

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

Mariano Soley-Guardia, Aleksandar Radosavljevic, Jhanine L. Rivera and Robert P. Anderson

Appendix S1 Methodological details and additional results.

Occurrence records and study region

Heteromys anomalus is found in north-western South America, including mainland areas of Colombia and Venezuela, as well as the islands of Margarita, Trinidad and Tobago. This species typically inhabits low-to-middle elevations (*c.* 0–1600 m a.s.l.), where it is associated with tropical evergreen and deciduous forests (Handley, 1976; Anderson, 2003; Anderson & Gutiérrez, 2009). We obtained a total of 215 unique occurrence records from Anderson (2003), Anderson & Gutiérrez (2009) and Anderson *et al.* (2012). We excluded from our analysis six records from the upper Río Magdalena valley (as well as that region itself), because the distribution there is poorly characterized owing to limited sampling. Including this area in the background region of our analysis would be likely to result in an inaccurate estimation of the species' niche because of an artefactual non-equilibrium distribution (i.e. comparable to the effects of dispersal limitations and unsuitable biotic contexts; Anderson & Raza, 2010; Peterson *et al.*, 2011, p. 125). In this way, we focus on the well-characterized populations of northern South America, representing the main distribution of the species. Geographic coordinates reported in the aforementioned studies were obtained from gazetteers (e.g. Paynter, 1982, 1997), exhaustive georeferencing using detailed maps of the region (according to locality descriptions on specimen tags, collectors' field catalogues and/or field notes), or a GPS in the field. Authors report an estimated maximum error of *< c.* 5 km (*c.* 2 minutes) for most records, with few having a maximum error of *c.* 10 km.

In order to reduce the potential for artefactual correlations between records and environmental variables that merely reflect sampling bias (Hortal *et al.*, 2008; Veloz, 2009),

we spatially filtered the occurrence records. Even though species records are naturally autocorrelated (owing to endogenous and exogenous factors; see Dormann, 2007), spurious autocorrelation may result from sampling biases (Yackulic *et al.*, 2013). Such autocorrelation is difficult to quantify but can be diminished through spatial filtering. Specifically, whenever there was a cluster of records, we removed the minimum necessary number to retain only those that were at least a fixed distance of 10 km apart from each other (Anderson & Raza, 2010; Radosavljevic & Anderson, 2014). We did this by measuring Euclidean distances between records within a cluster in ARCGIS[®] 9.2 (ESRI, Redlands, CA, USA), and assessing all possible solutions. When multiple co-optimal solutions were possible for a given cluster, we randomly chose one. This yielded a final total of 126 unique occurrence records used for model calibration.

We defined a study region that minimized the inclusion of large areas where the species has not been documented, probably owing to sampling bias, inappropriate biotic contexts or dispersal limitations (& Raza, 2010; Barve *et al.*, 2011; but note relatively small areas inhabited by the congeners *H. australis* and *H. catopterus* in this region; Anderson, 2003; Anderson & Gutiérrez, 2009). This consisted of a rectangular area encompassing all 126 records, the margins of which were at least 20 km away from the most peripheral record in each cardinal direction (rounding to the nearest even 0.5°). The region is defined by the following coordinates: 7.50°–11.50° N, 60.00°–77.00° W.

Model calibration

The potential complexity of MAXENT models depends upon the type of relationship the program is allowed to infer between the species' occurrence records and each environmental variable (feature classes), as well as on how close of a match is expected between that plausible relationship and the actual environment represented by each occurrence record used in calibration (controlled by regularization; Elith *et al.*, 2011). Each feature class j (e.g. linear, quadratic) has its own associated regularization coefficient (β_j) that is set to a default value but can be individually varied (e.g. Anderson & Gonzalez, 2011). Alternatively, the regularization coefficient can be simultaneously varied across all features via the regularization multiplier value (hereafter the 'regularization multiplier'; Phillips & Dudík, 2008). Higher regularization leads to stronger penalties for complex models (e.g. one inferring unnecessary relationships between the species and the variables), and as such determines the set of variables (and complexity of modelled responses) that constitute the

final model (Elith *et al.*, 2011). In this way, higher values of the regularization multiplier should result in less overfitting to noise and/or biases in the data, as well as a reduction in the environmental dimensionality of the model.

Even though the feature classes and their level of regularization can be easily manipulated, most studies employing MAXENT rely on the default settings. These settings allow the use of particular features at a fixed regularization (i.e. the regularization multiplier set to 1.0), both of which depend on the number of input occurrence records (Phillips & Dudík, 2008). These settings were chosen based on previous extensive tuning experiments but are unlikely to be optimal for every species, region or environmental dataset, and indeed several studies now caution about the important effect of varying settings in model output (e.g. Elith *et al.*, 2010; Anderson & Gonzalez, 2011; Warren & Seifert, 2011; Renner & Warton, 2013; Shcheglovitova & Anderson, 2013). For this reason, we determined settings resulting in optimal model complexity by conducting tuning experiments. These consisted of preliminary models built with settings differing in the flexibility of response allowed, as well as in the level of protection against complexity.

Specifically, we varied both the feature classes explored by the model, and their level of regularization through the regularization multiplier. We used two combinations of feature classes, one consisting of all available continuous features in MAXENT (linear, quadratic, hinge, product and threshold; suggested by default according to sample size and hereafter referred as the ‘complex set’), and one consisting of a subset of these (linear, quadratic and hinge only; hereafter ‘simple set’, also considered appropriate for the used sample size; Phillips & Dudík, 2008; see MAXENT ‘Help’ file). We varied the regularization multiplier in 0.5 increments for both sets of feature classes, using all possible values between 0.5–4.0. It is possible that varying additional settings can have an effect on the resulting model (e.g. number of background points; Renner & Warton, 2013). However, given the current lack of a framework in the ENM field for simultaneously evaluating optimal model complexity across all possible scenarios (e.g. the amount of background points needed can vary according to model complexity), we limit our analyses to variation of feature classes and regularization multiplier only. In so doing, all other MAXENT settings were kept at their default values (e.g. auto-generation of background points, maximum number of the latter = 10 000, maximum iterations = 500; Phillips *et al.*, 2006; Phillips & Dudík, 2008).

Model evaluation

To test the predictive performance of the aforementioned preliminary models, we divided the occurrence records into calibration and evaluation datasets. Currently, the available option for automated model evaluation in MAXENT consists of randomly splitting occurrence records into calibration and evaluation subsets ('test percentage' or 'replicates'). However, such evaluations are subject to inflated scores owing to lack of spatial independence between these subsets of records (Veloz, 2009; Hijmans, 2012). For this reason, we evaluated model performance using a previously suggested evaluation scheme, consisting of spatially independent sub-samples (Araújo & Rahbek, 2006; Jiménez-Valverde *et al.*, 2011; Peterson *et al.*, 2011, p. 161).

Specifically, we divided occurrence records into eight geographically structured 'bins', dissecting the study region longitudinally (following Radosavljevic & Anderson, 2014). Each bin consisted of 15 or 16 records differing in their areal extent. In each iteration of model building, seven bins were included in calibration and one bin was excluded for evaluation. In this way, each bin was used once for evaluation in each series of experiments (see below). Additionally, the geographic area corresponding to the bin held out for evaluation was 'masked out', with no background data drawn from it during calibration (*masked geographically structured approach* of Radosavljevic & Anderson, 2014) to follow principles of study region selection (Anderson & Raza, 2010; Barve *et al.*, 2011). The number of bins chosen allowed occurrence records to be split into subsamples of nearly equal sizes that still yielded a combined sample size considered appropriate for using the complex set of feature classes when excluding the evaluation bin (i.e. > 80; Phillips & Dudík, 2008). While further splitting the data into more bins would also fulfil these criteria, we chose a number of bins that delimited reasonable areal extents (i.e. avoiding the usage of overly narrow bins whose records would exhibit high spatial autocorrelation with those of adjacent bins and thus negate the benefit of our partitioning scheme). Borders between bins were placed at the midpoint between the western-most and eastern-most occurrence records of adjacent bins. Processing of bins and their corresponding environmental variables was carried in ARCGIS® 9.2 (ESRI, Redlands, CA, USA). The different combinations of the two sets of feature classes and the eight different regularization multipliers resulted in a total of 16 models per evaluation bin, for a total of 128 preliminary models.

In order to select model settings that minimized overfitting but also maintained good discriminatory ability, we evaluated model performance using two measures: omission rate

and the area under the curve of the receiver operating characteristic plot (AUC; Fielding & Bell, 1997; Peterson *et al.*, 2011, p. 171). Our first criterion, the omission rate, quantifies the proportion of evaluation records that are considered to be unsuitable by a model after applying a specific threshold that results in a binary prediction. Hence, when compared with the theoretically expected omission rate (determined by the used threshold), an observed omission rate allows detection of overfitting to the calibration data (Shcheglovitova & Anderson, 2013; Radosavljevic & Anderson, 2014). Complementarily, our secondary criterion, AUC, provides a measure of the discriminatory ability of the model at all possible thresholds. For presence-only models, AUC is a relative measure, with discriminatory capacity only being relative to other models built under standardized conditions (e.g. same species and study region; Lobo *et al.*, 2008; Peterson *et al.*, 2011, p. 174; Yackulic *et al.*, 2013). For this reason, AUC values were calculated over the totality of the study region for each of the eight iterations made with the different settings (Radosavljevic & Anderson, 2014). For calculating omission rates in evaluation bins, we used the lowest presence (= minimum training presence of MAXENT) and the 10th percentile (= 10 percentile training omission of MAXENT) thresholds to define suitable vs. unsuitable areas (Pearson *et al.*, 2007). The former thresholding rule considers the lowest suitability score assigned to any of the occurrence records used during calibration as the minimum value denoting suitability, with an expected omission rate near zero for an independent evaluation dataset. In contrast, to denote minimum suitability, the latter threshold uses the lowest value assigned to any of the 90% of the calibration records with the highest suitability, with an expected omission of approximately 10% of evaluation records. We averaged the evaluation scores across the eight evaluation bins for each combination of settings. To minimize overfitting, we selected the combination that yielded average omission rates closest to expected. If the omission rate was the same (or essentially so) for different regularization multipliers of the same set of feature classes, we chose the lowest regularization multiplier as optimal, as such values tend to result in higher discrimination. We then assessed whether the average AUC value of the selected combination of settings indicated a high discrimination capacity relative to the other models (i.e. whether those settings yielded the highest AUC value or essentially so, not differing from the highest one by more than 0.001).

Models and their evaluation scores were obtained using the MAXENT 3.3.1 interface and batch files of command-line arguments (see Radosavljevic & Anderson, 2014 for examples). As MAXENT does not calculate evaluation scores for records found in regions to

which models are projected (i.e. for ‘masked’ evaluation bins), these were obtained using command-line arguments. Whereas AUC values can be calculated directly through these arguments, calculating omission rates required extracting suitability values for evaluation records and comparing these with the thresholds of suitability established with the calibration data using a spreadsheet (Microsoft Excel[®] 2007). Settings leading to best-performing preliminary models were considered as optimal and used to calibrate a model with the complete set of 126 filtered occurrence records. However, we acknowledge that even though extensive, the present search to determine optimal model complexity was not exhaustive in parameter space (the development of such automated tools being in clear need for the field).

Optimal settings consisted of the simple set of feature classes with a regularization multiplier of 2.5 (Table S1). In general, average evaluation omission rates decreased and average evaluation AUC values increased with increasing regularization multiplier values for both the simple and complex sets of feature classes. Omission stabilized at a multiplier value of 2.5 but is expected to continue decreasing with higher levels of regularization as the model loses specificity (Phillips *et al.*, 2006). Despite an overall increase with higher regularization multiplier values, AUC varied relatively little (0.69–0.72).

Identification of *protruding spatially marginal localities*

To identify protruding spatially marginal (PSM) localities, we plotted occurrence records according to the logistic prediction value received in the model made with all records, and obtained habitat descriptions for the lowest-ranking ones. Inspection of the plot revealed two notable gaps in assigned suitability towards the lower end. As the strongest gap occurred between ranks 36–37 (Fig. 2a of the main paper), we decided to obtain descriptive habitat information for all records below this gap. Additionally, we also obtained habitat descriptions for the 18 records above this gap (in turn separated from the rest by a minor gap in suitability), to assess potential differences in habitat between these sets of records. In total, we analyzed the habitat present at the 54 localities represented by the records with the lowest ranks in suitability. Even though this subset was chosen based on the a priori hypothesis that the major gaps in suitability could represent major changes in habitat type, this was not the case (Fig. 2b of the main paper).

Since the 15 records below the lowest-ranking extensive forest corresponded to natural mosaics (or to localities for which we could not retrieve habitat information but found within the same vegetational zone as the mosaics), we considered them to represent PSM

localities (Fig. 2 of the main paper). Hence, these records were excluded from the second model. Of these, 13 have coordinates with an estimated maximum error of $< c. 5$ km, and only two had a maximum error of $c. 10$ km (i.e. approximately corresponding to a maximum of five or 10 pixels of the environmental dataset, respectively). The latter two were found within the llanos of Venezuela, which represent extensive lowlands with no abrupt changes in elevation. For this reason, we consider that potential georeferencing errors have a small role (if any) in the differences observed in models built including and excluding these records. For instance, the maximum temperature of the warmest month (contributing the most to model gain during internal iterations) varied little within a radius of 10 km for the two PSM localities with the highest potential georeferencing error (30.2–33.7; 35.3–35.5 °C; compare with values for non-PSM localities in Fig. 3 of the main paper).

Principal components analysis of environmental variability within the study region

In order to better understand the effect of PSM localities in model calibration, in the main paper we plot all 126 occurrence records in the environmental space defined by the two variables that contributed the most to the model made with such records (i.e. variables most responsible for gain during internal iterations of the calibration; Fig. 3 of the main paper). However, because the modelling framework was not aimed at understanding variable contribution but rather maximizing predictive power (Araújo & Guisan, 2006), we also plotted these records in the environmental space defined by the first two principal components of a principal components analysis (PCA) of environmental variability within the study region.

This analysis was based on a sample of 10 000 cells randomly selected from the study region. For each cell, we extracted the value of each of the 19 bioclimatic variables included in the model using ARCGIS[®] 9.2 (ESRI, Redlands, CA, USA). We standardized the corresponding values so each variable had a mean of zero and a standard deviation of one. We then performed a PCA in R with the ‘prcomp’ function using a covariance matrix (R Core Team, 2013). The first principal component explained 39% of the variation, consisting mostly of variables representing mean temperatures and their extremes (mean annual temperature; maximum and minimum temperature of the respective warmest and coldest months; and mean temperatures of warmest, coldest, driest and wettest quarters). The second principal component explained 28% of the variation and consisted primarily of the variables related to

precipitation (annual precipitation; precipitation during driest and wettest months; and precipitation during wettest, driest and warmest quarters; see Table S3).

We then calculated the scores on these two principal components for the cells represented by the 126 occurrence records of the species. Additionally, we calculated these values for the totality of cells denoting three areas: the whole study region, the fraction of the study region considered as suitable by the model built with all records, and the fraction considered suitable when excluding the subset of PSM localities (both proportions determined using the lowest presence threshold). We then plotted these data to compare the environmental space they represent (Fig. S1). From this figure, it is clear that records from PSM localities segregate from the rest (including those representing extensive forests). Additionally, the subset of lowest-ranking PSM localities excluded from the building of the second model differ the most in environmental space. Similar patterns are also evident in the simpler and more intuitive environmental space shown in the main paper (Fig. 3). Additionally, this analysis illustrates how the niche inferred using all records is generally broader than that inferred when PSM localities were excluded, especially towards the right of the graph (denoting warmer conditions with various levels of precipitation). However, the niche inferred when excluding the subset of PSM localities is broader within a small region of environmental space denoting cool and wet conditions (green circle in figure), indicating that this niche is not simply a subset of that inferred when using all records. Even though these differences in the estimated niche are small relative to the available environmental space, they can have a significant impact in geography, as shown in the different geographic predictions of these models (Fig. S2; also Fig. 4 of the main paper).

Additional models

After building the model that excluded the subset of 15 PSM localities from the calibration process, we generated additional models to assess alternative explanations for the differences observed between this model and that built with all records (i.e. controlling for non-standardized factors). First, in order to rule out the possibility that differences in model predictions were because of a smaller sample size when excluding PSM localities, we built a suite of models that eliminated the same number of records but in a random manner. Secondly, because it is possible that changing the number of records might alter which settings are optimal for modelling the species' niche (especially as removed ones correspond

to particular habitats), we performed tuning experiments (as previously detailed) for the dataset that excluded the subset of PSM localities.

Controlling for a smaller sample size in the model built excluding PSM localities.

In order to test whether the more restrictive prediction of the model built excluding PSM localities resulted from a smaller sample size, we built 100 control models excluding the same number of records from the original data set, but in a random manner (i.e. removing 15 records with a random seed each time). These models were made using the same settings as both the original model, and that excluding PSM localities. The fraction of the study region predicted as suitable by these models was as extensive as that of the model built with all records (Table S2, Fig. S3).

Controlling for the use of potentially non-optimal settings in the model built excluding PSM localities.

Changing the number (and nature) of records used to calibrate a model may result in a change in settings optimal for modelling a species' niche. For this reason, we re-tuned the model for the dataset that excluded PSM localities following the same procedure detailed above for the model built with all records. The number of geographically structured bins and criteria for delimiting them remained the same, but this time each bin had 13, 14 or 15 records. The number of records per bin was chosen in a way that minimized differences in sample sizes and allowed bins to be divided by the middle of the longitudinally closest records (i.e. other alternatives consistently resulted in bins sharing records with the same longitude). Best-performing settings in this case consisted of the same set of features as the original model ('simple set'), but with a lower regularization multiplier (2.0). The geographic prediction of this model was qualitatively equivalent to that of the model that excluded PSM localities but used settings deemed as optimal for the whole dataset (Table S2). For this reason, and for ease of interpretation, we restrict comparisons in the main paper to models built with same settings.

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. & Gonzalez, I., Jr (2011) Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecological Modelling*, **222**, 2796–2811.
- 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. & Rahbek, C. (2006) How does climate change affect biodiversity? *Science*, **313**, 1396–1397.
- 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.
- Dormann, C.F. (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, **16**, 129–138.
- Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

- 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.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Handley, C.O., Jr (1976) Mammals of the Smithsonian Venezuelan Project. *Brigham Young University Science Bulletin, Biological Series*, **20**(5), 1–91.
- Hijmans, R.J. (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, **93**, 679–688.
- 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.
- Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P. & Lobo, J.M. (2011) Use of niche models in invasive species risk assessments. *Biological Invasions*, **13**, 2785–2797.
- Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.
- Paynter, R.A., Jr (1982) *Ornithological gazetteer of Venezuela*. Museum of Comparative Zoology, Harvard University, Cambridge, MA.
- Paynter, R.A., Jr (1997) *Ornithological gazetteer of Colombia*, 2nd edn. Museum of Comparative Zoology, Harvard University, Cambridge, MA.
- 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., 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, 49. Princeton University Press, Princeton, NJ.
- 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.
- Radosavljevic, A. & Anderson, R.P. (2014) Making better MAXENT models of species distributions: complexity, overfitting, and evaluation. *Journal of Biogeography*, **41**, 629–643.
- R Core Team (2013) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Renner, I.W. & Warton, D.I. (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, **69**, 274–281.
- Shcheglovitova, M. & Anderson, R.P. (2013) Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. *Ecological Modelling*, **269**, 9–17.
- 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. & Seifert, S.N. (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, **21**, 335–342.
- 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.

Table S1 Average evaluation scores for preliminary MAXENT models of *Heteromys anomalus* in north-western South America, built with settings varying in the degree of complexity allowed. Scores represent average performance of each combination of settings across eight evaluation bins. Best-performing settings shown in bold. LQHTP denotes the ‘complex set’ of feature classes (linear, quadratic, hinge, threshold and product), whereas LQH denotes the ‘simple set’ (linear, quadratic and hinge only). Lower omission rates indicate better prediction of independent evaluation records (and hence, less overfitting to the calibration data). Higher values of the area under the curve of the receiver operating characteristic plot (AUC) denote better discriminatory ability by the model.

Used feature classes	Regularization multiplier value	Omission rate (lowest presence threshold)	Omission rate (10th percentile threshold)	AUC
LQHTP	0.5	0.27448	0.55104	0.69188
LQHTP	1	0.08594	0.27552	0.71911
LQHTP	1.5	0.06250	0.14896	0.71023
LQHTP	2	0.05469	0.14063	0.70819
LQHTP	2.5	0.03906	0.14844	0.70959
LQHTP	3	0.04688	0.12500	0.71512
LQHTP	3.5	0.03906	0.13281	0.71729
LQHTP	4	0.03906	0.12500	0.71679
LQH	0.5	0.11094	0.29844	0.71620
LQH	1	0.07813	0.18750	0.71526
LQH	1.5	0.06250	0.15625	0.72504
LQH	2	0.03906	0.13281	0.72270
LQH	2.5	0.01563	0.10938	0.72729
LQH	3	0.04688	0.10938	0.72749
LQH	3.5	0.00781	0.10156	0.72650
LQH	4	0.03906	0.10156	0.72775

Table S2 Fraction of the study region considered as suitable for *Heteromys anomalus* in north-western South America by MAXENT models built with different subsets of occurrence records (see text). Suitability was assigned according to two different thresholds. The subset of protruding spatially marginal (PSM) localities excluded consists of that represented by the records receiving lowest suitability values in the model in the top row.

Dataset used	MAXENT settings (feature classes/regularization multiplier value)	Fractional predicted area (lowest presence threshold)	Fractional predicted area (10th percentile threshold)
All records ($n = 126$)	Linear, quadratic and hinge/2.5	0.910	0.718
Excluding PSM localities ($n = 111$)	Linear, quadratic and hinge/2.5	0.712	0.467
Excluding PSM localities ($n = 111$) (control for optimality of settings)	Linear, quadratic and hinge/2.0	0.702	0.463
Excluding random records ($n = 111$; range for 100 models) (control for sample size)	Linear, quadratic and hinge/2.5	0.834–0.958	0.59–0.763

Table S3 Results of the principal components analysis (PCA) of environmental variability within the study region based on a covariance matrix of 19 bioclimatic variables. Variables were standardized prior to the analysis. Elements of the unit eigenvector are given for the first two principal components. Eigenvalues and the percentage of variation explained by each component are shown at the bottom.

Environmental variables	Unit eigenvector	
	PC1	PC2
Annual mean temperature	0.357	-0.073
Mean diurnal range	-0.004	0.163
Isothermality	-0.106	-0.199
Temperature seasonality	0.072	0.186
Maximum temperature of warmest month	0.363	-0.020
Minimum temperature of coldest month	0.344	-0.116
Temperature annual range	0.054	0.242
Mean temperature of wettest quarter	0.348	-0.083
Mean temperature of driest quarter	0.359	-0.077
Mean temperature of warmest quarter	0.361	-0.059
Mean temperature of coldest quarter	0.356	-0.079
Annual precipitation	-0.016	-0.381
Precipitation of wettest month	0.025	-0.318
Precipitation of driest month	-0.118	-0.355
Precipitation seasonality	0.160	0.261
Precipitation of wettest quarter	0.040	-0.304
Precipitation of driest quarter	-0.122	-0.366
Precipitation of warmest quarter	-0.145	-0.308
Precipitation of coldest quarter	0.136	-0.192
Eigenvalue	7.383	5.366
Percentage of variation explained	38.9 %	28.2 %

Figure S1 Environmental characterization of several entities according to the first two principal components of the principal components analysis (PCA) of environmental variability within the study region. The light-coloured cloud (delineated with dashed polygon) indicates the environment available within the totality of the study region. Overlaying it, the orange cloud denotes the subset of that environment considered as suitable for *Heteromys anomalus* by the original MAXENT model built with all records. The foremost green cloud indicates the subset of the study region considered as suitable by the MAXENT model built excluding the subset of protruding spatially marginal (PSM) localities. The lowest presence threshold was applied to delimit suitability in both models. Symbols represent records used to build the models (see legend). Note that most of the records considered to represent PSM localities and excluded from the calibration of the second model (hollow symbols) share a similarly hot and dry environmental space. Although the niche inferred when excluding PSM localities is mostly a subset of that inferred when using all records (not apparent owing to overprinting by green), it is also slightly broader towards the upper left of the plot (bright green circle).

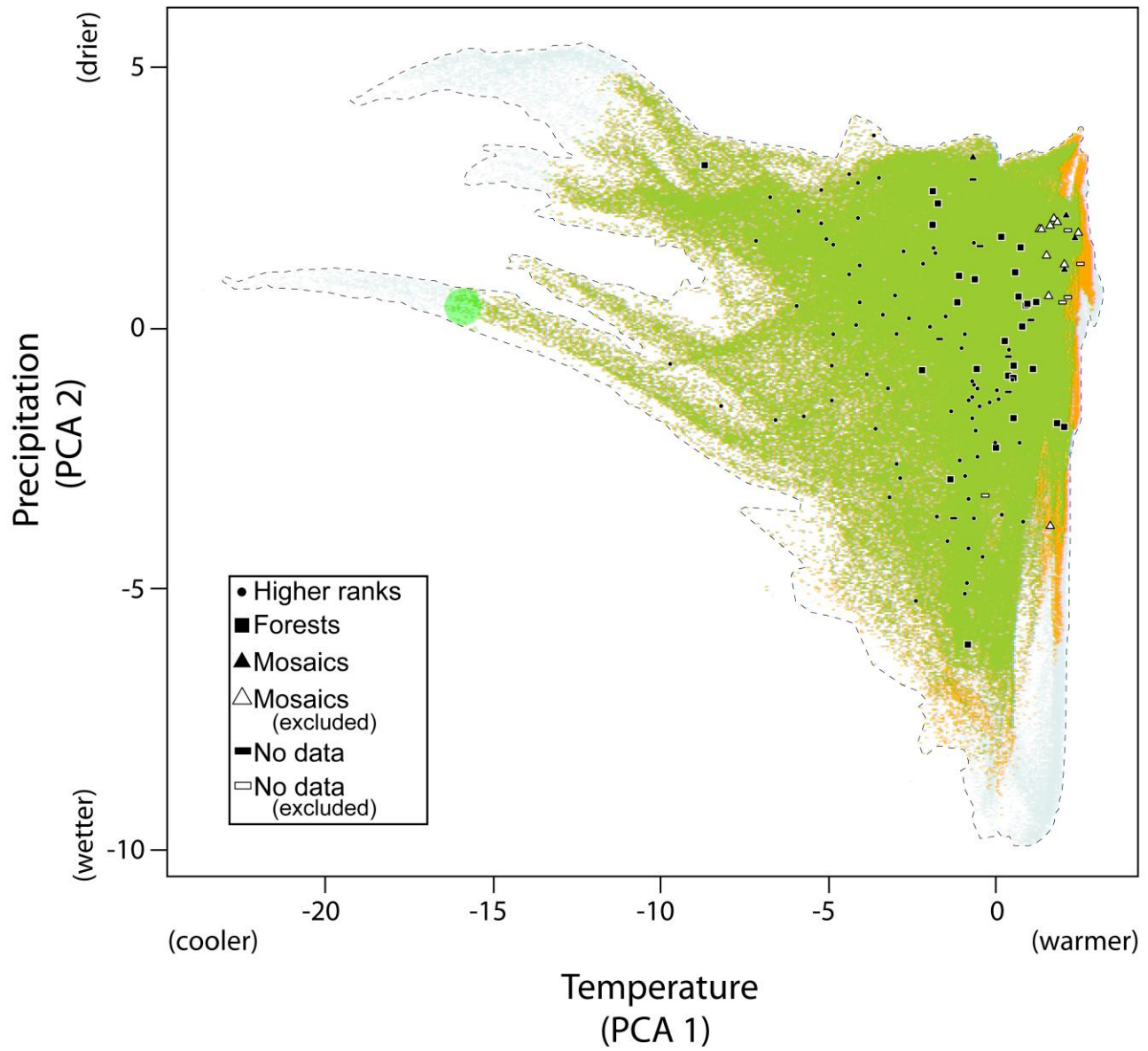


Figure S2 Differences in predicted suitability for *Heteromys anomalus* across north-western South America according to the MAXENT models calibrated with and without the lowest-ranking records representing protruding spatially marginal (PSM) localities. Values were obtained by subtracting the continuous prediction of the latter model from that of the former one on a cell-by-cell basis. Differences range between 0 and 0.43 out of a maximum suitability of 1.0 in each cell. Colours show magnitude and direction of the differences. In blue: cells assigned a higher suitability by the model made excluding PSM localities (differences of 0–0.073). In red: cells assigned a higher suitability by the model made with all records; pale red indicates differences of 0–0.25, dark red denotes differences of 0.25–0.43. Legend indicates the different subsets of records, and the habitat present at those examined in the present study.

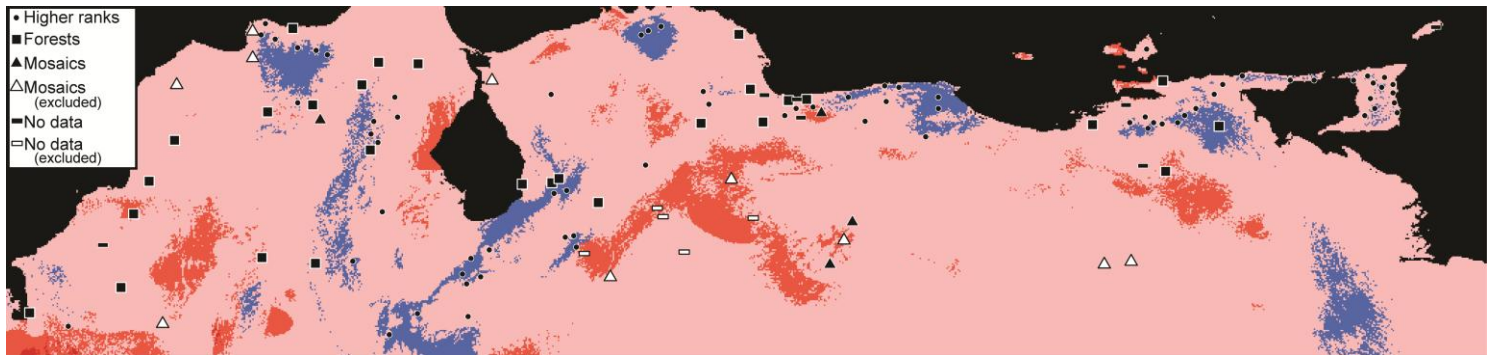


Figure S3 Histogram of the fraction of the study region predicted as suitable at the lowest presence threshold by the 100 models controlling for sample size (hollow columns). Overlaid on the histogram are the corresponding values for the model built with all records (black), and that built excluding protruding spatially marginal (PSM) localities (grey). Control models were calibrated excluding the same number of records as in the model built without PSM localities (15), but selected in a random manner each time. Note how the extent of the study region considered as suitable by most control models is very similar to that of the model built with all records. In contrast, this area is much smaller for the model built excluding the non-random subset of 15 records representing PSM localities.

