

1

Université de Montréal

2

Spatially continuous identification of beta diversity hotspots using species distribution models

3

4

By

5

Gabriel Dansereau

6

20147609

7

Département de sciences biologiques

8

Faculté des arts et des sciences

9

Advisory Committee Meeting

10

November 28, 2019

¹¹ **Contents**

¹²	Abstract	2
¹³	Introduction	3
¹⁴	Methods	5
¹⁵	1. Data Collection	5
¹⁶	2. Data Manipulation	6
¹⁷	3. SDM – The BIOCLIM Method	6
¹⁸	4. LCBD Calculation	7
¹⁹	5. Prediction Validity	8
²⁰	6. Alternative methods	9
²¹	7. Climate Change Scenarios and Temporal Beta Diversity	10
²²	Preliminary Results	10
²³	References	19

²⁴ **List of Tables**

²⁵	1 Distribution of the data in the eBird checklists for the countries used in the analyses	13
²⁷	2 WorldClim 2 climate variables used in the analyses	13

²⁸ **List of Figures**

²⁹	1 Single Species Distributions - Raw	15
³⁰	2 Single Species Distributions - SDM	15
³¹	3 Species Richness - Raw	16
³²	4 Species Richness - SDM	16
³³	5 LCBD values - Raw (transformed)	17
³⁴	6 LCBD values - SDM	17
³⁵	7 LCBD-richness relationship - Raw	18
³⁶	8 LCBD-richness relationship - SDM	19

³⁷ **Abstract**

³⁸ Beta diversity is an essential measure to describe the organization of biodiversity in space.
³⁹ The calculation of local contributions to beta diversity (LCBD), specifically, allows for the
⁴⁰ identification of sites with exceptional diversity within a region of interest, which is useful
⁴¹ for both community ecology and conservation purposes. However, beta diversity implies a
⁴² comparison among the sites of a given region, thus, its use is restricted to sites with known
⁴³ species composition, and to discontinuous spatial scales. We therefore propose a method
⁴⁴ to calculate LCBD indices on continuous scales for a whole region of interest, including
⁴⁵ unsampled sites. First, species distributions can be predicted on continuous scales using
⁴⁶ species distribution models (SDM). These models, such as the BIOCLIM method, use the
⁴⁷ environmental conditions at sampled sites to predict the presence or absence of each species at
⁴⁸ unsampled locations. Second, LCBD statistics can then be computed on the SDM predictions.
⁴⁹ We therefore show that it is possible to identify beta diversity hotspots on spatially continuous
⁵⁰ and extended scales. Our results confirm that LCBD values are related to species richness,
⁵¹ and that species-poor sites contribute most to beta diversity.

52 **Introduction**

53 Beta diversity, defined as the variation in species composition among sites in a geographic
54 region of interest (Legendre, Borcard, and Peres-Neto 2005), is an essential measure to
55 describe the organization of biodiversity in space. Total beta diversity within a community
56 can be partitioned into local contributions to beta diversity (LCBD) (Legendre and De Cáceres
57 2013), which allows for the identification of sites with exceptional species composition,
58 hence exceptional biodiversity. Such a method is useful for both community ecology and
59 conservation biology, as it highlights sites that are most important for their research or
60 conservation values. However, LCBD calculation methods require complete information
61 on community composition, such as a community composition matrix Y , thus they are
62 inappropriate for partially sampled or unsampled sites. To our knowledge, these methods
63 have mostly been applied on community data from sampled sites, hence on discontinuous
64 spatial scales, e.g. at intervals along a river stream (Legendre and De Cáceres 2013). This
65 raises the following questions: 1) could LCBD indices be extended to continuous spatial
66 scales, and 2) could this provide novel ecological insights in poorly sampled regions? We
67 aim to answer these questions by combining the LCBD calculation methods with predictive
68 biogeography approaches, and suggest that this would allow for the identification of hotspots
69 with high conservation value in poorly sampled regions.

70 Species distribution models (SDMs) already allow to make predictions on continuous spatial
71 scales, and these predictions could therefore be used to calculate LCBD indices. SDMs, also
72 known as bioclimatic envelope models (Araújo and Peterson 2012), aim to predict species
73 presence or absence based on previous observations of occurrence, and the environmental
74 conditions at which these were made (Poisot et al. 2019). Examples of uses include climate
75 change impact and invasion risk assessment, reserve selection and design, and discovery
76 of new populations (Araújo and Peterson 2012). This way, they generate novel ecological
77 insights for unsampled or lesser-known locations (Poisot et al. 2019), an approach yet to
78 be applied to the LCBD framework. We believe that a predictive approach such as this one
79 would bring a new perspective to biodiversity study and community ecology. By using SDMs,
80 we would be able to expand community information already available, and thus work on a
81 much larger community matrices than in typical LCBD studies, which might highlight new

82 diversity hotspots.

83 Climate and biodiversity data on extended spatial scales are increasingly available online.
84 For instance, the Worldclim 2.0 database (Fick and Hijmans 2017) provides interpolated
85 climate data for global land areas at very high spatial resolution, and the eBird platform
86 (Sullivan et al. 2009) provides a rapidly growing, citizen-contributed database of worldwide
87 bird observations. Both of these are commonly used in SDMs, and offer relevant information
88 on extended spatial scales. Therefore, we believe that these datasets could be used to predict
89 community composition and calculate LCBD indices on continuous spatial scales, and that
90 the result would be representative of the true community structure.

91 The predictive approach we suggest would be especially useful in poorly sampled regions, or
92 in regions with only sparse sampling. While it does not replace a full sampling within the
93 community, predictions and exploratory analyses do provide relevant ecological insights that
94 could be used in different ways. For instance, our method could help identify unsampled
95 sites with potential conservation value which should be targeted as soon as possible in future
96 studies. We also believe that it method could also be combined with IPCC climate change
97 scenarios, which provide projections for climate variables, in a way that would allow us to
98 model beta diversity changes with climate change and to identify the sites where the changes
99 in the community will be most important. Once again, the method would therefore prove
100 very relevant in an informative approach to suggest sites to prioritize for future conservation
101 and more structured research.

102 In this document, we cover in more details the methods that we suggest for this M.Sc. research
103 project. The preparation part of the project, including data collection and manipulation, has
104 already been done, and a workflow for the analyses, including code implementation, has been
105 defined as well. We also detail preliminary analyses and results intended as proof-of-concept
106 for the approach, which of course needs to be refined. Finally, we discuss methods that we
107 intend to use in future analyses, and whose feasibility is not as clearly stated.

108 **Methods**

109 **1. Data Collection**

110 We decided to focus our analyses on bird species and collected the data available on eBird for
111 the Warblers family. The complete database contains nearly 600 million observations. We
112 chose to focus specifically on the Warblers family, as it is a diverse group, popular among
113 birders, with over 30 million observations. Global citizen-contributed databases often present
114 additional challenges compared to conventional datasets due to their lack of structure, as well
115 as spatial and taxonomic biases (Johnston et al. 2019). For instance, there was a clear bias in
116 our data towards the United States, where there were much more observations and sampling
117 events (tbl. 1). However, eBird offers two advantages over other large scale datasets (Johnston
118 et al. 2019): 1) the data is structured as checklist and users can explicitly specify their
119 observations as “complete checklists” when all detected species were reported, which allows
120 to infer information on species absences, and 2) the dataset is semi-structured and checklists
121 are associated with metadata describing sampling effort, such as duration of search, distance
122 travelled and number of observers, which can be used as controls in the analyses. Hence,
123 model performance can be improved by inferring absences and subsampling checklists, while
124 spatial bias can be compensated by including effort covariates in the model. Therefore, we
125 believe the dataset can be appropriately used to achieve our objective of expanding measures
126 of exceptional biodiversity through space.

127 We collected the data available in the WorldClim 2 database (Fick and Hijmans 2017) for
128 North America, to which we decided to restrict our analyses. The WorldClim data consists of
129 spatially interpolated monthly climate data for global areas, available for resolutions from 10
130 arc-minutes to 30 arc-seconds (around 18 km² and 1 km² at the equator). Since the release of
131 the first version of the database in 2005 (Hijmans et al. 2005), it became the most common
132 source of climate data for SDM studies (Booth et al. 2014). The variables we used were
133 different measures of temperature and precipitation(tbl. 2), and they have been show to have
134 very high global cross-validation coefficients (> 0.99 and 0.86 respectively) (Fick and Hijmans
135 2017). We chose to use the coarser 10 arc-minutes resolution in our preliminary analyses,
136 as we believed it was sufficient for proof of concept of our method. However, Hijmans et
137 al. (2005) showed high within-grid cell variation in the 10 arc-minutes data, and therefore

138 recommended the use of the finer resolution, which hid less of the variation known to the
139 model. Given this, we might reconsider the resolution to use in our final analyses.

140 We chose to restrict our analyses to North America given the high amount of data available in
141 eBird. We believed it represented a suitable scale for our models, large enough to cover a lot
142 of variation in environmental variables and community structure, as well as phenomena
143 such as species migration. We also expected such extent of the spatial scale to cover for
144 imprecision in estimated species ranges.

145 **2. Data Manipulation**

146 WorldClim variables and eBird occurrence data are provided in different formats, so they
147 required some manipulations to be combined together. WorldClim variables are provided in
148 a 2-dimensional grid format, useful for large scale analyses and visualization, where each
149 cell or pixel has a size corresponding to the resolution of 10 arc-minutes. Each of the 19
150 variables forms a different grid. On the other hand, eBird records are occurrence-based, so
151 each entry in the dataset corresponds to an observation of a single species at a given time
152 and location. These entries can easily be matched to the 2D grid format of the WorldClim
153 variables through their spatial coordinates, which we found more useful for large scale
154 analyses and visualization. Hence, for each species, we matched all occurrences in eBird to
155 the grid format of the WorldClim variables, and then created a presence-absence community
156 matrix Y , taking all the grid cells as sites. At the 10 arc-minutes resolution, we obtained
157 39 024 sites with occurrences and 62 species in total. All data manipulations and further
158 analyses were realized in *Julia v1.2.0* (Bezanson et al. 2017) with the basic structure built
159 around the soon-to-be-released `SimpleSDMLayers.jl` package.

160 **3. SDM – The BIOCLIM Method**

161 We predicted species distributions using the BIOCLIM method (Nix 1986), a climate-envelope
162 model, considered a classic in the field. The method simply relates a species' distribution
163 to the ranges of bioclimatic variables at known locations (Booth et al. 2014). It has long
164 been outperformed by other methods (Elith et al. 2006), but it is still commonly used for
165 its simplistic approach and ease of understanding, as well as its simple relation to niche
166 theory (Booth et al. 2014; Hijmans et al. 2017). It is also primarily designed for presence-only

167 data. Despite that, we chose this method for our preliminary analyses as it was easier to
168 implement and because we believe it to be sufficient for proof-of-concept. We discuss possible
169 alternatives in the “Alternative Methods” section below.

170 The BIOCLIM method defines species potential ranges as a multidimensional environmental
171 hypervolume bounded by the minimum and maximum values for all occurrences (Franklin
172 2010). For each species, we established the percentile distribution of each environmental
173 variable at the known locations of occurrence (Hijmans et al. 2017). All sites were then
174 compared to those percentile distributions and given a score per variable according to
175 their ranking between 0.0 (1st percentile) and 1.0 (100th percentile). The median or 50th
176 percentile was considered the most suitable value of the variable, and values larger than
177 0.5 were subtracted from 1. Therefore, both tails are considered the same. The minimum
178 percentile score across all environmental variables was then selected as the predicted value
179 for each site. Values were multiplied by 2 and could therefore be interpreted as probabilities
180 of species occurrence (Hijmans et al. 2017). Predictions of 1 should be rare by definition,
181 as they require a perfectly median site on all variables, and values of 0 should be frequent,
182 since they occur whenever an environmental value is outside the range of the observed ones
183 (Hijmans et al. 2017).

184 The final step was to convert the probabilities into presence-absence data, so they could be
185 compared with the raw occurrence data. We transformed the probabilities into 0s and 1s
186 by converting all values greater than zero to one. Although it might tend to overestimate
187 species ranges, such a transformation is common in SDMs and can be accounted for during
188 result validation with specific methods (Franklin 2010). We also considered applying a
189 threshold, which would be determined by sensitivity analysis. In any case, converting into
190 presence-absence data allowed easier calculation of the richness and beta diversity metric.

191 4. LCBD Calculation

192 We calculated the LCBD statistics through the total variance of the matrix Y for both the raw
193 data and SDM predictions. Legendre and De Cáceres (2013) showed that LCBD coefficients
194 can be calculated directly through the total variance of matrix Y , or through a matrix of
195 dissimilarities among sampling units. We chose the first approach as it also allows to compute

196 species contributions to beta diversity (SCBD), which could also prove useful for conservation
197 purposes, but we did not investigate these for now. Before computing the LCBD statistics,
198 the presence-absence matrix Y had to be transformed in an appropriate way (Legendre and
199 De Cáceres 2013). We chose to apply the Hellinger transformation to the raw data and no
200 transformation on the SDM predictions for now, although we did not investigate these in
201 detail. The most appropriate transformation still needs to be determined, especially for
202 the SDM predictions. We then computed a matrix S of squared deviations from column
203 means and summed all the values of S to obtain the total sum of squares (SS) of the species
204 composition data (Legendre and De Cáceres 2013). LCBD coefficients are then computed as
205 $LCBD * i = SS_i / SS * Total$, where SS_i is the sum of squares of a sampling unit i . Finally, since
206 our matrix Y is very large, the LCBD coefficients are very small, so we scaled them to the
207 maximum value observed.

208 **5. Prediction Validity**

209 The exact way of testing the validity of the predictions remains to be determined, and will
210 also depend on the exact methods used to make the SDM predictions. A key element to note
211 is that both SDM predictions and LCBD values will have to be validated, hence they might
212 require different methods. Metrics that measure the accuracy of categorical or probabilistic
213 predictions in SDMs are well documented, in various forms. Some require absence data to test
214 against, and can be used on probabilistic predictions directly (area-under-curve, AUC) or after
215 a conversion of the predictions to binary presence-absence using a certain threshold (Kappa
216 index, measuring the difference between observed and chance agreement in a confusion
217 matrix) (Franklin 2010). Other methods are appropriate for presence-only data, such as the
218 Boyce Index. In any case, measuring prediction error is only one part of validation, and
219 finding appropriate data for evaluation is almost as essential (Franklin 2010), especially
220 since we also aim to describe community structure. Separating the data into a training and
221 testing dataset, with 70% and 30% of the observations for instance, is a possible approach
222 common in machine learning methods, although all of the available observations might be
223 needed in some cases (Franklin 2010). An interesting approach, suggested by Elith et al.
224 (2006) for SDMs, would be to find independent, well-structured presence-absence datasets
225 for validation, on which both SDM predictions and beta diversity metrics could be tested.

226 This approach has the advantage that the testing data is truly independent of the training
227 one, hence it could be used with certain tests of significance. Although it might not cover the
228 entire extent of the predictions in a single test, this method brings a closer comparison to the
229 way LCBD metrics are used in most studies, and it would provide interesting perspectives if
230 combined with other, full-extent validation methods.

231 **6. Alternative methods**

232 Many methods generally outperform BIOCLIM for the predictions, as shown by Elith et al.
233 (2006). In our case, better predictions will come by two different means: 1) approaches that
234 are better than BIOCLIM to model the relationship between species presence-absence (or even
235 abundance) and environmental variables, and 2) approaches that account for other drivers of
236 species distributions, such as ecological interactions and species migration. Machine learning
237 methods, especially, would be also be interesting alternatives to consider. MAXENT (Phillips,
238 Anderson, and Schapire 2006), another presence-only method, has come to be one of the most
239 widely used methods in SDM studies, often with WorldClim variables (Booth et al. 2014).
240 Similarly, Random Forests are simple to put in place, take into account both presence and
241 absence data, allow for quantification of the variables importance in explaining variation, and
242 offer intrinsic testing metrics (Franklin 2010). However, while those methods might return
243 more accurate predictions, they do not implicitly model other drivers of species distribution,
244 among which species interactions and functional niche. Integrating those factors might
245 prove more difficult given our dataset and our focus Warblers species, as no appropriate
246 information on their interaction is available, to our knowledge. Joint species distribution
247 models (JSDMs) might be an interesting way to encompass those, as they attempt to model
248 species co-occurrence, rather than the distribution of single species (Pollock et al. 2014).
249 A different taxonomic group and dataset with more details on interactions could also be
250 used. On the other hand, finding a method that could be applied to any taxonomic group,
251 especially those well represented in large citizen-contributed datasets, would be most useful
252 for research and conservation purposes.

253 **7. Climate Change Scenarios and Temporal Beta Diversity**

254 We aim to apply our method to environmental conditions from IPCC climate change scenarios,
255 first to model community compositions after climate change on continuous scales through
256 SDMs, and then to identify the sites where the community has changed in the most exceptional
257 ways. This can be seen from the variation in LCBD values, but also through temporal beta
258 diversity indices (TBI) (Legendre 2019). TBI indices allow to study changes in community
259 composition through time from repeated surveys at given sites. Whereas LCBD values
260 essentially measure the contribution to beta diversity of each site compared to all others,
261 TBI measure changes in community composition for a single site between two surveys, and
262 can also be decomposed into species losses and gains (Legendre 2019). Moreover, TBI can
263 be tested for significance using a permutation test. An approach similar to that of Legendre
264 and Condit (2019) would be most interesting to follow. First, they computed LCBD indices
265 and compared the sites that were significant for two surveys 30 years apart, highlighting
266 a swamp region where important changes seemed to have occurred. Then, they used TBI
267 indices to confirm the sites with significant changes, decompose these changes into losses and
268 gains, and identify the species that had changed the most. Such an approach could be highly
269 informative with our data, although the permutation tests and corrections to apply might
270 cause problems given the number of sites implied in our study.

271 The possibility of using climate change scenarios in the SDMs also needs to be assessed.
272 We did not try to download nor find the appropriate data for now. Interpolated variables
273 are sometimes different than those used in Worldclim 2.0. Therefore, the SDM models and
274 predictions might have to be different too, and potentially less reliable. Nonetheless, we
275 believe it will be possible to do some kind of time analysis linking beta diversity, climate
276 change and species distribution modelling, and that it could return highly informative results
277 for conservation purposes.

278 **Preliminary Results**

279 Our preliminary results mainly consisted of comparisons between the raw occurrence data
280 and the SDM predictions for the four following elements: single-species distribution (figs. 1,
281 2), species richness (figs. 3, 4), LCBD coefficients (figs. 5, 6), and the relationship between the

282 species richness and LCBD coefficients (figs. 7, 8). Two main results emerged from them: 1)
283 the models provided seemingly valid and relevant results for poorly sampled regions, both
284 expected species-poor and species-rich ones, and 2) there was an association between species
285 richness and LCBD coefficients partially confirming the relationship shown in other studies.

286 First, the example of the Yellow Warbler (*Setophaga petechia*), one of the most observed species,
287 showed that the single-species models predicted a broad distribution covering poorly sampled
288 areas, with notable patches of absence across the continent (figs. 1, 2). Likewise, species
289 richness, defined as the number of species present per site, showed a clear latitude gradient,
290 with the poorest sites to the North and the richest to the South (figs. 3, 4). A form of altitude
291 gradient could also be observed, with the Rockies and other mountains well delimited by their
292 lower values. In both cases, the results make intuitive sense and highlight the models ability
293 to predict species presence despite poor or no sampling. Mexico, for example, has much
294 sparser sampling and fewer observations, but the models predict Yellow Warblers presence in
295 most areas nonetheless, as well as higher species richness than on the highly sampled Atlantic
296 Coast, which make sense for a more southern location. We believe these to be valid insights
297 on poorly sampled locations, although we reckon that intuitive reasoning is not a proof of
298 anything, and that the model might be wrong in important ways. In any case, it highlights
299 the need for an appropriate method of validation, as well as a thoughtful consideration of
300 other factors such as species migration.

301 Second, our preliminary LCBD results seemed to confirm the association between species
302 richness and LCBD coefficients, while also being valid and insightful. They were however
303 harder to interpret given the use of the Hellinger transformation for the raw occurrence
304 data only. Raw occurrence data showed a negative relationship between species richness
305 and LCBD coefficients (fig. 7), as observed previously by Heino and Grönroos (2017), with
306 no clear geographic pattern (fig. 5). If anything, the highest values seemed to be at the
307 borders of the most sampled regions, where are located the sites in which fewer species
308 were observed (fig. 3). On the other hand, SDM predictions showed a quadratic form, with
309 the LCBD coefficients re-increasing after some threshold (fig. 8). The geographic pattern is
310 also clearer, with the highest values to the northern and southern extremes (fig. 6). Since
311 LCBD indices should highlight the most exceptional species composition, both species poor

312 or species rich, this result is no necessarily surprising. In a way, extremes and intermediate
313 values might be represented more evenly on the continuous scale of a SDM than in the raw
314 data. Although raw occurrence data results concorded with those of Heino and Grönroos
315 (2017), the different result with the SDM projections might show the importance of going
316 beyond occurrence data when using large but spatially biased databases such as eBird. Once
317 again, our results highlight the need for well-thought method and an investigation of the
318 appropriate transformation to use on the data.

319 Finally, one disappointing aspect of our method is that the result failed to identify patterns
320 on finer scales. The trends shown by the SDMs for both the species richness and LCBD
321 coefficients were large-scale, latitude-related patterns. Except for mountains, few exceptional
322 sites are actually shown in the middle of the landscape. While it might have been unrealistic
323 to expect such results from a coarse analysis like ours, it would be useful for conservation
324 purposes to be able the identify precise sites within smaller regions. This might be achieved
325 by using a finer resolution, which we should probably reconsider in light of these results, or by
326 using a different technique, such as training the models and predicting species distributions
327 on large scales, but computing and scaling LCBD values on finer local ones, which might
328 highlight regional differences in a new way.

Table 1: Distribution of the data in the eBird checklists for the countries used in the analyses

Country	Observations	Checklists	Species	Species per	Species per	Species per
				checklist	checklist	checklist
	(mean)	(median)	(maximum)			
US	19 206 453	7 840 526	56	2.450	2.0	34
CA	3 360 650	1 115 625	45	3.012	2.0	31
MX	407 227	147 599	61	2.759	2.0	21
Total	22 974 330	9 103 750	63	2.523	2.0	34

Table 2: WorldClim 2 climate variables used in the analyses

Variable	Description
1	Annual Mean Temperature
2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
3	Isothermality (BIO2/BIO7) (* 100)
4	Temperature Seasonality (standard deviation *100)
5	Max Temperature of Warmest Month
6	Min Temperature of Coldest Month
7	Temperature Annual Range (BIO5-BIO6)
8	Mean Temperature of Wettest Quarter
9	Mean Temperature of Driest Quarter
10	Mean Temperature of Warmest Quarter
11	Mean Temperature of Coldest Quarter
12	Annual Precipitation
13	Precipitation of Wettest Month
14	Precipitation of Driest Month
15	Precipitation Seasonality (Coefficient of Variation)
16	Precipitation of Wettest Quarter
17	Precipitation of Driest Quarter
18	Precipitation of Warmest Quarter

Variable	Description
19	Precipitation of Coldest Quarter

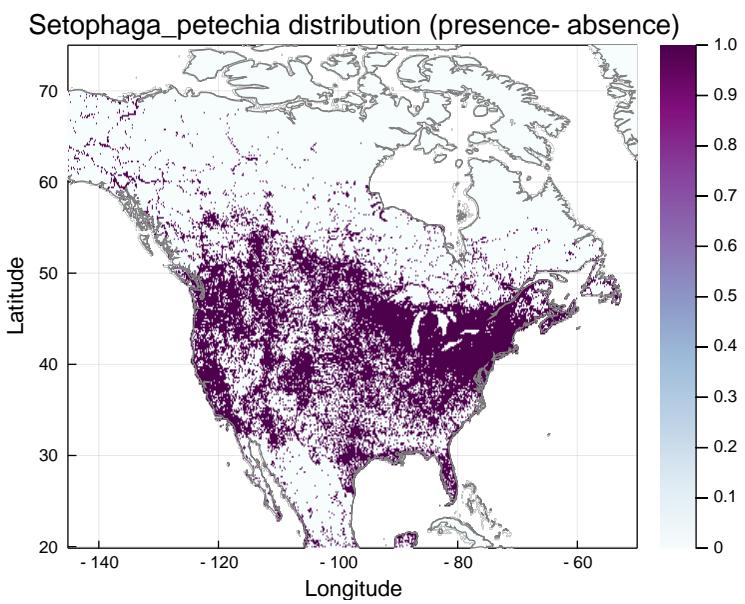


Figure 1: Single Species Distributions - Raw

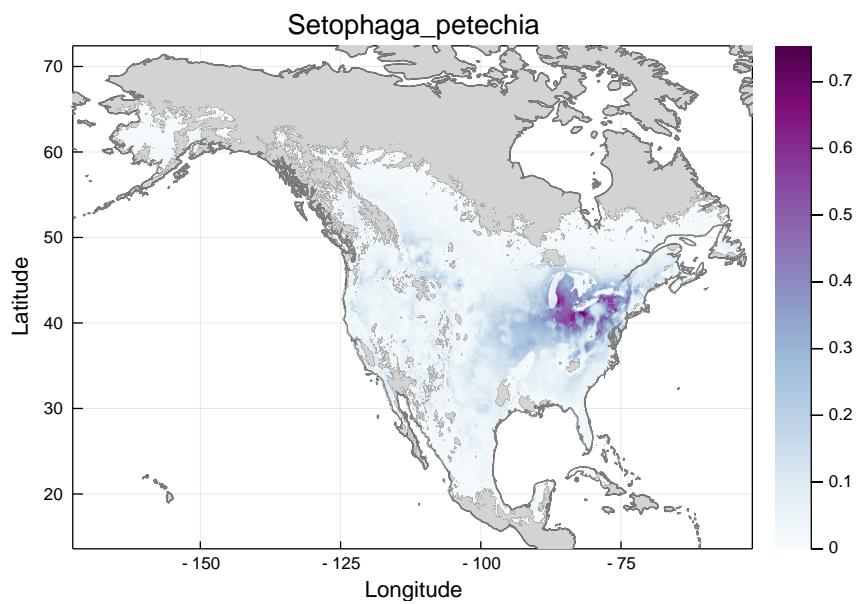


Figure 2: Single Species Distributions - SDM

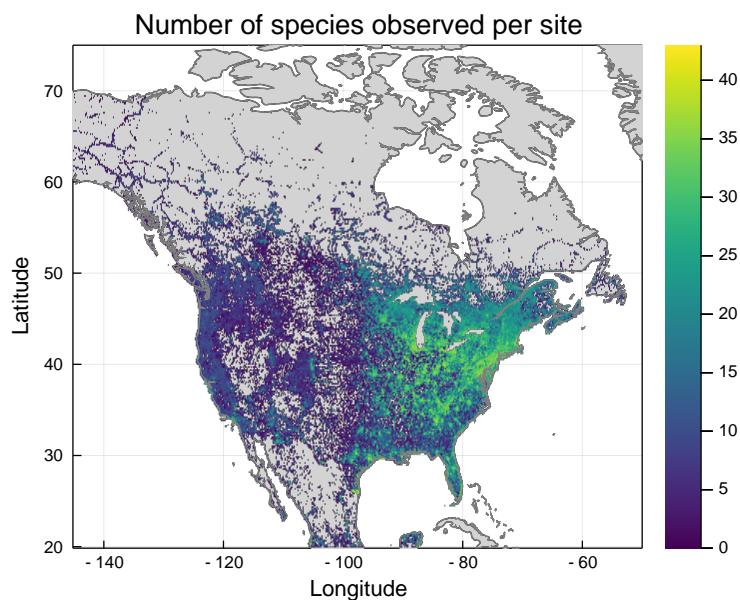


Figure 3: Species Richness - Raw

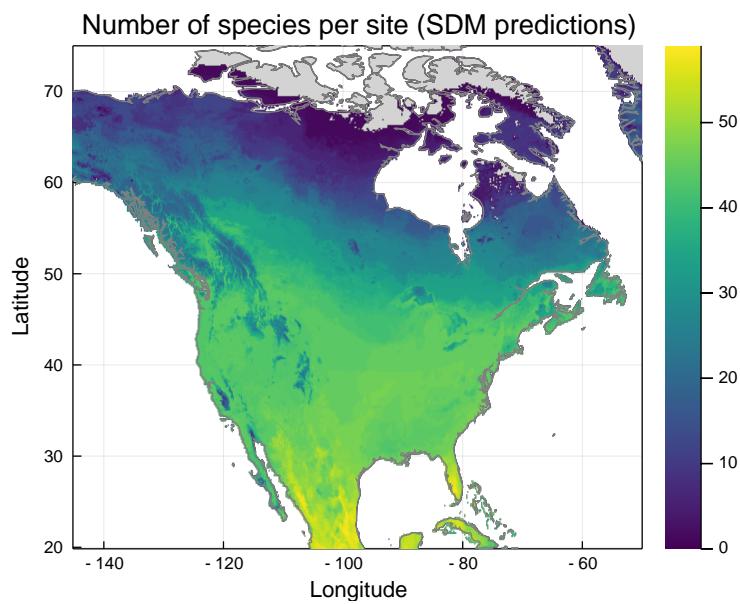


Figure 4: Species Richness - SDM

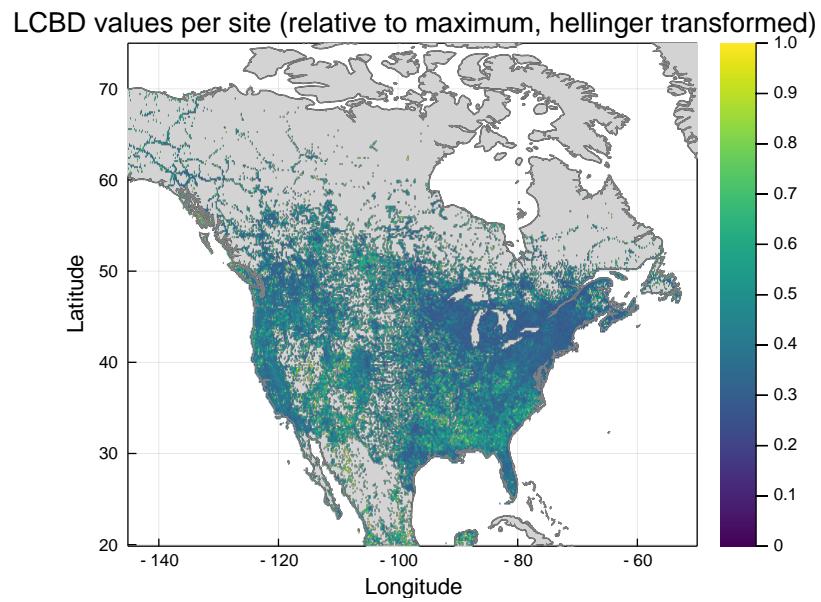


Figure 5: LCBD values - Raw (transformed)

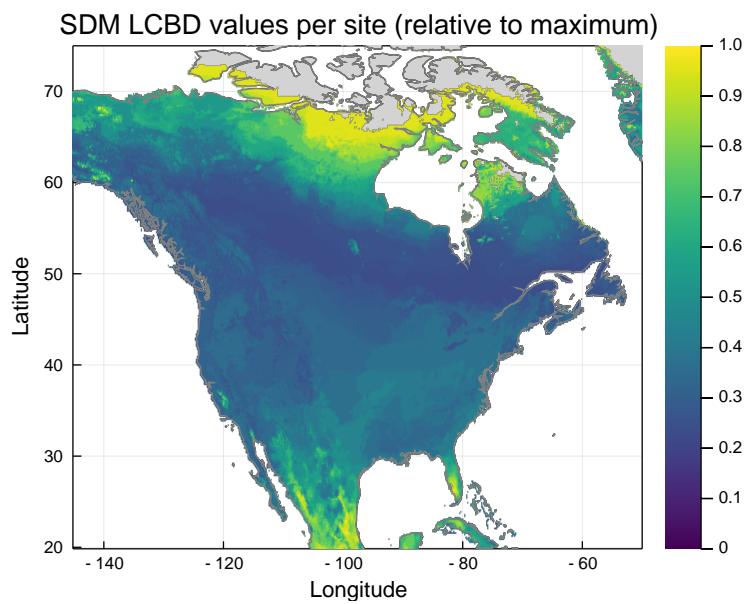


Figure 6: LCBD values - SDM

Relationship between LCBD (hellinger transformed) and species richness

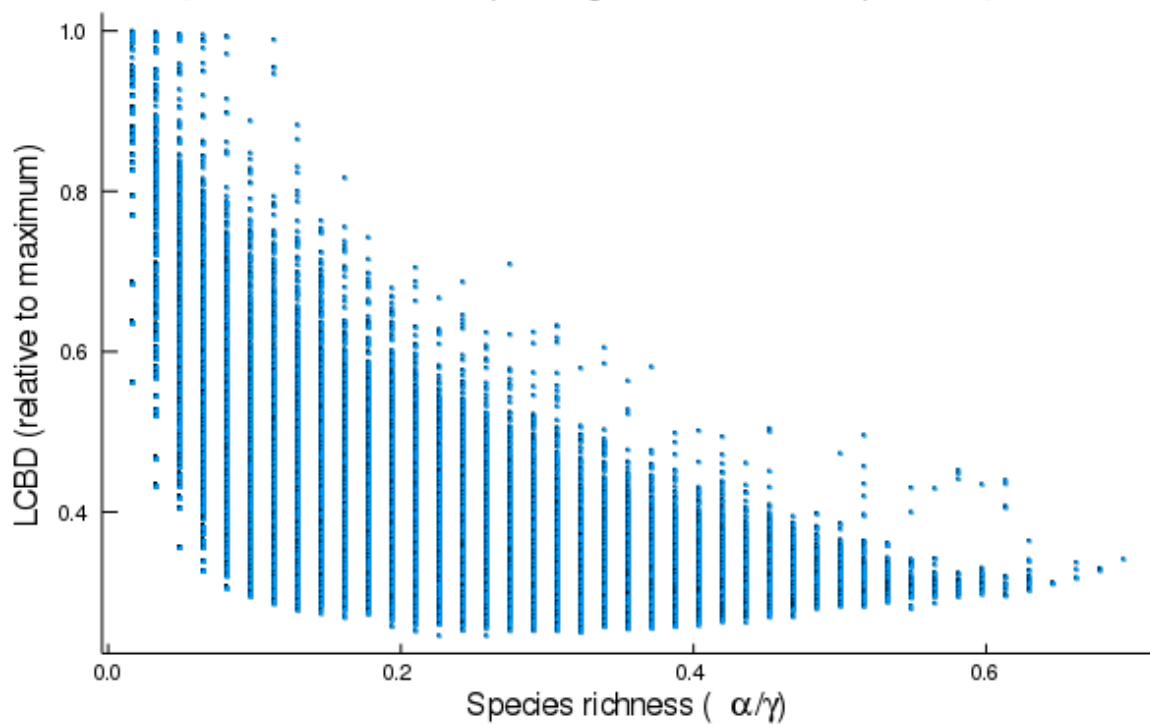


Figure 7: LCBD-richness relationship - Raw

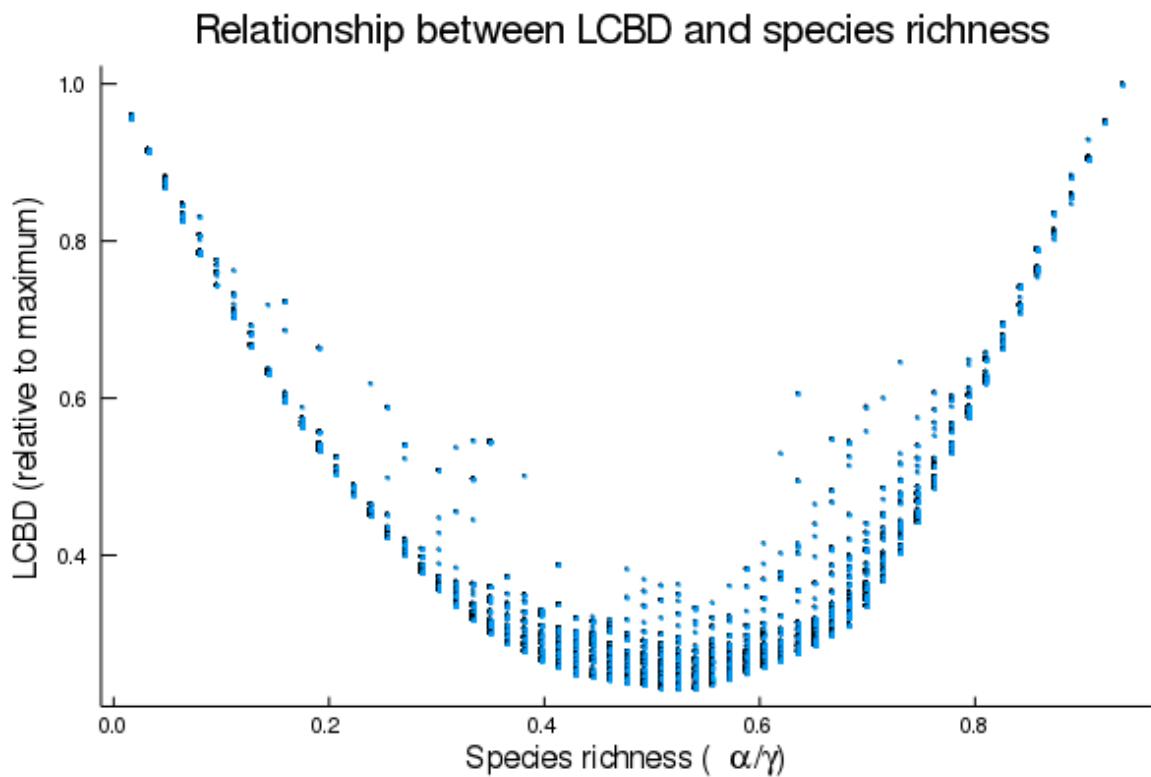


Figure 8: LCBD-richness relationship - SDM

329 References

- 330 Araújo, Miguel B., and A. Townsend Peterson. 2012. “Uses and Misuses of Bioclimatic
331 Envelope Modeling.” *Ecology* 93 (7): 1527–39. <https://doi.org/10.1890/11-1930.1>.
- 332 Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B. Shah. 2017. “Julia: A Fresh
333 Approach to Numerical Computing.” *SIAM Review* 59 (1): 65–98. <https://doi.org/10.1137/141000671>.
- 334
- 335 Booth, Trevor H., Henry A. Nix, John R. Busby, and Michael F. Hutchinson. 2014. “BIOCLIM:
336 The First Species Distribution Modelling Package, Its Early Applications and Relevance to
337 Most Current MaxEnt Studies.” *Diversity and Distributions* 20 (1): 1–9. <https://doi.org/10.1111/ddi.12144>.
- 338
- 339 Elith, Jane, Catherine H. Graham, Robert P. Anderson, Miroslav Dudík, Simon Ferrier, Antoine
340 Guisan, Robert J. Hijmans, et al. 2006. “Novel Methods Improve Prediction of Species’
341 Distributions from Occurrence Data.” *Ecography* 29 (2): 129–51. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>.
- 342

- 343 Fick, Stephen E., and Robert J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution
344 Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37 (12):
345 4302–15. <https://doi.org/10.1002/joc.5086>.
- 346 Franklin, Janet. 2010. "Moving Beyond Static Species Distribution Models in Support of
347 Conservation Biogeography: Moving Beyond Static Species Distribution Models." *Diversity*
348 and *Distributions* 16 (3): 321–30. <https://doi.org/10.1111/j.1472-4642.2010.00641.x>.
- 349 Heino, Jani, and Mira Grönroos. 2017. "Exploring Species and Site Contributions to Beta
350 Diversity in Stream Insect Assemblages." *Oecologia* 183 (1): 151–60. <https://doi.org/10.1007/s00442-016-3754-7>.
- 352 Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis. 2005.
353 "Very High Resolution Interpolated Climate Surfaces for Global Land Areas." *International*
354 *Journal of Climatology* 25 (15): 1965–78. <https://doi.org/10.1002/joc.1276>.
- 355 Hijmans, Robert J., Steven Phillips, John Leathwick, and Jane Elith. 2017. *Dismo: Species*
356 *Distribution Modeling*. <https://CRAN.R-project.org/package=dismo>.
- 357 Johnston, A., W. M. Hochachka, M. E. Strimas-Mackey, V. Ruiz Gutierrez, O. J. Robinson,
358 E. T. Miller, T. Auer, S. T. Kelling, and D. Fink. 2019. "Best Practices for Making Reli-
359 able Inferences from Citizen Science Data: Case Study Using eBird to Estimate Species
360 Distributions." *bioRxiv*, March, 574392. <https://doi.org/10.1101/574392>.
- 361 Legendre, Pierre. 2019. "A Temporal Beta-Diversity Index to Identify Sites That Have
362 Changed in Exceptional Ways in Space-Time Surveys." *Ecology and Evolution* 9 (6): 3500–
363 3514. <https://doi.org/10.1002/ece3.4984>.
- 364 Legendre, Pierre, Daniel Borcard, and Pedro R. Peres-Neto. 2005. "Analyzing Beta Diver-
365 sity: Partitioning the Spatial Variation of Community Composition Data." *Ecological*
366 *Monographs* 75 (4): 435–50. <https://doi.org/10.1890/05-0549>.
- 367 Legendre, Pierre, and Richard Condit. 2019. "Spatial and Temporal Analysis of Beta Diversity
368 in the Barro Colorado Island Forest Dynamics Plot, Panama." *Forest Ecosystems* 6 (1): 7.
369 <https://doi.org/10.1186/s40663-019-0164-4>.
- 370 Legendre, Pierre, and Miquel De Cáceres. 2013. "Beta Diversity as the Variance of Community

- 371 Data: Dissimilarity Coefficients and Partitioning." *Ecology Letters* 16 (8): 951–63. <https://doi.org/10.1111/ele.12141>.
- 372
- 373 Nix, Henry A. 1986. "A Biogeographic Analysis of Australian Elapid Snakes." *Atlas of Elapid*
- 374 *Snakes of Australia* 7: 4–15.
- 375 Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. "Maximum Entropy
- 376 Modeling of Species Geographic Distributions." *Ecological Modelling* 190 (3): 231–59.
- 377 <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- 378 Poisot, Timothée, Richard LaBrie, Erin Larson, Anastasia Rahlin, and Benno I. Simmons. 2019.
- 379 "Data-Based, Synthesis-Driven: Setting the Agenda for Computational Ecology." *Ideas in*
- 380 *Ecology and Evolution* 12 (July). <https://doi.org/10.24908/iee.2019.12.2.e>.
- 381 Pollock, Laura J., Reid Tingley, William K. Morris, Nick Golding, Robert B. O'Hara, Kirsten
- 382 M. Parris, Peter A. Vesk, and Michael A. McCarthy. 2014. "Understanding Co-Occurrence
- 383 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>.
- 386 Sullivan, Brian L., Christopher L. Wood, Marshall J. Iliff, Rick E. Bonney, Daniel Fink, and
- 387 Steve Kelling. 2009. "eBird: A Citizen-Based Bird Observation Network in the Biological
- 388 Sciences." *Biological Conservation* 142 (10): 2282–92. <https://doi.org/10.1016/j.biocon.2009.05.006>.