can bring a new perspective to
24 LCBD studies by filling in gaps in community composition data to perform analyses on broader
25 scales. Single-species SDMs aim at predicting the distribution of a species in unsampled loca-
26 tions based on information (such as environmental data) from sampled locations with reported
27 occurrences. Many approaches allow going from single-species SDMs to a whole community on
28 which to evaluate community-level metrics, yet their relevance has not been explicitly evaluated
29 for ecological uniqueness and LCBD indices. The most straightforward approach is stacked

30 distribution models (S-SDMs) (Ferrier and Guisan 2006; Guisan and Rahbek 2011). Single-
31 species SDMs are first performed separately, then combined to form a community prediction on
32 which community-level analyses can be applied. S-SDMs tend to overestimate species richness
33 (Dubuis et al. 2011; D'Amen et al. 2015; Zurell et al. 2020), which could result from threshold-
34 ing the probabilities into presence-absence data before stacking the species distributions (Cal-
35 abrese et al. 2014). Summing the occurrence probabilities without applying a threshold is an
36 alternative (Calabrese et al. 2014), but it may limit some analyses as it does not return species
37 identities for every site (Zurell et al. 2020), as is required with LCBD calculations. In com-
38 parison, joint species distribution models (JSDMs)(Pollock et al. 2014) try to improve predic-
39 tions by incorporating species co-occurrence or shared environmental responses into the mod-
40 els. However, these models do not always improve community-level predictions compared to
41 S-SDMs (Zurell et al. 2020). Spatially explicit species assemblage modelling (SESAM) (Guisan
42 and Rahbek 2011), hierarchical modelling of species communities (HMSC) (Ovaskainen et al.
43 2017), and Bayesian networks (BN) (Staniczenko et al. 2017) are other alternatives that could
44 yield better community predictions than S-SDMs. On the other hand, they add methodological
45 and computational overload, impeding their use for broad spatial extents. Moreover, their rele-
46 vance for community prediction is often validated against extensive work on species richness.
47 By comparison, ecological uniqueness and LCBD indices have rarely been used in predictive
48 frameworks. Therefore, S-SDMs may prove an appropriate first step to establish some prediction
49 baselines.

50 Combining LCBD indices with a predictive approach through SDMs will allow measuring
51 uniqueness over broader spatial extents, across continuous landscapes, and on a higher num-
52 ber of sites than what has previously been studied. LCBD scores have typically been used at
53 local or regional scales with relatively few sites (up to 60 sites on extents covering 10 km to 400
54 km, Legendre and De Cáceres 2013; da Silva and Hernández 2014; Heino et al. 2017; Heino
55 and Grönroos 2017). Some studies did use the measure over broader, near-continental extents
56 (Yang et al. 2015; Poisot et al. 2017; Taranu, Pinel-Alloul, and Legendre 2020), but the total
57 number of sites in these studies were relatively small (maximum 51 sites). Recent studies also
58 investigated LCBD and beta diversity on sites distributed in contiguous grids or as pixels, hence

59 uniform sampling intervals and no spatial gaps, but these did not cover large extents and a high
60 number of sites (up to 1250 sites and 6 km², Tan et al. 2017, 2019; Legendre and Condit 2019;
61 D’Antraccoli et al. 2020). Two recent studies have, however, adopted promising predictive ap-
62 proaches on regional extents. First, Niskanen et al. (2017) predicted LCBD values of plant
63 communities (and three other diversity measures) on a continuous scale and a high number of
64 sites (> 25,000) using Boosted Regression Trees (BRTs). However, they modelled the diversity
65 measures directly after calculating them on a smaller number of sampled sites. Second, Vascon-
66 celos, Nascimento, and Prado (2018) used ecological niche models (ENMs) to predict anurans
67 ecological niches according to actual and forecasted environmental conditions, then calculated
68 the LCBD values on the predictions to identify biodiversity hotspots. Using this approach, pre-
69 dicted LCBD values are calculated in a way closer to the original formulation. This development
70 of predictive techniques is exciting, especially as it could be pushed a step further to continental
71 extents, a higher number of sites, and more species occurrences using SDMs and massive data
72 sources. Still, it should be accompanied by an investigation of the determinant of ecological
73 uniqueness in such conditions.

74 Measuring ecological uniqueness from LCBD indices over broad spatial extents and spatially
75 continuous data also raises the question of which sites will be identified as exceptional and for
76 what reason. The method intends that sites stand out and receive a high LCBD score whenever
77 they display an exceptional community composition, be it a unique assemblage of species with
78 high conservation value or a community richer or poorer than others in the region (Legendre
79 and De Cáceres 2013). Both the original study and many of the later empirical ones have shown
80 a negative relationship between LCBD scores and species richness (Legendre and De Cáceres
81 2013; da Silva and Hernández 2014; Heino et al. 2017; Heino and Grönroos 2017), although
82 other studies observed both negative and positive relationships at different sites (Kong et al.
83 2017) or quadrats (Yao et al. 2021). Some studies showed that the direction of the relationship
84 is related to the percentage of rare species in the community (da Silva, Hernández, and Heino
85 2018; Yao et al. 2021). However, beta diversity and species rarity are both concepts that depend
86 on scale. For instance, total beta diversity increases with spatial extent (Barton et al. 2013) and
87 varies because of higher environmental heterogeneity and sampling of different local species

88 pools (Heino et al. 2015). Therefore, the LCBD-richness relationship and the effect of rare
89 species on LCBD values should be investigated over broad spatial extents, as they might not be
90 constant across scales.

91 Here, we examined whether species distribution models (SDMs) can be combined with local
92 contributions to beta diversity (LCBD) to assess ecological uniqueness over broader spatial ex-
93 tents. We also investigated the effect of scale changes on beta diversity quantification. We first
94 predicted species distributions on continental scales using extended occurrence data from eBird
95 and Bayesian additive regression trees (BARTs). We then quantified uniqueness with the LCBD
96 measure for both predicted and observed data. Next, we examined the site-wise difference us-
97 ing direct comparison, a spatial autocorrelation test, and generalized linear regression. We then
98 investigated the relationship between uniqueness and species richness for different regions and
99 scales and according to the proportion of rare species.

100 **Methods**

101 **Occurrence data**

102 We used occurrence data from eBird (Sullivan et al. 2009) downloaded through the eBird Basic
103 Data set from June 2019 (eBird Basic Dataset 2019). We restricted our analyses to the New
104 World warbler family (*Parulidae*) in North America (Canada, the United States, Mexico). eBird
105 is a semi-structured citizen science data set, meaning that observations are reported as checklists
106 of species detected in an observation run (Johnston et al. 2020). Observers can explicitly specify
107 that their checklist contains all species they could detect and identify during a sampling event, in
108 which case it is labelled as a “complete checklist.” Using complete checklists instead of regular
109 ones allows researchers to infer non-detections in locations where detection efforts occurred,
110 which offers performance gains in species distribution models (Johnston et al. 2020). Therefore,
111 we selected the data from the complete checklists only. Our final data set comprised 62 warbler
112 species and 22,974,330 observations from 9,103,750 checklists. Warblers are a diverse group
113 with many species, are popular among birders given their charismatic aspect, and are widely
114 distributed in various habitats across North America.

115 **Environmental data**

116 Our environmental data consisted of climatic data from WorldClim 2.1 (Fick and Hijmans 2017)
117 and land cover data from the Copernicus Global Land Service (Buchhorn et al. 2019). We
118 restricted these data to a spatial extent comprised between -145.0 and -50.0 degrees of longitude
119 and between 20.0 and 75.0 degrees of latitude. First, the WorldClim data consist of spatially
120 interpolated monthly climate data for global land areas. We used the standard BIOCLIM variables
121 (Booth et al. 2014) from WorldClim 2.1, which represent annual trends, ranges, and extremes
122 of temperature and precipitation, but selected only 8 out of the 19 ones to avoid redundancy
123 (bio1, bio2, bio5, bio6, bio12, bio13, bio14, bio15). We downloaded the data at a resolution of
124 10 arcminutes (around 18 km² at the equator), the coarsest resolution available, which should
125 mitigate potential imprecision in the eBird data regarding the extent of the sampled areas in
126 each observation checklist. Moreover, some studies have argued that coarser resolutions lead
127 to less overestimation of species richness and better identification of bird biodiversity hotspots
128 given the patchiness of observation data (Hurlbert and Jetz 2007). We acknowledge that using an
129 arcminutes-based resolution means that the surface area of our pixels will not be equal depending
130 on the latitude.

131 Second, the Copernicus data are a set of variables representing ten land cover classes (e.g., crops,
132 trees, urban areas) and measured as a percentage of land cover. The data are only available at
133 a finer resolution of 100 m. We coarsened them to the same ten arcminute resolution as the
134 WorldClim data by averaging the pixels' cover fraction values. We removed two variables (moss
135 and snow) from our predictive models as their cover fraction was 0% on all sites with warbler
136 observations.

137 **Species distribution models**

138 We converted the occurrence data to a presence-absence format compatible with community
139 analyses. We considered every pixel from our ten arcminutes environmental layers as a site and
140 then verified, for each species, if there was a single observation in every site. Finally, we recorded
141 the outcome as a binary value: present (1) if a species was ever recorded in a site and absent (0) if

142 it was not. Complete checklists help ensure that these zeros represent non-detections, rather than
143 the species not being reported; hence we considered them as absence data, similar to Johnston
144 et al. (2020), although we recognize that there exists a doubt on whether these truly represent
145 non-detections.

146 We predicted species distribution data on continuous scales from our presence-absence data
147 using Bayesian Additive Regression Trees (BARTs) (Chipman, George, and McCulloch 2010),
148 a classification and regression trees method recently suggested for species distribution modelling
149 (Carlson 2020). BARTs are based on an ensemble of trees, similarly to Boosted Regression Trees
150 and Random Forest, but follow a sum-of-trees model and a Bayesian framework. Trees are first
151 constrained as weak learners by priors regarding structure and nodes, then updated through an
152 iterative Bayesian backfitting Markov Chain Monte Carlo (MCMC) algorithm which ultimately
153 generates a posterior distribution of predicted classification probabilities (Chipman, George,
154 and McCulloch 2010; Carlson 2020). In the context of species distribution modelling, BARTs
155 showed high performance when compared to other predictive algorithms (Konowalik and Nosol
156 2021; Tytar and Baidashnikov 2021). We first performed BARTs separately for all species and
157 estimated the probability of occurrence in all the sites of our region of interest using the posterior
158 median. We then converted the results to a binary outcome according to the threshold that
159 maximized the True Skill Statistic (TSS) for each species, as suggested by Carlson (2020).

160 **Quantification of ecological uniqueness**

161 We used the method of Legendre and De Cáceres (2013) to quantify compositional unique-
162 ness from overall beta diversity for both the observed and predicted data. First, we assembled
163 the presence-absence data by site to form two site-by-species community matrices, one from
164 observed data, called Y (39,024 sites by 62 species), and one from predicted data, called \hat{Y}
165 (99,382 sites by 62 species). Next, we measured species richness per site as the number of
166 species present. Finally, we removed the sites without any species from the predicted matrix
167 \hat{Y} , for a new total of 85,526 sites (this was unnecessary for the observed matrix Y). We then
168 applied the Hellinger transformation to both matrices in order to compute beta diversity from
169 the community composition data (Legendre and De Cáceres 2013). We measured total beta di-

170 versity as the variance of each community matrix and calculated the local contributions to beta
171 diversity (LCBD), which quantify how much a specific site (a row in each matrix) contributes
172 to the overall variance in the community (Legendre and De Cáceres 2013). High LCBD values
173 indicate a unique community composition, while low values indicate a more common species
174 set. We note that our LCBD values, which add up to 1 because the values are divided by the
175 total sum-of-squares of the data matrix, were very low given the high number of sites in both
176 Y and \hat{Y} . However, the relative difference between the scores in one set matters more than the
177 absolute value to differentiate their uniqueness.

178 **Comparison of observed and predicted values**

179 We performed three verification to compare the richness and uniqueness estimates obtained
180 from our predicted distributions to those obtained with the eBird occurrence data. First, we
181 performed a direct comparison by subtracting the richness and LCBD estimates obtained from
182 Y (the observed data) from the estimates obtained from \hat{Y} (the predicted data). To do so, we
183 used the richness estimates as-is but modified the LCBD values to achieve a non-biased com-
184 parison, given that the values were initially calculated on sets of different lengths. Therefore,
185 we recomputed the LCBD scores only for the sites for which we had occurrences in both Y and
186 \hat{Y} , which mostly corresponded to the sites in Y , minus a few sites where the SDMs predicted no
187 species occurrence. We then plotted the richness and LCBD differences to examine their spatial
188 distributions. Second, we performed the modified t test from Clifford, Richardson, and Hemon
189 (1989) to assess the correlation between the observed and predicted estimates and test for spatial
190 association. We performed the test separately for the richness and the LCBD estimates. Third,
191 we performed Generalized Linear Models between the observed and predicted estimates and
192 plotted the deviance residuals to examine their spatial distribution. We used a negative binomial
193 regression with a log link function for the richness estimates and a beta regression with a logit
194 link function for the LCBD values, similar to Heino and Grönroos (2017) and Yao et al. (2021).

195 **Investigation of regional and scaling variation**

196 To investigate possible regional and scaling effects, we recalculated LCBD values on various
197 subregions at different locations and scales. First, we selected two subregions of equivalent size
198 (20.0 longitude degrees by 10.0 latitude degrees) with contrasting richness profiles and corre-
199 sponding to different ecoregions to verify if the relationship between species richness and LCBD
200 values was similar. The first subregion was in the Northeast (longitude between -80.0 and -60.0,
201 latitude between 40.0 and 50.0), was mostly species-rich (for both the observed and predicted
202 data), and corresponded to the Eastern Temperate Forests level I ecoregion (Commission for
203 Environmental Cooperation 1997). The second subregion was in the Southwest (longitude be-
204 tween -120.0 and 100.0, latitude between 30.0 and 40.0), was mostly species-poor, and covered
205 Mediterranean California, North American Deserts, Temperate Sierras, and Southern Semi-
206 Arid Highlands ecoregions (Commission for Environmental Cooperation 1997). Second, we
207 recalculated the LCBD indices at three different extents, starting with a focus on the Northeast
208 subregion and progressively extending the extent to encompass the Southwest subregion. We did
209 these two verifications with both the observed and predicted data but only illustrate the results
210 with the predicted data as both were qualitatively similar.

211 **Proportion of rare species**

212 We investigated the effect of the proportion of rare species in the community on the direction
213 of the relationship between species richness and LCBD values in our Northeast and Southwest
214 subregions. Following De Cáceres et al. (2012) and Yao et al. (2021), we classified species
215 as rare when they occurred in less than 40% of the sites in each subregion. We calculated the
216 proportion of rare species for every site. We then grouped the sites for both subregions depend-
217 ing on whether they were part of an ascending or a descending portion in the LCBD-richness
218 relationship. Given that the relationship sometimes displays a curvilinear form with a positive
219 quadratic term (Heino and Grönroos 2017; Tan et al. 2019), we separated the ascending and
220 descending portions based on the species richness at the site with the lowest LCBD value (us-
221 ing the median richness if there were multiple sites). This value corresponds to the inflection
222 point of the relationships. For example, the lowest LCBD value was $7.045e-05$ in the Northeast

223 subregion and the corresponding richness was 23. All the sites with more than 23 species were
224 assigned to the ascending portion, and all the sites with 23 species or fewer were assigned to the
225 descending portion. In the Southwest subregion, the lowest LCBD value and its corresponding
226 richness were 5.438e-05 and 12, respectively. We then mapped the ascending and descending
227 groups to view their spatial distribution. We also examined the distribution of the rare species
228 proportions in both groups using a kernel density estimation plot. Similar to our previous veri-
229 fication, we performed this analysis with both observed and predicted data but once again only
230 illustrate the results with the predicted data as both were qualitatively similar.

231 **Software**

232 We used *Julia v1.6.1* (Bezanson et al. 2017) for most of the project and *R v4.1.0* (R Core Team
233 2021) for some specific steps. We used the *Julia* package `SimpleSDMLayers.jl` (Dansereau and
234 Poisot 2021) as the basic framework for our analyses, to download the WorldClim 2.1 data, and to
235 map our results through the package's integration of `Plots.jl`. We also used `StatsPlots.jl` to
236 produce the kernel density estimation plots in our rare species analysis. We computed the LCBD
237 indices with our own function implemented in *Julia*, whose results were verified by comparison
238 to the `beta.div` function from the package `adespatial` (Dray et al. 2021) in *R*. We used the *R*
239 packages `auk` (Strimas-Mackey, Miller, and Hochachka 2018) to extract and manipulate eBird
240 data, `embarcadero` (Carlson 2020) to perform the BART models, `vegan` (Oksanen et al. 2019) to
241 apply the Hellinger transformations, and `SpatialPack` (Vallejos, Osorio, and Bevilacqua 2020)
242 to perform the modified *t* test (with the function `modified.ttest`) from Clifford, Richardson,
243 and Hemon (1989). We used `MASS` (Venables and Ripley 2002) and `betareg` (Cribari-Neto
244 and Zeileis 2010) to perform the negative binomial and beta regressions, respectively. We also
245 used `GDAL` (GDAL/OGR contributors 2021) to coarsen the Copernicus land cover data. All
246 the scripts required to reproduce the analyses are archived on Zenodo ([https://doi.org/10.
247 5281/zenodo.6024392](https://doi.org/10.5281/zenodo.6024392)).

248 **Results**

249 **Species distribution models generate relevant community predictions**

250 Species richness from observation data (Fig. 1a) was higher on the East Coast and lower on
251 the West Coast, with many unsampled patches in the North, South, and Central West. Richness
252 results from SDM data (Fig. 1b) displayed higher richness on the East Coast and sites with few or
253 no species up north and in the Central West. There was no clear latitudinal gradient in richness
254 but rather an East-West one. Landmarks such as the Rockies and croplands in the Central West
255 (which should be species-poor habitats) were notably visible on the maps, separating the East
256 and West. LCBD scores from observation data (Fig. 1c) were low on the East Coast and higher
257 on the border of sampled sites in the Central West. They were also higher in the North and in
258 the South where observations were sparser. Results from SDM predictions were qualitatively
259 similar (Fig. 1d), with lower LCBD values in the East and more unique sites in the Central West,
260 Central Mexico, and some Northern regions. There was no clear latitudinal gradient, and the
261 East-West contrast, while present, was less clear than on the richness maps.

262 [Figure 1 about here.]

263 The modified *t* test of Clifford, Richardson, and Hemon (1989) showed a high correlation be-
264 tween the observed and predicted estimates of richness and uniqueness, as well as a statistically
265 significant spatial association between the values. For species richness, the correlation coeffi-
266 cient was 0.777, the *F*-statistic was 20.007, and the p-value was 6.093e-04. For LCBD scores,
267 the correlation coefficient was 0.518, the *F*-statistic was 40.083, and the p-value was 5.528e-09.

268 The difference between the observed and predicted estimates (predicted richness - observed
269 richness and predicted LCBD - observed LCBD) showed opposite geographic distributions for
270 species richness and ecological uniqueness (Fig. 2). Predicted richness estimates were higher
271 than observed estimates on the East Coast, on the West Coast and in Mexico but were lower than
272 observed estimates in the Central West (Fig. 2a). Predicted LCBD estimates, on the other hand,
273 were lower than observed estimates on the East Coast and higher in the Central West (Fig. 2b).

274 Regression residuals showed similar geographic distributions to their corresponding difference
275 values (Fig. 3).

276 [Figure 2 about here.]

277 [Figure 3 about here.]

278 **Uniqueness displays regional variation as two distinct profiles**

279 The relationship between LCBD values and species richness displayed contrasting profiles in
280 species-rich and species-poor regions (Fig. 4). In the species-rich Northeast region , LCBD
281 scores displayed a mostly decreasing relationship with species richness, with a slightly curvi-
282 linear form and increase of values for very rich sites. The sites with the highest LCBD values
283 were the species-poor sites while the species-rich sites displayed scores. The Southwest subarea
284 showed a different relationship with a sharper initial decline and a larger increase as richness
285 reached 20 species. The sites with the highest LCBD values were the poorest in terms of species
286 richness, as in the Northeast region, but the species-rich sites were proportionally more unique
287 in the Southwest region. Total beta diversity was higher in the Southwest subregion (0.417) than
288 in the Northeast (0.179), indicating higher compositional differences between the sites.

289 [Figure 4 about here.]

290 **Uniqueness depends on the scale on which it is measured**

291 The LCBD-richness relationship showed important variation when scaling up and changing the
292 region's extent (Fig. 5). For smaller extents, starting with a species-rich region, the relationship
293 is well defined, mostly decreasing but notably curvilinear (with a lesser increase for richness
294 values higher than the median). However, as the extent increases and progressively reaches
295 species-poor regions, the relationship broadens, displays more variance, and loses its curvilinear
296 aspect while keeping a decreasing form. Total beta diversity was higher when increasing the
297 spatial extent, going from 0.121 to 0.284 and 0.687.

298 [Figure 5 about here.]

299 **Uniqueness depends on the proportion of rare species**

300 The proportion of rare species per site differed depending on the classification in the ascending or
301 descending portions of the LCBD-richness relationship (Fig. 6). The proportion of rare species
302 was higher in the sites corresponding to the ascending portions of the relationships (shown in 4)
303 than in the sites corresponding to the descending portions for both subregions. The classification
304 of the sites in the two portions showed a clear latitudinal gradient in the Northeast subregion,
305 while it was distributed in patches in the Southwest subregion (Fig. 6).

306 [Figure 6 about here.]

307 **Discussion**

308 Our results showed a decreasing relationship between species richness and LCBD values on
309 broad spatial extents (Fig. 5c) but also highlighted that the exact form of this relationship varies
310 depending on the region and the spatial extent on which it is measured. Our species-rich North-
311 east subregion (Fig. 4a) showed a decreasing relationship, very similar to previous studies, and
312 slightly curvilinear, as described by Heino and Grönroos (2017) and Tan et al. (2019). This
313 result for warbler species is in line with the original study on fish communities (Legendre and
314 De Cáceres 2013) and with following ones on insect metacommunities (da Silva and Hernández
315 2014; Heino et al. 2017; Heino and Grönroos 2017), dung beetles (da Silva, Hernández, and
316 Heino 2018; da Silva, Bogoni, and Heino 2020), aquatic beetles (Heino and Alahuhta 2019),
317 stream macroinvertebrates (Sor, Legendre, and Lek 2018), stream diatoms (Vilmi, Karjalainen,
318 and Heino 2017), multi-trophic pelagic food webs (phytoplankton, zooplankton, fish) (Taranu,
319 Pinel-Alloul, and Legendre 2020), temperate forest trees (Tan et al. 2019), mammals (da Silva,
320 Bogoni, and Heino 2020), wetland birds (de Deus et al. 2020), and various phylogenetic groups
321 (plants, lizards, mites, anurans, mesoinvertebrates) (Landeiro et al. 2018). However, it was
322 originally argued that the negative relationship was not general or obligatory (Legendre and De
323 Cáceres 2013). Different LCBD-richness relationships have also been observed, with both pos-
324 itive and negative relationships for different sites or taxonomic groups in some studies (Kong et

325 al. 2017; Teittinen et al. 2017), as well as a negative relationship with the number of common
326 species but a positive relationship with the number of rare species (Qiao et al. 2015).

327 Our results further show that the relationship may depend on the region's richness profile, as
328 the relationship was different in our species-poor Southwest subregion, with a sharper initial
329 decrease (Fig. 4b). Therefore, the curvilinear form may depend on how pronounced the contrast
330 is between the region's median richness and its richest ecologically feasible sites. The increasing
331 part of the curvilinear form for higher richness values was also more pronounced in our results
332 (Fig. 4a,b; Fig. 5c) than in previous studies (e.g., Tan et al. 2019), which reinforces the idea that
333 the relationship and its curvilinear form may vary depending on the region.

334 The variation in the LCBD-richness relationship when extending the study extent showed that
335 the uniqueness patterns highlighted are not necessarily the same depending on the scale on which
336 it is used (Fig. 5). The relationship progressively lost its clear definition and curvilinear form
337 as the East and West profiles merged, creating a new distinct profile with more variation and
338 no curvilinear form. Therefore, aggregating too many different sites might possibly mask some
339 patterns of uniqueness in species-rich sites. Total beta diversity, on the other hand, showed the
340 variation expected from previous studies, increasing with spatial extent (Fig. 5) (Barton et al.
341 2013; Heino et al. 2015). Its value was high at the continental scale (0.687) but lower than what
342 has been observed in some studies (e.g., 0.80 on macroinvertebrate communities in the Lower
343 Mekong Basin by Sor, Legendre, and Lek 2018).

344 Our results confirm that the proportion of rare species in the community may affect the direc-
345 tion of the relationship between species richness and ecological uniqueness (Fig. 6). da Silva,
346 Hernández, and Heino (2018) suggested that the proportion of rare and common species in the
347 communities determines whether the relationship will be negative, non-significant, or positive.
348 Yao et al. (2021) showed an association between the direction of the relationship and the pro-
349 portion of rare species, with sites with a lower proportion (between 60% and 75% in their case)
350 displaying a negative relationship and sites with a higher proportion (around 85%) showing a
351 positive one. Our results further show that sites associated with a positive relationship within a
352 curvilinear one tended to have a higher rare species proportion (Fig. 6). This also implies that
353 the proportion of rare species was higher in species-rich sites than in species-poor ones in both

354 our Northeast and Southwest subregions. Further work should attempt to disentangle the effects
355 of the rare species proportion and the region's richness profile.

356 Our results showed that SDM models provide richness and uniqueness predictions highly cor-
357 related to the occurrence data while filling gaps in poorly sampled regions (Fig. 1). The results
358 showed a statistically significant spatial association between predicted and observed estimates
359 despite correcting for autocorrelation using the modified *t*-test from Clifford, Richardson, and
360 Hemon (1989). A positive autocorrelation on large distances indicates aggregates or structures
361 repeating through space (Legendre and Fortin 1989). This is consistent with our results, as the
362 distribution of richness and uniqueness values was visibly spatially structured in both our ob-
363 served and predicted data (Fig. 1). Nonetheless, it is possible that the autocorrelation in the
364 predicted values could represent an artifact of the predictive models (capturing the spatial struc-
365 ture from the environmental variables, for example), and might not represent the true autocor-
366 relation expected for the uniqueness estimates. Further work could verify this by quantitatively
367 comparing the autocorrelation and spatial structures in the observed and predicted uniqueness
368 estimates.

369 Predicted values also tended to underestimate uniqueness in species-rich regions and overesti-
370 mate it in species-poor ones, with the opposite trend for species richness (Figs. 2, 3). Overpre-
371 diction of richness using S-SDMs was reported previously (Dubuis et al. 2011; D'Amen et al.
372 2015; Zurell et al. 2020). No comparable baseline exists for predictions of LCBD values, as
373 our study is the first to compare LCBD estimates from observed and predicted data in such a
374 way. However, some studies showed that LCBD distributions were spatially structured across
375 sampling sites (da Silva, Hernández, and Heino 2018). On the other hand, the spatial structure
376 in our results did not exactly concord with the one reported by Heino and Alahuhta (2019), who
377 showed a negative relationship between LCBD values and latitude for diving beetles commu-
378 nities in Northern Europe. In comparison, our LCBD scores increased both in the North and
379 South (Fig. 1), hence did not strictly increase with latitude, and also showed a clear East-West
380 gradient. Overall, our distribution results (Figs. 1, 2, 3) also have implications for conservation,
381 as they confirm that species richness and ecological uniqueness measured from LCBD values
382 may conflict and highlight different potential hotspots (Dubois, Proulx, and Pellerin 2020; Yao

383 et al. 2021), thus reinstating the need to protect both with complementary strategies.

384 Our predictions for regions with sparse sampling are of interest as they allow a quantitative
385 evaluation, however imperfect, for sites where we would otherwise have no information. Our
386 SDMs also offered relevant LCBD predictions using eBird, arguably one of the largest presence-
387 absence data sets available (when using its complete checklist system), showing the measure's
388 potential on such massive data. Together, the potential to generate uniqueness predictions in
389 new locations and through massive data opens new opportunities for LCBD analyses on ex-
390 tended spatial scales and on a broader diversity of taxons. An interesting way forward would
391 be to test these results using more advanced community assembling techniques than S-SDMs.
392 The use of SESAM (Guisan and Rahbek 2011) with probabilistic SDMs, probability ranking,
393 and species richness predictions as macroecological constraints returns high site-level prediction
394 accuracy (Zurell et al. 2020) and would be compatible with presence-absence LCBD calcula-
395 tions. The use of probabilistic stacks rather than binary ones (Calabrese et al. 2014) could also
396 constitute a novel way to calculate LCBD indices. Both these procedures should reduce the
397 richness deviation we observed, and it would be interesting to verify if this can also be the case
398 with LCBD values. An ensemble of SDM algorithms could also be used to capture a broader
399 range of possible outcomes for the LCBD predictions. However, given the high performance
400 of BARTs in model comparisons (Konowalik and Nosol 2021; Tytar and Baidashnikov 2021)
401 and the large extent we covered, we do not believe the changes to the LCBD gradients would be
402 strong enough to affect our results in a meaningful way.

403 This study showed how ecological uniqueness can be measured over broad spatial extents, in-
404 cluding for regions with sparse sampling, and how scale changes may affect beta diversity quan-
405 tification. It is the first study to assess the relevance of local contributions to beta diversity
406 calculated on the output of species distribution models. It is also the first to compare the re-
407 lationship between LCBD values and species richness for different regions and spatial extents.
408 First, our results showed that the negative relationship often observed between species richness
409 and LCBD scores can take different forms depending on the richness profile of the regions on
410 which it is measured. Therefore, species-rich and species-poor regions may display different
411 ways to be unique. Second, the negative relationship was not constant when varying the spatial

412 study extent and may be less clearly defined at broad scales when contrasting regional relation-
413 ships are present. The broad-scale uniqueness profile might then be completely distinct from
414 the regional profiles constituting it. Finally, species distribution models offer a promising way
415 to generate uniqueness predictions on broad spatial extents and could prove useful to identify
416 beta diversity hotspots in unsampled locations on large spatial scales, which could be important
417 targets for conservation purposes.

418 **Acknowledgments**

419 We acknowledge that this study was conducted on land within the traditional unceded territory
420 of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwinini-
421 wak nations. We received financial support from the Fonds de recherche du Québec - Nature
422 et technologie (FRQNT) and the Computational Biodiversity Science and Services (BIOS²)
423 NSERC CREATE training program. We thank Élise Filotas and Anne-Lise Routier for their
424 helpful comments on this manuscript.

425 **References**

- 426 Barton, Philip S., Saul A. Cunningham, Adrian D. Manning, Heloise Gibb, David B. Linden-
427 mayer, and Raphael K. Didham. 2013. “The Spatial Scaling of Beta Diversity.” *Global*
428 *Ecology and Biogeography* 22 (6): 639–47. <https://doi.org/10.1111/geb.12031>.
- 429 Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B. Shah. 2017. “Julia: A Fresh
430 Approach to Numerical Computing.” *SIAM Review* 59 (1): 65–98. <https://doi.org/10.1137/141000671>.
- 431
- 432 Booth, Trevor H., Henry A. Nix, John R. Busby, and Michael F. Hutchinson. 2014. “BIOCLIM:
433 The First Species Distribution Modelling Package, Its Early Applications and Relevance to
434 Most Current MaxEnt Studies.” *Diversity and Distributions* 20 (1): 1–9. <https://doi.org/10.1111/ddi.12144>.
- 435

436 Buchhorn, Marcel, Bruno Smets, Luc Bertels, Myroslava Lesiv, Nandin-Erdene Tsendbazar,
437 Martin Herold, and Steffen Fritz. 2019. "Copernicus Global Land Service: Land Cover
438 100m: Epoch 2015: Globe." Zenodo. <https://doi.org/10.5281/zenodo.3243509>.

439 Calabrese, Justin M., Grégoire Certain, Casper Kraan, and Carsten F. Dormann. 2014. "Stack-
440 ing Species Distribution Models and Adjusting Bias by Linking Them to Macroecological
441 Models." *Global Ecology and Biogeography* 23 (1): 99–112. <https://doi.org/10.1111/geb.12102>.

442

443 Carlson, Colin J. 2020. "Embarcadero: Species Distribution Modelling with Bayesian Additive
444 Regression Trees in R." *Methods in Ecology and Evolution* 11 (7): 850–58. <https://doi.org/10.1111/2041-210X.13389>.

445

446 Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. 2010. "BART: Bayesian
447 Additive Regression Trees." *Annals of Applied Statistics* 4 (1): 266–98. <https://doi.org/10.1214/09-AOAS285>.

448

449 Clifford, Peter, Sylvia Richardson, and Denis Hemon. 1989. "Assessing the Significance of the
450 Correlation Between Two Spatial Processes." *Biometrics* 45 (1): 123–34. <https://doi.org/10.2307/2532039>.

451

452 Commission for Environmental Cooperation. 1997. *Ecological Regions of North America*.
453 Commission for Environmental Cooperation. [http://www3.cec.org/islandora/en/item/
454 1701-ecological-regions-north-america-toward-common-perspective/](http://www3.cec.org/islandora/en/item/1701-ecological-regions-north-america-toward-common-perspective/).

455 Cribari-Neto, Francisco, and Achim Zeileis. 2010. "Beta Regression in R." *Journal of Statistical
456 Software* 34 (1): 1–24. <https://doi.org/10.18637/jss.v034.i02>.

457 D'Amen, Manuela, Anne Dubuis, Rui F. Fernandes, Julien Pottier, Loïc Pellissier, and Antoine
458 Guisan. 2015. "Using Species Richness and Functional Traits Predictions to Constrain As-
459 semblage Predictions from Stacked Species Distribution Models." *Journal of Biogeography*
460 42 (7): 1255–66. <https://doi.org/10.1111/jbi.12485>.

461 D'Antraccoli, Marco, Giovanni Bacaro, Enrico Tordoni, Gianni Bedini, and Lorenzo Peruzzi.
462 2020. "More Species, Less Effort: Designing and Comparing Sampling Strategies to Draft
463 Optimised Floristic Inventories." *Perspectives in Plant Ecology, Evolution and Systematics*

464 45: 125547. <https://doi.org/10.1016/j.ppees.2020.125547>.

465 Dansereau, Gabriel, and Timothée Poisot. 2021. “SimpleSDMLayers.jl and GBIF.jl: A Frame-
466 work for Species Distribution Modeling in Julia.” *Journal of Open Source Software* 6 (57):
467 2872. <https://doi.org/10.21105/joss.02872>.

468 De Cáceres, Miquel, Pierre Legendre, Renato Valencia, Min Cao, Li-Wan Chang, George Chuy-
469 ong, Richard Condit, et al. 2012. “The Variation of Tree Beta Diversity Across a Global
470 Network of Forest Plots.” *Global Ecology and Biogeography* 21 (12): 1191–1202. <https://doi.org/10.1111/j.1466-8238.2012.00770.x>.

472 Deus, Filipe Ferreira de, Karl-L. Schuchmann, Julia Arieira, Ana Silvia de Oliveira Tissiani,
473 and Marinêz Isaac Marques. 2020. “Avian Beta Diversity in a Neotropical Wetland: The
474 Effects of Flooding and Vegetation Structure.” *Wetlands* 40 (5): 1513–27. <https://doi.org/10.1007/s13157-019-01240-0>.

476 Dray, Stéphane, David Bauman, Guillaume Blanchet, Daniel Borcard, Sylvie Clappe, Guillaume
477 Guenard, Thibaut Jombart, et al. 2021. “Adespatial: Multivariate Multiscale Spatial Analy-
478 sis.” <https://CRAN.R-project.org/package=adespatial>.

479 Dubois, Raphaëlle, Raphaël Proulx, and Stéphanie Pellerin. 2020. “Ecological Uniqueness of
480 Plant Communities as a Conservation Criterion in Lake-Edge Wetlands.” *Biological Con-
481 servation* 243: 108491. <https://doi.org/10.1016/j.biocon.2020.108491>.

482 Dubuis, Anne, Julien Pottier, Vanessa Rion, Loïc Pellissier, Jean-Paul Theurillat, and Antoine
483 Guisan. 2011. “Predicting Spatial Patterns of Plant Species Richness: A Comparison of
484 Direct Macroecological and Species Stacking Modelling Approaches.” *Diversity and Dis-
485 tributions* 17 (6): 1122–31. <https://doi.org/10.1111/j.1472-4642.2011.00792.x>.

486 eBird Basic Dataset. 2019. *Version: EBD_relJun-2019*. Cornell Lab of Ornithology, Ithaca,
487 NY, USA.

488 Ferrier, Simon, and Antoine Guisan. 2006. “Spatial Modelling of Biodiversity at the Commu-
489 nity Level.” *Journal of Applied Ecology* 43 (3): 393–404. <https://doi.org/10.1111/j.1365-2664.2006.01149.x>.

490

491 Fick, Stephen E., and Robert J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution
492 Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37 (12):
493 4302–15. <https://doi.org/10.1002/joc.5086>.

494 GDAL/OGR contributors. 2021. *GDAL/OGR Geospatial Data Abstraction Software Library*.
495 Manual. Open Source Geospatial Foundation. <https://gdal.org>.

496 Guisan, Antoine, and Carsten Rahbek. 2011. "SESAM a New Framework Integrating Macroecological and Species Distribution Models for Predicting Spatio-Temporal Patterns of Species Assemblages." *Journal of Biogeography* 38 (8): 1433–44. <https://doi.org/10.1111/j.1365-2699.2011.02550.x>.

500 Guisan, Antoine, and Wilfried Thuiller. 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>.

503 Heino, Jani, and Janne Alahuhta. 2019. "Knitting Patterns of Biodiversity, Range Size and Body Size in Aquatic Beetle Faunas: Significant Relationships but Slightly Divergent Drivers." *Ecological Entomology* 44 (3): 413–24. <https://doi.org/10.1111/een.12717>.

506 Heino, Jani, Luis Mauricio Bini, Johan Andersson, Johannes Bergsten, Ulf Bjelke, and Frank Johansson. 2017. "Unravelling the Correlates of Species Richness and Ecological Uniqueness in a Metacommunity of Urban Pond Insects." *Ecological Indicators* 73: 422–31. <https://doi.org/10.1016/j.ecolind.2016.10.006>.

510 Heino, Jani, and Mira Grönroos. 2017. "Exploring Species and Site Contributions to Beta Diversity in Stream Insect Assemblages." *Oecologia* 183 (1): 151–60. <https://doi.org/10.1007/s00442-016-3754-7>.

513 Heino, Jani, Adriano S. Melo, Luis Mauricio Bini, Florian Altermatt, Salman A. Al-Shami, David G. Angeler, Núria Bonada, et al. 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): 1235–48. <https://doi.org/10.1002/ece3.1439>.

518 Hurlbert, Allen H., and Walter Jetz. 2007. "Species Richness, Hotspots, and the Scale Depen-
519 dence of Range Maps in Ecology and Conservation." *Proceedings of the National Academy*
520 *of Sciences* 104 (33): 13384–89. <https://doi.org/10.1073/pnas.0704469104>.

521 Johnston, A., W. M. Hochachka, M. E. Strimas-Mackey, V. Ruiz Gutierrez, O. J. Robinson, E.
522 T. Miller, T. Auer, S. T. Kelling, and D. Fink. 2020. "Analytical Guidelines to Increase the
523 Value of Citizen Science Data: Using eBird Data to Estimate Species Occurrence." *bioRxiv*,
524 574392. <https://doi.org/10.1101/574392>.

525 Kong, Heng, Mathieu Chevalier, Pascal Laffaille, and Sovan Lek. 2017. "Spatio-Temporal Vari-
526 ation of Fish Taxonomic Composition in a South-East Asian Flood-Pulse System." *PLOS*
527 *ONE* 12 (3): e0174582. <https://doi.org/10.1371/journal.pone.0174582>.

528 Konowalik, Kamil, and Agata Nosol. 2021. "Evaluation Metrics and Validation of Presence-
529 Only Species Distribution Models Based on Distributional Maps with Varying Coverage."
530 *Scientific Reports* 11 (1): 1482. <https://doi.org/10.1038/s41598-020-80062-1>.

531 Landeiro, Victor Lemes, Bárbarah Franz, Jani Heino, Tadeu Siqueira, and Luis Mauricio Bini.
532 2018. "Species-Poor and Low-Lying Sites Are More Ecologically Unique in a Hyperdiverse
533 Amazon Region: Evidence from Multiple Taxonomic Groups." *Diversity and Distributions*
534 24 (7): 966–77. <https://doi.org/10.1111/ddi.12734>.

535 Legendre, Pierre, Daniel Borcard, and Pedro R. Peres-Neto. 2005. "Analyzing Beta Diversity:
536 Partitioning the Spatial Variation of Community Composition Data." *Ecological Mono-*
537 *graphs* 75 (4): 435–50. <https://doi.org/10.1890/05-0549>.

538 Legendre, Pierre, and Richard Condit. 2019. "Spatial and Temporal Analysis of Beta Diversity
539 in the Barro Colorado Island Forest Dynamics Plot, Panama." *Forest Ecosystems* 6 (1): 7.
540 <https://doi.org/10.1186/s40663-019-0164-4>.

541 Legendre, Pierre, and Miquel De Cáceres. 2013. "Beta Diversity as the Variance of Community
542 Data: Dissimilarity Coefficients and Partitioning." *Ecology Letters* 16 (8): 951–63. <https://doi.org/10.1111/ele.12141>.

543 <https://doi.org/10.1111/ele.12141>.

544 Legendre, Pierre, and Marie-Josée Fortin. 1989. "Spatial Pattern and Ecological Analysis."
545 *Vegetatio* 80 (2): 107–38. <https://doi.org/10.1007/BF00048036>.

546 Niskanen, Annina K. J., Risto K. Heikkinen, Henry Väre, and Miska Luoto. 2017. “Drivers of
547 High-Latitude Plant Diversity Hotspots and Their Congruence.” *Biological Conservation*
548 212: 288–99. <https://doi.org/10.1016/j.biocon.2017.06.019>.

549 Oksanen, Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan
550 McGlinn, Peter R. Minchin, et al. 2019. “Vegan: Community Ecology Package.” <https://CRAN.R-project.org/package=vegan>.

552 Ovaskainen, Otso, Gleb Tikhonov, Anna Norberg, F. Guillaume Blanchet, Leo Duan, David
553 Dunson, Tomas Roslin, and Nerea Abrego. 2017. “How to Make More Out of Community
554 Data? A Conceptual Framework and Its Implementation as Models and Software.” *Ecology*
555 *Letters* 20 (5): 561–76. <https://doi.org/10.1111/ele.12757>.

556 Poisot, Timothée, Cynthia Guéveneux-Julien, Marie-Josée Fortin, Dominique Gravel, and Pierre
557 Legendre. 2017. “Hosts, Parasites and Their Interactions Respond to Different Climatic
558 Variables.” *Global Ecology and Biogeography* 26 (8): 942–51. <https://doi.org/10.1111/geb.12602>.

560 Poisot, Timothée, Richard LaBrie, Erin Larson, Anastasia Rahlin, and Benno I. Simmons. 2019.
561 “Data-Based, Synthesis-Driven: Setting the Agenda for Computational Ecology.” *Ideas in*
562 *Ecology and Evolution* 12. <https://doi.org/10.24908/iee.2019.12.2.e>.

563 Pollock, Laura J., Reid Tingley, William K. Morris, Nick Golding, Robert B. O’Hara, Kirsten M.
564 Parris, Peter A. Vesk, and Michael A. McCarthy. 2014. “Understanding Co-Occurrence by
565 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>.

568 Qiao, Xiujuan, Qianxi Li, Qinghu Jiang, Junmeng Lu, Scott Franklin, Zhiyao Tang, Qinggang
569 Wang, et al. 2015. “Beta Diversity Determinants in Badagongshan, a Subtropical Forest in
570 Central China.” *Scientific Reports* 5 (1): 17043. <https://doi.org/10.1038/srep17043>.

571 R Core Team. 2021. “R: A Language and Environment for Statistical Computing.” Vienna,
572 Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

573 Silva, Pedro Giovâni da, Juliano André Bogoni, and Jani Heino. 2020. "Can Taxonomic and
574 Functional Metrics Explain Variation in the Ecological Uniqueness of Ecologically-Associated
575 Animal Groups in a Modified Rainforest?" *Science of The Total Environment* 708: 135171.
576 <https://doi.org/10.1016/j.scitotenv.2019.135171>.

577 Silva, Pedro Giovâni da, and Malva Isabel Medina Hernández. 2014. "Local and Regional
578 Effects on Community Structure of Dung Beetles in a Mainland-Island Scenario." *PLOS*
579 *ONE* 9 (10): e111883. <https://doi.org/10.1371/journal.pone.0111883>.

580 Silva, Pedro Giovâni da, Malva Isabel Medina Hernández, and Jani Heino. 2018. "Disentangling
581 the Correlates of Species and Site Contributions to Beta Diversity in Dung Beetle Assem-
582 blages." *Diversity and Distributions* 24 (11): 1674–86. [https://doi.org/10.1111/ddi.](https://doi.org/10.1111/ddi.12785)
583 [12785](https://doi.org/10.1111/ddi.12785).

584 Sor, Ratha, Pierre Legendre, and Sovan Lek. 2018. "Uniqueness of Sampling Site Contributions
585 to the Total Variance of Macroinvertebrate Communities in the Lower Mekong Basin." *Eco-*
586 *logical Indicators* 84: 425–32. <https://doi.org/10.1016/j.ecolind.2017.08.038>.

587 Staniczenko, Phillip P. A., Prabu Sivasubramaniam, K. Blake Suttle, and Richard G. Pearson.
588 2017. "Linking Macroecology and Community Ecology: Refining Predictions of Species
589 Distributions Using Biotic Interaction Networks." *Ecology Letters* 20 (6): 693–707. <https://doi.org/10.1111/ele.12770>.

591 Strimas-Mackey, Matthew, Eliot Miller, and Wesley Hochachka. 2018. "Auk: eBird Data Ex-
592 traction and Processing with AWK." [https://cornelllabofornithology.github.io/](https://cornelllabofornithology.github.io/auk/)
593 [auk/](https://cornelllabofornithology.github.io/auk/).

594 Sullivan, Brian L., Christopher L. Wood, Marshall J. Iliff, Rick E. Bonney, Daniel Fink, and
595 Steve Kelling. 2009. "eBird: A Citizen-Based Bird Observation Network in the Biological
596 Sciences." *Biological Conservation* 142 (10): 2282–92. [https://doi.org/10.1016/j.](https://doi.org/10.1016/j.biocon.2009.05.006)
597 [biocon.2009.05.006](https://doi.org/10.1016/j.biocon.2009.05.006).

598 Tan, Lingzhao, Chunyu Fan, Chunyu Zhang, Klaus von Gadow, and Xiuhua Fan. 2017. "How
599 Beta Diversity and the Underlying Causes Vary with Sampling Scales in the Changbai Moun-
600 tain Forests." *Ecology and Evolution* 7 (23): 10116–23. <https://doi.org/10.1002/>

601 [ece3.3493](#).

602 Tan, Lingzhao, Chunyu Fan, Chunyu Zhang, and Xiuhai Zhao. 2019. “Understanding and
603 Protecting Forest Biodiversity in Relation to Species and Local Contributions to Beta Di-
604 versity.” *European Journal of Forest Research* 138 (6): 1005–13. [https://doi.org/10.](https://doi.org/10.1007/s10342-019-01220-3)
605 [1007/s10342-019-01220-3](https://doi.org/10.1007/s10342-019-01220-3).

606 Taranu, Zofia E., Bernadette Pinel-Alloul, and Pierre Legendre. 2020. “Large-Scale Multi-
607 Trophic Co-Response Models and Environmental Control of Pelagic Food Webs in Québec
608 Lakes.” *Oikos* n/a (n/a). <https://doi.org/10.1111/oik.07685>.

609 Teittinen, Anette, Jianjun Wang, Simon Strömgård, and Janne Soininen. 2017. “Local and
610 Geographical Factors Jointly Drive Elevational Patterns in Three Microbial Groups Across
611 Subarctic Ponds.” *Global Ecology and Biogeography* 26 (8): 973–82. [https://doi.org/](https://doi.org/10.1111/geb.12607)
612 [10.1111/geb.12607](https://doi.org/10.1111/geb.12607).

613 Tytar, V., and O. Baidashnikov. 2021. “Associations Between Habitat Quality and Body Size
614 in the Carpathian-Podolian Land Snail *Vestia turgida*: Species Distribution Model Selection
615 and Assessment of Performance.” *Zoodiversity* 55 (1): 25–40. [http://ojs.akademperiodyka.](http://ojs.akademperiodyka.org.ua/index.php/Zoodiversity/article/view/67)
616 [org.ua/index.php/Zoodiversity/article/view/67](http://ojs.akademperiodyka.org.ua/index.php/Zoodiversity/article/view/67).

617 Vallejos, R., F. Osorio, and M. Bevilacqua. 2020. *Spatial Relationships Between Two Geo-*
618 *referenced Variables: With Applications in r*. New York: Springer. [http://srb2gv.mat.](http://srb2gv.mat.utfsm.cl/)
619 [utfsm.cl/](http://srb2gv.mat.utfsm.cl/).

620 Vasconcelos, Tiago S., Bruno T. M. do Nascimento, and Vitor H. M. Prado. 2018. “Expected
621 Impacts of Climate Change Threaten the Anuran Diversity in the Brazilian Hotspots.” *Ecol-*
622 *ogy and Evolution* 8 (16): 7894–7906. <https://doi.org/10.1002/ece3.4357>.

623 Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth. New York:
624 Springer. <http://www.stats.ox.ac.uk/pub/MASS4/>.

625 Vilmi, Annika, Satu Maaria Karjalainen, and Jani Heino. 2017. “Ecological Uniqueness of
626 Stream and Lake Diatom Communities Shows Different Macroecological Patterns.” *Diver-*
627 *sity and Distributions* 23 (9): 1042–53. <https://doi.org/10.1111/ddi.12594>.

- 628 Yang, Jun, Frank A. La Sorte, Petr Pyšek, Pengbo Yan, David Nowak, and Joe McBride. 2015.
629 “The Compositional Similarity of Urban Forests Among the World’s Cities Is Scale Depen-
630 dent.” *Global Ecology and Biogeography* 24 (12): 1413–23. [https://doi.org/10.1111/
631 geb.12376](https://doi.org/10.1111/geb.12376).
- 632 Yao, Jie, Jihong Huang, Yi Ding, Yue Xu, Han Xu, and Runguo Zang. 2021. “Ecological
633 Uniqueness of Species Assemblages and Their Determinants in Forest Communities.” *Di-
634 versity and Distributions* 27 (3): 454–62. <https://doi.org/10.1111/ddi.13205>.
- 635 Zurell, Damaris, Niklaus E. Zimmermann, Helge Gross, Andri Baltensweiler, Thomas Sattler,
636 and Rafael O. Wüest. 2020. “Testing Species Assemblage Predictions from Stacked and
637 Joint Species Distribution Models.” *Journal of Biogeography* 47 (1): 101–13. [https://
638 doi.org/10.1111/jbi.13608](https://doi.org/10.1111/jbi.13608).

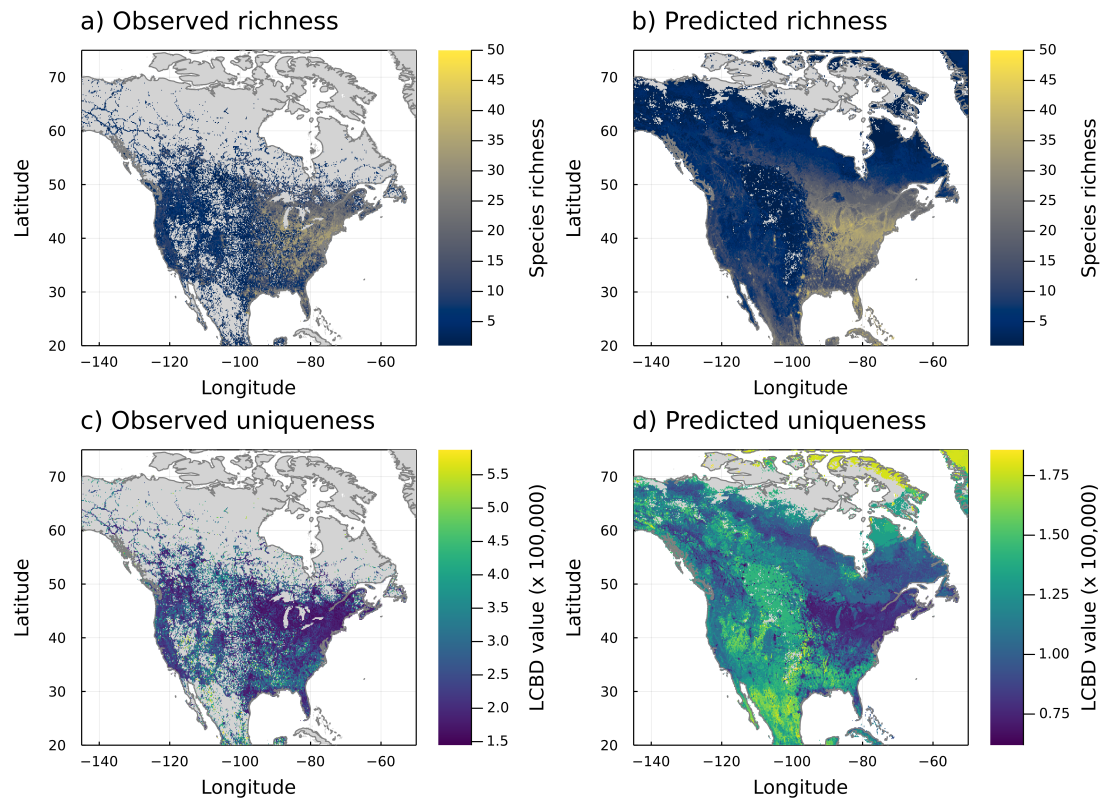


Figure 1: Comparison of species richness and LCBD scores from observed and predicted warbler occurrences in North America. Values were calculated for sites representing ten arcminute pixels. We measured species richness after converting the occurrence data from eBird (a) and the SDM predictions from our single-species BART models (b) to a presence-absence format per species. We applied the Hellinger transformation to the presence-absence data, then calculated the LCBD values from the variance of the community matrices separately for the occurrence data (c) and the SDM predictions (d). Areas in light grey (not on the colour scale) represent mainland sites with environmental data but without any warbler species present.

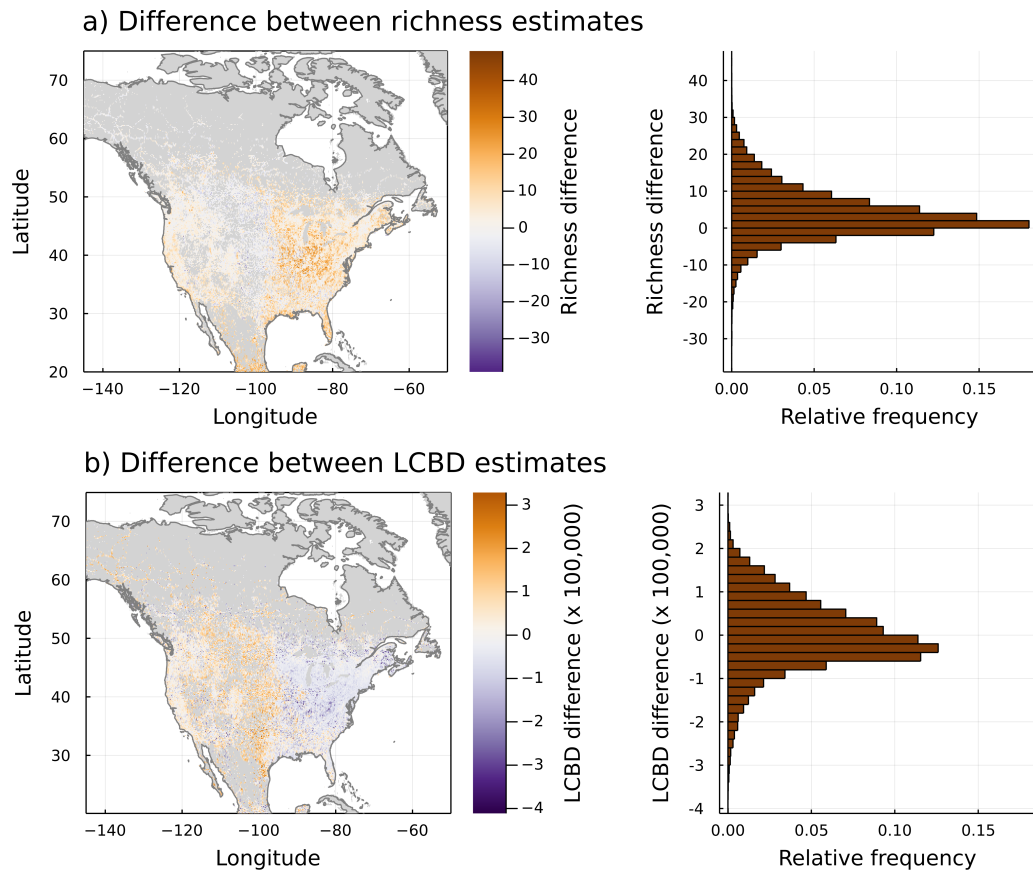


Figure 2: Comparison between observed and predicted estimates of species richness (a) and ecological uniqueness (b). The difference values represent the estimate from the predicted data set minus the estimate from the observed data set. LCBD values were recalculated for the same set of sites with observations in both data sets.

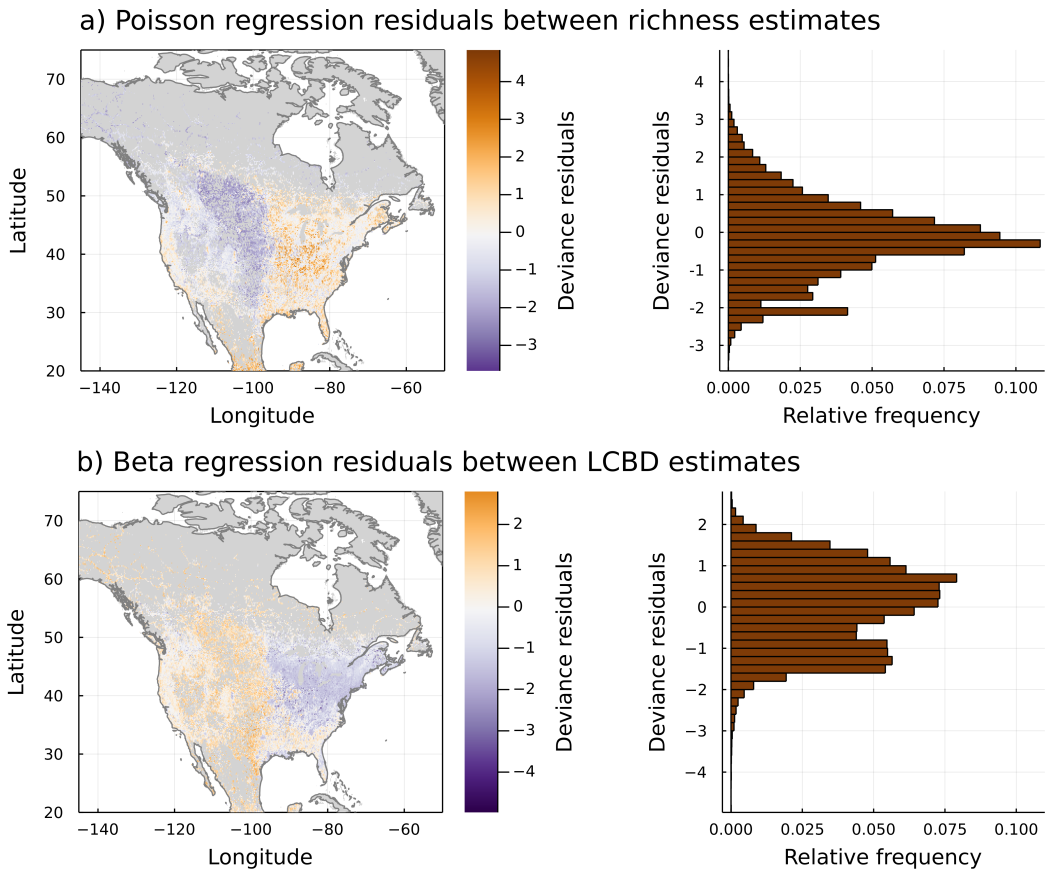


Figure 3: Comparison of the regression residuals between the observed and predicted estimates of species richness (a) and ecological uniqueness (b). The estimate from the predicted data set was used as the dependent variable and the estimate from the observed data set as the independent variable. A negative binomial regression with a log link function was used for species richness, and a beta regression with a logit link function was used for uniqueness. LCBD values were recalculated for the same set of sites with observations in both data sets.

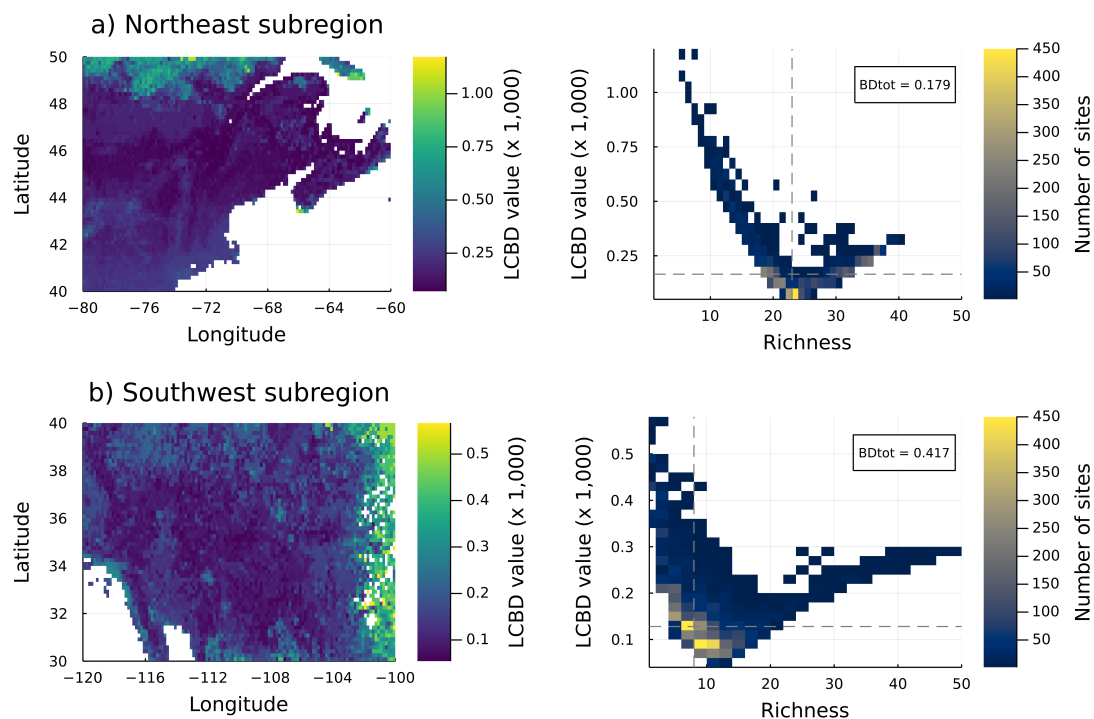


Figure 4: Comparison between a species-rich region (Northeast, a) and a species-poor one (Southwest, b) based on the SDM predictions for warbler species in North America. The left-side figures represent the LCBD scores for the assembled presence-absence predictions, calculated separately in each region. The colour scales are set to the respective range of LCBD scores to highlight the relative change within each region rather than compare the scores between both regions. The right-side 2-dimensional histograms represent the decreasing and slightly curvilinear relationship between LCBD values and species richness. The vertical and horizontal dashed lines respectively represent the median richness and LCBD value in each region, while BDotot represents the total beta diversity.

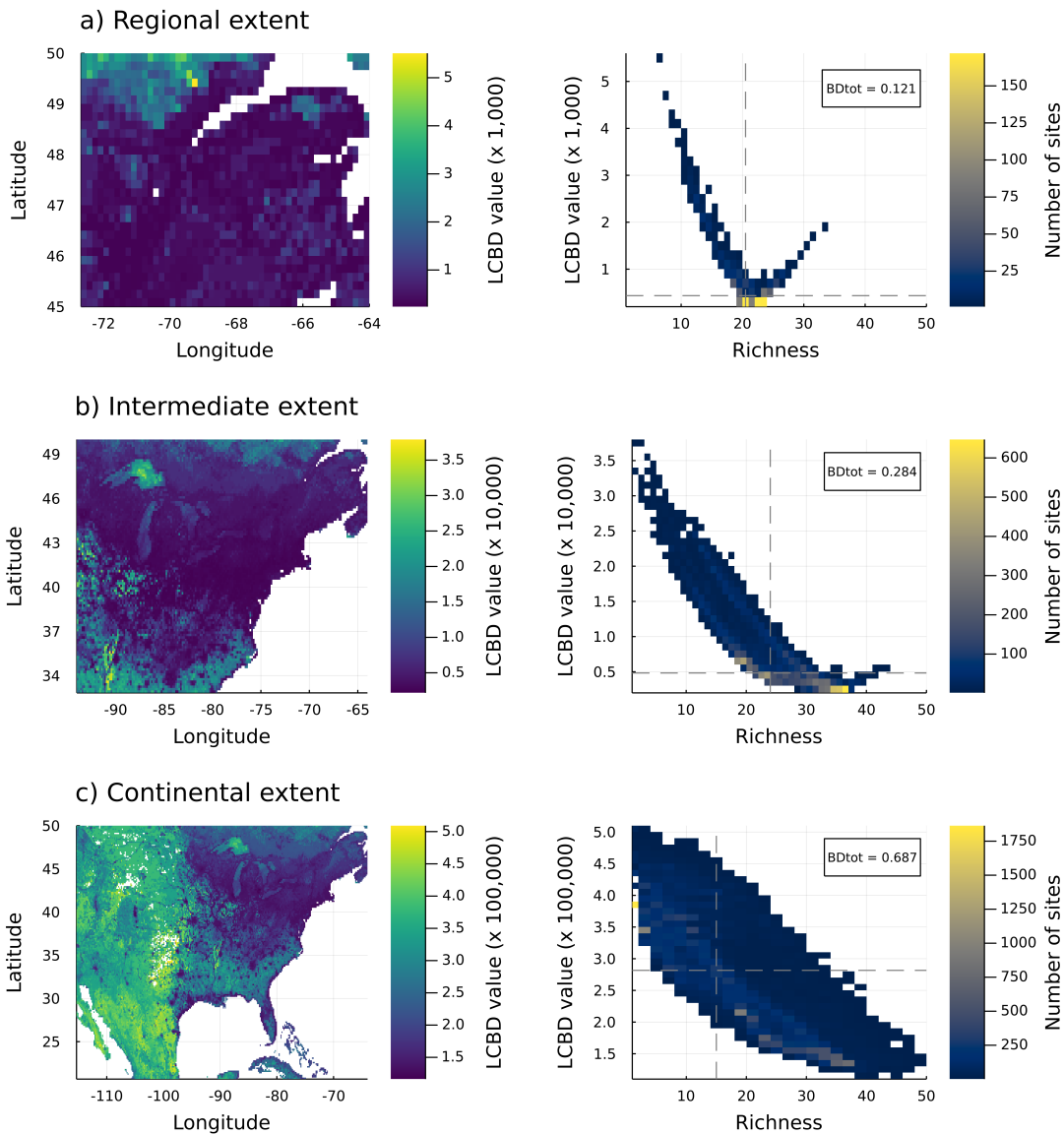


Figure 5: Effect of extent size on the relationship between site richness and LCB values based on the SDM predictions for warbler species in North America. The relationship progressively broadens and displays more variance when scaling up while total beta diversity increases. The LCB values were recalculated at each scale based on the sites in this region. The vertical and horizontal dashed lines respectively represent the median richness and LCB value in each region, while BDtot represents the total beta diversity.

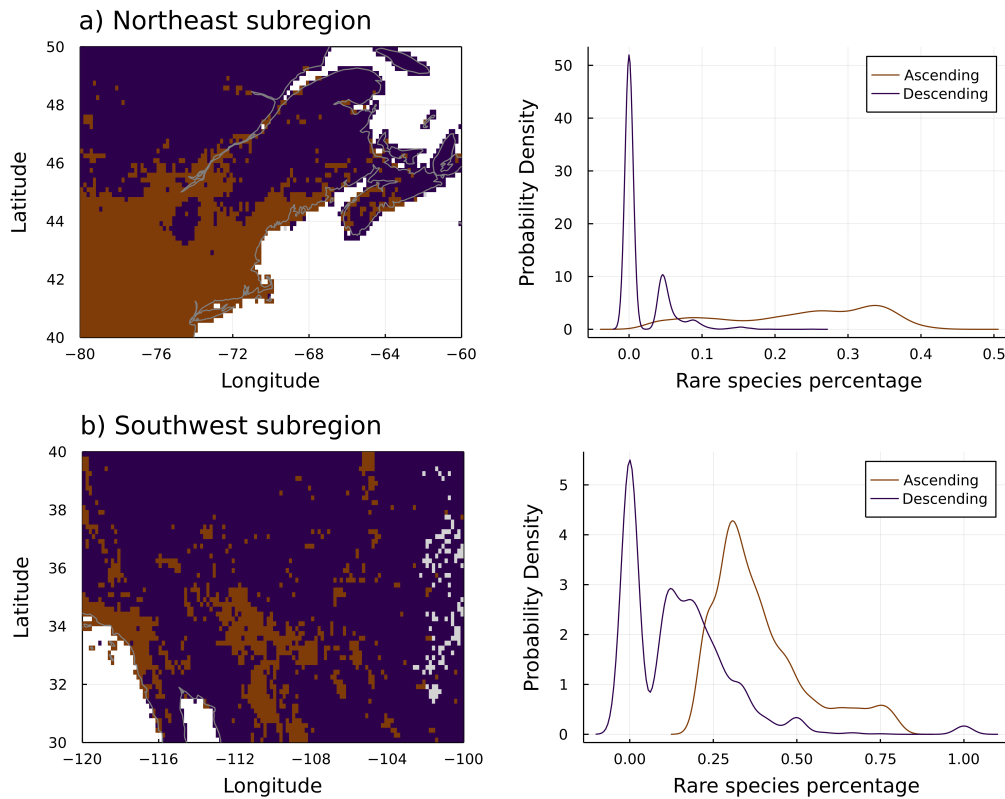


Figure 6: Proportion of rare species in the ascending and descending portions of the LCBD-richness relationship for the Northeast (a) and Southwest (b) subregions. The left side figures show the geographic distribution of the sites from each group. Sites were assigned to the ascending portion if their species richness was higher than the richness of the site with the lowest LCBD value, which corresponds to the inflection point of the right side figures of Fig. 4, and in the descending portion otherwise. The right side figures represent the kernel density estimation of the proportion of rare species in each group. Values on the y-axis are probability densities scaled so that the area under the curve equals one. Similarly, the area under the curve for a given range of values on the x-axis (proportions of rare species) represents the probability of observing a value in that range. Species were classified as rare when they occurred in fewer than 40% of the sites in the subregion. The proportion of rare species was then calculated for every site.