


Review

Niche Estimation Above and Below the Species Level

Adam B. Smith ^{1,*} William Godsoe,² Francisco Rodríguez-Sánchez,³ Hsiao-Hsuan Wang,⁴ and Dan Warren^{5,6}

Ecological niches reflect not only adaptation to local circumstances but also the tendency of related lineages to share environmental tolerances. As a result, information on phylogenetic relationships has underappreciated potential to inform ecological niche modeling. Here we review three strategies for incorporating evolutionary information into niche models: splitting lineages into sub-units, lumping across lineages, and partial pooling of lineages into a common statistical framework that implicitly or explicitly accounts for evolutionary relationships. We challenge the default practice of modeling at the species level, which ignores the process of niche evolution and erroneously assumes that the species is always the appropriate level for niche estimation. Progress in the field requires reexamination of how we assess models of niches versus models of distributions.

What Is the Appropriate Phylogenetic Level for Niche Estimation?

The **niche** (see [Glossary](#)) is integral to understanding distributions and patterns of biological diversity, the identification of critical habitat to mitigate the effects of global change, and the reconstruction evolutionary trajectories of clades. Although there are many ways to estimate niches, a common starting point is the correlative **ecological niche model** (ENM), which identifies environmental conditions suitable to a species by finding associations between locations where a species is known to be present and the environment at those sites [1]. Although most niche models are constructed at the species level, an increasing appreciation for the role of evolutionary processes, especially **niche conservatism** and **local adaptation**, has led to modeling of niches above and below the level of species [2–10]. These studies raise critical questions. For example, what are the risks of modeling a species as a single undifferentiated entity, ignoring ecotypes and local adaptation? How can we estimate fundamental ecological tolerances if species occupy a subset of inhabitable conditions? How can we exploit knowledge of hierarchical phylogenetic relationships when estimating niches above and below the level of the species?

The justification for considering evolutionary relationships when estimating the **fundamental niche** is straightforward. Rates of niche evolution vary enormously above and below the species level [11,12]. When niches show high levels of conservatism, close relatives are likely to respond similarly to environmental gradients [13]. By contrast, spatial heterogeneity in environments coupled with reduced gene flow can encourage local adaptation, leading to divergence in niches among closely related **lineages**. In view of these phenomena, there is little reason to assume *a priori* that species represent the most appropriate phylogenetic level for modeling. Nevertheless, the majority of niche estimation exercises are conducted at the species level, which assumes that data on species' current distributions adequately reflect

Highlights

Modeling niches at the species level disregards information about evolutionary processes above and below the level of species.

Species-level models ignore the potential for locally adaptive responses and assume that current distributions reflect the entire set of suitable conditions.

In many cases splitting taxa into sub-units and modeling each separately or lumping related taxa can improve niche estimates.

Partial pooling of lineages into a single multiresponse framework has under-utilized potential for niche estimation, especially when accounting for phylogenetic relationships.

The advancement of evolutionarily informed niche models will be accelerated by emphasizing their ability to generate biologically plausible niche envelopes over their ability to recreate geographic distributions, which can be shaped by factors unrelated to the fundamental niche.

¹Center for Conservation and Center for Conservation and Sustainable Development, Missouri Botanical Garden, Saint Louis, MO 63116, USA

²BioProtection Research Centre, Burns Building, Lincoln University, Lincoln 7647, Canterbury, New Zealand

³Department of Integrative Ecology, Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, 41092 Sevilla, Spain

⁴Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, TX 77843, USA

environmental tolerances and that each species responds to the environment as an undifferentiated entity [14,15]. These critical assumptions are rarely questioned in practice [13,16,17]. Nevertheless, their implications are substantial given the breadth of decisions that these models are used to support, such as the allocation of effort toward *in situ* versus assisted migration [4,18], the identification of corridors to facilitate climate-induced range shifts [19,20], and the focusing of attention on areas vulnerable to invasion by non-native species [21,22].

Here we classify the modeling strategies that have been used to address these concerns into three categories based on dominant evolutionary processes. First, in cases where niche evolution is rapid or local adaptation is suspected, ‘**splitting**’ lineages into distinct units (e.g., subspecies within a species) and creating separate niche models for each unit may best capture environmental relationships [2,4,6,7,9,23]. Second, in situations where related lineages display high levels of niche conservatism, ‘**lumping**’ presence records within a single model can more accurately represent a lineage’s niche [13,24–27]. Spanning the continuum between lumping and splitting is a third option: ‘**partial pooling**’ of lineages into a single model [28,29] that may or may not account for phylogenetic relationships (Box 1).

In the following section we review splitting and lumping, the two most commonly used methods for incorporating evolutionary processes into niche modeling. Although these methods show promise, each is based on strong assumptions of either local adaptation (splitting) or niche conservatism (lumping). In the subsequent section we address the more flexible approach of partial pooling, which has to date been largely overlooked as a tool for niche estimation. Finally, we call for insight into what constitutes a robust niche model (versus a model of geographic distribution) and how this will be key for advancement of the field.

Accounting for Niche Evolution by Splitting or Lumping

Splitting and lumping represent opposing strategies to challenge the assumption that the current distribution of a species is the best surrogate for its niche (Figure 1). Adaptation of populations to local circumstances appears to frequently occur [30–34]. In these cases a promising strategy is to divide species into locally adapted subpopulations and model each as a distinct entity. At the same time, niche conservatism also appears to be prevalent among

⁵Senckenberg Biodiversity and Climate Research Center (SBiK-F), Frankfurt am Main, Germany

⁶Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology, Okinawa, Japan

*Correspondence:
adam.smith@mobot.org (A.B. Smith).

Box 1. ‘Splitting’, ‘Lumping’, and ‘Partial Pooling’

In ecology the practice of combining samples goes by a diverse set of terms often used interchangeably, including ‘lumping’, ‘pooling’, ‘grouping’, ‘combining’, ‘integrating’, and ‘assimilating’, among others [13,42,88,110]. Depending on the context, any of these designations can mean either combining occurrences from different classes (e.g., species, genus) and treating them as if they belonged to a single class (as often done in paleoecology when identification of fossil records to the species level is difficult [43,44]) or combining occurrences in a single multiresponse model that respects sample class [76]. Here we use ‘lumping’ to refer to modeling occurrences of different lineages as if they were a single lineage, ‘splitting’ to fitting a model to each lineage independently, and ‘partial pooling’ to modeling occurrences of related lineages in a single mixed-effects or multiresponse model [28].

Confusingly, multiresponse models themselves go by different names in ecology, including ‘community-level’, ‘multi-species’, and ‘joint-species’ models [76]. A somewhat more focused set of methods is encompassed by the term ‘vegetation modeling’, which usually refers to the combination of taxa with similar functional forms to estimate community-level properties [45]. Regardless of name, all of these types of models are typically applied to collections of taxa irrespective of their phylogenetic relatedness [111] except in the broadest sense (e.g., all plants). Moreover, these models typically focus on the spatial distribution, diversity, assembly mechanisms, and functional traits of potentially co-occurring species [76,78,82,83]. By contrast, the goal of niche modeling is to estimate fundamental environmental tolerances [1,112], often of lineages that are allopatric and which may even comprise extant and extinct lineages. Our focus is on modeling niches of lineages that are relatively closely related regardless of whether they co-occur in time and space.

diverse taxa [12,35–38]. If niche overlap is high enough between related taxa, lumping them into a single, undifferentiated unit may provide a more robust depiction of environmental tolerances. Lumping can be especially useful for addressing underrepresentation of the niche arising from dispersal limitation [25], anthropogenic persecution [26,39], disturbance [40], biotic interactions [24,41], and limited sampling [13]. In some instances lumping is also necessitated by the intractability of differentiating taxa to the species level, such as when modeling fossil data [42–44] or plant functional types [45].

Although lumping and splitting have become prominent alternatives to species-level modeling only in recent years, each of these strategies has a history nearly as long as ecological niche modeling itself. Lumping of species into guilds, life forms, or vegetation types was especially common in the formative period of niche modeling and often motivated by the lack of detailed knowledge of individual species' distributions [46–48]. Some early niche modeling efforts also attempted to account for intraspecific variation by modeling subtaxa delineated on the basis of differences in allozymes, karyotype, morphology, and ecology [33,49–52].

Lumping and splitting tend to produce very different niche estimates (Box 2) [4]. Split models generally estimate narrower niches than their lumped counterparts [7,53], but splitting does not necessarily result in smaller estimates of occupied geographic area [14] or in projections that suggest greater vulnerability to anticipated climate change [7–9,23,53–55]. When distributional data is adequate, split models usually have higher apparent predictive accuracy than lumped models [8–10,15,54], but if the data are incomplete lumped models can outperform split models when niche overlap with related taxa is moderately high [13]. Split and lumped models for the same taxon are often sensitive to different environmental factors [2,4,6–8,54–56].

Both splitting and lumping require difficult decisions about the appropriate phylogenetic level (generally, population, species, genus, and so on) at which to lump or split. As a result, it is vital to scrutinize the evidence on which the decision to combine or divide occurrence data is based. A wide variety of data have been used to demarcate (or merge) subunits: neutral genetic markers [4,7,8,10,23,53–56] and adaptive genetic variation [2,6]; morphological, ecotypic, and phenological variation [7,8,14]; chemical diversity [7,56]; geographic contact and introgression [27]; community associations with other species [53]; and geographic isolation [55]. While most studies base divisions (or groupings) on more than one type of evidence, very few [8] test alternative schemes for dividing or combining lineages.

We advise against justifying splitting or lumping solely on the basis of differences between models or on traditional null model tests of niche overlap. Although splitting and lumping can increase model reliability under specific circumstances, it is challenging to compare models on an equal basis because of factors that are difficult to control across lumped or split units (e.g., range size and geographic extent [57], sample size [58], differences in environments available to each subdivision [59], autocorrelation between training and test sites that varies in strength depending on whether occurrences are combined or divided [60]). Similarly, variation in predictor importance across split models is not a reliable indicator of the robustness of division since models trained on biologically meaningless subdivisions of a range are often sensitive to different predictors [61,62]. Monte Carlo-based null model tests of niche overlap [63–65] offer a seemingly attractive method by which to evaluate the appropriateness of proposed subdivisions. These tests assess whether the degree of similarity between a pair of taxa is greater or less than what is expected by chance, but this is not the same as asking whether information from one subdivision informs niche estimates of another. For example, a pair of lineages could

Glossary

Ecological niche model (ENM): a statistical method that identifies environmental conditions conducive to the persistence of a taxon. As used here, ENMs estimate niches whereas SDMs estimate geographic distributions.

Fundamental niche: the set of environmental conditions in which a taxon can survive and reproduce ignoring biotic interactions.

Lineage: a set of genetically related populations that are typically monophyletic and typically assumed to be adaptive units. Lineages may form above, at, or below the level of the species.

Local adaptation: higher fitness of genotypes in their environment of origin relative to their fitness in other environments.

Lumping: as used here, modeling multiple lineages or sublineages together as a single statistical sample.

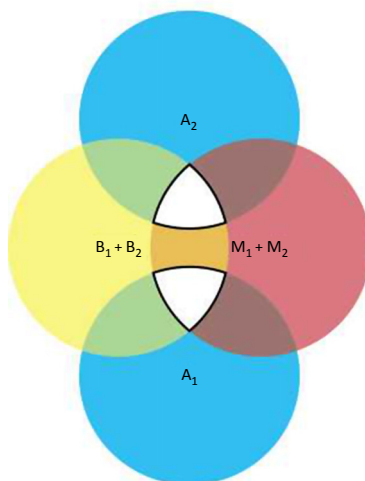
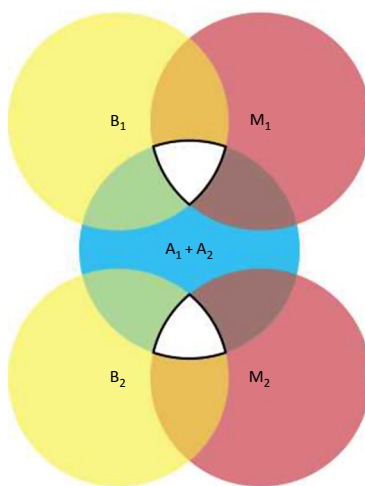
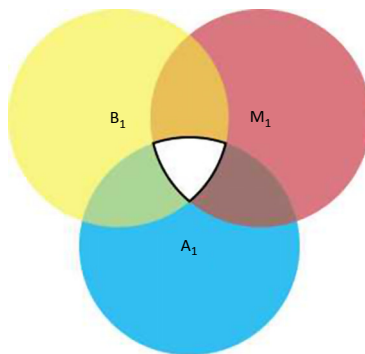
Niche: generally, the set of requirements necessary for the persistence of a lineage. As used in this review, 'niche' refers to favorable environmental conditions.

Niche conservatism: the tendency of closely related lineages to have similar ecological tolerances and preferences. In the current context we are specifically interested in environmental requirements.

Partial pooling: modeling multiple lineages simultaneously in a single statistical framework so that information is shared across lineages. There are many kinds of partial pooling models, which may or may not explicitly account for phylogenetic relationships.

Species distribution model (SDM): a statistical tool that estimates the environmental tolerances of a species. As used here, SDMs estimate geographic distribution while ENMs estimate niches.

Splitting: dividing lineages (genera, species, subspecies, etc.) into subunits and modeling each separately.



Trends in Ecology & Evolution

(See figure legend on the bottom of the next page.)

Box 2. Lumping versus Splitting

Niche models are commonly applied at the species level but this practice can obscure relevant information about other phylogenetic levels. To illustrate, we highlight work on wood partridges (genus *Dendrortyx*) endemic to Central America. The genus is divided into three species, two of which are further divided into subspecies based on morphological variation and geographic separation, although wide variation within and across populations makes taxonomic designation uncertain.

To explore this situation, Mota-Vargas and Rojas-Soto [14] used ecological niche models to estimate niches at three taxonomic levels (genus, species, and subspecies; Figure 1). They found striking differences between lumped and split niches. In particular, temperature was the most important variable in the genus-level model, whereas temperature and precipitation had equal influence in the species- and subspecies-level models. Niches at one level were as broad as, and sometimes broader than, combined niches estimated for the next-lower level. The greater breadth of the higher-level niche was generally – but not always – reflected in the geographic area estimated to be suitable. For example, the genus-level model predicted a suitable geographic area 39% larger than the sum of the species-level models, and the species-level model for *D. macroura* predicted an area 59% larger than the sum of its subspecies models. By contrast, the species-level model for *D. leucophrys* predicted a suitable area 29% smaller than the combination of its three subspecies models. Differences between split models were used to corroborate proposed locations of environmental barriers conducive to morphological (and thus taxonomic) differentiation.

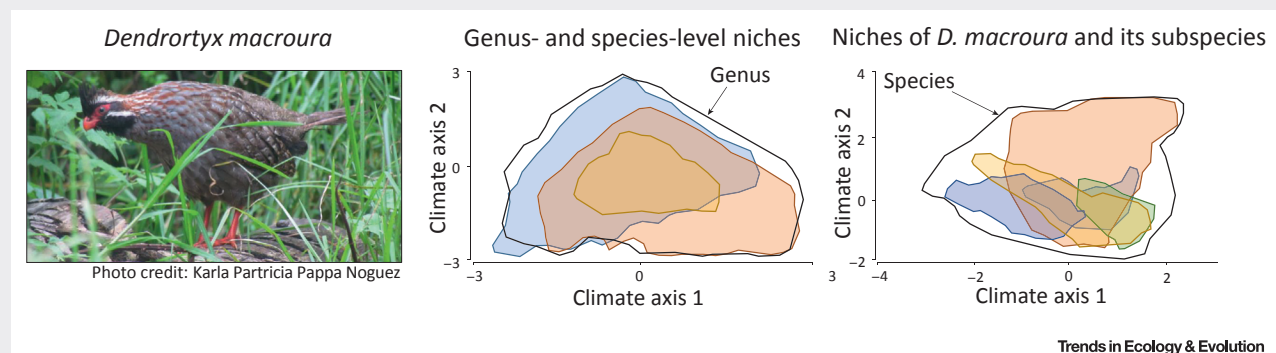


Figure 1. Lumped versus Split Niche Models. The wood partridge genus *Dendrortyx* comprises three species, two of which are further subdivided into several subspecies (left). Lumping species into a genus-level model results in broader niches than individual species-level models, which are indicated by the filled polygons in the middle panel. In turn, species-level models estimate broader niches than subspecies models, which are indicated by the filled polygons in the right panel (only results for *Dendrortyx macroura* are shown). The left and middle panels are adapted from Mota-Vargas and Rojas-Soto [14].

be more similar than expected given available habitats and yet still be sufficiently different in their environmental tolerances that their niches are best modeled by splitting.

In a related vein, an ongoing debate revolves around the prevalence of niche shifts in invaded portions of non-native species' ranges. Although current evidence suggests that niche shifts are rare [37] (and thus lumping across native and invaded ranges is usually warranted [66]), tests for niche shifts are robust only within shared environmental space [63] and thus do not

Figure 1. Choosing an Appropriate Niche Modeling Strategy. The 'biotic–abiotic–movement' (BAM) diagram is commonly used to conceptualize niche modeling (top) [1, 13, 109]. The set of suitable abiotic conditions (A_1) for a species is represented in blue. The goal of most niche modeling studies is to delineate the environmental conditions that fall within this region. However, taxa often occur only within a subset of the abiotically suitable environmental conditions, as their distributions are also limited to the set of locations that have favorable abiotic environments (A_1) and biotic interactions (B_1) and are reachable by dispersal (M_1). As a result, although the taxon can tolerate all conditions with A_1 , it will be possible to collect occurrence data only from the intersection of A_1 , B_1 , and M_1 , represented by the white Reuleaux triangle. In the middle panel, niche evolution is sufficiently slow that a pair of closely related taxa effectively have the same fundamental niche ($A_1 + A_2$). These taxa can therefore tolerate the same set of environments but differ in their realized niches due to differences in biotic interactions (B_1 and B_2) and dispersal (M_1 and M_2). In this case a lumped or partially pooled model will produce a better niche estimate than is possible for either taxon alone, as the joint distribution of both taxa does a much better job of sampling the set of suitable environments. In the bottom panel, a pair of taxa have diverged ecologically to the point where they have entirely nonoverlapping abiotic niches (A_1 and A_2). In this case a split model is advised because a partially pooled or lumped niche estimate is likely to be misleading, even if the taxa were restricted to the same set of biotic interactions ($B_1 + B_2$) and able to reach the same locations through dispersal ($M_1 + M_2$).

indicate whether niche shifts occur in novel environments as species expand into their invaded ranges.

Since the utility of lumping or splitting strongly depends on the degree of niche conservatism among taxa (Figure 1), whenever possible the justification for dividing (or combining) subdivisions should be based on evidence for (or lack of) selectively based niche differentiation among purported subunits. Such evidence includes results from progeny tests and common garden experiments [6,34,67,68] or association mapping of alleles with observed phenotypic differences across populations [2]. Secondly, splits can be based on aspects of life history or functional traits expected to experience strong selection [8], although it is important to keep in mind that phenotypic plasticity can obscure lines along which divisions or mergers should be based [7]. Similarly, neutral genetic variation may or may not correspond to selectively delineated subunits [2,7]. Of course, the utility of lumping or splitting will always depend on the amount and quality of distribution and environmental data. If the available data are unlikely to provide an accurate representation of sub niches, it may be preferable to lump them in a single model [13] at the expense of misrepresenting subtaxon niche variation.

Regardless of the basis for lumping and splitting, the relative risks of making an incorrect determination should be assessed carefully, especially if model outcomes are used to guide conservation decisions. Following Hällfors and colleagues [4], we recommend that decisions over whether to lump or split be based on the risks and benefits of assuming local adaptation or niche conservatism and that they follow, when possible, a least-regret approach. In some cases, this may involve applying management actions suggested by both lumping and splitting strategies [4]. Given the challenges, it may be tempting to absolve oneself of the decision by taking the traditional approach of modeling at the species level, but we emphasize that the hierarchical nature of evolutionary relationships means that this is nonetheless a decision to both lump (across subpopulations) and split (from other species) [14,15].

Accounting for Niche Evolution by Partial Pooling

An emerging alternative to lumping and splitting is partial pooling of multiple lineages simultaneously within a single joint model, which permits the modeling of each lineage's individual response to the environment while simultaneously 'borrowing' information across lineages. The wide variety of partial pooling algorithms makes tracing their history in ecological niche modeling difficult, but generally speaking, pooled niche models that do not explicitly account for phylogenetic relationships have a historical grounding in ordination methods [69,70]. By contrast, models that do explicitly incorporate phylogenetic information are a more recent addition undergoing rapid development [71,72].

We classify partial pooling models into two types. The first comprises 'phylogenetically naïve' partial pooling models which ignore phylogenetic relationships and thus are useful for cases when detailed phylogenetic information is lacking. A wide variety of algorithms can be used in this manner, from standard mixed effects [73] to hierarchical Bayesian models [28] to finite mixture and joint **species distribution models** (SDMs) [74–83]. In general, these types of models share in common an ability to differentiate the niches of abundant taxa while tending to pull rarer taxa toward a common response (Box 3). For example, finite mixture models, or so-called 'archetype' models, group taxa by their relationship to the environment (e.g., linear, unimodal) while allowing each taxon to have different coefficients describing its particular response [75,84,85]. The ability of partial pooling models to accommodate taxa with few occurrences is particularly advantageous in addressing the challenges of modeling data-poor taxa [86,87]. Nevertheless, the tendency of these models to push rare taxa toward the group

Box 3. Comparing Splitting, Lumping, and Partial Pooling Models

To illustrate the general behavior of each modeling approach, we simulated the niches of six related taxa (Figure I) across an artificial environmental gradient (hereafter, 'temperature') from 0 to 20 °C (see supplemental information online). The six taxa vary in their sensitivity to temperature according to their location in the phylogeny (green and yellow taxa being more warm adapted than bluer taxa). In this example, the splitting strategy [i.e., fitting six independent binomial generalized linear models (GLMs), one to each taxon] recovers different niches among taxa at the expense of misestimating relative suitability, especially at colder temperatures. The lumping strategy (fitting a single GLM to all taxa together) finds a common response while ignoring intertaxon differences. Partial pooling implemented with a binomial generalized linear mixed model (GLMM) that ignores phylogenetic relationships allows each taxon to respond differently while nonetheless drawing them toward a common response. Finally, a phylogenetically informed generalized linear mixed-effects model (PGLMM) implemented with phylogenetic 'attraction' causes closely related taxa to be more similar in their estimated response (compare with the phylogenetically naïve GLMM).

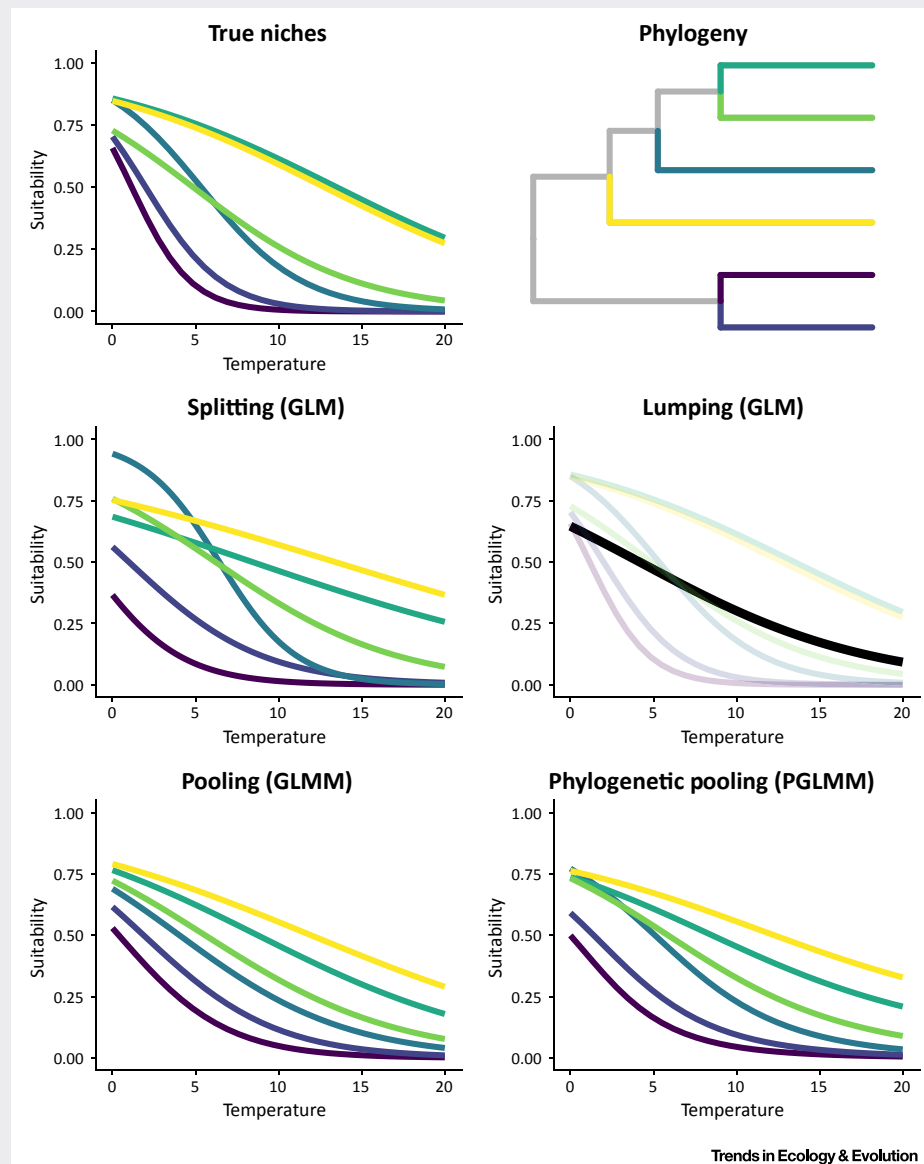


Figure I. Simulated Example Illustrating Different Modeling Approaches. Split, lumped, phylogenetically naïve, and phylogenetically informed partial pooling models depict different environmental relationships for the same set of related taxa.

mean risks biasing the responses of these same taxa [88], so the model structure and appropriate amount of pooling must be considered carefully.

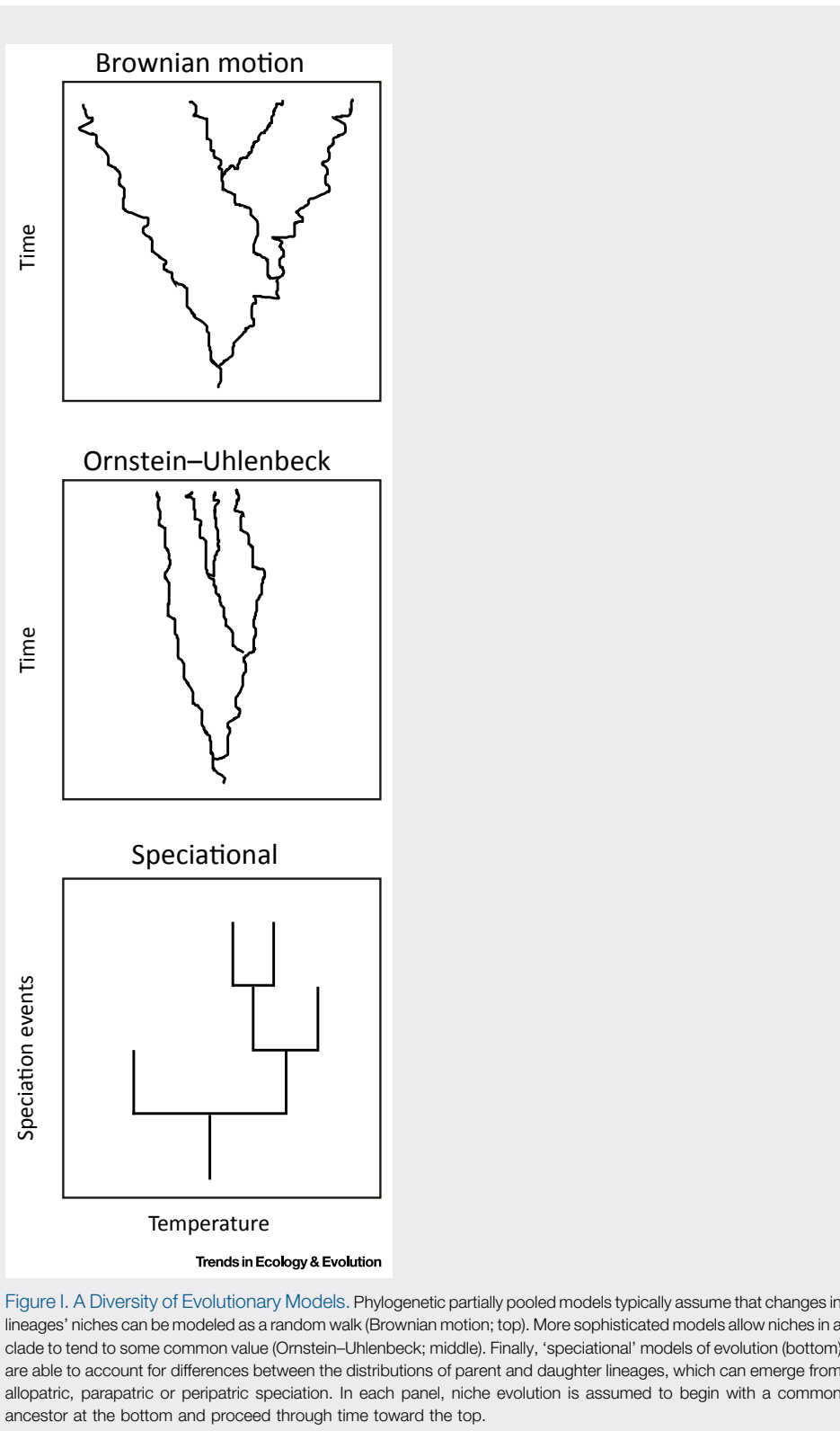
The second class of pooling methods comprises ‘phylogenetically informed’ partial pooling models, which explicitly account for evolutionary relationships. These models are also algorithmically diverse and thus account for evolutionary relationships in different ways. Some methods leverage information on phylogenetic distance either using variance–covariance matrices based on evolutionary distance [71,72,89] or by employing phylogenetic distance directly as a predictor [90,91]. Either of these approaches can account for either phylogenetic attraction (a less extreme form of lumping), in which related lineages tend to share environments, or phylogenetic repulsion (a less extreme form of splitting), in which closely related lineages inhabit dissimilar environments, possibly due to competition [71,92]. Other algorithms implicitly reflect the hierarchical nature of evolutionary relationships by using nested model structures [3,93] (e.g., by means of nested random effects [93] or more complicated approaches [3]). Mirroring evolutionary relationships within the model structure implicitly assumes attraction between lineages in the same subclade, since members of a subclade tend to have similar coefficients. As with phylogenetically naïve models, the degree of attraction (or repulsion) in phylogenetically informed models is partly determined by the data available for each lineage and so is likely to affect rare lineages more than abundant ones (Box 3).

Phylogenetically informed niche models are an area of active development. Recent advances aim to not only increase statistical robustness [94,95] but also account for multiple levels of relatedness (e.g., above and below the species level) [5], while controlling for imperfect detection [96] and spatial and temporal autocorrelation [92]. Other areas requiring attention include: the need for sophisticated cross-validation schemes that account for phylogenetic correlation [3,60]; uncertainties about the minimum number of taxa necessary for a robust model [28]; whether rare species do indeed respond to the environment similarly to more common species [88]; and reliance on simple models of evolution (Box 4).

Given the evolutionary basis on which all of biology is built, phylogenetically informed models offer great promise for the improvement of niche estimates [5,72,92]. First and foremost, these models have the potential to capture evolutionary processes operating at multiple phylogenetic levels [5]. We see this possibility as the ‘best of both worlds’ as it reflects the very real potential for niche differentiation and conservatism to operate simultaneously at different intensities across phylogenetic scales. Second, partial pooling models can accommodate rare lineages that would otherwise be difficult to model. In phylogenetically informed models, responses of rare lineages will also be influenced by their phylogenetic distance to other lineages, thereby at

Box 4. Models of Evolution

Phylogenetically informed partial pooling models typically assume that lineages that are more closely related tend to have more similar or dissimilar niches. This requires an explicit model of how the niche changes over time, but there are open questions about the best way to represent niche evolution. Existing approaches assume a random (Brownian motion) model of evolution where the environments occupied by a lineage diverge gradually over time with no directional trend (Figure 1, top). However, there may be situations where niche specialization is disfavored, say, if there are constraints in the degree to which a lineage can adapt to extreme temperatures. This leads to a constrained random walk (an Ornstein–Uhlenbeck model; Figure 1, middle), which is commonly used in macroevolution but less frequently in phylogenetically informed ecological niche modeling [71]. Macroevolutionary studies have developed approaches to account for rapid evolution during speciation, the simplest of which is to model divergence based on the number of speciation events separating lineages (versus divergence time between lineages), leading to a ‘speciation’ model of evolution (Figure 1, bottom; [3,113]). Allopatric speciation can alter lineages’ distributions even when they do not alter niches [59], an important effect that may be missed by current models of evolution.



least partially obviating the concern that niches of rare lineages could become incorrectly drawn toward the overall central tendency. Finally, the sharing of information across lineages can help guard against overfitting and thereby mistaking circumstantial differences in environments inhabited by lineages for adaptively driven niche differentiation [59].

Concluding Remarks: Toward Better Niche Estimates

Niche models are growing increasingly sophisticated in their ability to account for multiple facets of data. Nevertheless, very few modeling exercises account for phylogenetic relationships despite the foundational importance of evolution in shaping species' responses to the environment. Indeed, the practice of modeling at the individual species' level is so ubiquitous that the underlying assumptions around phylogenetic scale and relatedness are typically not discussed at all [15]. We see this as a missed opportunity. The inclusion of evolutionary information in the niche modeling process has the potential to produce better fundamental niche estimates and thereby greatly improve our understanding of niche evolution and provide a more robust basis for conservation decision-making.

While we advocate for greater use of partial pooling to incorporate evolutionary information into niche modeling, we also recognize that in some situations this approach may be neither desirable nor practical. For instance, phylogenetically informed models are difficult to construct, evaluate, and communicate and may not bring significant advantages over simpler approaches (i.e., nonphylogenetic models) for some datasets (e.g., cases with poorly resolved phylogenies or few species). Likewise, independent (split) species-level models may perform as well as partial pooling if taxa are well sampled. Finally, although partial pooling has the potential to ameliorate uncertainty in decisions over whether to lump or split units of analysis, practitioners must still make the decision on how to combine or divide occurrences used for modeling [43,44,97,98].

In a conservation context, the utility of splitting, lumping, and partial pooling will also be determined by the decision-making framework in which they are applied. For example, the splitting and partial pooling methods may be particularly apt for identifying the critical habitat of subspecific taxa under the United States Endangered Species Act, which allows the listing of individual vertebrate populations. However, most existing conservation plans have been developed around the species concept, so updating plans using methods that subdivide species may incur unacceptable costs if subspecific models and species-level models give different outcomes [99]. We urge caution when applying splitting or partial pooling models to species that have become genetically differentiated due to anthropogenic fragmentation since they may overstate differences that are likely to be important to dissolve through management action [100]. However, these models could also be used to guide efforts to merge anthropogenically differentiated populations by identifying populations that are most similar in their response to the environment and thereby reduce risks from managed genetic introgression.

In this review we have discussed methodological issues surrounding the inclusion of evolution in the modeling process, but the focus on geographic distributions creates cultural inertia in the field that must also be addressed, particularly with respect to model evaluation. The use of lumping, splitting, or partial pooling may in many cases result in niche estimates that are more biologically accurate, but which poorly estimate the current geographic distribution of the lineages included in the model. The current emphasis on model selection based solely on geographic predictions of niche models will therefore tend to reject biologically realistic models of niches in favor of overparameterized models that include spurious correlations with other processes (e.g., dispersal limitation, biotic interactions [101–103]). The move toward

evolutionarily informed models therefore requires a shift in how we think about and evaluate niche models: they must be assessed based on their biological plausibility rather than the geographic distributions we make by projecting them. Promising avenues include comparing models' ability to estimate known physiological limits [17,104–106] and the generation of unimodal response curves that match theoretical expectations [103,107,108].

Niche estimation has largely proceeded by assuming that neither anything above nor anything below the species level is relevant to understanding species' relationships with the environment. Each of the strategies we review here offers potential in addressing this shortcoming. Phylogenetically informed partial pooling holds substantial yet unrealized promise, but we anticipate that all of these methods will remain useful tools if outstanding issues can be addressed (see Outstanding Questions). Looking forward, we anticipate the development of flexible modeling frameworks that account for hierarchical evolutionary relationships above and below the species level, variable rates of niche evolution, and possibly uncertainty in lineage membership and the phylogeny itself.

Acknowledgments

This review emerged from the Investigative Workshop 'Species' Range Shifts in a Warming World' at the United States National Institute for Mathematical and Biological Synthesis (NIMBioS) (http://www.nimbios.org/workshops/WS_rangeshifts), sponsored by the National Science Foundation (NSF) through NSF Award #DBI-1300426, with additional support from the University of Tennessee, Knoxville. We thank Sean Hoban, Allan Strand, Andria Dawson, and Michelle Lawing for organization and NIMBioS for logistical and financial support. F.R.-S. was supported by a postdoctoral fellowship from the Spanish Ministry of Economy and Competitiveness (FPD-2013-16756), W.G. by the New Zealand Tertiary Education Commission CoRE grant to the Bio-Protection Research Centre, and A.B.S. by the Alan Graham Fund in Global Change. We thank Stephen J. Murphy for helpful comments on a draft version of the manuscript.

Supplemental Information

Supplemental information associated with this article can be found online at <https://doi.org/10.1016/j.tree.2018.10.012>.

References

- Peterson, A.T. *et al.* (2011) *Ecological Niches and Geographic Distributions* (MPB-49), Princeton University Press
- Banta, J.A. *et al.* (2012) Climate envelope modelling reveals intraspecific relationships among flowering phenology, niche breadth and potential range size in *Arabidopsis thaliana*. *Ecol. Lett.* 15, 769–777
- Godoy, B.S. *et al.* (2018) When phylogeny and ecology meet: modeling the occurrence of Trichoptera with environmental and phylogenetic data. *Ecol. Evol.* 8, 5313–5322
- Hällfors, M.H. *et al.* (2016) Addressing potential local adaptation in species distribution models: implications for conservation under climate change. *Ecol. Appl.* 26, 1154–1169
- Joly, S. *et al.* (2018) On the importance of accounting for intraspecific genomic relatedness in multi-species studies. *bioRxiv* Published May 14, 2018. <http://dx.doi.org/10.1101/321927>
- Koralewski, T.E. *et al.* (2015) Plants on the move: assisted migration of forest trees in the face of climate change. *For. Ecol. Manag.* 344, 30–37
- Maguire, K.C. *et al.* (2018) Intraspecific niche models for ponderosa pine (*Pinus ponderosa*) suggest potential variability in population-level response to climate change. *Syst. Biol.* 67, 965–978
- Marcen, A. *et al.* (2016) Tackling intraspecific genetic structure in distribution models better reflects species geographical range. *Ecol. Evol.* 6, 2084–2097
- Pearman, P.B. *et al.* (2010) Within taxon niche structure: niche conservatism, divergence and predicted effects of climate change. *Ecography* 33, 990–1003
- Serra-Varela, M. *et al.* (2015) Does phylogeographical structure relate to climatic niche divergence? A test using maritime pine (*Pinus pinaster* Ait.). *Glob. Ecol. Biogeogr.* 24, 1302–1313
- Kellermann, V. *et al.* (2012) Phylogenetic constraints in key functional traits behind species' climate niches: patterns of desiccation and cold resistance across 95 *Drosophila* species. *Evolution* 66, 3377–3389
- Peterson, A.T. (2011) Ecological niche conservatism: a time-structured review of evidence. *J. Biogeogr.* 38, 817–827
- Qiao, H. *et al.* (2017) Using data from related species to overcome spatial sampling bias and associated limitations in ecological niche modelling. *Methods Ecol. Evol.* 8, 1804–1812
- Mota-Vargas, C. and Rojas-Soto, O.R. (2016) Taxonomy and ecological niche modeling: implications for the conservation of wood partridges (genus *Dendrortyx*). *J. Nat. Conserv.* 29, 1–13
- Romero, D. *et al.* (2014) Uncertainty in distribution forecasts caused by taxonomic ambiguity under climate change scenarios: a case study with two newt species in mainland Spain. *J. Biogeogr.* 41, 111–121
- Qiao, H. *et al.* (2017) Accessible areas in ecological niche comparisons of invasive species: recognized but still overlooked. *Sci. Rep.* 7, 1213
- Qiao, H. *et al.* (2018) An evaluation of transferability of ecological niche models. *Ecography* Published online August 18, 2018. <http://dx.doi.org/10.1111/ecog.03986>
- Wang, H.-H. *et al.* (2015) Species distribution modelling for conservation of an endangered endemic orchid. *AoB Plants* 7, pii039

Outstanding Questions

How should lumped and split models be post-processed to obtain predictions at the appropriate level (usually species or populations) for stakeholders?

At what phylogenetic scale is phylogenetic level (i.e., corresponding roughly to genus, family, etc.) no longer informative about species-level niches?

Rare taxa can 'borrow strength' from common taxa in partial pooled frameworks, allowing the modeling of taxa that would erstwhile have too few presences for modeling in 'single-taxon' frameworks. However, the degree to which rare taxa truly conform to the responses of common species remains an open question. Ergo, do these methods mis-specify the niches of rare taxa and thereby risk misrepresenting their vulnerability to anticipated global change?

Niches could be more or less similar than expected by chance, but splitting or lumping, respectively, might still yield better niche estimates. How do we avoid conflating allopatry with dissimilarity in environmental tolerances?

Do phylogenetically informed partial pooling methods perform better than other methods when extrapolating to non-analog climates?

How can we perform model selection for lumping/splitting/pooling applied to the same set of lineages while controlling for confounding factors that vary according to the strategy employed (e.g., sample size, study region extent, range size)?

How do niches scale along phylogenetic hierarchies? Are constraints fundamentally imposed at certain levels and do lineages below that level adaptively 'wander' within those constraints?

How can we capture niche similarities and differences arising simultaneously from stabilizing selection and diversifying selection operating at different phylogenetic levels?

19. Alagador, D. *et al.* (2016) Climate change, species range shifts and dispersal corridors: an evaluation of spatial conservation models. *Methods Ecol. Evol.* 7, 853–866
20. Carroll, C. *et al.* (2018) Climatic, topographic, and anthropogenic factors determine connectivity between current and future climate analogs in North America. *Glob. Change Biol.* 24, 5318–5331
21. Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *Q. Rev. Biol.* 78, 419–433
22. Thuiller, W. *et al.* (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Glob. Change Biol.* 11, 2234–2250
23. Prasad, A.M. and Potter, K.M. (2017) Macro-scale assessment of demographic and environmental variation within genetically derived evolutionary lineages of eastern hemlock (*Tsuga canadensis*), an imperiled conifer of the eastern United States. *Biodivers. Conserv.* 26, 2223–2249
24. Alvarez-Martinez, J.M. *et al.* (2015) Can Eltonian processes explain species distributions at large scale? A case study with great bustard (*Otis tarda*). *Divers. Distrib.* 21, 123–138
25. Bush, A. *et al.* (2018) Truncation of thermal tolerance niches among Australian plants. *Glob. Ecol. Biogeogr.* 27, 22–31
26. Martínez-Freiria, F. *et al.* (2016) Contemporary niche contraction affects climate change predictions for elephants and giraffes. *Divers. Distrib.* 22, 432–444
27. Thorne, J.H. *et al.* (2013) Alternative biological assumptions strongly influence models of climate change effects on mountain gorillas. *Ecosphere* 4, 108
28. Gelman, A. and Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press
29. Qian, S.S. *et al.* (2010) On the application of multilevel modeling in environmental and ecological studies. *Ecology* 91, 355–361
30. Bocedi, G. *et al.* (2013) Effects of local adaptation and interspecific competition on species' responses to climate change. *Ann. N. Y. Acad. Sci.* 1297, 83–97
31. Lascoux, M. *et al.* (2016) Local adaptation in plants. *eLS* Published online February 15, 2016. <http://dx.doi.org/10.1002/9780470015902.a0025270>
32. Moran, E.V. and Ormond, R.A. (2015) Simulating the interacting effects of intraspecific variation, disturbance, and competition on climate-driven range shifts in trees. *PLoS One* 10, e0142369
33. Peterson, A.T. and Holt, R.D. (2003) Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. *Ecol. Lett.* 6, 774–782
34. Valladares, F. *et al.* (2014) The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecol. Lett.* 17, 1351–1364
35. Culumber, Z.W. and Tobler, M. (2018) Correlated evolution of thermal niches and functional physiology in tropical freshwater fishes. *J. Evol. Biol.* 31, 722–734
36. Mesquita, D.O. *et al.* (2016) Life-history patterns of lizards of the world. *Am. Nat.* 187, 689–705
37. Petitpierre, B. *et al.* (2012) Climatic niche shifts are rare among terrestrial plant invaders. *Science* 335, 1344–1348
38. Pie, M.R. *et al.* (2017) The evolution of climatic niches in squamate reptiles. *Proc. Biol. Sci.* 284, 20170268
39. Faurby, S. and Araújo, M.B. (2018) Anthropogenic range contractions bias species climate change forecasts. *Nat. Clim. Change* 8, 252–256
40. Tinner, W. *et al.* (2013) The past ecology of *Abies alba* provides new perspectives on future responses of silver fir forests to global warming. *Ecol. Monogr.* 83, 419–439
41. Mpakairi, K.S. *et al.* (2017) Missing in action: species competition is a neglected predictor variable in species distribution modelling. *PLoS One* 12, e0181088
42. Maguire, K.C. *et al.* (2016) Controlled comparison of species- and community-level models across novel climates and communities. *Proc. Biol. Sci.* 283, 20152817
43. Moreno-Amat, E. *et al.* (2017) Incorporating plant fossil data into species distribution models is not straightforward: pitfalls and possible solutions. *Quat. Sci. Rev.* 170, 56–68
44. Rodríguez-Sánchez, F. and Arroyo, J. *et al.* (2011) Cenozoic climate changes and the demise of Tethyan laurel forests: lessons for the future from an integrative reconstruction of the past. In *Climate Change, Ecology and Systematics* (Hodkinson, T.R., ed.), pp. 280–303, Cambridge University Press
45. Snell, R.S. *et al.* (2014) Using dynamic vegetation models to simulate plant range shifts. *Ecography* 37, 1184–1197
46. Bonn, A. and Schröder, B. (2001) Habitat models and their transfer for single and multi species groups: a case study of carabids in an alluvial forest. *Ecography* 24, 483–496
47. Box, E.O. (1981) Predicting physiognomic vegetation types with climate variables. *Vegetatio* 45, 127–139
48. Woodward, F.I. (1987) *Climate and Plant Distribution*, Cambridge University Press
49. Austin, M.P. *et al.* (1990) Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecol. Monogr.* 60, 161–177
50. Costa, J. *et al.* (2002) Ecologic niche modeling and differentiation of populations of *Triatoma brasiliensis* Neiva, 1911, the most important Chagas' disease vector in northeastern Brazil (Hemiptera, Reduviidae, Triatominae). *Am. J. Trop. Med. Hyg.* 67, 516–520
51. Kohlmann, B. *et al.* (1988) Environmental predictions and distributional limits of chromosomal taxa in the Australian grasshopper *Caledia captiva* (F.). *Oecologia* 75, 483–493
52. Thuiller, W. *et al.* (2004) Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* taxa. *Ecology* 85, 1688–1699
53. Cacciapaglia, C. and van Woesik, R. (2018) Marine species distribution modelling and the effects of genetic isolation under climate change. *J. Biogeogr.* 45, 154–163
54. Ikeda, D.H. *et al.* (2017) Genetically informed ecological niche models improve climate change predictions. *Glob. Change Biol.* 23, 164–176
55. Malone, S.L. *et al.* (2018) The future of subalpine forests in the southern Rocky Mountains: trajectories for *Pinus aristata* genetic lineages. *PLoS One* 13, e0193481
56. Shinneman, D.J. *et al.* (2016) Exploring climate niches of ponderosa pine (*Pinus ponderosa* Douglas ex Lawson) haplotypes in the western United States: implications for evolutionary history and conservation. *PLoS One* 11, e0151811
57. Smith, A.B. (2013) On evaluating species distribution models with random background sites in place of absences when test presences disproportionately sample suitable habitat. *Divers. Distrib.* 19, 867–872
58. Soutan, A. and Safi, K. (2017) The interplay of various sources of noise on reliability of species distribution models hinges on ecological specialisation. *PLoS One* 12, e0187906
59. Godsoe, W. (2010) Regional variation exaggerates ecological divergence in niche models. *Syst. Biol.* 59, 298–306
60. Roberts, D.R. *et al.* (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40, 913–929
61. Raes, N. (2012) Partial versus full species distribution models. *Nat. Conservacao* 10, 127–138
62. Titeux, N. *et al.* (2017) The need for large-scale distribution data to estimate regional changes in species richness under future climate change. *Divers. Distrib.* 23, 1393–1407
63. Broennimann, O. *et al.* (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. *Glob. Ecol. Biogeogr.* 21, 481–497
64. Nunes, L.A. and Pearson, R.G. (2017) A null biogeographical test for assessing ecological niche evolution. *J. Biogeogr.* 44, 1331–1343

65. Warren, D.L. *et al.* (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62, 2868–2883
66. Broennimann, O. and Guisan, A. (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biol. Lett.* 4, 585–589
67. Benito-Garzon, M. *et al.* (2013) Interspecific differences in tree growth and mortality responses to environmental drivers determine potential species distributional limits in Iberian forests. *Glob. Ecol. Biogeogr.* 22, 1141–1151
68. Mendola, M.L. *et al.* (2015) The role of ecotypic variation and the environment on biomass and nitrogen in a dominant prairie grass. *Ecology* 96, 2433–2445
69. Ramensky, L.G. (1929) К методике сравнительной обработки и систематизации списков растительности и других объектов, определяемых несколькими несходно действующими факторами (On methods for comparative analysis and ordering of plant lists and other objects determined by multiple factors), *Trudy Soveščaniya Geobotanikov-lugovedov*, 1–26 (in Russian)
70. ter Braak, C.J.F. (1994) Canonical community ordination. Part I: basic theory and linear methods. *Ecoscience* 1, 127–140
71. Ives, A.R. and Helmus, M.R. (2011) Generalized linear mixed models for phylogenetic analyses of community structure. *Ecol. Monogr.* 81, 511–525
72. Ives, A.R. and Garland, T. (2014) Phylogenetic regression for binary dependent variables. In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology* (Garamszegi, L.Z., ed.), pp. 231–261, Springer
73. Bolker, B.M. *et al.* (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* 24, 127–135
74. Clark, J.S. *et al.* (2017) Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecol. Monogr.* 87, 34–56
75. Hui, F.K.C. *et al.* (2013) To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models. *Ecology* 94, 1913–1919
76. Nieto-Lugilde, D. *et al.* (2018) Multiresponse algorithms for community-level modelling: review of theory, applications, and comparison to species distribution models. *Methods Ecol. Evol.* 9, 834–848
77. Ovaskainen, O. and Soininen, J. (2011) Making more out of sparse data: hierarchical modeling of species communities. *Ecology* 92, 289–295
78. Ovaskainen, O. *et al.* (2016) Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods Ecol. Evol.* 7, 428–436
79. Ovaskainen, O. *et al.* (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.* 20, 561–576
80. Pollock, L.J. *et al.* (2014) Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). *Methods Ecol. Evol.* 5, 397–406
81. Thorson, J.T. *et al.* (2016) Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. *Glob. Ecol. Biogeogr.* 25, 1144–1158
82. Warton, D.I. *et al.* (2015) So many variables: joint modeling in community ecology. *Trends Ecol. Evol.* 30, 766–779
83. Zhang, C. *et al.* (2018) Comparing the prediction of joint species distribution models with respect to characteristics of sampling data. *Ecography* 41, 1876–1887
84. Dunstan, P.K. *et al.* (2013) Finite mixture of regression modeling for high-dimensional count and biomass data in ecology. *J. Agric. Biol. Environ. Stat.* 18, 357–375
85. Hui, F.K. *et al.* (2015) Multi-species distribution modeling using penalized mixture of regressions. *Ann. Appl. Stat.* 9, 866–882
86. Breiner, F.T. *et al.* (2015) Overcoming limitations of modelling rare species by using ensembles of small models. *Methods Ecol. Evol.* 6, 1210–1218
87. Lomba, A. *et al.* (2010) Overcoming the rare species modelling paradox: a novel hierarchical framework applied to an Iberian endemic plant. *Biol. Conserv.* 143, 2647–2657
88. Pacifici, K. *et al.* (2014) Guidelines for a priori grouping of species in hierarchical community models. *Ecol. Evol.* 4, 877–888
89. Hadfield, J.D. and Nakagawa, S. (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J. Evol. Biol.* 23, 494–508
90. Morales-Castilla, I. *et al.* (2017) Combining phylogeny and co-occurrence to improve single species distribution models. *Glob. Ecol. Biogeogr.* 26, 740–752
91. Oke, O.A. *et al.* (2014) Integrating phylogenetic community structure with species distribution models: an example with plants of rock barrens. *Ecography* 37, 614–625
92. Kaldhusdal, A. *et al.* (2015) Spatio-phylogenetic multispecies distribution models. *Methods Ecol. Evol.* 6, 187–197
93. Bates, A.E. *et al.* (2010) Deep-sea hydrothermal vent animals seek cool fluids in a highly variable thermal environment. *Nat. Commun.* 1, 14
94. Bürkner, P.C. (2018) Advanced Bayesian multilevel modeling with the R package brms. *R J.* 10, 395–411
95. Pearce, W.D. *et al.* (2015) pez: phylogenetics for the environmental sciences. *Bioinformatics* 31, 2888–2890
96. Frishkoff, L.O. *et al.* (2017) Phylogenetic occupancy models integrate imperfect detection and phylogenetic signal to analyze community structure. *Ecology* 98, 198–210
97. Nogués-Bravo, D. *et al.* (2016) Amplified plant turnover in response to climate change forecast by Late Quaternary records. *Nat. Clim. Change* 6, 1115–1119
98. Veloz, S.D. *et al.* (2012) No-analog climates and shifting realized niches during the Late Quaternary: implications for 21st-century predictions by species distribution models. *Glob. Change Biol.* 18, 1698–1713
99. Malaney, J.L. and Cook, J.A. (2013) Using biogeographical history to inform conservation: the case of Preble's meadow jumping mouse. *Mol. Ecol.* 22, 6000–6017
100. Ralls, K. *et al.* (2018) Call for a paradigm shift in the genetic management of fragmented populations. *Conserv. Lett.* 11, e12412
101. Escobar, L.E. *et al.* (2018) Ecological niche modeling re-examined: a case study with the Darwin's fox. *Ecol. Evol.* 8, 4757–4770
102. Jarnevich, C.S. *et al.* (2017) Minimizing effects of methodological decisions on interpretation and prediction in species distribution studies: an example with background selection. *Ecol. Model.* 363, 48–56
103. Merow, C. *et al.* (2014) What do we gain from simplicity versus complexity in species distribution models? *Ecography* 37, 1267–1281
104. Austin, M.P. *et al.* (2009) Physiological responses and statistical models of the environmental niche: a comparative study of two co-occurring *Eucalyptus* species. *J. Ecol.* 97, 496–507
105. Martínez, B. *et al.* (2015) Combining physiological threshold knowledge to species distribution models is key to improving forecasts of the future niche for macroalgae. *Glob. Change Biol.* 21, 1422–1433
106. Rodríguez-Sánchez, F. and Arroyo, J. (2008) Reconstructing the demise of Tethyan plants: climate-driven range dynamics of *Laurus* since the Pliocene. *Glob. Ecol. Biogeogr.* 17, 685–695
107. Bell, D.M. and Schlaepfer, D.R. (2016) On the dangers of model complexity without ecological justification in species distribution modeling. *Ecol. Model.* 330, 50–59
108. Guevara, L. *et al.* (2018) Toward ecologically realistic predictions of species distributions: a cross-time example from tropical montane cloud forests. *Glob. Change Biol.* 24, 1511–1522

109. Soberón, J. and Nakamura, M. (2009) Niches and distributional areas: concepts, methods, and assumptions. *Proc. Natl. Acad. Sci. U. S. A.* 106 (Suppl. 2), 19644–19650
110. Scheffer, M. *et al.* (2018) Toward a unifying theory of biodiversity. *Proc. Natl. Acad. Sci. U. S. A.* 115, 639–641
111. Ferrier, S. and Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *J. Appl. Ecol.* 43, 393–404
112. Soberón, J. and Peterson, A.T. (2011) Ecological niche shifts and environmental space anisotropy: a cautionary note. *Rev. Mex. Biodivers.* 82, 1384–1355
113. Mooers, A.O. *et al.* (1999) Using phylogenies to test macroevolutionary hypotheses of trait evolution in cranes (Gruinae). *Am. Nat.* 154, 249–259