Research

Ecography 37: 130–136, 2014
doi: 10.1111/j.1600-0587.2013.00403.x

Subject Editor: Pedro Peres Neto. Accepted 3 July 2013

Inferring the similarity of species distributions using Species' Distribution Models

William Godsoe

W. Godsoe (godsoe@nimbios.org), Biological Sciences, Univ. of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand.

A common problem in ecology is our need to reliably compare information on the distributions of distinct species. Since it is not always possible to directly compare the distributions of two species, numerous papers now seek to compare the predictions of Species' Distribution Models (SDMs, estimates of the probability that two species are present, given environmental data). At present, it is not clear when these analyses of SDMs actually reproduce comparisons of species' distributions. I use analytic results and simulations to show that even if the SDMs for two species are identical; their distributions may still differ dramatically. I use this problem to motivate a new index to compare SDMs – an estimate of the Sørensen's similarity of the distributions of two species. I then provide a script to compute this estimator in R. Using simulations I show that this estimator provides a much stronger inference of the similarity of species distributions than previously proposed methods. This work clarifies the interpretation of comparisons of SDMs.

For more than a century, ecologists have attempted to find simple, reliable methods to measure the similarity of the distributions of different species (Forbes 1907, Dice 1945, Whittaker 1952). For our purposes, we can define a species' distribution as the set of geographic locations in which a species is present, and the similarity of species' distributions as a summary of the frequency with which two species occur in the same locations. Ideally, we could make such comparisons using error free data on the geographic distributions of each species across the same set of localities. However, in most instances it is not possible to obtain data from all species at each locality. As a result, many recent studies have instead estimated the probability that each of two species are present given environmental data, a procedure referred to hereafter as Species' Distribution Modeling (SDM); (Araujo et al. 2011, Wilson 2011, Jetz et al. 2012). It is then possible to compare the similarity of the predicted probability of presence generated for each SDM at each location. Unfortunately, we do not know when comparisons of SDMs provide a reliable indication of the similarity of the distributions of two species. In this paper I show that naive comparisons of SDMs are a poor surrogate for the similarity of species' distributions. I then develop a novel estimator for the similarity of species' distributions.

This paper deliberately focuses on the similarity of the distributions of two species, rather than the similarity of their niches (Peterson et al. 1999, Broennimann et al. 2012, Petitpierre et al. 2012). A species' niche can be thought of as the set of environments in which a species is found. The primary reason for focusing on species' distributions is concreteness; if we make an inference about the similarity of species' distributions we can go into the field and verify if our inference was correct. It is much harder to test if our inferences about the similarity of species' niches are correct. Moreover, there will be many applications for which an inference on the similarity of the distributions of multiple species (i.e. how frequently species co-occur) would be more appropriate than an inference on the similarity of their niches (i.e. the similarity of their response to some environmental variable). For example, it will be useful for a naturalist to know the likelihood that two species will occur in the same location just as an epidemiologist may want to know how frequently disease vectors and hosts occur in the same location. As scientists are in the process of organizing large biodiversity databases, and creating species' distribution models for thousands of species (Jetz et al. 2012), a rigorous understanding of how to derive information on the similarity of species' distributions will become ever more valuable. An interesting example of this type of approach is (Araujo et al. 2011) who examine the potential for hundreds of terrestrial vertebrate species to interact with one another across Europe. To do this, they generate SDMs to summarize information on the distributions of individual species. They then infer the potential for interactions among species by examining the overlap of these SDMs.

A key portion of the argument presented in this paper is that the similarity of the predictions of two SDMs can be a misleading surrogate for the similarity of two species' distributions as presented in classical quantitative ecology. To see this, it is useful to review what would constitute identical species' distributions and what would constitute identical SDMs. There is a long tradition of comparing the similarity

of the geographic distributions of two species (Whittaker 1952, 1956, Janson and Vegelius 1981). This is done by surveying many sites, and then summarizing the overlap in the locations in which the two species are found. In this tradition, two species are deemed identical when 'both species always occur together' (Janson and Vegelius 1981). In practice this means that when we compute the similarity of two species using methods such as Sørensen similarity or Jaccard's similarity the distributions of two species would only be identical if every site that contained one species also contained the other. Note that using current terminology from the SDM literature, such a comparison of the similarity of the locations in which two species occur is a comparison in geographic space, as opposed to a comparison of the environments occupied by the two species (Soberon and Nakamura 2009, Peterson et al. 2011).

Many existing comparisons of SDMs focus on obtaining predictions for the chance that each species is found at each location in a study area of interest, and then quantify the similarity of these predictions. Typically this is done by using a distance metric such as Schoener's distance, Hellinger distance or Bray–Curtis distance to compute the differences between the predictions of each SDM across all locations (Rödder and Engler 2011) and then using 1 minus this distance as a measure of the similarity of the SDMs. Using this approach the distance between the predictions of the two SDMs will be 0 when the predicted probabilities of presence of the two SDMs are identical at each location (Legendre and Legendre 2012).

We might of course hope that if the SDMs of two species are identical, the distributions of the two species should likewise be identical. However, in Fig. 1, I sketch why this need not be the case. In this figure I present a conceptual example of two species that infrequently co-occur, and hence have dissimilar distributions. However both species are more likely to occur in warm climates than in cold climates. As illustrated, SDMs fit individually to the distribution of each species would be quite similar (i.e. the predicted probability of presence for each species at each location can be the same).

In this paper I develop a formal mathematical argument to demonstrate why identical SDMs frequently imply dissimilar species' distributions. Recognizing this, I then derive and analyze an alternative index of the similarity of SDMs – an estimate of the Sørensen similarity of species' distributions. Using simulations, I show that this index produces strong predictions on the similarity of species' distributions, while a direct analysis of the similarity of SDMs does not.

Figure 1. An example of why similar Species' Distribution Models (SDMs) need not imply similar species' distributions. This figure presents a schematic of temperature gradients across the eastern United States, with light portions of the map indicating warmer climates and dark regions indicating colder climates. This map depicts presences for two species (circles for species 1, diamonds for species 2), both of which specialize in warmer climates. A SDM for species 1 would infer that species 1 is more likely to be found in warm climates, just as a SDM for species 2 would infer that species 2 is more likely to occur in warm climates. Even if these SDMs are similar, the two species rarely co-occur (the two species co-occur at only a single site; indicated by an arrow). As such, the species have dissimilar distributions, but similar SDMs.

The model

Why identical SDMs imply dissimilar species' distributions

I base my analysis on a simple interpretation of SDMs: they create models to predict whether a species will be present in the set of locations found within a geographic region G. This assumption is particularly useful for a large class of presence/absence SDMs such as generalized linear models (GLMs) and boosted regression trees (BRTs) that assume that presence/absence data is binomially distributed. These models can be interpreted as a method to infer P_{ij}, the probability that species i is present in location j; (Crawley 2005). Note that for a sufficiently large dataset the estimate of the probability that our species is present in location j will approach the true probability that our species is present.

A common alternative to fitting presence/absence SDMs is to fit a model that distinguishes presences and pseudoabsences (a surrogate for absence points constructed by randomly sampling points within a study area). Some methods that use presences and pseudoabsences such as the popular Maxent algorithm also produce logistic output reminiscent of the output of a presence/absence GLM (Phillips and Dudik 2008). However, the interpretation of these presence/pseudoabsence models is more complex as these methods produce a biased (incorrect) estimate on the prevalence of each species (Phillips et al. 2009). GLM and BRT SDMs fit using presence/pseudoabsence data also produce logistic output, but with similar caveats to Maxent models.

I will show that computing the distance between the predictions of SDMs for two species is frequently an inappropriate way to measure the similarity of species' distributions. To do this I will use a new proof to

demonstrate that for a broad range of parameter values, species with non-identical distributions can produce SDMs that are identical. SDMs are identical when the predictions are identical everywhere ($P_{1j} = P_{2j}$ for all locations in j). To simplify notation, I will denote the probability that either species is present at location j as P_j. The probability that both species are present is thus P_j^2. To say that the distributions are identical implies that the presence of one species guarantees that both species will be present. We lose this guarantee whenever the probability that both species are present is less than the chance of observing one species (whenever $P_j^2 < P_j$). This will be true any time that P_j is greater than 0 but less than 1. Combining these observations, we should expect 2 identical SDMs to imply dissimilar species' distributions whenever $0 < P_j < 1$. In this case, any method that deems two SDMs to be identical when the SDMs furnish identical predictions will give an incorrect picture of the similarity of the distributions of the two species. Distance metrics in general, and Hellinger Distance and Schoener's distance in particular meet this criteria (Legendre and Legendre 2012).

Alternative index: expected shared presences

Rather than directly comparing the similarity of the predictions of the SDMs, I propose that we use the predictions generated by SDMs to estimate the similarity of the distributions of each species. To do this, we need to specify the biological property we wish to estimate. One of the most pertinent questions is how frequently would either species encounter the other. We could formalize this problem with the question: what fractions of presences come from locations where the species co-occur?

$$\underset{similarity}{Sørensen} = \frac{presences\ from\ locations\ with\ both\ taxa}{presences\ for\ species\ 1 + presences\ for\ species\ 2} \quad (1)$$

In a recent paper (Godsoe 2012) I explain why this quantity is equal to Sørensen's similarity index. Sørensen's similarity requires presence/absence data and so we cannot compute it directly from the predicted probability that each species is present.

In Supplementary material Appendix 1, I derive an estimator for the Sørensen similarity of species' distributions, using the output of presence/absence SDMs to compute the Expected fraction of Shared Presences (ESP):

$$ESP = \frac{2\sum_j P_{1j}P_{2j}}{\sum_j \left(P_{1j} + P_{2j}\right)} \quad (2)$$

I propose using this index to compare the similarity of the predictions of SDMs, or conversely 1 minus this value as a measure of the dissimilarity of species' distributions. Equation (2) can be thought of as a generalization of Sørensen's similarity that ranges between 0 and 1 (Supplementary material Appendix 1). In Supplementary material Appendix 2 I provide an R script to calculate this function and it will be made available in a subsequent release of the program ENMtools (Warren et al. 2010).

Simulations

Using simulated data, I test our ability to estimate the similarity of the distributions of two species using ESP, and two representative metrics: Schoener's D, and Hellinger Distance. I also test our ability to model the similarity of species' distributions by using a threshold to convert the continuous output of the SDM into predicted presences, then measuring the similarity of these predictions.

I focus on simulations where the true probability that each species is present is described by the logistic function (Godsoe 2010, Meynard and Kaplan 2012, 2013). In a logistic model the probability of presence depends on a value (η) that encapsulates the effects of environmental variables on a species' distribution. In turn, I transform η to obtain a probability that a species is present given information on predictor variables, using a logistic transformation:

$$P(present) = \frac{1}{1 + e^{-\eta}} \quad (3)$$

I focus on the effect of a single artificial environmental variable; hereafter referred to as elevation. Simulations used 1000 locations across a range of simulated elevations from 0 to 1000 m and were divided into two rounds. In the first round, the probability of presence for each species increased with elevation, though the rate at which it increased could vary between species. In this round of simulations $\eta = a + b \times elevation$ where $a = -5, -4, -3, -2, -1$ and $b = 0.002, 0.004, 0.006, 0.008, 0.010$. These parameter values change the prevalence of each species from approximately 20/1000 sites for the parameter values $a = -5$, $b = 0.002$ to approximately 900/1000 sites when $a = -1$, $b = 0.01$. This simulation round also allowed for substantial differences in the overlap between pairs of species. For some parameter values more than half of the environmental gradient was suitable to a species (Fig. 2B), while for others, only high elevation sites had a high probability of being suitable. In the second round the probability of presence peaked at intermediate elevations, in these simulations $\eta = a + b \times elevation + c \times elevation^2$ where $a = -5$, $b = 0.022, 0.024, 0.026, 0.028, 0.030$, $c = -0.0000350, -0.0000365, -0.0000375, -0.00003875, -0.0000400$. These parameters allow for a species to be present at intermediate elevations, but the precise location of this peak and the prevalence of each species are free to vary (Fig. 2). In this round, the prevalence ranged from approximately 35/1000 sites for the parameter values $b = 0.022$, $c = -0.0000400$ to approximately 350/1000 sites when $b = 0.03$, $c = -0.0000350$. In each round of simulations there were 25 combinations of parameter values for each of two species, giving a total of $25^2 = 625$ simulations per round.

Since I focus on SDM methods that assume presences are binomially distributed, I likewise simulate each presence by taking a draw from a binomial distribution where the probability of presence is given by Eq. (3). This was done in R (R Development Core Team).

I fit SDMs using the GLM algorithm in three ways. 1) fitting a model using all presences and all absences. 2) using all presences and all available locations as pseudoabsences (1000 locations with elevations ranging from 1 to 1000 m).

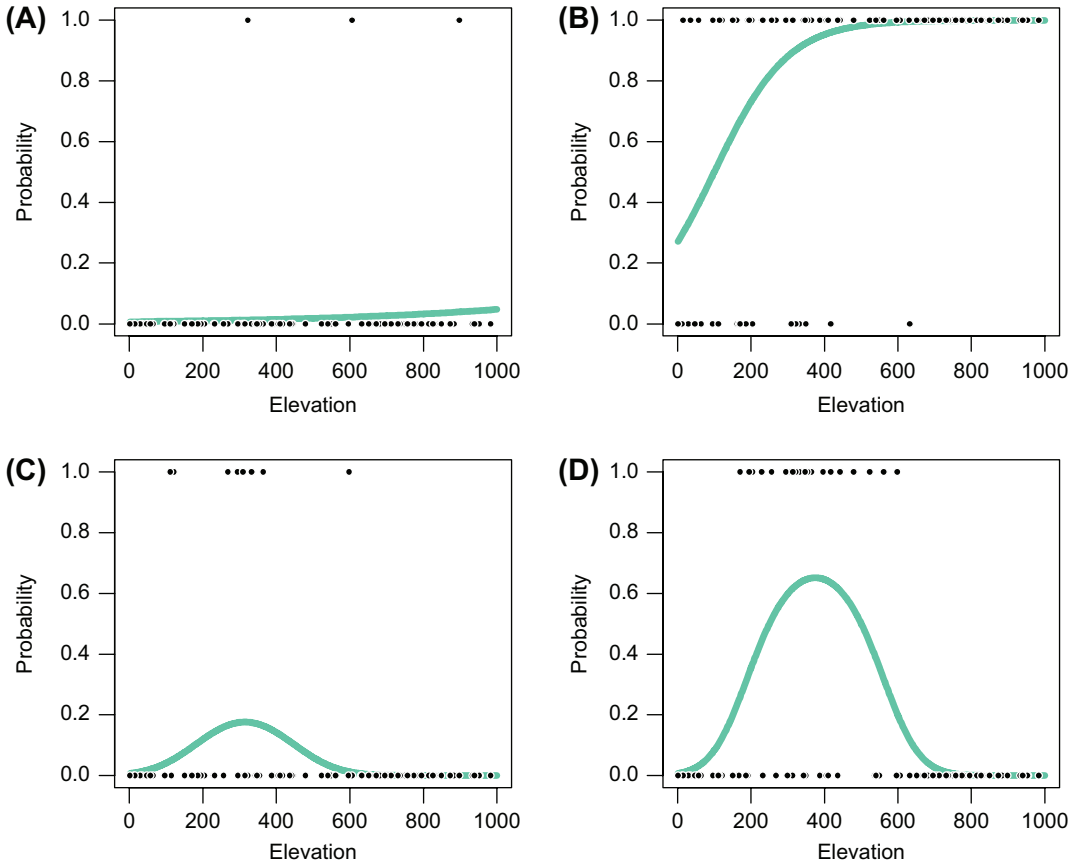

Figure 2. Representative illustrations of species distributions simulated using parameter values from the current study. In these panels the green lines represent the probability of presence generated by a logistic function while the black points represent observations of the species' distribution with points at 1 representing presences and points at 0 representing absences. In (A) and (B) the probability of presence increases with elevation. However, in (A) the probability of presence increases only slightly and so this species is rare, even at high elevations. In contrast the probability that species B is present increases dramatically with elevation. As a result this species is present at nearly all the high elevation sites. In (C) and (D) the probability of presence has peaks at intermediate elevations, however these species differ in their prevalence. In panel (C) the simulated species is quite rare, even at intermediate elevations while the simulated species in panel (D) is more common at intermediate elevations. Parameter values: (A) $a = -5$, $b = 0.002$ (B) $a = -1$, $b = 0.010$ (C) $a = -5$, $b = 0.022$, $c = 0.000035$ (D) $a = -5$, $b = 0.03$, $c = 0.00004$.

3) using all presences and an equal number of pseudoabsences (randomly selected background points). Note that if two species were to have identical distributions, and we fit our models using GLM and with the same environmental data, then we should expect the probability of presence inferred for each to be identical. With the BRT algorithm the SDMs need not be identical, but we should still expect them to be similar (Elith et al. 2008).

I then contrasted three potential approaches for measuring the similarity of the resulting SDMs. First, I estimated the similarity of the distributions using ESP as described in Eq. 2. Second, I tested the common practice of converting the raw output generated by SDMs into predicted presences and predicted absences with a threshold. To do this I applied the sensitivity = specificity threshold implemented in the Presence Absence R package (Liu et al. 2005, Freeman 2007) to distinguish presences from absences in each SDM. This method finds a threshold such that there are an equal number of false presences and false absences and is a frequently recommended way to use information from a Receiver Operating Curve plot to select thresholds (Liu et al. 2005). Third, I measured the similarity of the SDMs using two commonly recommended distance metrics, Schoener's D and Hellinger Distance (Warren et al. 2008, Rödder and Engler 2011).

Finally, I computed how well each metric of the similarity of SDMs did at predicting the Sørensen similarity of the species' distributions by computing the coefficient of determination (R^2). In total this resulted in 15 000 separate ways to make comparisons using SDMs (4 summary statistics × 3 methods of fitting SDMs × 2 simulation rounds × 625 parameter combinations).

Results of simulations

Using presence/absence data, ESP was a strong predictor of the similarity of the two species' distributions (Fig. 3A; Table 1). When using presences and a constant number of pseudoabsences ESP exaggerated the similarity of species' distributions, while there was no correlation between ESP and the similarity of species' distributions if the SDMs were fit using presences and an equal number of pseudoabsences. At best, the distance metrics and the overlap of thresholded

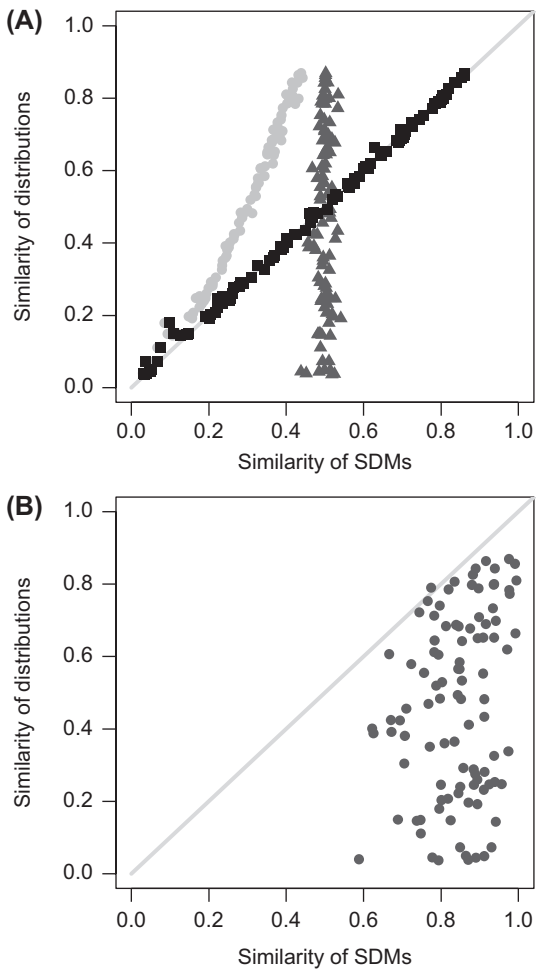

Figure 3. Our ability to estimate the (Sørensen) similarity of the distributions of two simulated species versus the similarity of SDMs inferred using four different methods. (A) Expected shared presences (the formula in Eq. 2) is tightly correlated with Sørensen's similarity of the species' distributions when fitting SDMs with presence/absence data (black squares). When using presences and all available locations as pseudoabsences, the expected shared presences estimator (light grey circles) underestimates the similarity of species distributions. When using presences and an equal number of pseudoabsences (dark grey triangles) there was no correlation between the similarity of species' distributions and the similarity of the SDMs. (B) Comparisons of SDMs using Schoener's D overestimate the similarity of species' distribution. For clarity, this figure summarizes a random sample of size 100 from simulation round 1 in which probability of presence increases with elevation.

output were weakly correlated with the similarity of species' distributions, with R^2 values generally less than 0.2 (Table 1). Schoener's D over-estimated the similarity of species' distributions (Fig. 3B). The same was true of Hellinger Distance and a comparison of the similarity of thresholded outputs, two methods whose predictions were poorly correlated with the Sørensen similarity of species' distributions (Table 1).

A common concern is that some properties of SDMs change depending on the region being sampled (Lobo 2010, Barbet-Massin 2012). To ensure that my results are robust to the study region chosen, I developed a modified simulation in which I extended the environmental gradient by 1000 m (such that it spans from 1:2000 m). The performance of ESP using presence/absence data and each of the two variants of the presence pseudoabsence models were virtually unchanged. These simulations are available upon request.

Discussion

There is an ongoing struggle to better integrate SDMs into ecological theory (Holt 2009, Soberon and Nakamura 2009). My results demonstrate that an incautious comparison of SDMs will provide a misleading indication of the similarity of species' distributions. To remedy this problem I suggest a specific change in the method used to compare SDMs. Below, I discuss how this work affects the interpretation of SDMs comparisons, the applicability of my conclusions to other analyses of the similarity of species' niches and the implications of this work for ecology.

Intuitively, we might expect that similar SDMs imply similar species' distributions. As I have shown, this is not the case for models of the probability that species are present. Unless these models predict the probability of presence with certainty ($P_{ij} \in 0,1$) the appropriate interpretation of the output of SDMs is that we are unsure whether a species will be present or not, and it is even less likely that two such species will be present at the same location. Most available metrics ignore this distinction. As a result, current methods that directly measure the similarity of SDMs can be a poor surrogate for the similarity of species' distributions (Table 1). However, ESP provides a strong surrogate for the similarity of species' distributions.

The distinction between comparisons of SDMs and comparisons of species' distributions that I present can be relevant across many spatial scales. For this distinction to matter, all we need to show is that the probability of presence is between 0 and 1 at the spatial scale of interest. If there is any doubt as to whether this problem is relevant for a particular set of presences/absence SDMs fit using methods such as BRT or GLM, simply examine the predicted probabilities of presence at each location. ESP will be useful if the predicted probabilities of presence are commonly between 0 and 1. Biologically, ask if it should be possible in principle to predict every presence of each species at the scale of interest using only the environmental variables used to fit the SDM. To take an example from my own work, I have previously tried to predict the distribution of Joshua trees *Yucca brevifolia* across a large spatial scale the Mojave Desert in the southwestern United States, using climate data (Godsoe et al. 2009). We know that the distribution of this species depends on non-climatic variables such as soil composition (Rowlands 1978), but it is much more difficult to incorporate this information into SDMs. As a result, using available climate data, we should not expect to be able to predict the distribution of Joshua tree with complete certainty, even at relatively large spatial scales. If we can't expect to perfectly predict the distribution with complete certainty using SDMs, then we shouldn't expect raw comparisons of SDMs to produce an unambiguous surrogate for comparison of the similarity of species' distributions.

To derive my measure of the similarity of two SDMs I have used the fact that, all else being equal, the probabilities

Table 1. A comparison of our ability to infer the similarity of two species' distributions simulated using Eq. 3 and parameter values described in the main text. Each entry describes the coefficient of determination (R^2) between a summary statistic of the similarity of two SDMs and the Sørensen similarity of the distributions of two species. Rows represent summary statistics described in the main text while columns represent separate ways to fit SDMs.

	Presence/ absence	Presence/constant number of pseudoabsences	Presence equal number of pseudoabsences
Simulation round 1: probability of presence increases monotonically with elevation			
Estimated shared presences	0.997	0.976	0.001
Schoener's D of SDM output	0.065	0.093	0.042
Hellinger D of SDM output	0.033	0.086	0.04
Overlap of threshold output	0.182	0.098	0.061
Simulation round 2: probability of presence peaks at intermediate elevations			
Estimated shared presences	0.951	0.942	0.000
Schoener's D of SDM output	0.171	0.176	0.155
Hellinger D of SDM output	0.194	0.171	0.177
Overlap of threshold output	0.141	0.146	0.135

of presence for each species are assumed to be independent using popular algorithms such as BRT and GLM. I would like to emphasize that my goal is to judge comparisons of SDMs by their ability to reproduce comparisons of species' distributions, not to advocate the assumption of independence. This assumption arises naturally when we convert the output of SDMs such as BRTs or GLMs back into species' distributions. Thus, when we compare two SDMs, and we do nothing to explicitly model co-occurrence probability we should acknowledge our implicit assumption that co-occurrences are independent.

A crucial assumption of ESP is that we use SDMs that accurately infer the probability that each species is present. In simulations ESP produced a strong estimate of the similarity of species' distributions when using data derived from presence/absence SDMs, a method likely to produce predicted probabilities of presence that are close to the true probability of presence.

It is much harder to infer the probability of presence accurately using presence/pseudoabsence data (Phillips et al. 2009). I would suggest caution when using ESP with most pseudoabsence methods, but see Royle et al. (2012). ESP produced an imperfect estimate of the similarity of species' distributions when using presence/pseudoabsence data. None of the methods tested produced a strong estimate of the similarity of species' distributions when using presences and an equal number of pseudoabsences, however ESP was particularly poor estimator and I do not recommend using ESP when comparing SDMs fit with presences and an equal number of pseudoabsences. ESP's estimate of the similarity of two SDMs fit using presences and a constant number of pseudoabsences was correlated with Sørensen's similarity of the distributions. Note that these simulations assumed consistent sampling and perfect detection probabilities for both species, assumptions that may be optimistic in practice. Changing the number of pseudoabsences may change the estimate provided by ESP. For example, increasing the number of pseudoabsences will decrease the proportion of presences used by the SDM. In turn, this could induce a downward bias in the predicted distribution similarity. Extending the study region to encompass a broader range of conditions did little to affect the strength ESP. This suggests that our ability to infer the similarity of species distributions may be less sensitive to changes in the region of interest than is our ability to distinguish presences from absences or make inferences about a species' niche (Lobo 2010, Barbet-Massin 2012).

In this paper I have deliberately avoided an exhaustive survey of methods that could be used to compare species' distributions or SDMs. In view of the large number of metrics that could potentially be used to compare the similarity of species' distributions or the similarity of SDMs (Hubalek 1982, Legendre and Gallagher 2001, Wilson 2011, Legendre and Legendre 2012) I believe that a such a survey would cause more confusion than it would alleviate. Instead I have outlined a general argument for why identical SDMs frequently imply dissimilar species' distributions. This argument applies to the metrics that are most commonly used by biogeographers such as Schoner's D and Hellinger Distance (Warren et al. 2008, Rödder and Engler 2011), and I have illustrated the crucial predictions of this argument with simulations, specifically that analyses of SDMs using these metrics are poorly correlated with the similarity of species' distributions (Table 1), and that they provide over-estimate the similarity of species distributions (Fig. 3A).

Arguably the most biologically meaningful way to summarize information on the distributions of to species is to determine how often they co-occur. However, other approaches are possible. For example, one might determine if two species co-occur at a rate different from what we would observe under the null expectation that co-occurrence is governed by chance alone (Ovaskainen et al. 2010). Hubalek (1982) provides an explicit contrast of these two related problems and argues in favor of measures of co-occurrence such as Sørensen's similarity. Note that (Ovaskainen et al. 2010) used a Bayesian approach to fit a multivariate logistic model across multiple species. Either Bayesian methods or single species SDMs (Elith et al. 2006) may offer advantages in some applications, however, regardless of the method of statistical inference employed I would argue that it makes sense to make inferences about the frequency of co-occurrence using methods akin to those advocated here.

An increasing number of applications in ecology compare statistical inferences generated from data on species' distributions (Peterson et al. 1999, Broennimann et al. 2007, Rödder

and Engler 2011). However there are still substantial ambiguities in the interpretation of these analyses. My work shows 1) that there is a strong link between comparisons of SDMs and comparisons of distributions and 2) existing measures of the similarity of SDMs obscure this link. The similarity index I propose provides a simple biologically interpretable measure of the similarity of species' distributions. As I have shown, this measure accurately predicts differences in species' distributions that are obscured by commonly applied similarity metrics. Together this helps to forge a better connection between seemingly abstract comparisons of species' distribution models and the species' distributions that they purport to represent.

Acknowledgements – H. Buckley, N. Baker, J. Tylianakis, H. Chapman, E. Moltchanova and R. Gamlen-Greene provided helpful comments. A. Guissan provided a detailed, thoughtful and critical assessment of a previous draft. WG was supported by a post-doctoral fellowship at the National Inst. for Mathematical and Biological Synthesis, an institute sponsored by the National Science Foundation, The U.S. Dept of Homeland Security, the U.S. Dept of Agriculture through the National Science Foundation Award EF-0832858, with additional support from the Univ. of Tennessee, Knoxville.

References

Araujo, M. B. et al. 2011. Using species co-occurrence networks to assess the impacts of climate change. – Ecography 34: 897–908.

Barbet-Massin, M. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? – Methods Ecol. Evol. 3: 327–338.

Broennimann, O. et al. 2007. Evidence of climatic niche shift during biological invasion. – Ecol. Lett. 10: 701–709.

Broennimann, O. et al. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. – Global Ecol. Biogeogr. 21: 481–497.

Crawley, M. J. 2005. Statistics an introduction using R. – Wiley.

Dice, L. R. 1945. Measures of the amount of ecologic association between species. – Ecology 26: 297–302.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – Ecography 29: 129–151.

Elith, J. et al. 2008. A working guide to boosted regression trees. – J. Anim. Ecol. 77: 802–813.

Forbes, S. A. 1907. On the local distribution of certain Illinois fishes: an essay in statistical ecology. – Illinois State Laboratory of Natural History.

Freeman, E. 2007. PresenceAbsence: an r package for presence–absence model evaluation. – USDA Forest Service, Rocky Mountain Research Station.

Godsoe, W. 2010. Regional variation exaggerates ecological divergence in niche models. – Syst. Biol. 59: 298–306.

Godsoe, W. 2012. Are comparisons of species distribution models biased? Are they biologically meaningful? – Ecography 35: 769–779.

Godsoe, W. et al. 2009. Divergence in an obligate mutualism is not explained by divergent climatic factors. – New Phytol. 183: 589–599.

Holt, R. D. 2009. Bringing the hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. – Proc. Natl Acad. Sci. USA 106: 19659–19665.

Hubalek, Z. 1982. Coefficients of association and similarity, based on binary (presence–absence) data: an evaluation. – Biol. Rev. 57: 669–689.

Janson, S. and Vegelius, J. 1981. Measures of ecological association. – Oecologia 49: 371–376.

Jetz, W. et al. 2012. Integrating biodiversity distribution knowledge: toward a global map of life. – Trends Ecol. Evol. 27: 151–159.

Legendre, P. and Gallagher, E. D. 2001. Ecologically meaningful transformations for ordination of species data. – Oecologia 129: 271–280.

Legendre, P. and Legendre, L. 2012. Numerical ecology. – Elsevier.

Liu, C. et al. 2005. Selecting thresholds of occurence in the prediction of species distributions. – Ecography 28: 385–393.

Lobo, J. M. 2010. The uncertain nature of absences and their importance in species distribution modelling. – Ecography 33: 103–114.

Meynard, C. N. and Kaplan, D. M. 2012. The effect of a gradual response to the environment on species distribution modeling performance. – Ecography 35: 499–509.

Meynard, C. N. and Kaplan, D. M. 2013. Using virtual species to study species distributions and model performance. – J. Biogeogr. 40: 1–8.

Ovaskainen, O. et al. 2010. Modeling species co-occurence by multivariate logistic regression generates new hypotheses on fungal interactions. – Ecology 9: 2514–2512.

Peterson, A. T. et al. 1999. Conservatism of ecological niches in evolutionary time. – Science 285: 1265–1267.

Peterson, A. T. et al. 2011. Ecological niches and geographic distributions. – Princeton Univ. Press.

Petitpierre, B. et al. 2012. Climatic niche shifts are rare among terrestrial plant invaders. – Science 335: 1344–1348.

Phillips, S. J. and Dudik, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. – Ecography 31: 161–175.

Phillips, S. J. et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. – Ecol. Appl. 19: 181–197.

Rödder, D. and Engler, J. O. 2011. Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. – Global Ecol. Biogeogr. 20: 915–927.

Rowlands, P. G. 1978. The vegetation dynamics of the Joshua tree (*Yucca brevifolia* Engelm.) in the southwestern United States of America. – PhD thesis, Univ. of California, Riverside.

Royle, J. A. et al. 2012. Likelihood analysis of species occurence probability from presence-only data for modeling species distributions. – Methods Ecol. Evol. 3: 545–554.

Soberon, J. and Nakamura, M. 2009. Niches and distributional areas: concepts, methods, and assumptions. – Proc. Natl Acad. Sci. USA 106: 19644–19650.

Warren, D. L. et al. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. – Evolution 63: 2868–2883.

Warren, D. L. et al. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. – Ecography 33: 607–611.

Whittaker, R. H. 1952. A study of summer foliage insect communities in Great Smoky Mountains National Park. – Ecol. Monogr. 22: 2–44.

Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. – Ecol. Monogr. 26: 1–80.

Wilson, P. D. 2011. Distance-based methods for the analysis of maps produced by species distribution models. – Methods Ecol. Evol. 2: 623–633.

Supplementary material (Appendix ECOG-00403 at <www.oikosoffice.lu.se/appendix>). Appendix 1–2.