Accounting for shifts in the frequency of suitable environments when testing for niche overlap

William Godsoe[1,2]* and Bradley S. Case[3]

[1]Bio-Protection Research Centre, Lincoln University, PO box 85084, Lincoln 7647, New Zealand; [2]Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand; and [3]Department of Informatics and Enabling Technologies, Faculty of Environment Society & Design, Lincoln University, PO Box 85084, Canterbury 7647, New Zealand

Summary

1. Organismal biologists need to detect and anticipate the effects of niche evolution. An increasing number of studies use information on changes in species' distribution to test for niche evolution. Typically, these studies will test to see whether the range of environments occupied by a species is similar in each biogeographic region.

2. Niche evolution can change the occurrence probability (the frequency with which we observe a species) rather than the range of environments occupied. Using simulation, we test whether 13 previously implemented methods, and a novel approach (Expected Shared Presences), can identify shifts in occurrence probability.

3. We show that even the best previously implemented methods detect shifts in the range of environments occupied, but provide poor inferences for shifts in occurrence probability. The expected shared presences approach provides a strong estimate of niche evolution due to shifts in niche overlap and shifts in occurrence probability.

4. This work indicates that previous studies may have missed a substantial source of niche evolution. We argue that rigorous tests of niche overlap must account for shifts in both the range of environments occupied and the frequency of suitable environments.

Key-words: niche overlap, species invasion, ecological niche model, occurrence probability

Introduction

A fundamental problem in ecology is our need to predict how species will respond to novel environmental conditions. As a result, many studies now test for shifts in a species' ecology among distinct geographic regions. A key conceptual framework for these studies is the idea of niche overlap (Peterson, Soberon & Sanchez-Cordero 1999; Broennimann *et al.* 2007; Warren, Glor & Turelli 2008) (Table 1). For our purposes, the niche is defined as the set of environments suitable to a species (Hutchinson 1957). If a species occurs in two geographic regions, then niche overlap is the similarity of the niche of the species in two regions. Low niche overlap would imply dramatic changes in the ecology of the species among regions. This could be due to several factors, including evolution, or phenotypic plasticity. Additionally, the set of environments found in each region may be so dissimilar that it would be difficult to evaluate niche overlap with observations of the species' distributions (Godsoe 2010b, 2012; Barve *et al.* 2011). The niche overlap framework can also be used to compare the ecology of two separate species, but for the sake of simplicity, we will only describe the similarity of populations of a single species that occur in distinct geographic regions. The niche conservatism hypothesis states that changes in niche overlap are sufficiently slow to allow us to use information on the occurrence of a species in one region to predict occurrences in another region. In spite of the promise of niche overlap studies, the relationship between our observations of a species' distribution and a full description of its ecological niche is at best indirect, making it unclear when comparisons of the environments occupied by a taxon in two regions can be used to demonstrate niche conservatism. Here, we show that many tests of niche conservatism using observational data on species' distributions can obscure one of the most important signals of niche shifts – changes in the probability that environments are suitable to a species.

It is useful, at this point, to outline common features of the statistical tests that have been proposed to identify niche conservatism. Steps in common to such tests are to (i) use information on the distribution of a species to delimit the environments occupied in each region; (ii) select a set of environments that can be used to evaluate the 'response' of a species to the environment in each region; and (iii) compare the similarity of the response; if the response in each region is similar, infer strong niche overlap.

To describe the response of a species to the environment, ecologists summarize the environmental conditions in which the species is most likely to occur. Many environmental variables can potentially be used, but it is generally not possible to fit a model that includes all of the variables that matter to a particular species. As a result, models usually contain inherent uncertainty. One of the best known ways to use this information is to fit an ecological niche model (ENM), here defined as a statistical model that seeks to use information on the presences (and possibly absences) of a species to predict which

*Corresponding author. E-mail: william.godsoe@canterbury.ac.nz

environments will be suitable to a species and which will not. Since it is often difficult to obtain reliable absence data, many ENM methods will use pseudoabsences – a random selection of points collected throughout a study region – instead of actual absences.

A subject of ongoing controversy is step (2), determining which set of environments should be used to compare the similarity of ENMs. Some researchers will compare the 'response' of two species by projecting the response of each species onto the set of locations found in some geographic region (hereafter a comparison in geographic space) (Peterson, Soberon & Sanchez-Cordero 1999; Warren, Glor & Turelli 2008). Others make comparisons in environmental space by creating axes that summarize environmental variation and projecting our predictions of the frequency of presences onto these environmental axes (Broennimann *et al.* 2012). In the light of this difference of opinion, we make comparisons in both geographic and environmental space.

Many metrics have been used to compare the similarity of species responses in two regions. A common approach is to modify an ecological distance metric to compute a score that varies between 0 and 1, with a score of 1 indicating that the response of the two species is identical across the set of environments used in step 2, and a score of 0 indicating complete differences in the niches of the two taxa. We modify this definition slightly by requiring that two niches are only identical if every environment suitable to one species is also suitable to the other. This is a natural extension of the goals of previous papers to determine whether the niches of two species are effectively indistinguishable (niche equivalency) (Graham *et al.* 2004; Warren, Glor & Turelli 2008). This definition also has an elegant empirical meaning; if ecologists were to observe a sample of locations (Hooper *et al.* 2008), they would deem two taxa to have identical niches if every environment that was suitable for one species was also suitable for the other. At present, it is not clear that analyses of biogeographic data using currently available distance metrics satisfy this criterion. Tests of niche overlap must use a limited number of environmental variables when computing a species' response to the environment. As a result, niche shifts can occur along both measured and unmeasured axes. The signal along measured axes is reasonably well studied (Holt & Gomulkiewicz 1997; Kirkpatrick & Barton 1997). A shift in the range of temperatures that are suitable to a species should result in a shift in the range of temperatures in which a species is present. We will refer to such a shift as a change in niche overlap. Less attention has been given to the consequences of shifts in unmeasured environmental variables. Unmeasured environmental variables can alter the occurrence probability of a species (the frequency with which a species is observed in a given set of circumstances). This can occur even in the absence of changes in niche overlap (Godsoe 2010a). Consider a species that has requirements for both temperature (a variable that is easy to measure at large biogeographic scales) and soil pH (a variable that is difficult to measure at large biogeographic scales). If the species uses a wide range of soil pH, then most locations with an appropriate temperature will be suitable for that species; consequently, its

occurrence probability will be high. However, if the species requires a narrow range of pH, then even locations with an ideal temperature will frequently be unsuitable because an unmeasured facet of the environment (pH) makes the site unsuitable, and as a result the probability that a species will be present is much lower. Thus, a shift in the range of pH that is suitable to the species in Fig. 1 alters occurrence probability. In its native range, this species will rarely be present even in ideal temperatures (Fig. 1b), while in its invasive range, it always occurs at ideal temperatures (frequency = 1; Fig. 1d).

Though many studies have identified conditions under which the similarity of niches can be inferred (Elith & Graham 2009; Godsoe 2010b, 2012; Barve *et al.* 2011; Rödder & Engler 2011; Soberon & Peterson 2011), these studies have not tested our ability to detect shifts in occurrence probability. A recent study investigated the best way to use information from species distribution models (SDMs) to infer the similarity of the distributions of two species that occur within the same geographic region (Godsoe 2014). This work suggested that information on occurrence probability is easy to misinterpret when comparing SDMs. However, it was unclear if this conclusion applies to inferences on niche overlap. It is particularly difficult to learn about niches using information about species distributions (Soberon & Peterson 2005; Godsoe 2010b; Broennimann *et al.* 2012).

Here, we test whether available methods accurately infer shifts in niche overlap, including shifts in occurrence probability. Using a virtual species approach, we simulate both types of niche shifts. We then test our ability to detect these shifts in niche overlap using standardized implementations of 13 existing methods (Broennimann *et al.* 2012) and a new measure of niche similarity based on the expected shared presences (ESP) estimator proposed by Godsoe (2014). We illustrate the difference between ESP and traditional methods using data on the distribution of southern beech (*Nothofagus* spp.) in New Zealand.

Materials and methods

SIMULATING VIRTUAL SPECIES

We assume that a species will only be present if two conditions are met; first, the environment at a given location is a part of a species' niche and, second, the species has had an opportunity to disperse to that location. These assumptions represent a substantial, but valuable, simplification. These assumptions are similar to those applied in other virtual species studies and can be justified as a special case of more complex ecological models in which dispersal is weak (Levin 1974). Due to this model formulation, changes in the frequency of environments that are a part of a species' niche change the frequency of observed presences.

We simulated the distributions of two groups of organisms (i = 1,2). These could represent separate populations, subspecies or closely related species that live in separate biogeographic regions. The first group of organisms occurs in region 1 and the second group occurs in region 2. We simulated region 1 using climate data for North America, and simulated region 2 using similar data for Europe. To define the climatic niche of both groups, we assumed that each required an elliptical combination of mean annual temperature (t) and mean annual

Fig. 1. An illustration of why a niche shift along an unmeasured environmental axis change changes the occurrence probability of a species. (a) The grey rectangle represents the niche of a species projected onto environmental space. The niche of this species depends on two environmental variables: temperature, which is easy to measure and soil PH that is difficult to measure. Note that even if the temperature is ideal for this species, there is still a 50% chance that the PH is unsuitable. (b) A map in geographic space with a temperature gradient from cold (white) to warm (dark grey) temperatures. When we project the niche of the species in (a) onto (b), some locations are suitable (solid circles), while other locations are unsuitable (open circles). The species illustrated in (c) uses a similar range of temperatures as the species in (a). However, species (c) has evolved to use a broader range of PH. As a result, when we project its distribution into geographic space (d), many more locations at intermediate temperatures are suitable. In spite of the substantial niche differences between species (a) and (c), the range of temperatures occupied by each species is identical; instead, the change in PH requirements alters occurrence probability.
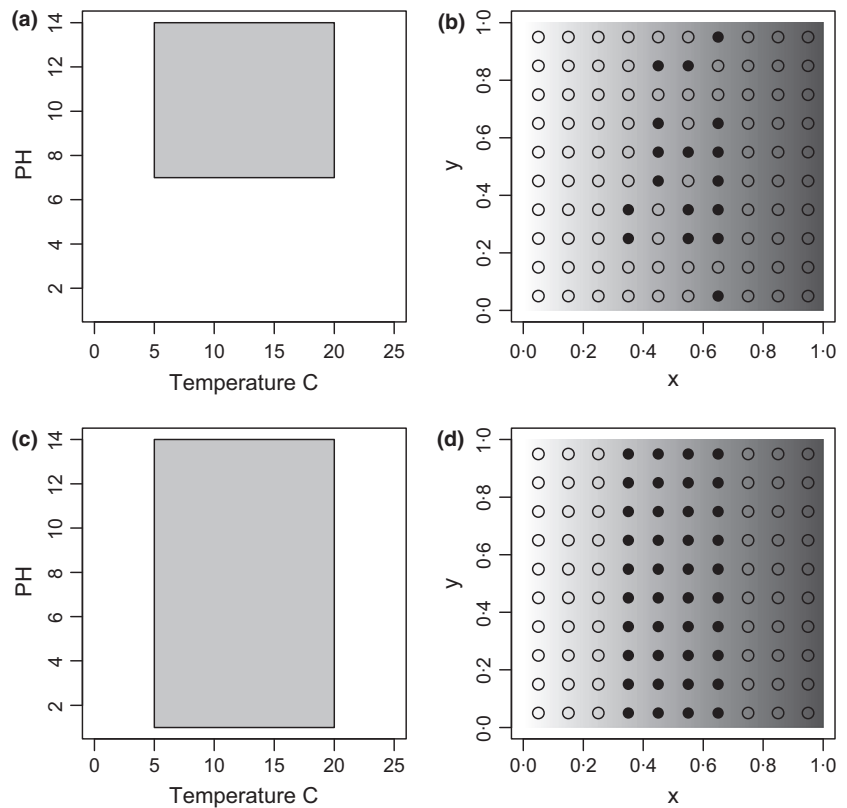
precipitation (p). The elliptical niche is simple to implement and follows previous simulation studies (Soberon & Peterson 2011). Such a niche can be described with the equation:

$$\left(\frac{t - t_i}{\alpha}\right) + \left(\frac{r - r_i}{\beta}\right) < 1, \qquad \text{eqn 1}$$

where t_i is the average temperature suitable to species i, α denotes the difference between the average temperature suitable to species i and the maximum temperature suitable to species i. The average amount of precipitation suitable to group i is r_i, while β is the difference between the average amount of precipitation suitable to group i, and the maximum precipitation value suitable to group i. The following parameter values were common to all simulations: $r_i = 800$, $\alpha = 300$, $\beta = 4\cdot5$ and $t_2 = 2\cdot5$. We obtained climate data from the Worldclim data set (Hijmans *et al.* 2005).

With this simulation framework, we can independently manipulate shifts in niche overlap and occurrence probability. To simulate shifts in temperature overlap, we manipulated the average temperature suitable to species 1 ($t_i = 2\cdot5$, 5, 7·5, 10 and 12·5). This changes the organism from being a specialist in cold climates (Fig. 2a) to being a specialist in warm climates (Fig. 2b). To simulate shifts in occurrence probability, we manipulated the probability that an environment with a suitable climate is suitable to group 1 (hereafter $p_{1,sc}$), allowing for the possibility that a location with a favourable climate is unsuitable because some non-climatic facet of the environment is unsuitable. To do this, we selected occurrence probabilities $p_{1,sc}$ that produce up to a 10-fold change in occurrence probability. The highest value for p_1 was 1, indicating that all sites with suitable climates were suitable (Fig. 2a). The lowest value for $p_{1,sc}$ was 0·1, indicating that only 1 in 10 locations with a favourable climate was suitable (Fig. 2c). The probability that group 2 was present in environments with suitable climates (p_2) was 1 in every simulation, implying that whenever the probability of occurrence was

>0, the species was present. Together, the simulations used five values of t_1, five values of $p_{1,sc}$ and five replicates for a total of 125 simulated species pairs. These combinations allow us to study a variety of forms of niche overlap. For example, a combination with $t_i = 5$ and $p_{1,sc} = 1$ produce slight differences in the temperature overlap of the two taxa, but no change in occurrence probability.

QUANTIFYING OUR ABILITY TO INFER NICHE SIMILARITY

We measured niche similarity using Sørensen's similarity index:

$$\text{Sørensen} = 2a/(2a + b + c). \qquad \text{eqn 2}$$

where a represents locations that are suitable to both groups, b represents observations that are suitable only to group 1 and c represents observations that are only suitable to group 2. Note that this metric ranges from 0 to 1, with 1 indicating niche equivalency (environments suitable to species 1 are always suitable to species 2), while 0 indicates that every observation suitable to one species will be unsuitable to the other.

We quantified simulated niche similarity in both geographic and environmental space. To compute the similarity of the distributions in environmental space, we computed the frequency of observations that are suitable to each species individually, and the frequency of observations that is suitable to both species. Using this information, we analytically derived the Sørensen similarity of the niches of the two taxa (Appendix S1). We also computed Schoener's D in environmental space, assuming we have perfect knowledge on which environments are a part of the niche of each species (Appendix S1). We ignored Schoener's D in subsequent calculations, because it was strongly correlated with Sørensen similarity (Fig. S3; $R^2 = 0\cdot973$). Manipulating the average temperature suitable to species i changes the Sørensen similarity from 1

Table 1. Terminology and notation used in the current manuscript

Term	Definition
Actual similarity	The true similarity of the niches of two taxa
a	The frequency with which we observe species 1 and species 2
b	The frequency with which we observe only species 1
c	The frequency with which we observe only species 2
ENM	A statistical method to use information on the distribution of a species to make inferences about its niche
Estimated niche similarity	The similarity of the niches of two species as inferred by a statistical method
ESP	An estimator for the Sørensen's similarity of occurrences for two species based on information on occurrence probability derived from statistical models
Niche	The set of environments that are suitable to a species
Niche conservatism	The niche conservatism hypothesis states that changes in the environmental requirements of a species are sufficiently slow to allow us to use information on the occurrence of a species in one region to predict occurrences in another region
Niche overlap	The overlap in the range of climates suitable to two species
Occurrence probability	The probability that we observe a species under a given set of circumstances
p_{ij}	The probability that a particular location, denoted location j, is suitable for a species
$p_{i,sc}$	The probability that an environment with a suitable climate will be suitable for species i
Presence/ absence	A model fit using presences and absences
Presence only	A model fit using only presences examples include PCA-env
Presence/ pseudoabsence	A model fit using presences and a random sample of available locations as a surrogate for absences examples include maximum entropy
r	Precipitation
r_i	The average precipitation suitable for species i
SDM	A statistical method to use information on the distribution of a species to make inferences about where it will be found
t	Temperature
t_i	The average temperature suitable to species i
α	The difference between the average temperature suitable to species i and the maximum temperature suitable to species i
β	The difference between the average amount of precipitation suitable to species i and the maximum precipitation value suitable to species i

(complete overlap) to 0 (no overlap; Table S1). Manipulating only the prevalence changes the similarity from 1 to 0·181 (Table S2).

To compare overlap in geographic space, we simulated a virtual species with a niche like group 1, across locations available to group 2 (i.e. Europe). We then computed the Sørensen similarity of the niches of species 1 and species 2 across all locations in Europe.

Our ability to predict niche overlap was quantified using each of the statistical methods below. First, we tested for a relationship between simulated niche similarity and detected niche similarity using the coefficient of determination R^2, representing the proportion of variation in simulated niche overlap that can be explained using each method to detect niche overlap. Next, we applied a more stringent test, the average absolute difference between simulated niche similarity and detected niche similarity (Δ_{abs}), a method that will only give a perfect score (0) if the detected similarity is precisely the same as the true similarity.

DETECTING OVERLAP USING STATISTICAL METHODS

For comparison with previous studies, we implement the majority of the methods tested in (Broennimann et al. 2012), see Table 2. Each method used presence-only data from a random selection of 4000 observation points in the continent of interest. One major class of methods uses principle component analysis (PCA) to summarize variation across many variables, and then applies kernel density estimators to quantify the relative density of presences. The similarity of these predictions is then measured on either environmental space (PCA-env) or geographic space (PCA-occ). In addition to these methods, we implemented three presence/pseudoabsence ENM approaches, boosted regression trees, generalized linear models and random forest. We also implemented several ordination methods, between-occ (Doledec & Chessel 1987), within-occ (Doledec & Chessel 1987), within-env, linear discriminant analysis, and ecological niche factor analysis (Hirzel et al. 2002). We present performance metrics for the kernel density estimation methods and ENFA using two calibration options, either calibrated using climate data from both Europe and North America, or calibrated in North America, but projected to Europe. We used the Schoener's D statistic to measure the similarity of the output produced by each of these methods.

Compatibility issues prevented us from implementing two tests in Broennimann et al. (2012). The omission of one ENM algorithm – maximum entropy – is unlikely to radically change the conclusions of the current paper; maximum entropy's performance was only slightly better than that of other ENM algorithms in Broennimann et al. (2012), and maximum entropy produces a biased estimate of occurrence probability (Phillips et al. 2009). The omission of one multivariate method, multidimensional scaling, is of little consequence as this approach performed poorly in Broennimann et al. (2012).

Godsoe (2014) provides a way to estimate the Sørensen similarity of two sets of observations using information on occurrence probability. This uses observations of j locations of interest (in our case, all of the locations in Europe for which we have climatic observations). It then uses information on the occurrence probability for each species at each location (hereafter p_{1j} for species 1 and p_{2j} for species 2) to compute the expected number of shared presences (ESP) for two species using the formula:

$$\text{Expected shared presences} = \frac{2\sum_j P_{1j}P_{2j}}{\sum_j (P_{1j} + P_{2j})}. \qquad \text{eqn 3}$$

Here, we extend this approach to compute the Sørensen similarity of the niches of two species, using the estimates of occurrence probability for species 1 and species 2 derived from our presence/absence ENM.

To implement ESP in our virtual species study, we estimated occurrence probability for each species by fitting a presence/absence ENM using the boosted regression tree algorithm (BRT). As implemented for ENMs, BRT is a machine learning method that seeks to predict the probability of presence for a species using a weighted combination of the predictions of a series of decision trees. We fit BRT using code described by Elith, Leathwick & Hastie (2008). The contribution of each subsequent decision tree is down-weighted by setting a learning rate (in our case 0·01). Each BRT model used presence/absence observations from a random selection of 4000 locations throughout the
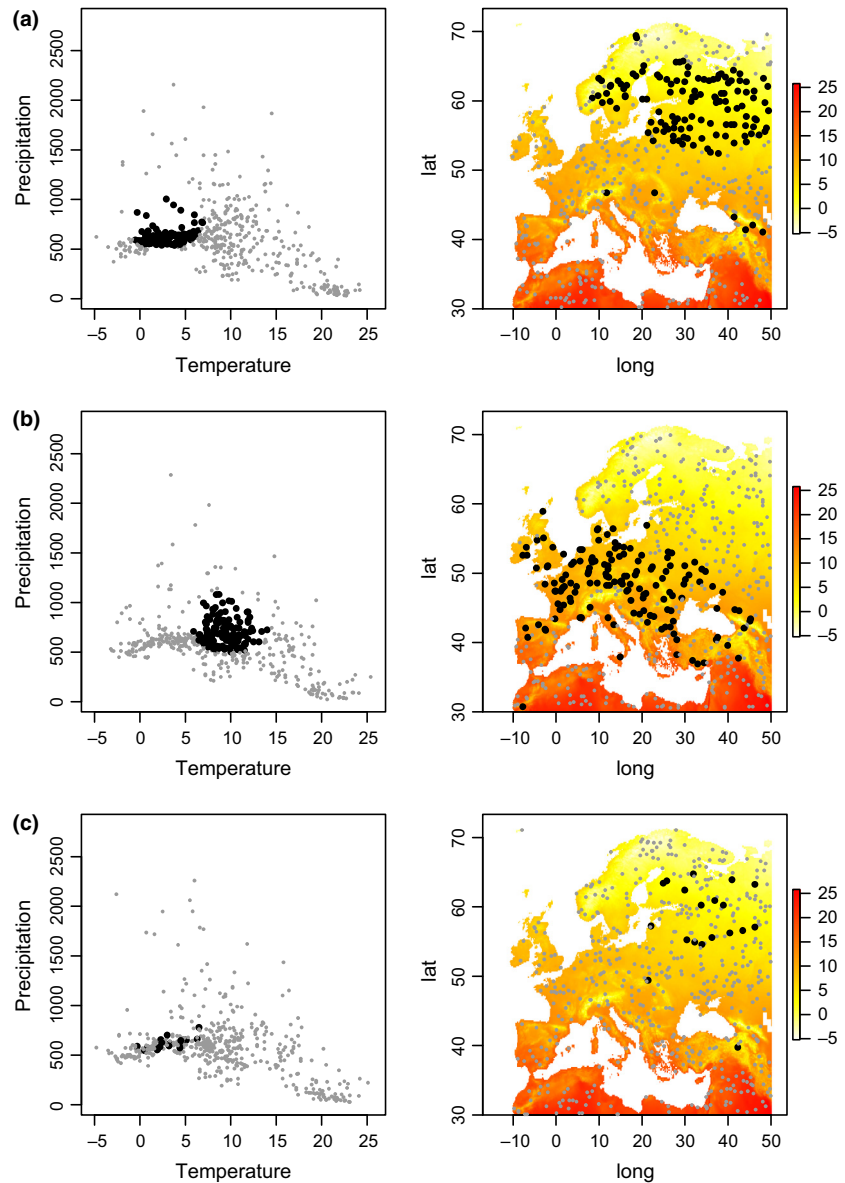
Fig. 2. An illustration of the two forms of niche shifts simulated in the current manuscript, with the left column portraying the data in environmental space (a plot of temperature vs. precipitation, with black points representing suitable environments and grey points representing unsuitable) and the right column portraying niches in geographic space (predictions overlain on Europe, with black points indicating suitable locations and grey points indicating unsuitable locations, map colour indicates temperature). (a) Illustrates a species whose niche encompasses cold environments with a moderate amount of precipitation. The species depicted in (b) uses a similar range of precipitation to that required in A, but this species requires warmer temperatures. The species depicted in (c) uses a similar range of precipitation and temperatures to A but has more restrictive environmental requirements and hence is found in fewer locations. For clarity, a random sample of 500 points is used on each plot. $t_i = 2.5$ in a&c and $t_i = 10$ in B. In C, $p_{i,sc} = 0.1$ in other panels $p_{i,sc} = 1$.

biogeographic region of interest (North America for species 1, Europe for species 2).

AN EMPIRICAL EXAMPLE OF THE APPLICATION OF ESP

When we have reliable presence/absence data at an appropriate scale, the difference between ESP and more traditional methods can be stark. We illustrate this difference by investigating niche overlap between southern beech trees in two separate geographic regions on the South Island of New Zealand.

Southern beech forests are a major element of New Zealand's forested ecosystems, comprising about 70% of the forested land area (Wardle 1984). In high elevation, low temperature, beech-dominated forests, silver beech (*Nothofagus menziesii*) and mountain beech (*N. solandri* vs. *cliffortioides*) dominate (Wardle 1984). There are broad similarities in the habitats used by the two species (Wardle 1991). We used New Zealand Land Cover Database version 3.3 (https://lris.scinfo.org.nz/layer/401-lcdb-v33-land-cover-database-version-33/), a GIS data set that delineates major land cover and vegetation types based on satellite image analysis to identify

forested regions across beech-dominated mountainous regions in the South Island of New Zealand. We then selected high elevation beech forests as those coinciding with areas characterized by mean annual temperatures <9°C, as identified using available climate data sets (Wratt *et al.* 2006). We count beech as present, so long as forests are present within this delimited region. All observations were conducted at a grid cell resolution of 2 km × 2 km.

Beech forests in the South Island of New Zealand can be divided into two regions separated by an iconic biogeographic anomaly, the 'Westland beech gap (WBG)', a region of 100–150 km where beech are absent. We divided this data set into two subsections one to the north of the WBG and one to the south.

Here, we seek to quantify the niche overlap by comparing the ENM of beech trees to the north of the WBG to predictions of an ENM generated from beech trees south of the WBG. We inferred the probability of presence for each region using BRT, with mean annual temperature as our predictor variable. In each region, we fit the model using a Bernoulli distribution, used fivefold cross-validation, a training fraction of 0·5, and after inspecting plots of cross-validated error rate, we elected to use 3000 decision trees in each model.

Table 2. A summary of the performance of methods used to infer niche similarity. To measure the performance of each method, we use coefficient of determination (R^2), interpreted as the proportion of variability in the Sørensen similarity of the niches of the two species that can be explained using each method to compute similarity

Method	Coefficient of determination (R^2)	Mean absolute difference (Δ_{abs})
Presence/absence ESP	0·978	0·042
PCA-occ-F	0·720	0·120
PCA-occ-T	0·718	0·110
PCA-env F	0·727	0·120
PCA-env T	0·726	0·120
Within-occ F	0·561	0·360
Between-group	0·664	0·130
Within-env	0·457	0·240
A_LDA	0·652	0·270
A_ENFA F	0·511	0·270
ENFA T	0·257	0·270
GLM	0·652	0·290
BRT	0·732	0·240
RF	0·481	0·390

Note that, for this measure, high scores indicate superior performance. In addition, we present Δ_{abs}, the average of the absolute value of the difference between Sørensen similarity and the inferred distance computed by each method, computed as average ($|\text{similarity}_i - \text{Sørensen similarity}_i|$) where i represents each of the 125 simulated data sets used. Note that, for this measure, scores close to zero indicate superior performance.

Results

In our simulations, 15 198 locations were suitable for species 2; the number of locations suitable for species 1 varied from 507 to 12 876. The Sørensen similarity of the niches of the two taxa was roughly the same in both geographic and environmental space ($R^2 = 0.955$). As a result, we present results in geographic space. In each case, the Sørensen similarity ranged between 0 and 1, though the average similarity was slightly higher in geographic space than it was in environmental space (0·271 vs. 0·224).

Broennimann *et al.* (2012) discuss the relative utility of the methods implemented in their study. The rankings they present are broadly consistent with the rankings in the current simulation study (Table 2). The kernel density estimator methods they propose (PCA-occ and PCA-env) performed well. Each of the four implementations of their kernel density estimators showed a small mean absolute difference (Δ_{abs}) and a high coefficient of determination (R^2) for the relationship between the Sørensen similarity of the niches of the two species and the distance inferred by each method. Presence/pseudoabsence ENMs and previously described presence-only methods produced weaker inferences. We focus subsequent interpretation on the kernel density estimation methods, since the simulations presented here confirm their strengths relative to other existing methods. Since all four kernel methods produced nearly identical inferences, we will restrict subsequent discussions on PCA-env-F, the preferred choice of Broennimann *et al.* (2012).

Though the kernel density estimation methods, such as PCA-env-F, outperformed previously implemented approaches, their absolute performance was much worse than that reported in Broennimann *et al.* (2012). This difference is easily explained by contrasting the effects of changes in temperature overlap and occurrence probability. Focusing only on changes in temperature overlap, PCA-env-F provided strong inferences on the similarity of niches (Fig. 3b; grey circles $R^2 = 0.989$, $\Delta_{abs} = 0.089$). When focusing only on shifts in occurrence probability (to do this, set $t_1 = t_2 = 2.5$), PCA-env-F is a poor predictor of niche similarity (Fig. 3b; grey triangles $R^2 = 0.365$, $\Delta_{abs} = 0.241$).

ESP fit using the presence/absence BRT model produced far better inferences of the Sørensen similarity of species' distributions than any other method (Fig. 3a). The similarity inferred using this method was tightly correlated with actual similarity of the niches of the two species ($R^2 = 0.978$), indicating that it explained roughly 25% more of the variance than any of the other methods tested (Table 2). The mean absolute difference
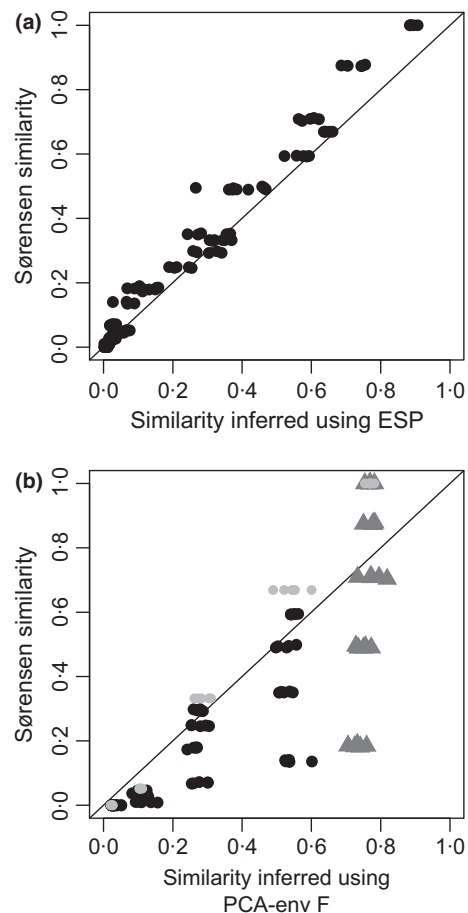

Fig. 3. Scatter plots of our ability to predict the Sørensen similarity of the niches of two species vs. inferred niche overlap. In panel (a), niche overlap is inferred by computing presence/absence ecological niche model (ENM)s with the boosted regression tree (BRT) algorithm then computing expected shared presences (ESP), Panel (b) uses the PCA-env method. Grey circles indicate points exhibiting only a shift in niche overlap. Grey triangles indicate points exhibiting only a shift in occurrence probability. Simulations exhibiting both a shift in niche overlap and a shift in occurrence probability are in black.

between inferred niche similarity and true niche similarity was roughly 2/3 smaller than the next best Δ_{abs} (Table 2).

In our *Nothofagus* data set, ESP provides a different measure of niche overlap than we would obtain using traditional metrics such as Schoener's D. The predictions of the model for trees north of the WBG are similar to predictions of the model derived from the south of the WBG (Fig. S4). As a result, the Schoener's D score is essentially identical (D = 0·97). In contrast, the ESP result infers that the occurrence probabilities of beech in the two areas are far from identical (ESP = 0·54). This discrepancy reflects how we interpret uncertainty in our models when we use ESP. Temperature data provide insufficient information to anticipate where trees will be found. This is easy to see in Fig. S4, as the predicted probability of presence rarely exceeds 0·6, meaning that we can be no more than 60% certain that any one site will be suitable to either tree type. ESP reflects this uncertainty and concludes that temperature data are insufficient to demonstrate strong niche overlap. It is still possible that an ENM including more variables in addition to temperature, or that an ENM fit using data collected at a different grid resolution, could lead to greater confidence in our predictions. This could increase the niche overlap inferred by ESP.

Discussion

Inferences about niche evolution permeate macroecology. For example, they are now being used to predict the distributions of invasive species (Petitpierre *et al.* 2012), model the effects of climate change and explore macroevolutionary changes in the distributions of many taxa. Many recent studies have focused on improving the rigour of our inferences of niche evolution (Elith & Graham 2009; Barve *et al.* 2011; Broennimann *et al.* 2012; Godsoe 2012), but less effort has been devoted to understanding how information on unmeasured environmental variables may bias our conclusions. Below, we consider how our work can be used to improve the rigour of analyses of niche overlap.

The simulations and analysis of the WBG suggest that future studies of niche overlap need to be designed to accurately estimate occurrence probabilities at the scale of interest to biogeographers. Traditional analyses of presence-only data provide a misleading picture of occurrence probability (Elith & Leathwick 2009; Phillips *et al.* 2009). The consequences of this well-known bias are clear in our study where even the best available presence-only methods obscured substantial shifts in occurrence probability (Fig. 3b). The poor performance of presence-only methods in the current study is particularly important, as it contradicts the common claim that presence-only methods provide a more realistic picture of a species' niche than do presence/absence data sets (Peterson, Soberon & Sanchez-Cordero 1999; Jiménez-Valverde, Lobo & Hortalk 2008; Broennimann *et al.* 2012). When possible, we recommend that studies of niche overlap use large, reliable presence/absence data sets, such as the large remotely sensed data set of *Nothofagus* observations. When this is not practical, an alternative approach may be to use presence/pseudoabsence

methods that have been specially tailored to infer occurrence probability (Royle *et al.* 2012).

We must then ensure that we are correctly interpreting information on occurrence probabilities. We have tried to show how traditional metrics can have difficulty detecting changes in occurrence probability (Fig. 3b; grey triangles $R^2 = 0.365$). The ESP metric we present is designed to estimate one particular way to measure niche overlap, the Sørensen's similarity of the niches of two species. In the light of the large number of similarity metrics available (Hubalek 1982), we believe that an exhaustive analysis based on all metrics would confuse more than it would clarify. Future studies may develop estimators for metrics of niche overlap other than Sørensen similarity.

Consistent with previous simulation analyses of niche overlap (Elith & Graham 2009; Barve *et al.* 2011; Broennimann *et al.* 2012; Godsoe 2012), we have used a simple model of dispersal, where each group can disperse to every location within the continent that it resides. Even with this simple model, one must think carefully about the role of dispersal. It is well known that analyses of niche overlap are prone to error when a taxon in one biogeographic region can disperse to locations with suitable environmental conditions that are unavailable in another region (Elith & Graham 2009; Godsoe 2010b; Barve *et al.* 2011). There is considerable opportunity to improve existing methods by better distinguishing the effects of environmental suitability, dispersal and stochasticity occurrence probability. One way to do this would be to further meld studies of niche overlap with investigations of population dynamics (Schurr *et al.* 2012).

Conclusions

Given the importance of inferring niche overlap in modern biogeography, it is essential that we use methods that can detect substantial niche changes. Using a review of existing hypotheses on niche shifts (Hierro, Maron & Callaway 2005) and conclusions drawn from niche theory (Godsoe 2010a, 2012; Godsoe & Harmon 2012), we have argued that tests of niche conservatism must explicitly account for the effects of occurrence probability on the distribution of each species. We have shown that some of the best available methods fail to detect substantial differences in occurrence probability, though the ESP method resolves this problem. This work suggests that there are substantial opportunities to detect signals of niche evolution obscured by existing biogeographic analyses.

Acknowledgements

Data accessibility

In addition, the publically available data sets mentioned in the text, and methods implemented using R scripts from Broennimann *et al.* (2012), we provide R

scripts to simulate data sets and perform novel analyses using ESP as online supporting information.

References

Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberon, J. & Villalobos, F. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, **222**, 1810–1819.

Broennimann, O., Treier, U.A., Muller-Scharer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*, **10**, 701–709.

Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., Petitpierre, B., Pellissier, L., Yoccoz, N.G. *et al.* (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, **21**, 481–497.

Doledec, S. & Chessel, D. (1987) Seasonal succession and spatial variables in freshwater environments. 1. Description of a complete 2-way layout by projection of variables. *Acta Oecologica, Oecologia Generalis*, **8**, 403–426.

Elith, J. & Graham, C.H. (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32**, 66–77.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *The Annual Review of Ecology, Evolution and Systematics*, **40**, 677–697.

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Godsoe, W. (2010a) I can't define the niche but I know it when I see it: a formal link between statistical theory and the ecological niche. *Oikos*, **119**, 53–60.

Godsoe, W. (2010b) Regional variation exaggerates ecological divergence in niche models. *Systematic Biology*, **59**, 298–306.

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

Godsoe, W. (2014) Inferring the similarity of species distributions using species' distribution models. *Ecography*, **37**, 130–136.

Godsoe, W. & Harmon, L.J. (2012) How do species interactions affect species distribution models? *Ecography*, **35**, 769–779.

Graham, C.H., Ron, S.R., Santos, J.C., Schneider, C.J. & Moritz, C. (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, **58**, 1718–1793.

Hierro, J.L., Maron, J.L. & Callaway, R.M. (2005) A biogeographical approach to plant invasions: the importance of studying exotics in their introduced and native range. *Journal of Ecology*, **93**, 5–15.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological-niche factor analysis: how to compute habitat suitability maps without absence data? *Ecology*, **83**, 2027–2036.

Holt, R.D. & Gomulkiewicz, R. (1997) How does immigration influence local adaptation? a reexamination of a familiar paradigm. *The American Naturalist*, **149**, 563–572.

Hooper, H.L., Connon, R., Callaghan, A., Fryer, G., Yarwood-Buchanan, S., Biggs, J., Maund, S.J., Hutchinson, T.H. & Sibly, R. (2008) The ecological niche of *Daphnia magna* characterized using population growth rate. *Ecology*, **89**, 1015–1022.

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

Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 415–427.

Jiménez-Valverde, A., Lobo, J.M. & Hortalk, J. (2008) Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, **14**, 885–890.

Kirkpatrick, M. & Barton, N.H. (1997) Evolution of a species' range. *The American Naturalist*, **150**, 1–23.

Levin, S.A. (1974) Dispersion and population interactions. *The American Naturalist*, **108**, 207–228.

Peterson, A.T., Soberon, J. & Sanchez-Cordero, V. (1999) Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265–1267.

Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C. & Guisan, A. (2012) Climatic niche shifts are rare among terrestrial plant invaders. *Science*, **335**, 1344–1348.

Phillips, S.J., Dudik, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J.R. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

Rödder, D. & Engler, J.O. (2011) Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. *Global Ecology and Biogeography*, **20**, 915–927.

Royle, J.A., Chandeler, R.B., Yackulic, C. & Nichols, J.D. (2012) Likelihood analysis of species occurrence probability from presence-only data for modeling species distributions. *Methods in Ecology and Evolution*, **3**, 545–554.

Schurr, F.M., Pagel, J., Cabral, J.S., Groeneveld, J., Bykova, O., O'Hara, R.B. *et al.* (2012) How to understand species' niches and range dynamics: a demographic research agenda for biogeography. *Journal of Biogeography*, **39**, 2146–2162.

Soberon, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1–10.

Soberon, J. & Peterson, A.T. (2011) Ecological niche shifts and environmental space anisotropy: a cautionary note. *Revista Mexicana de Biodiversidad*, **82**, 1348–1355.

Wardle, J. (1984) *The New Zealand Beeches, Ecology, Utilization and Management*. New Zealand Forest Service.

Wardle, P. (1991) *Vegetation of New Zealand*. The Blackburn Press, New Jersey, USA.

Warren, D.L., Glor, R.E. & Turelli, M. (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **63**, 2868–2883.

Wratt, D.S., Tait, A., Griffiths, G., Espie, P., Jessen, M., Keys, J. *et al.* (2006) Climate for crops: integrating climate data with information about soils and crop requirements to reduce risks in agricultural decision-making. *Meteorological Applications*, **315**, 305–315.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Additional derivations and methods.

Table S1. Summarizes manipulations of niche overlap used in simulations in the current manuscript.

Table S2. Summarizes how the manipulations of the prevalence of species 1 in the simulations change the similarity of the niche overlap of the two species in e-space.

Figure S1. An illustration of the fundamental niche of a species as modeled in the current manuscript.

Figure S2. An illustration of the overlap between the niches of two species.

Figure S3. A scatterplot of Schoener's D versus Sorensen similarity.

Figure S4. A comparison of the probability of presence inferred for Nothphagus.