

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 27, 2019

¹¹ **Contents**

¹²	Abstract	2
¹³	Introduction	3
¹⁴	Methods	4
¹⁵	1. Data Collection	4
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	9
²²	Preliminary Results	10
²³	References	17

²⁴ **List of Tables**

²⁵	1 WorldClim 2 climate variables used in the analyses	5
---------------	--	---

²⁶ **List of Figures**

²⁷	1 Single Species Distributions - Raw	13
²⁸	2 Single Species Distributions - SDM	13
²⁹	3 Species Richness - Raw	14
³⁰	4 Species Richness - SDM	14
³¹	5 LCBD values - Raw (transformed)	15
³²	6 LCBD values - SDM	15
³³	7 LCBD-richness relationship - Raw	16
³⁴	8 LCBD-richness relationship - SDM	17

³⁵ **Abstract**

³⁶ Beta diversity is an essential measure to describe the organization of biodiversity in space.
³⁷ The calculation of local contributions to beta diversity (LCBD), specifically, allows for the
³⁸ identification of sites with exceptional diversity within a region of interest, which is useful
³⁹ for both community ecology and conservation purposes. However, beta diversity implies a
⁴⁰ comparison among the sites of a given region, thus, its use is restricted to sites with known
⁴¹ species composition, and to discontinuous spatial scales. We therefore propose a method
⁴² to calculate LCBD indices on continuous scales for a whole region of interest, including
⁴³ unsampled sites. First, species distributions can be predicted on continuous scales using
⁴⁴ species distribution models (SDM). These models, such as the BIOCLIM method, use the
⁴⁵ environmental conditions at sampled sites to predict the presence or absence of each species at
⁴⁶ unsampled locations. Second, LCBD statistics can then be computed on the SDM predictions.
⁴⁷ We therefore show that it is possible to identify beta diversity hotspots on spatially continuous
⁴⁸ and extended scales. Our results confirm that LCBD values are related to species richness,
⁴⁹ and that species-poor sites contribute most to beta diversity.

50 **Introduction**

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

68 Species distribution models (SDMs) already allow to make predictions on continuous spatial
69 scales which could be used to calculate LCBD indices. These methods, also known as
70 bioclimatic envelope models (Araújo and Peterson 2012), aim to predict species presence or
71 absence based on observation of occurrences at known locations (Poisot et al. 2019). This way,
72 they generate novel ecological insights, and represent an approach yet to be applied to LCBD.
73 We believe that such an approach of generating novel ecological insights for unsampled or
74 lesser-known locations could be an interesting new perspective in the study. Through them,
75 we would be able to expand community information already available, and thus work on a
76 much larger community matrix than in typical LCBD studies.

77 Appropriate data to expand measures of exceptional biodiversity through space is increasingly
78 available online. For instance, the Worldclim 2.0 database (Fick and Hijmans 2017) provides
79 interpolated climate data for global land areas at very high spatial resolution, and the eBird

80 platform (Sullivan et al. 2009) provides a growing citizen-contributed database of worldwide
81 bird observations. Both of these are commonly used in SDMs, and offer relevant information
82 on extended spatial scales. Hence, we believe that we could use them to predict community
83 composition and calculate LCBD indices on continuous spatial scales, and that the result
84 would be representative of the true community structure.

85 The predictive approach we suggest would be especially useful in poorly sampled regions,
86 or in regions with only sparse sampling. While it doesn't replace a full sampling within the
87 community, it does provide relevant ecological insights. For instance, the method could help
88 identify unsampled sites with potential conservation value which should be targeted as soon
89 as possible in future studies. We also believe that our method could also be combined with
90 IPCC climate change scenarios, which provide projections for climate variables, in a way that
91 would allow us to model beta diversity changes with climate change and to identify the sites
92 where the changes in the community will be most important. Again, this method would be
93 more relevant as an informative approach to suggest sites to prioritize for future conservation
94 and more structured research.

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

101 **Methods**

102 **1. Data Collection**

103 We decided to focus our analyses on bird species and collected the data available on eBird
104 for the Warblers family. The complete database contains nearly 600 million observations,
105 and presents two main advantages over other large scale datasets (Johnston et al. 2019): 1)
106 data is structured as checklist and users can explicitly specify their observations as "complete
107 checklists" when all detected species were reported, which allows to infer information on
108 species absences, 2) the dataset is semi-structured and checklists are associated with metadata

describing sampling effort, such as duration of search, distance travelled and number of observers, which can be used as controls in the analyses. We chose to focus specifically on the Warblers family, as it is a diverse group, popular among birders, with over 30 million observations.

We decided to restrict our analyses to North America and collected climate data available in the WorldClim 2 database (Fick and Hijmans 2017). We believe North America represents a suitable scale, large enough to cover a lot of variation in environmental variables and community structure, as well as phenomenons such as species migration. We also expect such extent of the spatial scale to cover for imprecision in estimated species ranges. The WorldClim data consists of spatially interpolated monthly climate data for global areas, available for resolutions from 10 arc-minutes to 30 arc-seconds. The variables used are provided in Table 1, and consists of different measures of temperature and precipitation. We chose to use the coarser 10 arc-minutes resolution in our analyses, again to cover for imprecision, and because we believe it is sufficient for proof of concept.

Table 1: 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

Variable	Description
15	Precipitation Seasonality (Coefficient of Variation)
16	Precipitation of Wettest Quarter
17	Precipitation of Driest Quarter
18	Precipitation of Warmest Quarter
19	Precipitation of Coldest Quarter

¹²³ **2. Data Manipulation**

¹²⁴ WorldClim variables and eBird occurrence data are provided in different formats, so they
¹²⁵ require some manipulation to be combined together. WorldClim variables are provided
¹²⁶ in a 2-dimensional grid format, useful for large scale analyses and visualization, where
¹²⁷ each cell or pixel corresponds to the resolution of 10 arc-minutes. Each of the 19 variables
¹²⁸ forms a different grid. On the other hand, eBird records are occurrence-based, so each
¹²⁹ entry in the dataset corresponds to an observation of a single species at a given time and
¹³⁰ location. These entries can easily be matched to the 2D grid format of the WorldClim variables
¹³¹ through their spatial coordinates, which we found more useful for large scale analyses and
¹³² visualization. Hence, for each species, we matched all occurrences in eBird to the grid format
¹³³ of the WorldClim variables, and then created a presence-absence community matrix Y , taking
¹³⁴ all the grid cells as sites. At the 10 arc-minutes resolution, we obtained 39 024 sites with
¹³⁵ occurrences and 62 species. We also applied the Hellinger transformation on the raw presence-
¹³⁶ absence data, although the most appropriate method remains to be determined, especially
¹³⁷ since the data has to be compared with the SDM predictions. All data manipulations and
¹³⁸ further analyses were realized in *Julia v1.2.0* (Bezanson et al. 2017) with the basic structure
¹³⁹ built around the soon-to-be-released `SimpleSDMLayers.jl` package.

¹⁴⁰ **3. SDM – The BIOCLIM Method**

¹⁴¹ We used the BIOCLIM method (Nix 1986) to predict species distributions, which is a climate-
¹⁴² envelope model, considered a classic in the field, that simply relates a species distribution
¹⁴³ to the ranges of bioclimatic variables at observed locations (Booth et al. 2014). It has long
¹⁴⁴ been outperformed by other methods (Elith et al. 2006), but it is still commonly used for its

145 simplistic approach and ease of understanding, as well as its simple relation to niche theory
146 (Booth et al. 2014; Hijmans et al. 2017). It is also a method designed for presence-only data,
147 which does not require information on absences, nor take them into account if provided (as in
148 our case). Despite that, we chose this method for our preliminary analyses as it was easier to
149 implement and because we believe it to be sufficient for proof-of-concept. We discuss possible
150 alternatives in the “Alternative Methods” section below.

151 Briefly, the BIOCLIM method defines species potential range as a multidimensional envi-
152 ronmental hypervolume bounded by the minimum and maximum values of all presences
153 (Franklin 2010). For each species, the algorithm establishes the percentile distribution of
154 the values of each environmental variables at the known locations of occurrences (Hijmans
155 et al. 2017). All sites are then compared to those percentile distributions and given a score
156 for each variable according to their ranking between 0.0 (1st percentile) and 1.0 (100th
157 percentile). The median or 50th percentile is considered as the most suitable location, and
158 both tails (e.g. 10th and 90th percentile) are not distinguished, the values larger than 0.5
159 being subtracted from 1. The minimum percentile score across all environmental variables is
160 selected as the prediction value for each site and multiplied by 2 so values are between 0 and
161 1 (Hijmans et al. 2017). It should be noted that the limiting variable is thus not necessarily the
162 same for all sites. Values of 1 are rare, as it would mean a perfectly median site on all variables,
163 and values of 0 are frequent, since they are assigned whenever an environmental value is
164 outside the range of the observed ones (Hijmans et al. 2017). Finally, before calculating
165 richness or beta diversity metrics, we transformed the predictions back to a presence-absence
166 format, where all predictions greater than one are considered as presence. This might tend to
167 overestimate species ranges and create some sort of border effect, but we believe the effects
168 will be mitigated given the spatial extent and coarse scale of our study.

169 4. LCBD Calculation

170 We calculated the LCBD statistics through the total variance of the matrix Y for both the raw
171 data and SDM predictions. Legendre and De Cáceres (2013) showed that LCBD coefficients
172 can be calculated directly through the total variance of matrix Y , or through a matrix of
173 dissimilarities among sampling units. We chose the first approach as it also allows to compute
174 species contributions to beta diversity (SCBD), which could also prove useful for conservation

purposes, but we did not investigate these for now. Before computing the LCBD statistics, the presence-absence matrix Y had to be transformed in an appropriate way, as mentioned earlier. We chose to apply the Hellinger transformation to the raw data and no transformation on the SDM predictions for now, as the most appropriate one still needs to be determined. We then computed a matrix S of squared deviations from column means and summed all the values of S to obtain the total sum of squares (SS) of the species composition data (Legendre and De Cáceres 2013). LCBD coefficients are then computed as $LCBD_i = SS_i/SS_{Total}$, where SS_i is the sum of squares of a sampling unit i . Finally, since our matrix Y is very large, the LCBD coefficients are very small, so we scaled them to the maximum value observed.

5. Prediction Validity

The exact way of testing the validity of the predictions remains to be determined, and will also depend on the exact methods used to make the SDM predictions. A key element to note is that both SDM predictions and LCBD values will have to be validated, hence they might require different methods. Metrics that measure the accuracy of categorical or probabilistic predictions in SDMs are well documented, in various forms. Some require absence data to test against, and can be used on probabilistic predictions directly (area-under-curve, AUC) or after a conversion of the predictions to binary presence-absence using a certain threshold (Kappa index, measuring the difference between observed and chance agreement in a confusion matrix) (Franklin 2010). Other methods are appropriate for presence-only data, such as the Boyce Index. In any case, measuring prediction error is only one part of validation, and finding appropriate data for evaluation is almost as essential (Franklin 2010), especially since we also aim to describe community structure. Separating the data into a training and testing dataset, with 70% and 30% of the observations for instance, is a possible approach common in machine learning methods, although all of the available observations might be needed in some cases (Franklin 2010). An interesting approach, suggested by Elith et al. (2006) for SDMs, would be to find independent, well-structured presence-absence datasets for validation, on which both SDM predictions and beta diversity metrics could be tested. This approach has the advantage that the testing data is truly independent of the training one, hence it could be used with certain tests of significance. Although it might not cover the entire extent of the predictions in a single test, this method brings a closer comparison to the

205 way LCBD metrics are used in most studies, and it would provide interesting perspectives if
206 combined with other, full-extent validation methods.

207 **6. Alternative methods**

208 Other methods could possibly outperform BIOCLIM for the predictions, as shown by Elith
209 et al. (2006). Better predictions will come by two different means: 1) approaches that
210 are better than BIOCLIM to model the relationship between species presence-absence (or
211 even abundance) and environmental variables, and 2) approaches that account for other
212 drivers of species distributions, such as ecological interactions and species migration. The
213 most obvious alternative to BIOCLIM is MAXENT (Phillips, Anderson, and Schapire 2006),
214 another presence-only method that has come to be one of the most widely used methods.
215 Machine learning methods would be also be interesting alternatives that have been proven to
216 outperform BIOCLIM (Franklin 2010). Random Forests, especially are simple methods to put
217 in place, allow for quantification of the variables importance in explaining variation, and offer
218 intrinsic testing metrics. Neural networks could also be an interesting alternative. However,
219 while those methods might return more accurate predictions, they do not implicitly model
220 other drivers of species distribution, among which species interactions and functional niche.
221 Integrating those factors might prove more difficult given our dataset and our focus Warblers
222 species, as no appropriate information on their interaction is available, to our knowledge.
223 Joint species distribution models (JSDMs) might be an interesting way to encompass those, as
224 they attempt to model species co-occurrence, rather than the distribution of single species
225 (Pollock et al. 2014). A different taxonomic group and data datasets could also be used with
226 more details on interactions could also be used, though having a method that can be applied
227 to any taxonomic group would be more interesting. Yet, such an approach might prove to be
228 beyond the scope of the present research.

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

230 We aim to apply our method to environmental conditions from climate change scenarios, first
231 to model community compositions after climate change on continuous scales through SDMs,
232 and then to identify the sites where the community has changed in the most exceptional ways.
233 This can be done through LCBD values, but also through temporal beta diversity indices

234 (TBI) (Legendre 2019), which allow to study changes in community composition through
235 time from repeated surveys at given sites. Whereas LCBD values essentially measure the
236 contribution to beta diversity of each site compared to all other ones, TBI measure changes in
237 community composition for a single site between two surveys, and can also be decomposed
238 into species losses and gains. Moreover, TBI can be tested for significance using a permutation
239 test. An approach similar to that of Legendre and Condit (2019) would be most interesting to
240 follow: they first computed LCBD indices and compared the sites that were significant for two
241 surveys 30 years apart, highlighting a swamp region where important changes seemed to have
242 occurred, and then used TBI indices to confirm the sites with significant changes, decompose
243 those into losses and gains and identify the species that had changed the most. Such an
244 approach could be highly informative with our data, although the permutation tests and
245 corrections to apply might cause problems given the number of sites that would be implied
246 in our study. The possibility of using climate change scenarios in the SDMs also needs to be
247 investigated in more details. We did not try to download nor find the appropriate data for
248 now, but we found that the interpolated variables are sometimes different than those used in
249 Worldclim 2.0. The SDM models and predictions might therefore be slightly different than
250 those used for the LCBD calculations, and potentially less reliable. Nonetheless, we believe
251 it will be possible to do some kind of time analysis linking beta diversity, climate change
252 and species distribution modelling, and that it could return highly informative results for
253 conservation purposes.

254 Preliminary Results

255 Our preliminary results mainly consisted of comparisons between the raw occurrence data
256 and the SDM predictions for the four following elements: single-species distribution (figs. 1,
257 2), species richness (figs. 3, 4), LCBD coefficients (figs. 5, 6), and the relationship between the
258 species richness and LCBD coefficients (figs. 7, 8). Two main results emerged from them: 1)
259 the models provided seemingly valid and relevant results for poorly sampled regions, both
260 expected species-poor and species-rich ones, and 2) there was a strong association between
261 species richness and LCBD coefficients confirming the relationship shown in other studies.

262 First, the example of the Yellow Warbler (*Setophaga petechia*), one of the most observed species,
263 showed that the single-species models predicted a broad distribution covering poorly sampled

264 areas, with notable patches of absence across the continent (figs. 1, 2). Likewise, species
265 richness, defined as the number of species present per site, showed a clear latitude gradient,
266 with the poorest sites to the North and the richest to the South (figs. 3, 4). A form of altitude
267 gradient could also be observed, with the Rockies and other mountains well delimited by their
268 lower values. In both cases, the results make intuitive sense and highlight the models ability
269 to predict species presence despite poor or no sampling. Mexico, for example, has much
270 sparser sampling and fewer observations, but the models predict Yellow Warblers presence in
271 most areas nonetheless, as well as higher species richness than on the highly sampled Atlantic
272 Coast, which make sense for a more southern location. We believe these to be valid insights
273 on poorly sampled locations, although we reckon that intuitive reasoning is not a proof of
274 anything, and that the model might be wrong in important ways. In any case, it highlights
275 the need for an appropriate method of validation, as well as a thoughtful consideration of
276 other factors such as species migration.

277 Second, our preliminary LCBD results seemed to confirm the association between species
278 richness and LCBD coefficients, while also being valid and insightful. They were however
279 harder to interpret given the use of the Hellinger transformation for the raw occurrence
280 data only. Raw occurrence data showed a negative relationship between species richness and
281 LCBD coefficients (fig. 7), as observed previously by Heino and Grönroos (2017), with no
282 clear geographic pattern (fig. 5). If anything, the highest values seemed to be at the borders
283 of the most sampled regions, which are about where the sites with the less species observed
284 are located (fig. 3). On the other hand, SDM predictions showed a quadratic form, with the
285 LCBD coefficients re-increasing after some threshold (fig. 8). The geographic pattern is also
286 clearer, with the highest values to the northern and southern extremes (fig. 6). We suggest
287 that this result makes sense, as LCBD indices should highlight the most exceptional species
288 composition, and these are both well and continuously represented in a SDM. Although raw
289 occurrence data results concorded with those of Heino and Grönroos (2017), the “border
290 effect” and the difference with SDM projections might show the importance of going beyond
291 occurrence data when using large databases but spatially biased databases such as eBird.
292 Once again, our results highlight the need for well-thought method and an investigation of
293 the appropriate transformation to use on the data.

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

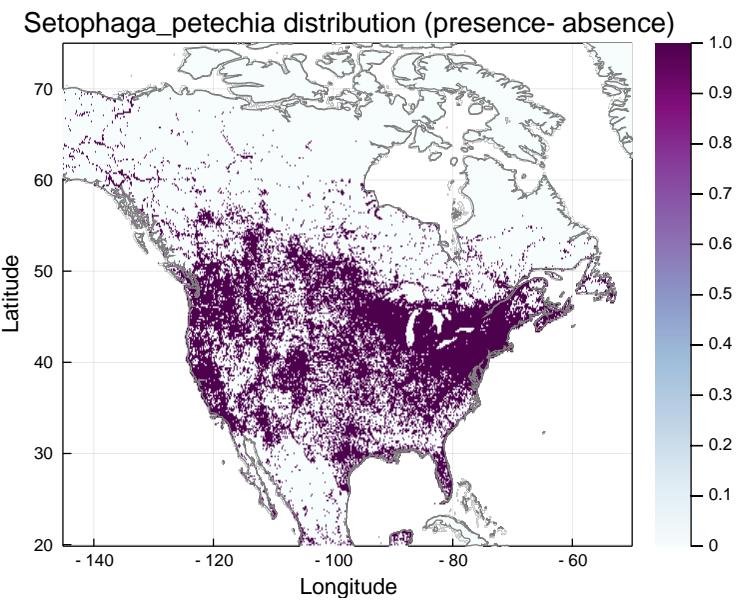


Figure 1: Single Species Distributions - Raw

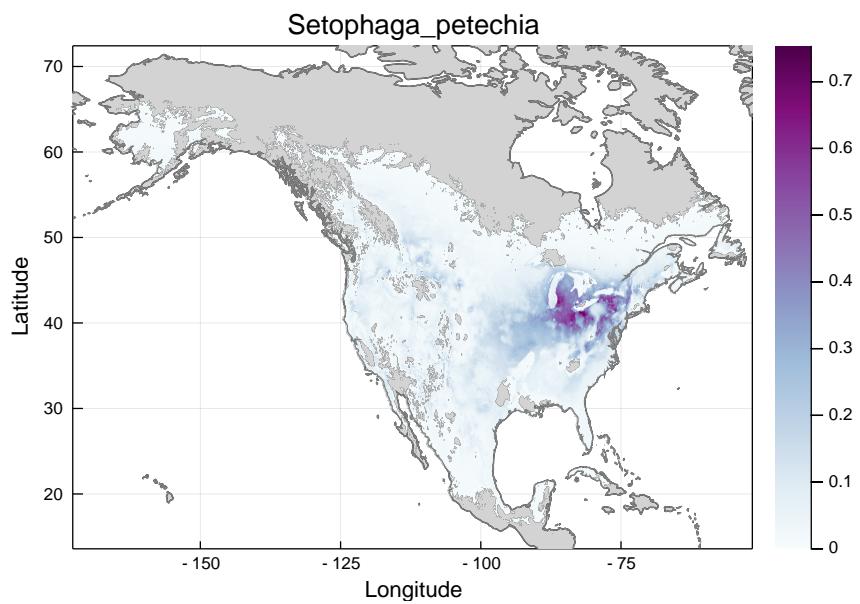


Figure 2: Single Species Distributions - SDM

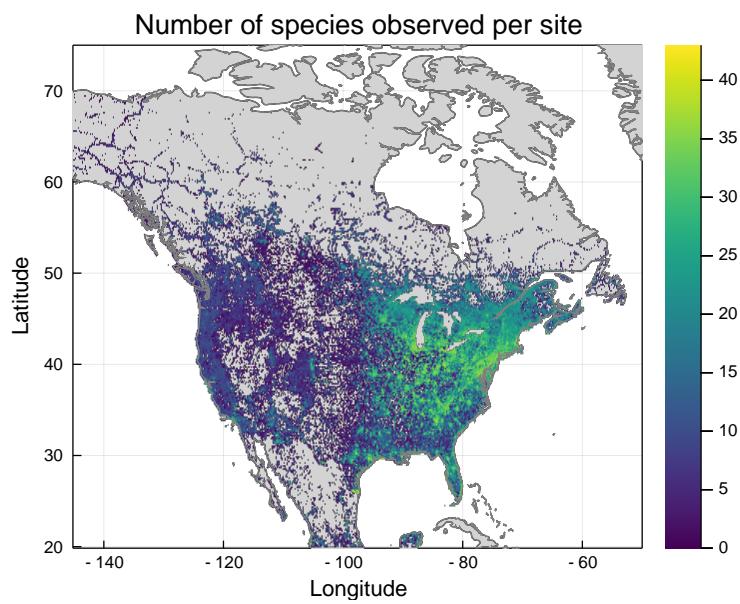


Figure 3: Species Richness - Raw

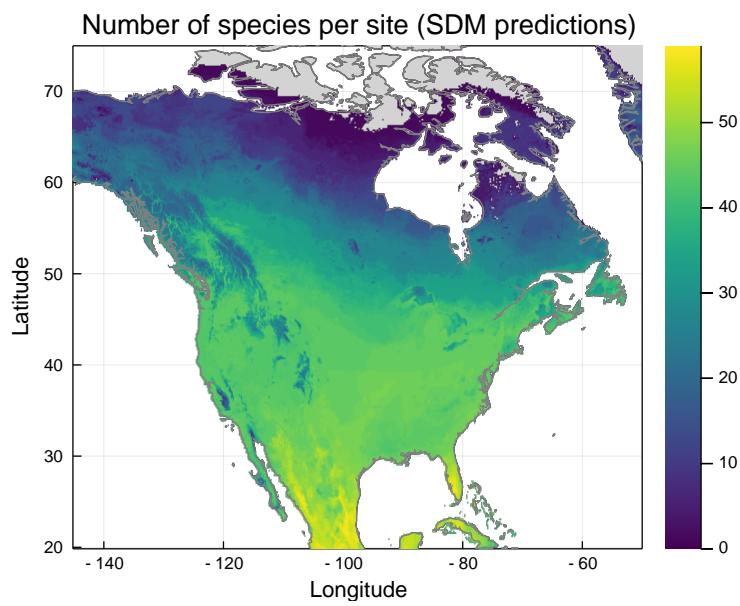


Figure 4: Species Richness - SDM

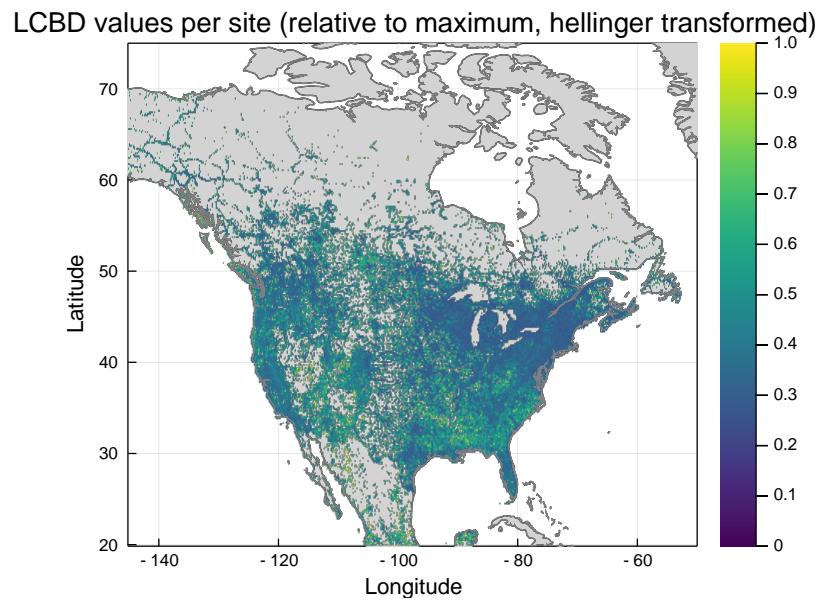


Figure 5: LCBD values - Raw (transformed)

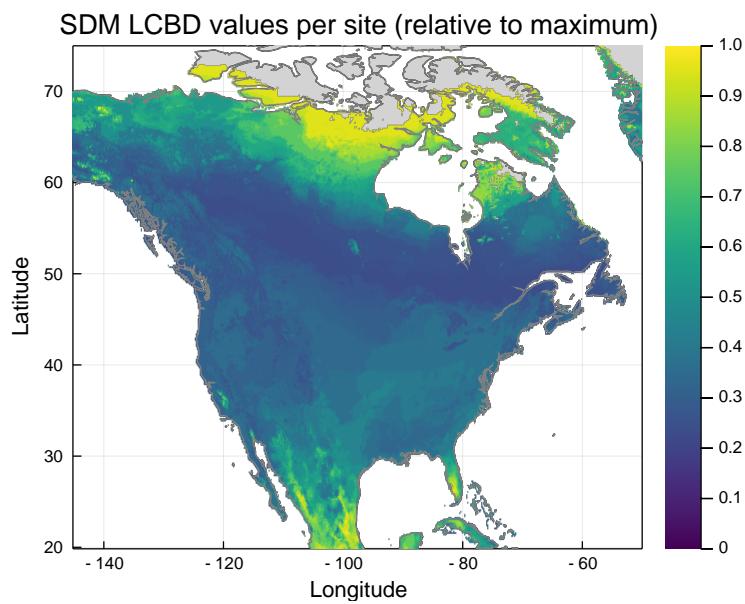


Figure 6: LCBD values - SDM

Relationship between LCBD (hellinger transformed) and species richness

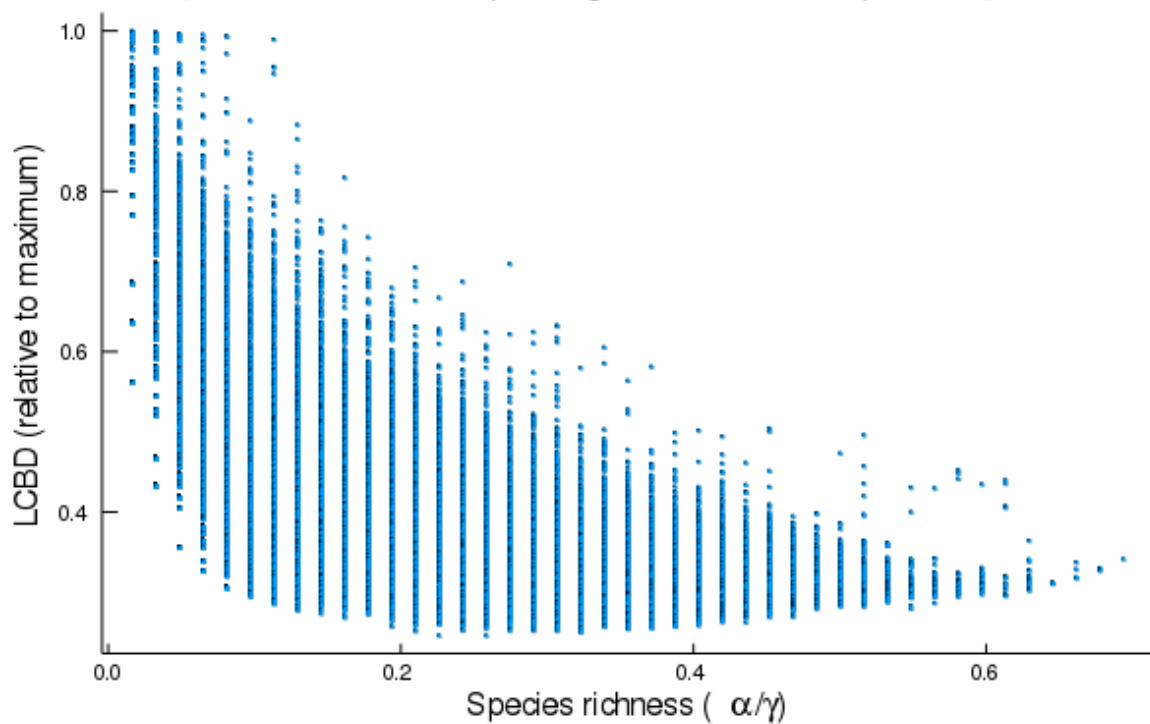


Figure 7: LCBD-richness relationship - Raw

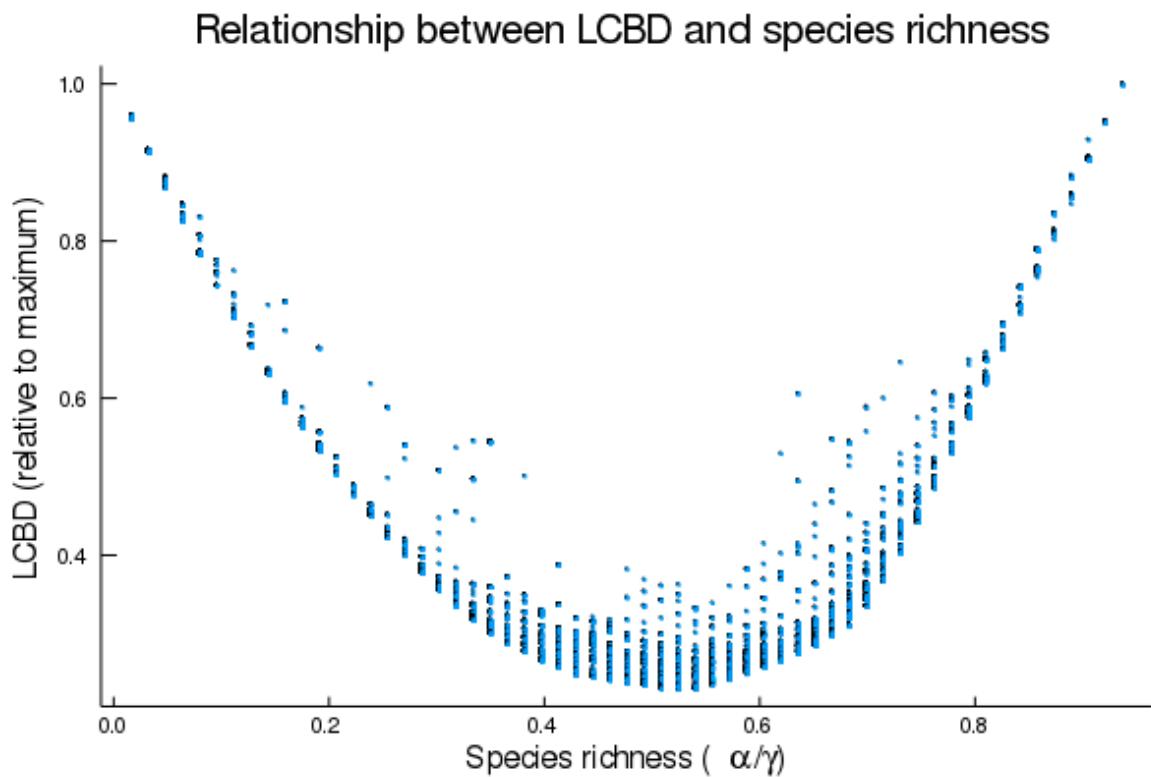


Figure 8: LCBD-richness relationship - SDM

304 **References**

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

- 318 Fick, Stephen E., and Robert J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution
319 Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37 (12):
320 4302–15. <https://doi.org/10.1002/joc.5086>.
- 321 Franklin, Janet. 2010. "Moving Beyond Static Species Distribution Models in Support of
322 Conservation Biogeography: Moving Beyond Static Species Distribution Models." *Diversity*
323 and *Distributions* 16 (3): 321–30. <https://doi.org/10.1111/j.1472-4642.2010.00641.x>.
- 324 Heino, Jani, and Mira Grönroos. 2017. "Exploring Species and Site Contributions to Beta
325 Diversity in Stream Insect Assemblages." *Oecologia* 183 (1): 151–60. <https://doi.org/10.1007/s00442-016-3754-7>.
- 327 Hijmans, Robert J., Steven Phillips, John Leathwick, and Jane Elith. 2017. *Dismo: Species*
328 *Distribution Modeling*. <https://CRAN.R-project.org/package=dismo>.
- 329 Johnston, A., W. M. Hochachka, M. E. Strimas-Mackey, V. Ruiz Gutierrez, O. J. Robinson,
330 E. T. Miller, T. Auer, S. T. Kelling, and D. Fink. 2019. "Best Practices for Making Reli-
331 able Inferences from Citizen Science Data: Case Study Using eBird to Estimate Species
332 Distributions." *bioRxiv*, March, 574392. <https://doi.org/10.1101/574392>.
- 333 Legendre, Pierre. 2019. "A Temporal Beta-Diversity Index to Identify Sites That Have
334 Changed in Exceptional Ways in Space-Time Surveys." *Ecology and Evolution* 9 (6): 3500–
335 3514. <https://doi.org/10.1002/ece3.4984>.
- 336 Legendre, Pierre, Daniel Borcard, and Pedro R. Peres-Neto. 2005. "Analyzing Beta Diver-
337 sity: Partitioning the Spatial Variation of Community Composition Data." *Ecological*
338 *Monographs* 75 (4): 435–50. <https://doi.org/10.1890/05-0549>.
- 339 Legendre, Pierre, and Richard Condit. 2019. "Spatial and Temporal Analysis of Beta Diversity
340 in the Barro Colorado Island Forest Dynamics Plot, Panama." *Forest Ecosystems* 6 (1): 7.
341 <https://doi.org/10.1186/s40663-019-0164-4>.
- 342 Legendre, Pierre, and Miquel De Cáceres. 2013. "Beta Diversity as the Variance of Community
343 Data: Dissimilarity Coefficients and Partitioning." *Ecology Letters* 16 (8): 951–63. [//doi.org/10.1111/ele.12141](https://doi.org/10.1111/ele.12141).
- 345 Nix, Henry A. 1986. "A Biogeographic Analysis of Australian Elapid Snakes." *Atlas of Elapid*

- 346 *Snakes of Australia* 7: 4–15.
- 347 Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. “Maximum Entropy
348 Modeling of Species Geographic Distributions.” *Ecological Modelling* 190 (3): 231–59.
349 <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- 350 Poisot, Timothée, Richard LaBrie, Erin Larson, Anastasia Rahlin, and Benno I. Simmons. 2019.
351 “Data-Based, Synthesis-Driven: Setting the Agenda for Computational Ecology.” *Ideas in
352 Ecology and Evolution* 12 (July). <https://doi.org/10.24908/iee.2019.12.2.e>.
- 353 Pollock, Laura J., Reid Tingley, William K. Morris, Nick Golding, Robert B. O’Hara, Kirsten
354 M. Parris, Peter A. Vesk, and Michael A. McCarthy. 2014. “Understanding Co-Occurrence
355 by Modelling Species Simultaneously with a Joint Species Distribution Model (JSDM).”
356 *Methods in Ecology and Evolution* 5 (5): 397–406. <https://doi.org/10.1111/2041-210X.12180>.
- 358 Sullivan, Brian L., Christopher L. Wood, Marshall J. Iliff, Rick E. Bonney, Daniel Fink, and
359 Steve Kelling. 2009. “eBird: A Citizen-Based Bird Observation Network in the Biological
360 Sciences.” *Biological Conservation* 142 (10): 2282–92. <https://doi.org/10.1016/j.biocon.2009.05.006>.