

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/261222865>

The effect of spatially marginal localities in modelling species niches and distributions

Article in *Journal of Biogeography* · March 2014

DOI: 10.1111/jbi.12297

CITATIONS

10

READS

160

4 authors, including:



[Robert P. Anderson](#)

City College of New York

52 PUBLICATIONS 10,993 CITATIONS

SEE PROFILE

All content following this page was uploaded by [Robert P. Anderson](#) on 24 August 2014.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

ORIGINAL
ARTICLE

The effect of spatially marginal localities in modelling species niches and distributions

Mariano Soley-Guardia^{1,2*}, Aleksandar Radosavljevic^{1†}, Jhanine L. Rivera¹ and Robert P. Anderson^{1,2,3}

¹Department of Biology, City College, City University of New York, New York, NY 10031, USA, ²The Graduate Center, City University of New York, New York, NY 10016, USA, ³Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, New York, NY 10024, USA

ABSTRACT

Aim We introduce and evaluate the potential effect of spatially marginal localities (specifically those protruding into unsuitable regions), in overestimating species niches and distributions when using ecological niche models (ENMs).

Location North-western South America.

Methods We built an ENM for the Caribbean spiny pocket mouse (*Heteromys anomalus*) using MAXENT and climatic variables. This species typically inhabits extensive tropical forests but can also range into drier habitats through patches of mesic vegetation. We ranked occurrence records according to the suitability value they received, and retrieved habitat information from collectors' field notes and the literature to determine whether those receiving lower values correspond to spatially marginal localities protruding into unsuitable regions. We then built a model excluding a subset of such localities and compared its geographic and environmental prediction with that of the original model.

Results Models differed substantially in their estimates of suitability. The original model resulted in an overly extensive prediction, considering as suitable hot and dry regions dominated by xerophytic vegetation. Records receiving the lowest suitability values in this model corresponded mainly to captures in patches of mesic forest surrounded by thorn scrub or savannas. The model calibrated without such records restricted suitability mostly to regions characterized by the typical habitat of the species.

Main conclusions When it is not possible to use variables that are more proximal or have finer resolutions, we recommend building complementary models that, together, can provide a more realistic estimate of the species' niche and corresponding geographic distribution. Jointly interpreting these models, researchers may better differentiate between areas harbouring typical habitat and those where the species can be found only if locally favourable conditions exist. Such a distinction is of relevance for a wide range of applications relying on inferences obtained from ENMs.

Keywords

Distribution, ecological niche model, environmentally marginal, *Heteromys anomalus*, MAXENT, mosaic, north-western South America, range, spatially marginal, species distribution model.

*Correspondence: Mariano Soley-Guardia, Department of Biology, City College, City University of New York, Marshak Science Building, Room J-526, 160 Convent Ave, New York, NY 10031, USA.

E-mail: msoley@gc.cuny.edu

†Current address: Plant Biology and Conservation, Northwestern University, Evanston, IL 60208, USA.

INTRODUCTION

Understanding species niches and their associated geographic distributions has long been a major interest in biology (e.g. Grinnell, 1917; Hutchinson, 1957; Brown, 1984), providing

insights into the natural history of taxa and, more importantly, underpinning ecological, biogeographical and evolutionary studies (Sagarin *et al.*, 2006). Accordingly, researchers have developed methods that associate species occurrence records with digital environmental data (Guisan

& Zimmermann, 2000; Peterson *et al.*, 2011). Several names have been given to these methods, but we refer to them as ecological niche models (ENMs), emphasizing their underlying niche assumption (Peterson *et al.*, 2011; Anderson, 2012; Warren, 2012).

Most correlative ENMs [e.g. Genetic Algorithm for Rule-set Prediction (GARP), generalized linear models (GLMs), Multivariate Adaptive Regression Splines (MARS), maximum entropy (MAXENT)], are generated by contrasting the environmental conditions where a species occurs with either (1) the conditions where it is absent or undocumented (presence/absence or presence/pseudo-absence), or (2) the conditions corresponding to a random sample of the study region (presence/background; Peterson *et al.*, 2011, p. 102). According to how closely they match the modelled relationships (i.e. the niche), cells within a gridded output surface are then assigned a relative value of suitability for the species (Peterson *et al.*, 2011, pp. 51, 97). However, the quality of resulting ENMs depends on that of the input data (Lozier *et al.*, 2009; Warren, 2012), as well as adherence to general and species-specific assumptions (Peterson *et al.*, 2011; Anderson, 2013).

One of the most important assumptions when building ENMs is that occurrence records represent source rather than sink habitat (Phillips *et al.*, 2006; Warren, 2012). Sink habitats harbour environmental conditions that do not correspond to the species' fundamental niche (Pulliam, 1988, 2000), hence their incorporation exaggerates niche breadth and potential geographic distributions. Inclusion of records representing sink habitat can be avoided by excluding obvious outliers from model calibration (e.g. Mateo *et al.*, 2013). However, inclusion of sink habitat in the model might also occur inadvertently owing to issues regarding variable choice and resolution, particularly along the margins of species distributions. This can occur even when records represent true sources, meaning that simply removing outliers may be insufficient. We first outline the nature of this issue and then introduce a novel approach to address it. In so doing, we call attention to a more cautious and realistic use of ENMs and propose a way to improve these valuable tools.

Environmentally versus spatially marginal localities

Virtually all species experience a variety of environments across their range, causing unequal abundances throughout, as some of the conditions more closely approach the optimum for the species than others (Whittaker, 1956; Brown, 1984; Brown *et al.*, 1995; Guo *et al.*, 2005). In the absence of highly negative biotic interactors (and lacking strong dispersal constraints), species should usually be most abundant in areas that exhibit optimal environmental conditions (i.e. most suitable habitat). Concomitantly, the species should decline in number the further conditions depart from this optimum, until it cannot persist (references above). Extirpation occurs due to limits in the physiological tolerances of

the focal species (regarding survival and/or reproduction), in those of its positive biotic interactors (e.g. resources, facilitators), or both (Gaston, 2003). We refer to localities (or habitat) harbouring conditions along the margins of the species' niche (i.e. barely satisfying abiotic and/or general biotic requirements) as *environmentally marginal* (Fig. 1a).

In these instances, environmentally marginal conditions are thought to cause the limits to species distributions (but see Wisz *et al.*, 2013). As such, these conditions are bound to be found along (i.e. create) the spatial margins of species ranges (Gaston, 2003). For this reason, we refer to environmentally marginal localities also as *spatially marginal*. However, while we expect environmentally marginal localities to always be spatially marginal, the converse will not necessarily be true. Instead, this will depend on how gradual or sharp the decline in habitat suitability is (e.g. Lomolino *et al.*, 2006, p. 77). Moreover, spatially (and potentially environmentally) marginal localities should not be considered to occur only on the margins of the species' overall range as is generally implied (especially by the 'abundant centre' or 'central-marginal' hypothesis; e.g. Hengeveld & Haack, 1982; Sagarin & Gaines, 2002). Instead, these localities should occur along every margin of each area occupied by a deme or population, including 'internal' margins surrounding unsuitable conditions within the local range (Fig. 1b; see also Brown *et al.*, 1996; Guo *et al.*, 2005).

The problem of spatially marginal localities in practice

Inclusion of sink habitat in the model becomes more likely at spatial margins for two main reasons. First, these margins represent contact zones between source and sink habitat, making discrimination conceptually and empirically challenging (Kawecki, 2008). Second, even if spatial margins represent source habitat, at coarse resolutions (or when lacking information on proximal variables; Austin, 2002) they might provide an environmental signal characterizing the surrounding sink habitat instead. It is this latter issue that we explore in the present study.

Environmental variables used in ENM have a resolution corresponding to summaries of values across space and time (e.g. 1 km² cells representing mean annual precipitation). This results in a diversity of local conditions represented by a single value (e.g. mean or median). For this reason, local environments allowing persistence at a given site might not be accurately represented by the variables examined (Fig. 1c). This is likely to be exacerbated when values are obtained by interpolation over large distances (e.g. Hijmans *et al.*, 2005). Such data reality may not pose a serious problem near the core of a deme's distribution, because even though local environments might be omitted, the predominant environmental conditions still characterize source habitat. However, it can become a serious issue at spatial margins, where suitable habitat transitions to unsuitable. In this way, records

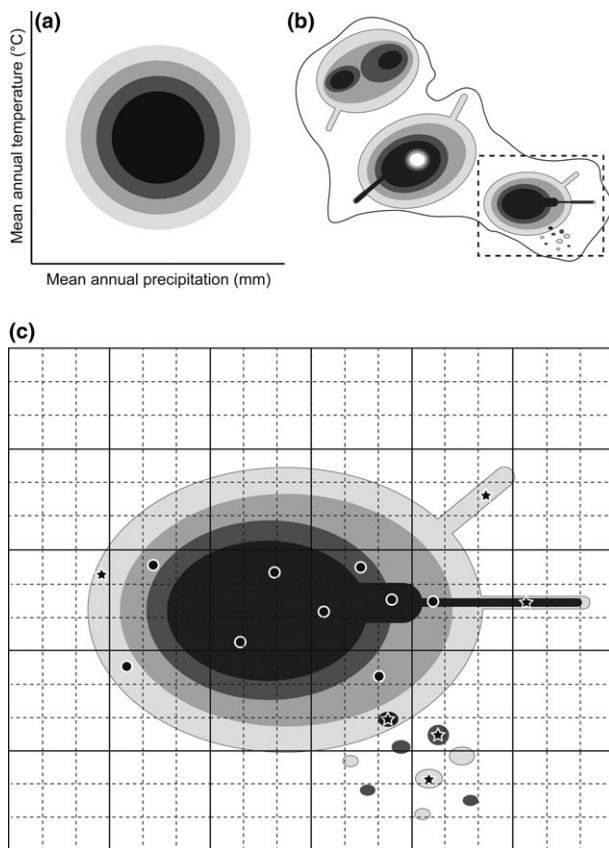


Figure 1 Hypothetical fundamental niche of a species and its projection in geography, showing differences between *environmental* and *spatial* marginality, as well as the potential effect of the latter in correlative ecological niche models. (a) Two-dimensional fundamental niche, with suitability declining from the dark centre to the pale environmentally marginal conditions. (b) Projection of the niche shown in panel (a) onto geographic space, with progressively darker areas denoting increasingly higher-quality source habitat and sinks shown in white. The outline roughly indicates the geographic range of the species. Even though the fundamental niche has a defined shape, its representation in geography is subject to the configuration of the landscape. In this way, highly suitable conditions may exist along spatial margins of the species' distribution and even protrude into the surrounding sink habitat. (c) Close up of the local range of the deme enclosed by the dashed rectangle in panel (b). Grids indicate two different resolutions at which environmental values are summarized to model the species' niche. Symbols represent hypothetical occurrences for the species. Circles indicate records that provide accurate information regarding the species' niche at both resolutions. Stars indicate records that result in exaggeration of the niche at one or both resolutions (due to summarizing of environmental values). This issue may arise in any heterogeneous area. However, note that records in heterogeneous areas near the core of the local range (certain circles) will still provide values that correctly reflect the species' overall niche breadth (albeit potentially affecting modelled response curves). In contrast, records found at heterogeneous spatial margins can result in incorporation of environmental values characterizing sink habitat. All spatially marginal localities can suffer from this issue; however, those protruding into sink habitat are much more likely to do so because of their smaller extent and higher contact with sinks.

from such margins might result in incorporation of environmental values characterizing the surrounding sink habitat, hence overestimating the species' niche and the corresponding geographic areas suitable for it (Fig. 1c). This possibility becomes more likely when suitable habitat protrudes extensively into unsuitable areas (either as contiguous strips or disjunct patches), increasing contact between sources and sinks. We refer to localities representing such conditions as *protruding spatially marginal* (PSM).

Objectives

We explore the effects that a particular case of spatially marginal localities might have in correlative ENMs: that in which suitable habitat protrudes into sink habitat (PSM localities). We focus on a forest-dwelling rodent that can inhabit small areas of gallery forest in otherwise unsuitable matrices of savannas or xerophytic thorn scrub: the Caribbean spiny pocket mouse, *Heteromys anomalus* (Thompson, 1815) (Rodentia: Heteromyidae). As an example, we use MAXENT, a presence-background modelling method (Phillips *et al.*, 2006) commonly applied to infer species niches and their distributions, and limit our analyses to abiotic predictor variables. However, the approach we use can be applied to any other modelling method, including presence-absence ones (Guisan & Zimmermann, 2000), models of species occupied distributional areas (species distribution models *sensu stricto*; Peterson *et al.*, 2011, p. 41; Anderson, 2012), and models that explicitly incorporate biotic factors (e.g. Lira-Noriega *et al.*, 2013).

Specifically, we used the following approach. We first built an ENM using an extensive dataset of occurrence records, which we then ranked according to the suitability value they received. We identified the set of records with lowest values (environmentally marginal according to the model) and retrieved descriptive habitat information from field notes and relevant literature to determine whether they corresponded to PSM localities. Finally, we built a second model excluding from the calibration process records representing PSM localities, and compared its prediction with that of the original model. We predict that: (1) records receiving a low suitability value will correspond to PSM localities, consisting of natural mosaics where the species' typical mesic habitat intermixes with those characterizing hot and dry regions; (2) at PSM localities, the species was captured within local patches of mesic habitat; and (3) if PSM localities actually provide the model with a misleading environmental signal (one typically characterizing sinks), a model made with such localities will overestimate the environmental conditions typically inhabited by the species (compared with a model made without them) – i.e. in this case the former will consider as suitable extensive hot and dry regions typically characterized by non-mesic vegetation where the species has not been captured despite extensive sampling in the region (Handley, 1976; Anderson *et al.*, 2012).

MATERIALS AND METHODS

Occurrence records, environmental variables and study region

The Caribbean spiny pocket mouse, *Heteromys anomalus*, is found in north-western South America, where it is associated with tropical evergreen and deciduous forests existing at c. 0–1600 m a.s.l. However, it is also known to range into drier regions (e.g. characterized by savannas and thorn scrub) through gallery forests (Handley, 1976; Anderson, 2003; Anderson & Gutiérrez, 2009), making it an appropriate candidate for this study. Occurrence records were obtained from the literature (previously verified by taxonomic experts and georeferenced with a precision similar to the resolution of the environmental data). To reduce the potential for spurious environmental correlations due to sampling bias (Hortal *et al.*, 2008; Veloz, 2009), the 215 unique occurrence records were spatially filtered, yielding a final total of 126 records (details in Appendix S1 in Supporting Information).

For environmental data, we used the bioclimatic variables available from the WorldClim project at a 30 arc-second resolution (c. 1 km² at the equator; Hijmans *et al.*, 2005). These variables provide climatic information that is likely to be relevant for this species at the scale employed (Luoto *et al.*, 2007; Peterson *et al.*, 2011, p. 89), and have proven successful in modelling the distribution of other Neotropical rodents (e.g. Anderson & Raza, 2010). However, their relatively coarse resolution makes them vulnerable to the potential issues associated with PSM localities. As we are interested in building a predictive rather than an explanatory model (Araújo & Guisan, 2006), we used the complete set of 19 bioclimatic variables aiming to maximize predictive accuracy under a machine learning approach (Breiman, 2001). Even though these variables can suffer from correlation, we take advantage of MAXENT's regularization to reduce unnecessary dimensionality resulting in overly complex models (see Elith *et al.*, 2011), while avoiding subjective *a priori* decisions regarding variable choice and relevance. Specifically, we carried spatially independent evaluations for preliminary models varying in the complexity allowed, reducing the likelihood of overfitting to noise or bias (see 'Model calibration' below and Appendix S1).

We constrained our models to a study region that encompassed the occurrence records but avoided inclusion of large areas where the species might not be present as a result of unsuitable biotic contexts or dispersal limitations. Specifically, we defined it as a rectangle encompassing the 126 filtered records, delimited by the nearest even 0.5° that was at least 20 km away from the most peripheral record in each cardinal direction (Appendix S1). This is more likely to yield response curves unbiased by factors not explicitly incorporated into the model, producing a more realistic approximation of the species' existing fundamental niche, and corresponding abiotically suitable area within the study region (Anderson & Raza, 2010; Barve *et al.*, 2011; Anderson, 2013).

Model calibration

We generated models using MAXENT 3.3.1 (Phillips *et al.*, 2006). This program has performed well in comparisons of modelling techniques that do not require recorded absences and allows fitting of complex responses (Elith *et al.*, 2006, 2011). To approximate optimal model complexity (affecting both dimensionality and flexibility of the modelled responses), we conducted tuning experiments. These consisted of preliminary models evaluated using spatially independent subsets (Radosavljevic & Anderson, 2014; Appendix S1). After determining settings resulting in best predictive performance (Table S1 in Appendix S1), we built a model using all 126 filtered records.

Identification of protruding spatially marginal localities

To identify records corresponding to PSM localities, we used a two-step process. The first consisted of identifying records considered environmentally marginal by the original MAX-ENT model. We ranked the 126 occurrence records according to the logistic prediction value they each received (Phillips & Dudík, 2008; Elith *et al.*, 2011). We then plotted logistic prediction value versus rank, investigating changes in slope or presence of gaps potentially denoting major changes in environmental quality.

For the second step, we gathered descriptive habitat information for a subset of records spanning the lower-suitability end of the plot, using collectors' field notes and published literature (Appendix S1). We also used vegetation maps (PdVSA, 1992; IGAC, 2003) to clarify insufficient or ambiguous descriptions (e.g. if the information pertained only to the specific trapping site but did not describe the surrounding habitat). We then classified each examined record into two broad categories: (1) extensive tropical forests (prior to human intervention) that should not suffer from issues of spatial marginality with the variables employed; or (2) natural mosaics of mesic and xeric habitat denoting PSM localities (where strips or patches of tropical forests exist within savannas or thorn scrub). Whenever possible, we also secured habitat information specific to the trapline (transect) where the species was captured within the mosaic.

Testing predictions

We tested prediction 1 – records receiving low suitability values correspond to PSM localities – in several ways. First, we visually assessed whether records corresponding to localities characterized by natural mosaics (i.e. PSM localities) did indeed receive low suitability values. Then, we tested whether the suitability values assigned to these mosaics were overall lower than those assigned to extensive forests by performing a Mann–Whitney *U*-test (one-tailed hypothesis). Test statistics were calculated using a spreadsheet (Microsoft

Excel® 2007), and critical values were obtained from Zar (1999; pp. App89–App100). Additionally, to gain more insight into the environmental characteristics shared by PSM localities, we plotted all 126 occurrence records in the environmental space defined by the two variables most important to the model (i.e. resulting in higher gain during internal iterations of model calibration). Because this represents an *ad hoc* analysis (the modelling framework was not aimed at understanding variable contribution), we also plotted records in the environmental space defined by the two principal components of a principal components analysis (PCA) of climatic variability within the study region (Appendix S1).

Regarding prediction 2 – at PSM localities, the species was captured within local patches of mesic habitat – the number of records for which we were able to find trapline habitat information was very small, precluding the application of statistical analyses. For this reason, we limit the test of this prediction to inspection of the observed findings.

To test prediction 3 – PSM localities provide the model with a misleading environmental signal, resulting in a model that overestimates the environmental conditions typically inhabited by the species, compared with a model built without them – we calibrated an additional model excluding a subset of PSM localities. As a conservative approach, we only excluded those PSM localities receiving suitability values lower than that assigned to the least-suitable record representing an extensive forest. For standardization purposes and ease of comparison, this model was built using the same settings as the original model (see Appendix S1 for a model built with optimal settings for this reduced dataset and resulting in qualitatively equivalent conclusions). We then compared the predictions of the two models with each other and with vegetation maps of the region (PdVSA, 1992; IGAC, 2003). To facilitate interpretation, we converted the continuous suitability surfaces of both models into binary predictions of suitable versus unsuitable areas. To do so, we used the lowest presence thresholding rule (= minimum training presence of MAXENT), where the lowest suitability score assigned to an occurrence record is considered as the minimum value denoting suitability. However, to avoid oversimplification, we also applied the 10th percentile thresholding rule (= 10 percentile training omission of MAXENT), which uses the lowest value assigned to any of the 90% of the records with highest scores to denote suitability (Pearson *et al.*, 2007). Differences in continuous outputs are shown in Appendix S1. To assess the potential effect of differing sample sizes when comparing both models, we also compare their predictions with those of 100 models built by randomly withholding the same number of records excluded in the second model (Appendix S1). Finally, we compared the environmental space considered as suitable by the models built with and without the subset of PSM localities. To do this, we plotted the corresponding estimated abiotically suitable areas on the first two principal components of the PCA (Appendix S1).

RESULTS

Identification of protruding spatially marginal localities and their natural history

By plotting occurrence records according to their suitability and obtaining habitat descriptions for the lowest-ranking ones, we detected that several of them represented PSM localities. We were able to obtain general habitat information for 42 of the 54 records analysed. Of these, 28 correspond to originally extensive evergreen or deciduous forests, and 14 correspond to natural mosaics (PSM localities) characterized by patches of forests intermixed with savannas and/or xerophytic thorn scrub (Appendix S2).

Testing predictions

In agreement with the first prediction, most of the lowest-ranking records corresponded to PSM localities (Fig. 2). Specifically, the least-suitable extensive forest corresponded to rank 16, below which all records corresponded to either natural mosaics (10), or localities for which we could not retrieve descriptive habitat information (5) but located within the same vegetational zone where most of the mosaics were found (savannas of the llanos; PdVSA, 1992). Records with ranks 17–21 corresponded to either mosaics or extensive forests, whereas higher ranks corresponded almost

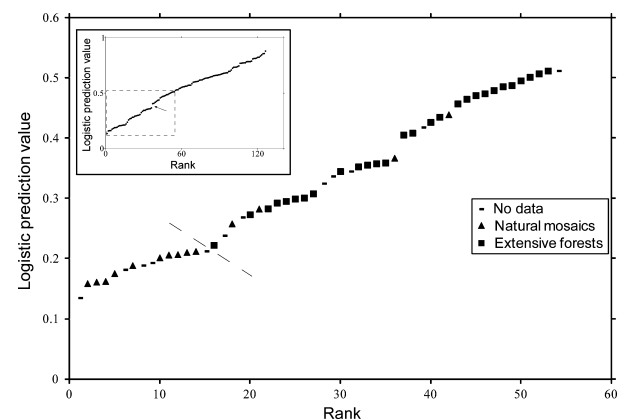


Figure 2 Occurrence records of *Heteromys anomalus* in north-western South America, ranked according to the logistic prediction values given by the original MAXENT model. Higher ranks denote more-suitable environments. Inset shows ranking of all 126 records, with arrow pointing to strongest gap in suitability at the lower end of the graph (used to choose the set of records analysed; Appendix S1). The dashed rectangle encompasses the 54 lowest-ranking records for which we aimed to gather descriptive habitat information (detailed in main figure). Notice predominance of natural mosaics (i.e. protruding spatially marginal localities; triangles) at the lower end of the plot and that of extensive forests (squares) at higher suitability values. Records below the lowest-ranking extensive forest (dashed line) correspond to those excluded from the calibration of the second model.

entirely to extensive forests (including those spanning the major gap in suitability; Appendices S1 & S2). Overall, suitability values assigned to natural mosaics were significantly lower than those assigned to extensive forests ($U = 359$, $n_1 = 14$, $n_2 = 28$, $P < 0.0005$). When visualized in environmental space, most records considered to represent PSM localities clearly differed from the rest (Fig. 3, and Fig. S1 in Appendix S1).

Confirming the second prediction, seven of the eight records representing PSM localities for which we secured trapline information corresponded to captures within the local patches of mesic forest (Appendix S2). At least in two instances, collectors also sampled the habitats surrounding the mesic forests. August (1984) sampled dry 'médano' habitat (characterized by xerophytic trees and shrubs), as well as seasonally flooded savannas, but only captured *Heteromys anomalus* within deciduous gallery forests, where flooding during the wet season is not extensive. Soriano & Clulow (1988) sampled seasonally flooded and non-flooded savannas, but only captured *H. anomalus* within gallery forests. The only instance where the species was not captured in a forest patch corresponded to a record along hedge-like borders of 'mayas' (terrestrial bromeliads) and thorny scrub along trails (Osgood, 1912). While we are not able to statistically evaluate these results, they suggest that within such mosaics, *H. anomalus* utilizes the portion that most closely resembles its typical mesic habitat.

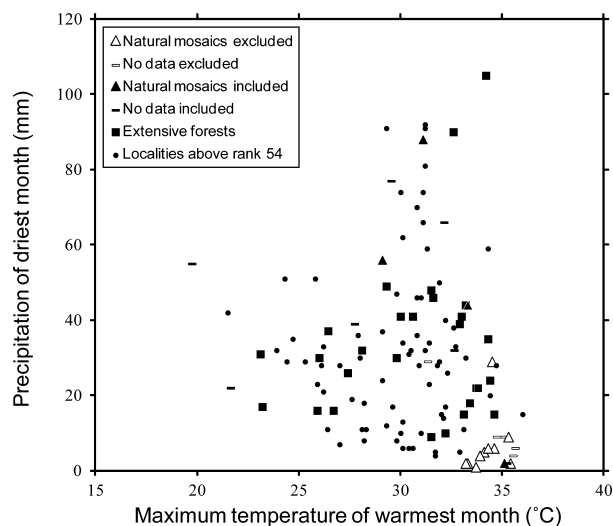


Figure 3 Two-dimensional environmental characterization of occurrence records of *Heteromys anomalus* in north-western South America. The environmental variables correspond to those contributing the most to model gain during internal iterations of the original MaxEnt model. Note how records excluded from the calibration of the second MaxEnt model (hollow symbols) differ from those that were included in the calibration of both models (solid symbols). Excluded records denote a similarly dry environmental space, characterized by the highest temperatures and lowest precipitation during the most extreme months.

In accordance with the third prediction, the model built with all records was unrealistically overpredictive when compared with that built excluding the 15 lowest-ranking records considered to represent PSM localities (Fig. 4, and Fig. S2 in Appendix S1). Overall, the model built with all records was much more permissive in its estimate of suitability, considering 91% of the study region as suitable at the lowest threshold (in contrast to 71% by the model built excluding PSM localities; Table S2 in Appendix S1). These differences stem from the fact that whereas both models considered regions characterized by mesic habitat as suitable, the one made with all records also considered suitable hot and dry regions typically exhibiting non-mesic habitat. These regions include much of the coastal lowlands of north-western Colombia and Venezuela, as well as the extensive llanos of central Venezuela (Fig. 4a). They are characterized either by xerophytic thorn scrub and littoral vegetation (coastal lowlands), or by extensive savannas (llanos; PdVSA, 1992; IGAC, 2003). While several of these areas were still considered as suitable even at the stricter threshold, the model built without the subset of PSM localities considered most such areas unsuitable under the lenient threshold (Fig. 4b). The overpredictive nature of the original model remained even when applying a thresholding rule corresponding to the number of PSM localities excluded from the second model (c. 12%, not shown). Models built by randomly removing the same number of records excluded in the second model were as overpredictive as the model built with all records (Fig. 4c, and Fig. S3 in Appendix S1). When comparing the estimated abiotically suitable areas in environmental space, it is clear that the model built with all records considers a broader range of environmental conditions as suitable for the species (i.e. a broader niche), specifically those characterizing warmer drier areas (Fig. S1). The possibility that such an environmental signal is the result of georeferencing errors seems unlikely, as these were small for almost all excluded PSM localities, and relevant variables showed little variability around them (Appendix S1).

DISCUSSION

Protruding spatially marginal localities and their effects on ecological niche models

As expected, *Heteromys anomalus* has been collected at localities consisting of vast expanses of its typical mesic habitat (Handley, 1976; Anderson & Gutiérrez, 2009) as well as at localities where such habitat protrudes into drier regions (PSM localities), especially through riparian strips of gallery forest (Appendix S2). Localities characterized by the typical habitat of the species were given a strong prediction value by the model, whereas PSM localities generally received a much lower value. Although striking and statistically significant, this pattern was not perfect. This could simply reflect the challenging task of obtaining a model that perfectly discriminates among records of differing environmental quality. At least two other alternatives exist. First, the environmental

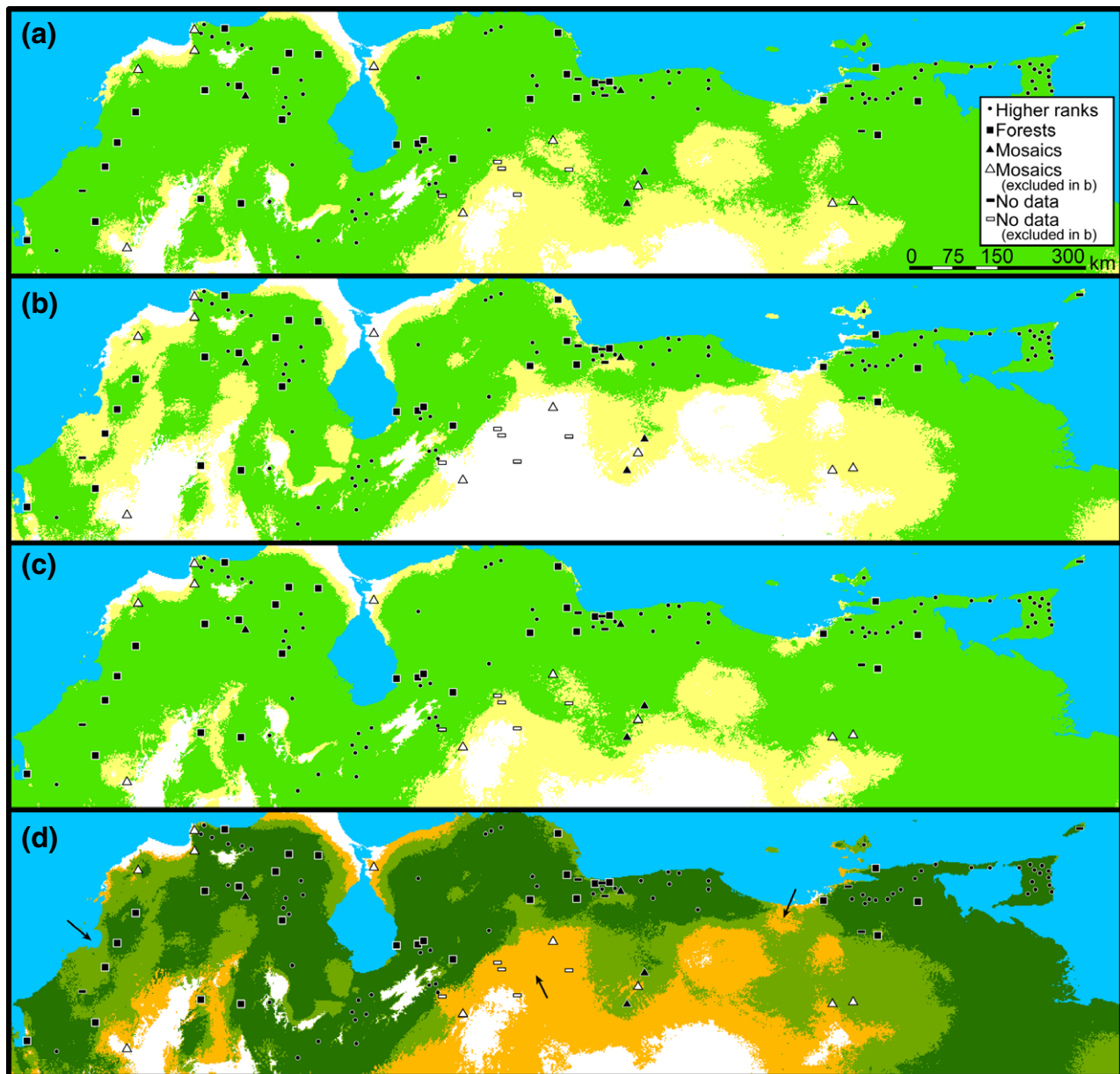


Figure 4 Abiotically suitable area for *Heteromys anomalus* in north-western South America modelled with MAXENT using different subsets of occurrence records. (a) Results of the original model, calibrated with all records. (b) Results of the model calibrated without the 15 records that ranked the lowest in panel (a), considered to represent protruding spatially marginal (PSM) localities. (c) One of 100 control models built by randomly excluding the same number of records as in panel (b) to control for sample size effects (see Appendix S1). In panels (a)–(c), lime green indicates suitable areas at the 10th percentile threshold, whereas yellow denotes additional areas suitable only under the lowest presence threshold. (d) Superimposition of models shown in panels (a) and (b). Greens indicate areas considered as suitable by the model made excluding PSM localities (dark green specifies suitability at the 10th percentile threshold; light green at the lowest presence threshold). Orange portrays additional areas considered as suitable only by the model made using all records (under the lowest presence threshold). Symbols indicate records used to build the models, specifying the habitat present at those analysed in the present study. Hollow symbols correspond to the PSM localities excluded in panel (b). Arrows indicate several noteworthy regions whose suitability depends upon the occurrence dataset used in calibration (left to right: coastal lowlands of north-western Colombia; Venezuelan llanos; and Depresión de Unare). These regions are characterized either by xerophytic thorn scrub or extensive savannas, and they are considered unsuitable (or suitable only at the lowest threshold) by the model built excluding PSM localities.

transition might be gradual along the region where higher ranking PSM localities were found. Second, extensive forests with lower suitability might markedly differ from the rest of forested localities in important environmental aspects.

Detailed ecological studies would be necessary to determine unequivocally whether this species only utilizes the mesic portion of the mosaic of habitats characterizing PSM localities. However, to our knowledge the species has never

been captured in homogeneous expanses of savannas or thorn scrub despite extensive sampling of small non-volant mammals in the region likely to detect its presence in such habitat (Handley, 1976; Rossi *et al.*, 2010; Anderson *et al.*, 2012; but see Utrera *et al.*, 2000, for captures in savannas associated with agriculture). For these reasons, we consider that these mosaics are mostly (or only) suitable for the species because of the patches of mesic habitat they harbour.

In the same manner, high resolution climatic data or proximal variables (e.g. wetness rather than precipitation; Austin, 2002) would be necessary to determine whether the patches of forest in PSM localities are truly environmentally marginal (as suggested by the original model). When visualized in the environmental space defined by the coarse climatic information included in the model, most PSM localities appear as unusually hot and seasonally dry (Fig. 3). However, it is possible that local environmental conditions within mesic patches at such localities approximate those of extensive forests. If so, in reality, these localities would be spatially but not environmentally marginal. In any case, with the coarse climatic variables used, the model made including PSM localities (but not that excluding them) considered extensive regions characterized by hot and dry conditions as suitable, regardless of whether they include patches of mesic habitat. For this reason, we consider that environmentally marginal or not, with the currently employed variables, these small patches of forest within a drier matrix provide the model with environmental values that correspond instead to the predominant surrounding sink habitat – where the species is not known to persist.

The strong collective signal provided by PSM localities is clear when comparing the prediction of the model built excluding a subset of these localities with that of models built excluding the same number of records but in a random manner. The overpredictive nature of the latter models indicates that rather than a mere sample size effect, the narrower prediction of the model built excluding PSM localities results from a reduction in the range of environmental values considered part of the niche (Fig. 3, Fig. S1). Simply applying a thresholding rule that accounts for the amount of records in the dataset representing PSM localities does not lead to the same result, as such records were still used to model the niche. Thus, without a finer resolution or inclusion of proximal variables, PSM localities will most likely lead to a substantial exaggeration of the species' existing fundamental niche and the area abiotically suitable for it within the study region (Fig. 4, Fig. S1).

We acknowledge that with current information it is not possible to ascertain whether hot xeric regions are unsuitable for *Heteromys anomalus* strictly because of their abiotic conditions (i.e. warmer and drier, as assumed in the present study), the lack of appropriate biotic contexts (e.g. constant seed productivity), or both. If hot xeric regions were unsuitable strictly because of inappropriate biotic contexts, this would imply that the abiotic tolerances estimated when including all records are not unrealistically broad, but rather

are biotically distorted (Anderson, 2013). Indeed, physiological experiments often reveal tolerances to temperatures higher than those present in the habitats that species occupy in nature (and thus higher than those inferred via ENMs; Araújo *et al.*, 2013). However, we find this possibility unlikely in the present study, as tolerance towards higher temperatures strongly depends upon water availability, and living within hot and xeric conditions requires complex behavioural and physiological adaptations typically not present in species found in mesic areas (Bartholomew, 1963). Therefore, we consider that inclusion of PSM localities in ENMs can overestimate tolerances for at least certain axes of the species' niche. Moreover, the abiotic conditions characterized by PSM localities consistently correlate with vast expanses of unsuitable habitat throughout the study region (i.e. extensive thorn scrub and savannas). We consider that a model deeming such extensive areas of unsuitable habitat as suitable definitely represents an overestimation of the areas the species could potentially inhabit.

Implications and proposed approach

While the scope of this study was limited to a single species and specific type of spatial marginality (mesic habitat protruding into hot and dry regions), given the wealth of species dependent upon patches of locally suitable habitat at spatial margins in general (Gaston, 2003), this issue is likely to be of broad relevance to the field of correlative ENM. However, its pervasiveness and intensity will depend upon the idiosyncrasies of the study system and modelling approach. For example, rare (or hard to survey) species are difficult to encounter even in optimal habitat, making detection in PSM localities still more unlikely and thus less of a concern. Detection probabilities can also be lower at PSM localities for abundant species if local conditions cause highly dynamic population patterns (e.g. Yackulic *et al.*, 2013). Finally, the specific effect of PSM localities will probably depend upon the environmental dimensionality of the study and modelling algorithm used. As the issue pertains to data documenting species presences, it should be of concern to other modelling techniques, including those relying on pseudo-absences or real absence data. However, the magnitude of the effect will differ according to the particularities of the technique; for instance, algorithms vary in their sensitivity towards environmental 'outliers' (Olden *et al.*, 2008).

Overall, we predict that the issue presented in this study will be of relevance in systems where: (1) the species depends upon patches of particular habitat in parts of its range; (2) these patches are not detected by the environmental variables used; and (3) such patches are scarce at least in certain parts of the range. Considering together models made including and excluding PSM localities would represent a fruitful approach in such instances. For example, superimposition of the prediction from the model made excluding PSM localities on top of that of the model calibrated with all records provides a better understanding of what modelled suitability

implies in different areas (Fig. 4d). Areas predicted as suitable by the model made excluding PSM localities represent those likely to harbour the typical habitat of the species. Complementarily, areas predicted as suitable only by the model made with all records indicate regions characterized by unsuitable habitat but where the species might exist within suitable patches if the appropriate local environmental conditions are present. Such a joint interpretation provides a more realistic depiction of the abiotically suitable area and potential patterns of population connectivity.

The approach of jointly interpreting ENMs made with and without PSM localities is likely to hold relevance in several areas of ecology and evolutionary biology. For example, ENMs are commonly used to infer past refugia and potential patterns of population connectivity tested with genetic data (e.g. Galbreath *et al.*, 2011; Lorenzen *et al.*, 2011). Jointly considering models depicting the effect of PSM localities, researchers could obtain more realistic inferences of past demographic scenarios, discriminating between the most relevant refugia, and areas that could have harboured suitable conditions, potentially acting as corridors. ENMs are also used to study niche evolution and speciation (e.g. Kozak & Wiens, 2006; McCormack *et al.*, 2010). However, overestimating species niches and distributions when unwittingly including PSM localities might result in biased conclusions regarding niche similarity and geographic connectivity.

This approach can also aid researchers using ENMs to address more practical applications. For example, joint predictions can help conservation biologists pinpoint areas harbouring a species' typical habitat, avoiding investment in regions where suitable conditions will rarely be met. Similarly, accurately discriminating these two sorts of areas can be crucial when assessing risks associated with invasive species or spreading zoonotic diseases (higher in the species' typical habitat; Sax & Brown, 2000; Peterson, 2006).

Future directions: sources, sinks and spatial margins

The approach implemented in this study demonstrates how by inspecting the modelling output, researchers can uncover the presence of PSM localities and identify those most problematic in a system. However, multiple alternative approaches can be employed for this goal. For instance, PSM localities might be uncovered by applying techniques used to identify outliers (e.g. Mahalanobis distances with respect to the centroid of the data; Mateo *et al.*, 2013), as well as by comparison of records in multidimensional space (PCA; e.g. Petitpierre *et al.*, 2012).

Additionally, this study focused specifically on the effect of PSM localities in ENMs. However, the role of spatially marginal localities in general remains to be tested, not only within this sub-discipline, but within broader biological fields. Much empirical research is still needed to elucidate the ecological and evolutionary roles of populations existing at such localities (Kawecki, 2008). For instance, certain localities can function either as sources or sinks (e.g. 'leaky sinks' of Holt & Gaines,

1992), contingent upon stochastic factors (Jansen & Yoshimura, 1998; Gonzalez & Holt, 2002; Guo *et al.*, 2005). Such factors might be especially important at spatially marginal localities, where potentially lower-quality environmental conditions or smaller areal extents can result in lower carrying capacities (Holt *et al.*, 2005). However, even localities acting as sinks can have important demographic consequences, including inflationary effects on abundance of adjacent sources, buffering of perturbations within sources, and even constituting metapopulations on their own (Roy *et al.*, 2005; Matthews & Gonzalez, 2007; and references above).

Finally, spatially marginal localities probably allow species to extract novel resources from unsuitable habitats. Such resources will usually be unavailable to a species except where unsuitable habitats intermix with those that the species is adapted to, bolstering its persistence in a given area. More broadly, the ability of a species to make partial use of sink habitat could be considered an additional aspect of its niche (*sink fishing*). This capacity will influence for how long (and consequently over what distances) the species can endure unfavourable conditions. Such issues also hold relevance for movement across sink habitats and colonization of suitable patches. Overall, we hope that this contribution not only highlights the relevance of spatially marginal localities for many practical applications, but also fosters theoretical and empirical advances regarding the study of geographic distributions.

ACKNOWLEDGEMENTS

We are grateful to the many who provided information on collecting localities, including Marisol Aguilera, Salvador Bohrer, Carlos Ibáñez, Omar Linares, José Ochoa-G., Magaly Ojeda and Javier Sánchez. The Smithsonian Institution provided access to the highly detailed data resulting from the visionary Smithsonian Venezuelan Project (SVP). Steven Phillips helped troubleshoot batch files for MAXENT; Lisa Manne provided feedback on statistical design; Diego Alvarado-Serrano and John Robinson offered guidance with R scripts. Babatunde Alade, Robert Boria, Peter Galante, Maria Gavrutenko, Samuel Glickman, Mark Maraj, Darla Thomas and Lucas Warton assisted with data management. James Brown, Ana Carnaval, Eliécer Gutierrez and Christopher Raxworthy provided encouraging discussions and suggestions that improved the manuscript. Robert Boria, Michael Hickerson, David Lohman and Robert Voss read preliminary versions of the manuscript. Finally, we are grateful to Rubén Mateo, Miguel Araújo and an anonymous referee for critical revision. This work was funded by National Science Foundation grants DEB-0717357 and DEB-1119915 (to R.P.A.), including a Research Experiences for Undergraduates supplement to J.L.R. J.L.R. obtained additional support from the City College Academy for Professional Preparation. M.S.G. received funding from the Graduate Center of the City University of New York and a travel award from the American Society of Mammalogists.

REFERENCES

- Anderson, R.P. (2003) Taxonomy, distribution, and natural history of the genus *Heteromys* (Rodentia: Heteromyidae) in western Venezuela, with the description of a dwarf species from the Península de Paraguaná. *American Museum Novitates*, **3396**, 1–43.
- Anderson, R.P. (2012) Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*, **1260**, 66–80.
- Anderson, R.P. (2013) A framework for using niche models to estimate impacts of climate change on species distributions. *Annals of the New York Academy of Sciences*, **1297**, 8–28.
- Anderson, R.P. & Gutiérrez, E.E. (2009) Taxonomy, distribution, and natural history of the genus *Heteromys* (Rodentia: Heteromyidae) in central and eastern Venezuela, with the description of a new species from the Cordillera de la Costa. *Systematic mammalogy: contributions in honor of Guy G. Musser* (ed. by R.S. Voss and M.D. Carleton). *Bulletin of the American Museum of Natural History*, **331**, 33–93.
- Anderson, R.P. & Raza, A. (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography*, **37**, 1378–1393.
- Anderson, R.P., Gutiérrez, E.E., Ochoa-G., J., García, F.J. & Aguilera, M. (2012) Faunal nestedness and species–area relationship for small non-volant mammals in “sky islands” of northern Venezuela. *Studies on Neotropical Fauna and Environment*, **47**, 157–170.
- Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.
- Araújo, M.B., Ferri-Yáñez, F., Bozinovic, F., Marquet, P.A., Valladares, F. & Chown, S.L. (2013) Heat freezes niche evolution. *Ecology Letters*, **16**, 1206–1219.
- August, P.V. (1984) Population ecology of small mammals in the llanos of Venezuela. *Special Publications of the Museum, Texas Tech University*, **22**, 71–104.
- Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.
- Bartholomew, G.A. (1963) The roles of physiology and behaviour in the maintenance of homeostasis in the desert environment. *Symposia of the Society for Experimental Biology*, **18**, 7–29.
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J. & Villalobos, F. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, **222**, 1810–1819.
- Breiman, L. (2001) Statistical modeling: the two cultures. *Statistical Science*, **16**, 199–231.
- Brown, J.H. (1984) On the relationship between abundance and distribution of species. *The American Naturalist*, **124**, 255–279.
- Brown, J.H., Mehlman, D.W. & Stevens, G.C. (1995) Spatial variation in abundance. *Ecology*, **76**, 2028–2043.
- Brown, J.H., Stevens, G.C. & Kaufman, D.M. (1996) The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, **27**, 597–623.
- Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.
- Galbreath, K.E., Cook, J.A., Eddingsaas, A.A. & DeChaine, E.G. (2011) Diversity and demography in Beringia: multilocus tests of paleodistribution models reveal the complex history of arctic ground squirrels. *Evolution*, **65**, 1879–1896.
- Gaston, K.J. (2003) *The structure and dynamics of geographic ranges*. Oxford University Press, New York.
- Gonzalez, A. & Holt, R.D. (2002) The inflationary effects of environmental fluctuations in source–sink systems. *Proceedings of the National Academy of Sciences USA*, **99**, 14872–14877.
- Grinnell, J. (1917) The niche-relationships of the California Thrasher. *The Auk*, **34**, 427–433.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Guo, Q., Taper, M., Schoenberger, M. & Brandle, J. (2005) Spatial-temporal population dynamics across species range: from centre to margin. *Oikos*, **108**, 47–57.
- Handley, C.O., Jr (1976) Mammals of the Smithsonian Venezuelan Project. *Brigham Young University Science Bulletin, Biological Series*, **20**(5), 1–91.
- Hengeveld, R. & Haeck, J. (1982) The distribution of abundance. I. Measurements. *Journal of Biogeography*, **9**, 303–316.
- 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.
- Holt, R.D. & Gaines, M.S. (1992) Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. *Evolutionary Ecology*, **6**, 433–447.
- Holt, R.D., Keitt, T.H., Lewis, M.A., Maurer, B.A. & Taper, M.L. (2005) Theoretical models of species' borders: single species approaches. *Oikos*, **108**, 18–27.
- Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M. & Baselga, A. (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, **117**, 847–858.
- Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, **22**, 415–427.

- IGAC (2003) *Atlas de Colombia*, 5th edn. Instituto Geográfico Agustín Codazzi, Colombia.
- Jansen, V.A.A. & Yoshimura, J. (1998) Populations can persist in an environment consisting of sink habitats only. *Proceedings of the National Academy of Sciences USA*, **95**, 3696–3698.
- Kawecki, T.J. (2008) Adaptation to marginal habitats. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 321–342.
- Kozak, K.H. & Wiens, J.J. (2006) Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution*, **60**, 2604–2621.
- Lira-Noriega, A., Soberón, J. & Miller, C.P. (2013) Process-based and correlative modeling of desert mistletoe distribution: a multiscalar approach. *Ecosphere*, **4**, 99.
- Lomolino, M.V., Riddle, B.R. & Brown, J.H. (2006) *Biogeography*, 3rd edn. Sinauer Associates, Sunderland, MA.
- Lorenzen, E.D., Nogués-Bravo, D., Orlando, L. et al. (2011) Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, **479**, 359–364.
- Lozier, J.D., Aniello, P. & Hickerson, M.J. (2009) Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. *Journal of Biogeography*, **36**, 1623–1627.
- Luoto, M., Virkkala, R. & Heikkinen, R.K. (2007) The role of land cover in bioclimatic models depends on spatial resolution. *Global Ecology and Biogeography*, **16**, 34–42.
- Mateo, R.G., de la Estrella, M., Felicísimo, A.M., Muñoz, J. & Guisan, A. (2013) A new spin on a compositionalist predictive modelling framework for conservation planning: a tropical case study in Ecuador. *Biological Conservation*, **160**, 150–161.
- Matthews, D.P. & Gonzalez, A. (2007) The inflationary effects of environmental fluctuations ensure the persistence of sink metapopulations. *Ecology*, **88**, 2848–2856.
- McCormack, J.E., Zellmer, A.J. & Knowles, L.L. (2010) Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation? insights from tests with niche models. *Evolution*, **64**, 1231–1244.
- Olden, J.D., Lawler, J.J. & Poff, N.L. (2008) Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology*, **83**, 171–193.
- Osgood, W.H. (1912) Mammals from western Venezuela and eastern Colombia. *Field Museum of Natural History Publications, Zoological Series*, **10**, 33–66.
- PdVSA (1992) *Imagen de Venezuela. Una visión espacial*. Petróleos de Venezuela S.A. Instituto de Ingeniería, Venezuela.
- Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Peterson, A.T. (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102–117.
- Peterson, A.T. (2006) Ecologic niche modeling and spatial patterns of disease transmission. *Emerging Infectious Diseases*, **12**, 1822–1826.
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological niches and geographic distributions*. Monographs in Population Biology, Vol. 49. Princeton University, Princeton, NJ.
- 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. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.
- Phillips, S.J., Anderson, R.P. & Shapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Pulliam, H.R. (1988) Sources, sinks, and population regulation. *The American Naturalist*, **132**, 652–661.
- Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.
- Radosavljevic, A. & Anderson, R.P. (2014) Making better MAXENT models of species distributions: complexity, overfitting, and evaluation. *Journal of Biogeography*, **41**, 629–643.
- Rossi, R.V., Voss, R.S. & Lunde, D.P. (2010) A revision of the didelphid marsupial genus *Marmosa*. Part 1. The species in Tate's 'Mexicana' and 'Mitis' sections and other closely related forms. *Bulletin of the American Museum of Natural History*, **334**, 1–83.
- Roy, M., Holt, R.D. & Barfield, M. (2005) Temporal autocorrelation can enhance the persistence and abundance of metapopulations comprised of coupled sinks. *The American Naturalist*, **166**, 246–261.
- Sagarin, R.D. & Gaines, S.D. (2002) The 'abundant centre' distribution: to what extent is it a biogeographical rule? *Ecology Letters*, **5**, 137–147.
- Sagarin, R.D., Gaines, S.D. & Gaylord, B. (2006) Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends in Ecology and Evolution*, **21**, 524–530.
- Sax, D.F. & Brown, J.H. (2000) The paradox of invasion. *Global Ecology and Biogeography*, **9**, 363–371.
- Soriano, P.J. & Clulow, F.V. (1988) Efecto de las inundaciones estacionales sobre poblaciones de pequeños mamíferos en los llanos altos occidentales de Venezuela. *Ecotrópicos*, **1**, 3–10.
- Utrera, A., Duno, G., Ellis, B.A., Salas, R.A., de Manzione, N., Fulhorst, C.F., Tesh, R.B. & Mills, J.N. (2000) Small mammals in agricultural areas of the western llanos of Venezuela: community structure, habitat associations, and relative densities. *Journal of Mammalogy*, **81**, 536–548.
- Veloz, S.D. (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, **36**, 2290–2299.
- Warren, D.L. (2012) In defense of 'niche modeling'. *Trends in Ecology and Evolution*, **27**, 497–500.

- Whittaker, R.H. (1956) Vegetation of the Great Smoky Mountains. *Ecological Monographs*, **26**, 1–80.
- Wisz, M.S., Pottier, J., Kissling, W.D. *et al.* (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.
- Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H. & Veran, S. (2013) Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution*, **4**, 236–243.
- Zar, J.H. (1999) *Biostatistical analysis*, 4th edn. Prentice Hall, Upper Saddle River, NJ.

SUPPORTING INFORMATION

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

- Appendix S1** Methodological details and additional results.
- Appendix S2** Habitat descriptions for collecting localities analysed in the present study.

BIOSKETCH

Mariano Soley-Guardia is interested in niche theory and methodological improvements leading to better estimates of both species niches and their geographic distributions. He is also involved in the integration of such inferences with molecular tools to address questions in the interface between ecology and evolution.

The Anderson lab at City College (<http://web.sci.ccny.cuny.edu/~anderson/>) is interested in the improvement and application of correlative ecological niche models. We believe these highly valuable tools can be further improved (conceptually and methodologically) so that they parallel the quality of inferences obtained from other disciplines.

Author contributions: M.S.G., R.P.A. and A.R. conceived the ideas detailed in the paper, which the first two later refined; M.S.G. built the models and performed the statistical analyses; J.L.R. and A.R. gathered most of the data concerning habitat descriptions, which was completed and classified by M.S.G. with input from R.P.A.; M.S.G. wrote the manuscript and prepared the figures with constant feedback from A.R. and R.P.A.

Editor: Miguel Araújo