

1 *Title:* Exploring the spatially explicit predictions of the Maximum Entropy Theory of Ecology

2 *Authors:* D.J. McGlinn^{1,2,3}, X. Xiao^{1,2}, J. Kitzes⁴, E.P.White^{1,2}

3 ¹*Biology Department, Utah State University, Logan, UT 84341 USA;*

4 ²*Ecology Center, Utah State University, Logan, UT 84341 USA;*

5 ³*Biology Department, College of Charleston, SC 29401;*

6 ⁴*Energy and Resources Group, University of California, Berkeley, CA 94720 USA;*

7 *Email addresses:* mcglinndj@cofc.edu; xiao@weecology.org; jkitzes@berkeley.edu;

8 ethan@weecology.org;

9 *Keywords:* beta-diversity, community similarity, distance decay, maximum entropy, species
10 turnover, vegetation

11 *Short running-title:* Spatially Explicit METE

12 *Corresponding author:* DJM

13 *Number of words in abstract:* 288

14 *Number of words in main body:* 5090

15 *Number of references:* 50

16

17 **Abstract**

18 **Aim:**

19 The Maximum Entropy Theory of Ecology (METE) is a unified theory of biodiversity that
20 attempts to simultaneously predict patterns of species abundance, size, and spatial structure. The
21 spatial predictions of this theory have repeatedly performed well at predicting diversity patterns
22 across scales. However, the theoretical development and evaluation of METE has focused on
23 predicting patterns that ignore inter-site spatial correlations. As a result the theory has not been
24 evaluated using one of the core components of spatial structure. We develop and test a semi-
25 recursive version of METE's spatially explicit predictions for the distance decay relationship of
26 community similarity and compare METE's performance to the classic random placement model
27 of completely random species distributions. This provides a better understanding and stronger
28 test of METE's spatial community predictions.

29 **Location:**

30 New world tropical and temperate plant communities.

31 **Methods:**

32 We analytically derived and simulated METE's spatially explicit expectations for the Sorensen
33 index of community similarity. We then compared the distance decay of community similarity of
34 16 mapped plant communities to METE and the random placement model.

35 **Results:**

36 The version of METE we examined was successful at capturing the general functional form of
37 empirical distance decay relationships, a negative power function relationship between
38 community similarity and distance. However, the semi-recursive approach consistently over-

39 predicted the degree and rate of species turnover and yielded worse predictions than the random
40 placement model.

41 **Main conclusions:**

42 Our results suggest that while METE's current spatial models accurately predict the spatial
43 scaling of species occupancy, and therefore core ecological patterns like the species-area
44 relationship, its semi-recursive form does not accurately characterize spatially-explicit patterns
45 of correlation. More generally, this suggests that tests of spatial theories based only on the
46 species-area relationship may appear to support the underlying theory despite significant
47 deviations in important aspects of spatial structure.

48

49

50 **Introduction**

51 Community structure can be characterized using a variety of macroecological relationships
52 such as the species-abundance, body size, and species spatial distributions. Increasingly
53 ecologists have recognized that many of these macroecological patterns are inter-related, and
54 progress has been made toward unifying the predictions of multiple patterns using theoretical
55 models (Storch *et al.*, 2008; McGill, 2010). One approach to predicting suites of macroecological
56 patterns are process-based models such as niche and neutral dispersal models, which have the
57 potential to provide biological insight into the process structuring ecological systems (Adler *et*
58 *al.*, 2007). Alternatively, a new class of constraint-based models suggest that similar patterns
59 may be produced by different sets of processes because the form of the predicted pattern is due
60 to the existence of statistical constraints rather than directly reflecting detailed biological
61 processes (Frank, 2009, 2014; McGill & Nekola, 2010; Locey & White, 2013).

62 The Maximum Entropy Theory of Ecology (METE) is a recent attempt to explain a
63 number of ecological patterns from the statistical constraint perspective (Harte *et al.*, 2008, 2009;
64 Harte, 2011; Harte & Newman, 2014). METE uses the principle of entropy maximization, that
65 the most likely distribution is the one with the least information (i.e., the one closest to the
66 uniform distribution) subject to a set of constraints (i.e., prior information), to predict
67 distributions of species abundance, body size, and spatial structure. A frequentist perspective on
68 the Maximum Entropy modeling approach is that every possible configuration of a system is
69 equally likely; therefore, the probability of a particular distribution is directly proportional to the
70 number of configurations that distribution is compatible with (Harte, 2011; Harte & Newman,
71 2014). The distribution with the largest number of compatible system configurations is the
72 predicted most likely state of the system. In contrast to detailed biological models of community

73 assembly, METE has no free parameters and only requires information on total community area,
74 total number of individuals, total number of species, and total metabolic rate of all individuals to
75 generate its predictions.

76 There is strong empirical support for METE's predictions for the species abundance
77 distribution and patterns related to the spatial distribution of individuals and species (Harte *et al.*,
78 2008, 2009; Harte, 2011; White *et al.*, 2012a; Xiao *et al.*, 2013; McGlinn *et al.*, 2013; Newman
79 *et al.*, 2014). Specifically, METE has been successful at predicting spatially implicit patterns of
80 community structure such as the species spatial abundance distribution and the species-area
81 relationship (Harte *et al.*, 2008, 2009; McGlinn *et al.*, 2013). It has even been proposed that the
82 METE spatial predictions yield a widely applicable universal species-area relationship (Harte *et*
83 *al.*, 2009, 2013, but see Šizling *et al.*, 2011, 2013). However all of METE's spatial predictions
84 that have been tested focus on spatially implicit patterns that ignore spatial correlations. As a
85 result the theory has not been evaluated using one of the core components of spatial structure.
86 This is due in part to the fact that METE's spatial correlation predictions have not been fully
87 derived.

88 The most commonly studied ecological pattern that relies on these spatial correlations is
89 the distance decay relationship (DDR) in which the similarity of species composition decreases
90 with distance (Nekola & White, 1999). The DDR provides a spatially-explicit, community-level
91 characterization of intra-specific aggregation patterns including correlations in space (Plotkin &
92 Muller-Landau, 2002; Palmer, 2005; Morlon *et al.*, 2008; McGlinn & Palmer, 2011), and
93 predicting the DDR is an important area of future development for METE because the DDR is
94 necessary to accurately extrapolate community patterns to unsampled areas (Harte, 2011).

95 Here we explore METE's spatially explicit predictions for the DDR by developing
96 analytical and simulation based solutions and comparing them to empirical data. We build on the
97 Hypothesis of Equal Allocation Probabilities (HEAP, Harte et al. 2005, Harte 2007) using an
98 approach that combines elements of both a non-recursive and recursive version of METE
99 (McGlenn et al. 2013). We test those predictions using data from 16 spatially explicit plant
100 communities and compare METE's performance to the classic Random Placement Model (RPM)
101 in which individuals are randomly placed on the landscape (Coleman, 1981). Our approach
102 provides a stronger evaluation of the performance of this model and whether it can explain
103 patterns of spatial structure in the absence of detailed biological processes.

104 **Methods**

105 METE has thus far been used to derive the probability that a random cell on a landscape will
106 be occupied by a given number of individuals (i.e., the intra-specific spatial abundance
107 distribution). Predictions for this distribution have been based either on recursively subdividing
108 an area in half or on predicting species abundances directly at smaller scales (Harte, 2011;
109 McGlenn *et al.*, 2013). In addition to the spatial abundance distribution, the DDR requires a
110 prediction for the correlations in abundance among neighboring cells, which has proven difficult
111 to derive for METE (Harte 2011).

112 *Developing METE's Spatially Explicit Predictions*

113 METE's spatial predictions depend on two conditional probability distributions which are
114 computed using independent applications of MaxEnt:

- 115 1) the species abundance distribution (SAD), $\Phi(n | S_0, N_0)$, the probability that a species has
116 abundance n in a community with S_0 species and N_0 individuals, and

117 2) the intra-specific spatial abundance distribution, $\Pi(n | A, n_0, A_0)$, the probability that n
118 individuals of a species with n_0 total individuals are located in a random quadrat of area A
119 drawn from a total area A_0 .

120 The METE prediction for Φ is calculated using entropy maximization with constraints on the
121 average number of individuals per species (N_0/S_0) and the maximum number of individuals N_0
122 for a given species, which yields a truncated log-series abundance distribution (Harte *et al.*, 2008;
123 Harte, 2011). The spatially implicit Π distribution is solved for using entropy maximization with
124 constraints on the average number of individuals per unit area (n_0/A_0) and the maximum number
125 of individuals n_0 of a given species. Although METE requires information on total metabolic rate
126 to derive its predictions, the exact value that this constraint takes has no influence on Φ and Π
127 (Harte *et al.*, 2009; Harte, 2011).

128 Previous studies have downscaled (or upscaled) METE's predictions using recursive and
129 non-recursive approaches. Here we develop a spatially explicit approach to downscaling
130 METE's predictions that combines elements of both approaches and builds off an existing
131 theoretical framework for modeling the DDR. With the recursive version of METE, Φ and Π are
132 solved for at each successive halving or bisection of A_0 until the area of interest is reached. After
133 each bisection, Φ and Π are calculated and used to derive predicted values of average S and N at
134 that scale which provide updated constraints for the next bisection (Harte et al. 2009).

135 Alternatively, a non-recursive approach can be used in which, Φ and Π at the spatial grain of
136 interest can be solved for directly from the constraints placed at A_0 (Harte et al. 2008). A semi-
137 recursive approach is also possible in which Π is recursively downscaled but Φ is not. The semi-
138 recursive predictions of METE have not been previously examined but this model builds directly
139 on the existing theoretical derivations of the DDR by Harte (2007) for the Hypothesis of Equal

140 Allocation Probabilities (HEAP). In Appendix A, Fig. A1 and A2 we examine how the semi-
141 recursive formulation of METE differs from a previous examination of the METE recursive and
142 non-recursive SARs (McGlenn et al. 2013), and in Appendix B we develop the analytical
143 derivations of the semi-recursive formulation of the DDR.

144 In the semi-recursive formulation of the DDR, multi-cell correlations emerge from the
145 spatially nested application of a recursive bisection scheme in which individuals are randomly
146 placed in the left or right half of a cell at each bisection (Fig. 1). Biologically, this can be thought
147 of as a sequentially dependent colonization rule in which individuals randomly choose to occupy
148 the left or right side of an area depending on the existing number of individuals in each half
149 (Harte et al. 2005, Harte 2007, and Conlisk et al. 2007). Our version of METE assumes that for
150 a single bisection there is an equal likelihood for every possible spatial configuration of
151 indistinguishable individuals (Eq. B1). Multi-cell spatial correlations emerge from this approach
152 because the two cells that are formed from a common parent cell are adjacent to one another and
153 are likely to be more similar in abundance than other cells on the landscape (Fig. 1). This
154 approach has three important and inter-related limitations: 1) At each stage in the bisection
155 algorithm, information about the cells surrounding the parent cell is ignored when determining
156 allocations within the parent cell, 2) between-cell distance is defined in reference to an artificial
157 bisection scheme which does not have a one-to-one correspondence with physical distance, and 3)
158 the correlation between cells does not decrease smoothly with physical distance. Alternative
159 approaches have been proposed for deriving the DDR for METE based on computing the single-
160 cell Π distribution at two or more scales and then using the scaling of this marginal distribution
161 to infer the probabilities of a given spatial configuration of abundance (Harte 2011). However,
162 these approaches have yet to yield predictions for the DDR.

163 The analytical forms of the semi-recursive formulation (Appendix B) are time-intensive
164 to compute due to the multiple levels of recursion, ignore patterns of abundance (i.e., are
165 formulated only in terms of presence-absence), and are not exact. An alternative approach to
166 deriving semi-recursive METE predictions for the DDR is to use a spatially-explicit simulation.

167 *Spatially Explicit METE Simulation*

168 To simulate semi-recursive METE's spatial predictions, the equal probability rule (Eq.
169 B1) that METE assumes when total area is halved is recursively applied starting at the anchor
170 scale A_0 and progressively bisecting the area until the finest spatial grain of interest is achieved
171 (Fig. 1). Abundance in the simulation model can be parameterized using an observed SAD or
172 using a random realization of the METE SAD given the values of S_0 and N_0 . Once the
173 abundances of the species are assigned, each species is independently spatially distributed.
174 Because the equal probability rule requires that there is an equal probability of 0 to n_0 individuals
175 occurring on the left or right side of the total area A_0 , the number of individuals in the left side
176 can be set as a draw from a discrete random uniform distribution between 0 and n_0 and the
177 remaining number of individuals are placed on the right hand side.

178 *Datasets*

179 We used a database of 16 spatially explicit and contiguous community datasets compiled
180 by McGlinn et al. (2013) to evaluate the DDR predictions of recursive METE (Table 1). All of
181 the sites were terrestrial, woody plant communities with the exception of the serpentine grassland
182 dataset which covered a terrestrial, herbaceous plant community. In the woody plant
183 communities, all stems were recorded that were at least 10 mm in diameter at breast height (i.e.,
184 1.4 m from the ground) with the exception of the Oosting and Cross Timbers sites where the
185 minimum diameter was 20 and 25 mm respectively. Recursive METE only generates predictions

186 for bisections of total area; therefore, we restricted our analysis to square or rectangular areas
187 with a length-to-width ratio of 2:1. Two of the sites had irregular plot designs: Sherman and
188 Cocoli. At these sites we partitioned the datasets into two 2:1 rectangles and analyzed each half
189 independently and then averaged the results (see Supplemental Information: Fig. S1 in McGlinn
190 et al. 2013). See McGlinn et al. (2013) for additional information on site selection criteria, and
191 in particular their Supplemental Table 1, which provides a more complete description of the
192 datasets used in our analysis.

193 *Data Analysis*

194 We compared the fit of METE with and without the observed SAD and the random
195 placement model (RPM) to the empirical DDRs. The METE predictions represented averages of
196 the abundance-based Sørensen index across 200 simulated communities. The abundance-based
197 RPM predictions were generated by distributing the observed number of individuals of each
198 species randomly in space and then computing the average abundance-based Sørensen index
199 across 500 permutations (Morlon *et al.*, 2008).

200 The DDR is sensitive to the choice of the spatial grain of comparison (Nekola & White,
201 1999); so, we examined the DDR at several spatial grains for each dataset. We examined spatial
202 grains resulting from 3-13 bisections of A_0 . To ensure that the samples at a given grain were
203 square we only considered odd numbers of bisections when A_0 was rectangular and even
204 numbers of bisections when A_0 was square. To ensure the best possible comparison between the
205 observed data and METE and to avoid detecting unusual spatial artefacts in the METE predicted
206 patterns we employed the “user rules” of Ostling et al. (2004) such that samples at a specific
207 grain (i.e., level of bisection) were only compared if they were separated by a specific line of
208 bisection (i.e., a given separation order, Fig. 1 and Appendix A, Fig. A3). This approach was

209 taken rather than the standard method of constructing the DDR from all possible pairwise sample
210 comparisons without reference to an imposed bisection scheme. We computed geographic
211 distance by averaging the distance between all the compared samples compared at a given
212 separation order. For the Crosstimbers study site we were not able to examine the DDR based on
213 the METE SAD because of difficulty in generating random realizations of the METE SAD
214 needed for the community simulator when S_0 is less than approximately 10. Typically averages
215 of community similarity are used to examine the geometry of the DDR; however, in some cases
216 the distribution of the similarity metric may be strongly skewed and therefore we computed both
217 averages and medians of community similarity at each separation order.

218 We used weighted least squares (WLS) regression to account for differences in the
219 number of pairwise comparisons at different spatial lags (there are many more comparisons at
220 short lags) when fitting the power and exponential models of the DDR (Venables & Ripley,
221 2002). We examined the power model and exponential models because they are the simplest
222 statistical models of the DDR, and it was recently suggested that at fine spatial scales the DDR
223 should be best approximated by a power model (Nekola & White, 1999; Nekola & McGill,
224 2014).

225 We checked that our results were consistent with the results provided in previous studies
226 (Harte, 2007, Fig. 6.7 and 6.8, 2011, Fig. 4.1), and that the DDR generated by the community
227 simulator closely agreed with the analytical solution Eq. B5 (Appendix B, Fig. B1). The code to
228 recreate the analysis is provided as Appendix D and at the following publicly available
229 repository: <http://dx.doi.org/10.6084/m9.figshare.978918>.

230 **Results**

231 In general, the semi-recursive METE distance decay relationship (DDR) provided a poor
232 fit to the empirical DDR (Figs. 2 and 3). The average and median community similarity results
233 were highly correlated ($r = 0.98$) and generated qualitatively similar results (Appendix A Figs.
234 A5 and A8); therefore, we focus on the results based on averaging similarity. While the METE
235 DDRs exhibited the general functional form of the empirical DDRs, an approximately power-law
236 decrease in similarity with distance, they typically had lower intercepts and steeper slopes than
237 the empirical DDRs (Fig. 2, Appendix A, Fig. A4 and A6). Both the empirical and METE
238 predicted DDR were better approximated by power rather than exponential models (Appendix A,
239 Fig. A6). METE converged towards reasonable predictions at fine spatial grains; however, this is
240 to be expected because at these scales similarity in both the observed and predicted patterns must
241 converge to zero due to low individual density (grey points in Fig. 3A,B). This is because when
242 individual density is low the probability of samples sharing species decreases rapidly simply due
243 to chance. The RPM is known to be a poor model for distance decay because it does not exhibit a
244 decrease in similarity with distance. However, it fit the empirical DDR slightly better than
245 METE (Figs. 2 and 3).

246 The METE DDR was not strongly influenced by the choice of using the observed or the
247 METE SAD (Figs. 2 and 3A,B). The METE SAD typically yielded a DDR with a slightly lower
248 intercept with the exception of the four tropical sites where it produced DDRs with slightly
249 higher intercepts. In general, we did not observe strong consistent differences between the
250 habitat types (Fig. 2, Appendix A, Fig. A7).

251 Our formulation of a semi-recursive METE produced SARs that generally agreed (i.e.,
252 within the 95% CI) with the recursive and non-recursive formulations of METE (Harte et al.

253 2009); however, it did appear that the semi-recursive approach systematically deviated towards
254 lower richness at fine spatial scales which is consistent with predicting stronger patterns of
255 spatial aggregation compared to the other formulations of METE (Appendix A, Fig. A1 and A2).

256 **Discussion**

257 The semi-recursive METE distance decay relationship (DDR) was well approximated by
258 a decreasing power function, and thus consistent with the general form of empirical DDRs, but it
259 provided a poor fit to empirical data. Specifically, the slope and the intercept of this power
260 function deviate substantially from empirical data resulting in a poor fit. These deviations
261 contrast with a number of studies showing that the theory successfully predicts both the Π
262 distribution and the SAR (Harte *et al.*, 2008, 2009; Harte, 2011; McGlinn *et al.*, 2013; but see
263 Šizling *et al.*, 2011). Both Π and the SAR are influenced by the spatially explicit pattern of
264 intraspecific aggregation but neither pattern reflects inter-quadrat correlations and therefore they
265 represent coarse metrics of spatial structure. The combination of a well fit SAR and a poorly fit
266 DDR suggests that the current version of METE accurately characterizes average occupancy, but
267 fails to characterize the spatial relationships among cells (McGeoch & Gaston, 2002; Storch *et*
268 *al.*, 2003; McGlinn & Hurlbert, 2012; Nekola & McGill, 2014).

269 These results only apply directly to the particular HEAP-based semi-recursive version of
270 the spatial METE theory, which represents a middle ground in terms of approach between Harte
271 *et al.* (2008) and Harte *et al.* (2009). Other approaches to deriving the METE DDR may perform
272 better than the semi-recursive approach if they can be developed. It has been suggested that there
273 is no *a priori* reason to prefer one version of the theory and that the best way to choose among
274 the different versions is empirically (Haegeman & Etienne, 2010; Harte, 2011). However, the
275 traditionally defined recursive and non-recursive versions of METE have shortcomings with

276 respect to how their assumptions and predictions are scaled, and the semi-recursive approach we
277 defined is limited by its dependence on an artificial bisection scheme. Specifically the recursive
278 approach predicts that the SAD has the same functional form, a truncated log-series, at all scales.
279 This is problematic because SADs are typically not scale-invariant if, as METE predicts, species
280 display intraspecific spatial aggregation (Green & Plotkin, 2007; Šizling *et al.*, 2009). The non-
281 recursive approach does not suffer from this problem because the SAD is only solved for at the
282 anchor scale; however, Haegemann and Etienne (2010) found that the non-recursive predictions
283 for a multi-cell generalization of the Π distribution were scale-inconsistent. The semi-recursive
284 approach does not suffer from this shortcoming because its multi-cell form (see Eq. 2.2 in
285 Conlisk *et al.*, 2007) is only defined over the set of bisections that are consistent with a landscape
286 in which n_0 individuals are distributed (see Appendix C for proof). However, the set of bisections
287 is artificial and multi-cell correlations only emerge from this approach in reference to bisection
288 distance rather than directly to physical distance between cells such that cells have equal
289 magnitude of correlation regardless of their physical distance if they have equivalent separation
290 orders (see Conlisk *et al.*, 2007 for a critique of distances defined by separation indices). An
291 important future direction for METE is to attempt to develop spatial multi-cell predictions using
292 approaches that avoid these shortcomings and the two approaches suggested by Harte (2011) for
293 deriving the METE DDR may provide a useful starting point for future development.

294 Our results suggest that semi-recursive METE differs from spatial patterns observed in
295 nature. This deviation could indicate that the emergent statistical approach to modeling spatial
296 structure is incorrect, with specific biological processes such as dispersal limitation or
297 environmental filtering directly controlling spatial correlation (Condit *et al.*, 2002; Gilbert &
298 Lechowicz, 2004; Karst *et al.*, 2005; Seidler & Plotkin, 2006; Chase, 2007; McGlinn & Palmer,

299 2011). Alternatively it could mean that while the general idea underlying the theory is valid, the
300 specific formulation is wrong. For example it could be that the approaches outlined by Harte
301 (2011) that are more sophisticated in how they handle spatial correlations will be more
302 appropriate or that a generalized version of this kind of recursive approach like that developed by
303 Conlisk et al. (2007) in which the degree of aggregation is a tunable parameter will capture the
304 reality of biological systems more precisely. However, process-, and constraint-based models
305 should not necessarily be treated as mutually exclusive. For example other process-based
306 theories make power-law like predictions for the form of the DDR. In fact, it has recently been
307 suggested that at fine spatial scales most theories will make predictions that are approximately
308 power-law in nature (Nekola & McGill, 2014). This means that simply noting power-law like
309 DDR relationships does not provide a strong method for differentiating among theories. In fact,
310 had we simply looked for power-law like behavior we would have concluded that the semi-
311 recursive METE was consistent with empirical data. However, one of the properties that makes
312 METE such a strong theory is that it makes specific predictions for precise parameters as well as
313 general forms of empirical relationships. This allows it to be more rigorously compared to data
314 and to other theories that predict different parameters values for a similar general form of the
315 DDR (e.g., neutral theory).

316 Our results mirror those of Xiao et al. (2013) and Newman et al. (2014) evaluating the
317 non-spatial aspects of METE. All three studies show that when evaluating the theory using
318 multiple patterns simultaneously some of the predictions perform well and some perform poorly.
319 It is inherently difficult for theories to predict large numbers of patterns simultaneously, which is
320 why evaluating theory in this way provides stronger tests than evaluating single patterns (McGill,
321 2003; McGill *et al.*, 2006). General theories like METE that make multiple predictions are

322 therefore both easier to evaluate and also more broadly useful since they allow a large number of
323 patterns to be predicted from a relatively small amount of information. Because there are many
324 patterns to evaluate it is also more likely that deviations from theory will be identified (White *et*
325 *al.*, 2012). In some cases these deviations may indicate that the theory is fundamentally unsound,
326 but in others it may suggest modifications to the theory to address the observed deviations
327 (White *et al.*, 2012). Whether METE can be modified to address the observed deviations from
328 empirical data remains to be seen. In the case of the DDR, despite its generality, there are a
329 limited number of models that attempt to predict the DDR from first principles (Chave & Leigh,
330 2002; Condit *et al.*, 2002; Zillio *et al.*, 2005; Harte, 2007, 2011; Nekola & McGill, 2014), which
331 means that it may be worth pursuing the METE approach further.

332 METE is one of several general theories in ecology that make many predictions for many
333 aspects of ecological community structure based on only a small amount of information. Our
334 analysis of the semi-recursive formulation of METE's spatially explicit prediction for the DDR
335 suggests that this form of the theory over-predicts the strength of spatial correlation. These
336 results coupled with studies of the species-area relationship suggest that semi-recursive METE
337 accurately predicts the scaling of species occupancy but not spatial correlation. More generally,
338 our results demonstrate that tests of spatial theories that focus solely on the species-area
339 relationship and related patterns are only evaluating part of the spatial pattern, the distribution of
340 occupancy among cells. Evaluating these theories using the DDR in addition to the SAR will
341 help identify cases where the theories are correctly identifying some aspects of spatial structure,
342 but not others, and thus yield stronger tests of the underlying theory. In some cases this will
343 require extending the theory to make additional predictions, but this effort will provide both
344 more testable and more usable theories.

345 **Acknowledgements**

346 We thank John Harte for extensive conversations on the use of maximum entropy in ecology.
347 Robert K. Peet provided data for the oak-hickory and old field pine forests. Jessica Green
348 provided data for the serpentine grassland data. José Arévalo provided data for the oak woodland
349 data. The BCI forest dynamics research project was made possible by National Science
350 Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-
351 0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-
352 9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, support
353 from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the
354 John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Small World
355 Institute Fund, and numerous private individuals, and through the hard work of over 100 people
356 from 10 countries over the past two decades. The plot project is part the Center for Tropical
357 Forest Science, a global network of large-scale demographic tree plots. The Luquillo
358 Experimental Forest Long-Term Ecological Research Program, supported by the U.S. National
359 Science Foundation, the University of Puerto Rico, and the International Institute of Tropical
360 Forestry. This research was supported by a CAREER grant from the U.S. National Science
361 Foundation to EPW (DEB-0953694).

362 **References**

363 Adler, P.B., HilleRisLambers, J. & Levine, J.M. (2007) A niche for neutrality. *Ecology Letters*,
364 **10**, 95–104.
365 Arévalo, J.R. (2013) Spatial analysis and structure of a cross-timber stand in the TallGrass
366 Prairie Preserve (Pawhuska, Oklahoma). *Journal of Forestry Research*, **24**, 47–52.

- 367 Chase, J.M. (2007) Drought mediates the importance of stochastic community assembly.
368 *Proceedings of the National Academy of Sciences of the United States of America*, **104**,
369 17430–17434.
- 370 Chave, J. & Leigh, E.G. (2002) A spatially explicit neutral model of beta-diversity in tropical
371 forests. *Theoretical Population Biology*, **62**, 153–168.
- 372 Coleman, B.D. (1981) On random placement and species-area relations. *Mathematical*
373 *Biosciences*, **54**, 191–215.
- 374 Condit, R. (1998) *Tropical Forest Census Plots*, Springer-Verlag and R. G. Landes Company,
375 Berlin, Germany and Georgetown, Texas.
- 376 Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B., Nunez, P., Aguilar, S.,
377 Valencia, R., Villa, G., Muller-Landau, H.C., Losos, E. & Hubbell, S.P. (2002) Beta-
378 diversity in tropical forest trees. *Science*, **295**, 666–669.
- 379 Condit, R., Salomon Aguilar, Hernandez, A., Perez, R., Lao, S., Angehr, G., Hubbell, S.P. &
380 Foster, R.B. (2004) Tropical Forest Dynamics across a Rainfall Gradient and the Impact
381 of an El Niño Dry Season. *Journal of Tropical Ecology*, **20**, 51–72.
- 382 Conlisk, E., Bloxham, M., Conlisk, J., Enquist, B. & Harte, J. (2007) A new class of models of
383 spatial distribution. *Ecological Monographs*, **77**, 269–284.
- 384 Frank, S.A. (2014) Generative models versus underlying symmetries to explain biological
385 pattern. *Journal of Evolutionary Biology*, **27**, 1172–1178.
- 386 Frank, S.A. (2009) The common patterns of nature. *Journal of Evolutionary Biology*, **22**, 1563–
387 1585.

- 388 Gilbert, B. & Lechowicz, M.J. (2004) Neutrality, niches, and dispersal in a temperate forest
389 understory. *Proceedings of the National Academy of Sciences of the United States of*
390 *America*, **101**, 7651–7656.
- 391 Gilbert, G.S., Howard, E., Ayala-Orozco, B., Bonilla-Moheno, M., Cummings, J., Langridge, S.,
392 Parker, I.M., Pasari, J., Schweizer, D. & Swope, S. (2010) Beyond the tropics: forest
393 structure in a temperate forest mapped plot. *Journal of Vegetation Science*, **21**, 388–405.
- 394 Green, J.L., Harte, J. & Ostling, A. (2003) Species richness, endemism and abundance patterns:
395 tests of two fractal models in a serpentine grassland. *Ecology Letters*, **6**, 919–928.
- 396 Green, J.L. & Plotkin, J.B. (2007) A statistical theory for sampling species abundances. *Ecology*
397 *Letters*, **10**, 1037–1045.
- 398 Haegeman, B. & Etienne, R.S. (2010) Entropy maximization and the spatial distribution of
399 species. *The American Naturalist*, **175**, E74–E90.
- 400 Harte, J. (2011) *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and*
401 *Energetics*, Oxford University Press Inc., New York.
- 402 Harte, J. (2007) *Toward a mechanistic basis for a unified theory of spatial structure in*
403 *ecological communities at multiple spatial scales. Scaling Biodiversity Ecological*
404 *Reviews*. (ed. by D. Storch, P.A. Marquet, and J.H. Brown), pp. 101–126. Cambridge
405 University Press, Cambridge.
- 406 Harte, J., Kitzes, J., Newman & Rominger, A.J. (2013) Taxon Categories and the Universal
407 Species-Area Relationship: A Comment on Šizling et al., “Between Geometry and
408 Biology: The Problem of Universality of the Species-Area Relationship.” *The American*
409 *Naturalist*, **181**, 282–287.

- 410 Harte, J. & Newman, E.A. (2014) Maximum information entropy: a foundation for ecological
411 theory. *Trends in Ecology & Evolution*, **29**, 384–389.
- 412 Harte, J., Smith, A.B. & Storch, D. (2009) Biodiversity scales from plots to biomes with a
413 universal species-area curve. *Ecology Letters*, **12**, 789–797.
- 414 Harte, J., Zillio, T., Conlisk, E. & Smith, A.B. (2008) Maximum entropy and the state-variable
415 approach to macroecology. *Ecology*, **89**, 2700–2711.
- 416 Hubbell, S.P., Condit, R. & Foster, R.B. (2005) Barro Colorado Forest Census Plot Data. URL
417 <http://ctfs.arnarb.harvard.edu/webatlas/datasets/bci>.
- 418 Hubbell, S.P., Foster, R.B., O'Brien, S.T., Harms, K.E., Condit, R., Wechsler, B., Wright, S.J. &
419 de Lao, S.L. (1999) Light-gap disturbances, recruitment limitation, and tree diversity in a
420 neotropical forest. *Science*, **283**, 554–557.
- 421 Karst, J., Gilbert, B. & Lechowicz, M.J. (2005) Fern community assembly: The roles of chance
422 and the environment at local and intermediate scales. *Ecology*, **86**, 2473–2486.
- 423 Locey, K.J. & White, E.P. (2013) How species richness and total abundance constrain the
424 distribution of abundance. *Ecology Letters*, **16**, 1177–1185.
- 425 McDonald, R.I., Peet, R.K. & Urban, D.L. (2002) Environmental correlates of oak decline and
426 red maple increase in the North Carolina piedmont. *Castanea*, **67**, 84–95.
- 427 McGeoch, M.A. & Gaston, K.J. (2002) Occupancy frequency distributions: patterns, artefacts
428 and mechanisms. *Biological Reviews*, **77**, 311–331.
- 429 McGill, B. (2003) Strong and weak tests of macroecological theory. *Oikos*, **102**, 679–685.
- 430 McGill, B.J. (2010) Towards a unification of unified theories of biodiversity. *Ecology Letters*, **13**,
431 627–642.

- 432 McGill, B.J., Maurer, B.A. & Weiser, M.D. (2006) Empirical evaluation of neutral theory.
433 *Ecology*, **87**, 1411–1423.
- 434 McGill, B.J. & Nekola, J.C. (2010) Mechanisms in macroecology: AWOL or purloined letter?
435 Towards a pragmatic view of mechanism. *Oikos*, **119**, 591–603.
- 436 McGlinn, D.J. & Hurlbert, A.H. (2012) Scale dependence in species turnover reflects variance in
437 species occupancy. *Ecology*, **93**, 294–302.
- 438 McGlinn, D.J. & Palmer, M.W. (2011) Quantifying the influence of environmental texture on the
439 rate of species turnover: evidence from two habitats. *Plant Ecology*, **212**, 495–506.
- 440 McGlinn, D.J., Xiao, X. & White, E.P. (2013) An empirical evaluation of four variants of a
441 universal species–area relationship. *PeerJ*, **1**, e212.
- 442 Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., Valencia, R. &
443 Green, J.L. (2008) A general framework for the distance-decay of similarity in ecological
444 communities. *Ecology Letters*, **11**, 904–917.
- 445 Nekola, J.C. & McGill, B.J. (2014) Scale dependency in the functional form of the distance
446 decay relationship. *Ecography*.
- 447 Nekola, J.C. & White, P.S. (1999) The distance decay of similarity in biogeography and ecology.
448 *Journal of Biogeography*, **26**, 867–878.
- 449 Newman, E.A., Harte, M.E., Lowell, N., Wilber, M. & Harte, J. (2014) Empirical tests of within-
450 and across-species energetics in a diverse plant community. *Ecology*.
- 451 Ostling, A., Harte, J., Green, J.L. & Kinzig, A.P. (2004) Self-similarity, the power law form of
452 the species-area relationship, and a probability rule: Reply. *American Naturalist*, **163**,
453 627–633.

- 454 Palmer, M.W. (2005) Distance decay in an old-growth neotropical forest. *Journal of Vegetation*
455 *Science*, **16**, 161–166.
- 456 Palmer, M.W., Peet, R.K., Reed, R.A., Xi, W. & White, P.S. (2007) A multiscale study of
457 vascular plants in a North Carolina piedmont forest. *Ecology*, **88**, 2674.
- 458 Peet, R.K. & Christensen, N.L. (1987) Competition and Tree Death. *BioScience*, **37**, 586–595.
- 459 Plotkin, J.B. & Muller-Landau, H.C. (2002) Sampling the species composition of a landscape.
460 *Ecology*, **83**, 3344–3356.
- 461 Seidler, T.G. & Plotkin, J.B. (2006) Seed dispersal and spatial pattern in tropical trees. *PLoS*
462 *Biology*, **4**, 2132–2137.
- 463 Šizling, A.L., Kunin & Storch, D. (2013) Taxon-and-Area Invariances, Maximum Entropy, and
464 the Species-Area Relationship: A Reply to Harte et al. *The American Naturalist*, **181**,
465 288–290.
- 466 Šizling, A.L., Kunin, W.E., Šizlingová, E., Reif, J. & Storch, D. (2011) Between Geometry and
467 Biology: The Problem of Universality of the Species-Area Relationship. *The American*
468 *Naturalist*, **178**, 602–611.
- 469 Šizling, A.L., Storch, D., Reif, J. & Gaston, K.J. (2009) Invariance in species-abundance
470 distributions. *Theoretical Ecology*, **2**, 89–103.
- 471 Storch, D., Šizling, A.L. & Gaston, K.J. (2003) Geometry of the species-area relationship in
472 central European birds: testing the mechanism. *Journal of Animal Ecology*, **72**, 509–519.
- 473 Storch, D., Šizling, A.L., Reif, J., Polechová, J., Šizlingová, E. & Gaston, K.J. (2008) The quest
474 for a null model for macroecological patterns: geometry of species distributions at
475 multiple spatial scales. *Ecology Letters*, **11**, 771–784.
- 476 Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S*, 4th edn. Springer.

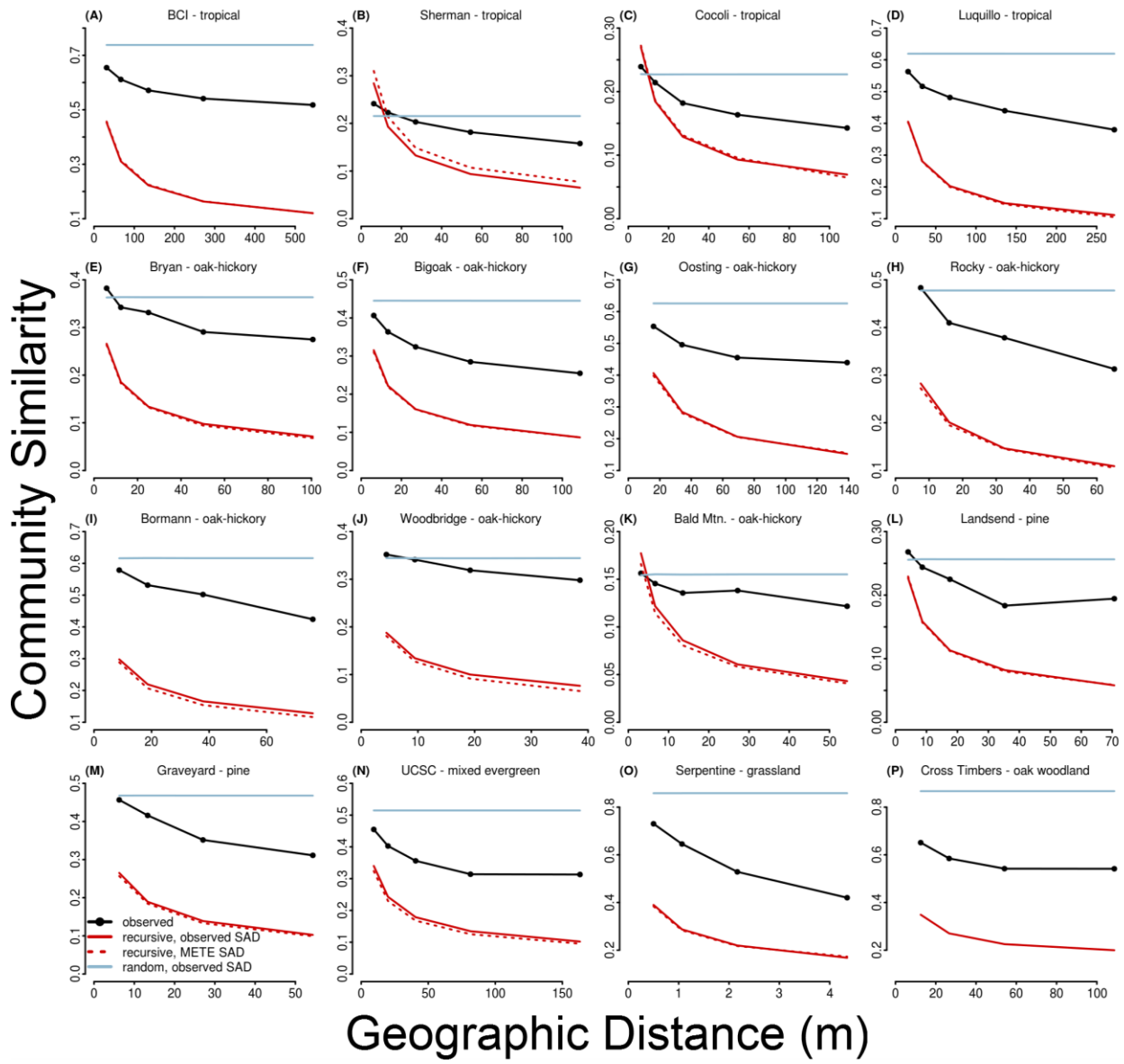
- 477 White, E.P., Katherine M. Thibault & Xiao, X. (2012a) Characterizing species abundance
478 distributions across taxa and ecosystems using a simple maximum entropy model.
479 *Ecology*, **93**, 1772–1778.
- 480 White, E.P., Xiao, X., Isaac, N.J.B. & Sibly, R.M. (2012b) *Methodological tools. Metabolic*
481 *Ecology*, Wiley-Blackwell.
- 482 Xi, W., Peet, R.K., Decoster, J.K. & Urban, D.L. (2008) Tree damage risk factors associated
483 with large, infrequent wind disturbances of Carolina forests. *Forestry*, **81**, 317–334.
- 484 Xiao, X., McGlenn, D.J. & White, E.P. (2013) A strong test of the Maximum Entropy Theory of
485 Ecology. *arXiv*.
- 486 Zillio, T., Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2005) Spatial scaling in model
487 plant communities. *Physical Review Letters*, **95**.
- 488 Zimmerman, J.K., Everham, III, E.M., Waide, R.B., Lodge, D.J., Taylor, C.M. & Brokaw, N.V.L.
489 (1994) Responses of tree species to hurricane winds in subtropical wet forest in Puerto
490 Rico: implications for tropical tree life histories. *Journal of Ecology*, **82**, 911–922.
- 491 **Biosketch**
- 492 Weecology is founded on the belief that better communication and collaboration between
493 empirical and quantitative scientists is necessary for tackling many of the big scientific
494 challenges in ecology. The purpose of Weecology is to facilitate collaborative research through a
495 variety of mechanisms including shared resources, expertise, and web-based collaborative tools.
496 Our hope is help scientists (and ourselves) collaborate and communicate across disciplinary
497 divides and generate higher-quality novel research as a result.

498 Tables:

499 *Table 1.* Summary of the habitat type and state variables of the vegetation datasets. The state
 500 variables are total area (A_0), total abundance (N_0) and total number of species (S_0). A_{\min} and A_{\max}
 501 are the finest and coarsest areas (m^2) examined. Data were collected on woody forest plants with
 502 the exception of the serpentine site which contained herbaceous grassland plants.

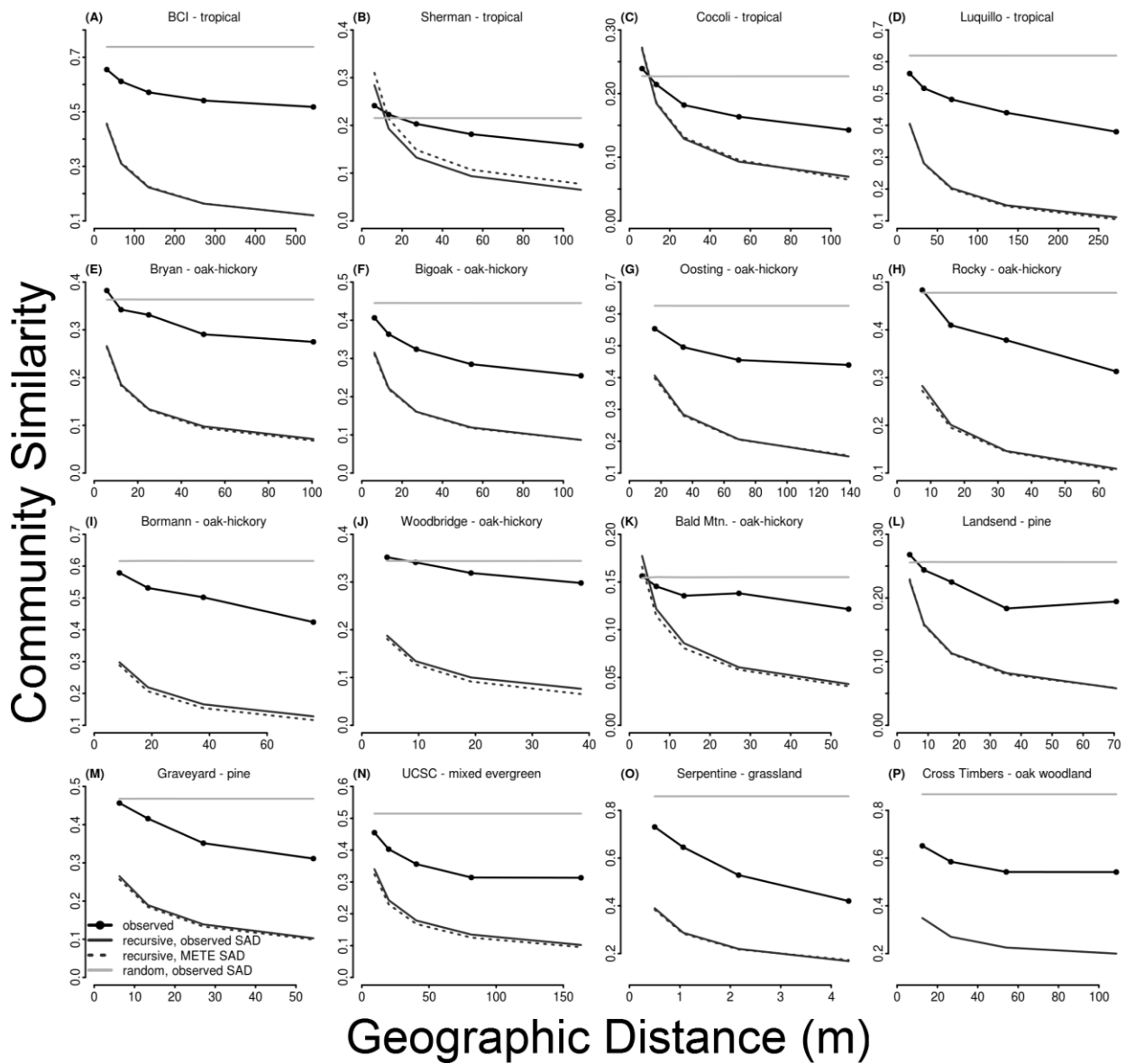
Site name	Habitat type	Ref	A_{\min}	A_{\max}	A_0	N_0	S_0
BCI	tropical	¹⁻³	61.0	62500	500000	205096	301
Sherman	tropical	⁴	2.4	625	20000	7623	175
Cocoli	tropical	⁴	2.4	625	20000	4326	139
Luquillo	tropical	⁵	15.3	15625	125000	32320	124
Bryan	oak-hickory	⁶⁻⁸	2.1	535	17113	3394	48
Big Oak	oak-hickory	⁶⁻⁸	2.4	625	20000	5469	40
Oosting	oak-hickory	⁹	16	4096	65536	8892	39
Rocky	oak-hickory	⁶⁻⁸	3.5	900	14400	3383	37
Bormann	oak-hickory	⁶⁻⁸	4.8	1225	19600	3879	30
Wood Bridge	oak-hickory	⁶⁻⁸	1.2	315	5041	758	19
Bald Mtn.	oak-hickory	⁶⁻⁸	2.4	156	5000	669	17
Landsend	old field, pine	⁶⁻⁸	1.0	264	8450	2139	41
Graveyard	old field, pine	⁶⁻⁸	2.4	625	10000	2584	36
UCSC	mixed-evergreen	¹⁰	5.4	1406	45000	5885	31
Serpentine	serpentine	¹¹	0.3	4	64	37182	24
Cross Timbers	oak woodland	¹²	9.8	2500	40000	7625	7
Ranges			0.3-61.0	4-62500	64-500000	669-205096	7-301

503 ^{1.} Condit (1998), ^{2.} Hubbell et al. (1999), ^{3.} Hubbell et al. (2005), ^{4.} Condit et al. (2004), ^{5.}
 504 Zimmerman et al. (1994), ^{6.} Peet and Christensen (1987), ^{7.} McDonald et al. (2002), ^{8.} Xi et al.
 505 (2008), ^{9.} Palmer et al. (2007), ^{10.} Gilbert et al. (2010), ^{11.} Green et al. (2003), ^{12.} Arévalo (2013)



515

516



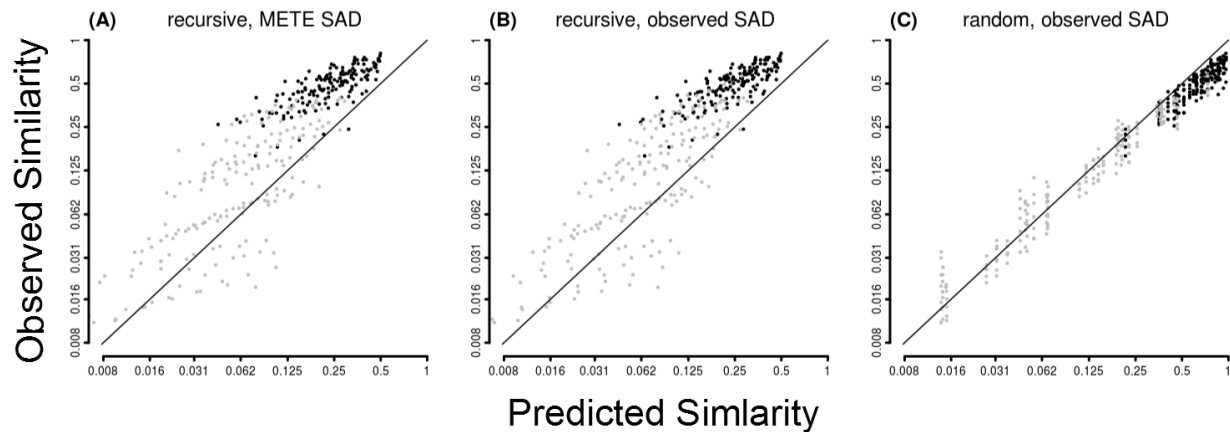
517

518 *Fig 2.* The observed (black line with dots) and predicted distance decay relationships (METE:
 519 dark grey lines, solid for the observed SAD, dashed for the METE SAD; random placement:
 520 light grey line) for each site at a single spatial grain. Community similarity represents the
 521 average of the abundance-based Sørensen index for each spatial lag. The spatial grain displayed
 522 was taken at either 8 or 9 bisections of the total area depending on whether the total extent was a
 523 square or a rectangle respectively. Geographic distance was calculated as the average physical

524 distance between the samples compared at given separation order (see Methods and Fig.1 for

525 additional information).

526



527

528 Fig 3. The log-log transformed one-to-one plots of the predicted and observed abundance-based
529 Sørensen similarity values for the three models across all distances and spatial grains. The solid
530 line is the one-to-one line. The grey points represent values from spatial grains in which the
531 average individual density was low (i.e., less than 10 individuals) and thus both the observed and
532 predicted similarities must be close to zero simply because of a sampling effect.