



## Informing conservation units: barriers to dispersal for the yellow anaconda

Francine Kershaw<sup>1</sup>, Tomás Waller<sup>2</sup>, Patricio Micucci<sup>2</sup>, Juan Draque<sup>2</sup>, Mariano Barros<sup>2</sup>, Emilio Buongermini<sup>2</sup>, Richard G. Pearson<sup>3</sup> and Martín Mendez<sup>4,5\*</sup>

<sup>1</sup>Department of Ecology, Evolution and Environmental Biology, Columbia University, 1200 Amsterdam Avenue, New York, NY 10027, USA, <sup>2</sup>Fundación Biodiversidad, Juncal 754 Piso 7 #60, Buenos Aires, C1062ABD Argentina, <sup>3</sup>Center for Biodiversity and Conservation, American Museum of Natural History, 79th Street and Central Park West, New York, NY 10024, USA, <sup>4</sup>Sackler Institute for Comparative Genomics, American Museum of Natural History, 79th Street and Central Park West, New York, NY 10024, USA, <sup>5</sup>Wildlife Conservation Society, 2300 Southern Boulevard, Bronx, New York, NY 10460, USA

\*Correspondence: Martín Mendez, Wildlife Conservation Society, 2300 Southern Boulevard, Bronx, New York 10460, USA. E-mail: mmendez@wcs.org

### ABSTRACT

**Aim** Fine-scale population structure is often unaccounted for in the delineation of conservation units, potentially compromising long-term species persistence. Identifying biogeographic and environmental drivers of population boundaries is therefore of key conservation concern. We aimed to explore barriers to dispersal for the harvested yellow anaconda (*Eunectes notaeus*) using an ecological niche model. Our secondary aim was to test the relative geographic and environmental contributions of a multisource occurrence data set in species range predictions.

**Location** Paraguay River drainage, central South America.

**Methods** We developed an ecological niche model for the yellow anaconda using Maxent and a multisource species occurrence data set. Following nine iterations of model development, nine environmental variables were selected for model inclusion. We used the models to identify potential barriers to dispersal and employed jackknifing to identify the primary environmental variables that best explain barrier presence. We assessed the geographic and environmental overlap of models built with each data subset.

**Results** Characterization of suitable habitat was found to be most powerful in northern Argentina and southern Paraguay. A persistent barrier to dispersal was identified in northern Argentina and corresponded to the presence of dry Cambisol soils. Data subsets were found to contribute different information to the final model in terms of geographic and environmental space.

**Main Conclusions** Ecologically meaningful barriers to dispersal support recent genetic hypotheses of population subdivision. These barriers should be considered when delineating species management units to ensure sustainable harvest levels. Multisource data sets may produce more powerful niche predictions and represent a useful resource for data-poor species. Further, model results should be interpreted alongside complementary analyses for more effective conservation strategies.

### Keywords

Biogeographic barriers, conservation units, ecological niche model, multiple data sources, South America, yellow anaconda.

### INTRODUCTION

Rapid anthropogenic change is driving biodiversity loss at an unprecedented rate, resulting in measurable impacts to ecosystem diversity and function (Butchart *et al.*, 2010). Risk of global and local extinction disproportionately impact species with narrow climatic and ecological requirements, as well as

limited dispersal ability (Williams *et al.*, 2007; Svenning *et al.*, 2009; Carvalho *et al.*, 2010). These factors may also subdivide species into demographically isolated populations with greater sensitivity to demographic fluctuations, including those resulting from anthropogenic disturbance (Awise, 2000; Amato *et al.*, 2009). Loss of evolutionary lineages unique to certain areas may occur if population structure is

not accounted for in the delineation of species management units; in the long-term, this has the potential to compromise the persistence of the species (Amato *et al.*, 2009; Hekkala *et al.*, 2011). Consequently, to ensure the sustainability of any management programme, it is key to identify and preserve distinct populations and suitable habitat (Noss *et al.*, 1997). It then becomes essential to characterize environmental agents with the potential to promote and constrain dispersal, and which therefore drive population structure (Amato *et al.*, 2009).

The yellow anaconda (*Eunectes notaeus*) is one of four extant species of anaconda in South America and is the largest snake species inhabiting Argentina (Waller *et al.*, 2007). In general, reptiles are under particular threat due to their high physiological sensitivity to ambient temperature (Bennet, 1990; Mitchell & Janzen, 2010; McConnachie *et al.*, 2011), low dispersal ability and obligate adaptation of some species to freshwater habitat (McMenamin *et al.*, 2008; Pauwels *et al.*, 2008). Representative of this group, *Eunectes* spp. are highly dependent upon inland freshwater habitats (Pizzatto *et al.*, 2007), and the known range of *E. notaeus* primarily encompasses the Paraguay River drainage, from the Pantanal Region in Bolivia, Paraguay and Brazil, to north-eastern Argentina where it reaches its southernmost distribution (Waller *et al.*, 2007).

In addition, reptiles are known to suffer intensive extraction for a variety of uses, chiefly the food and skin trades (Cox *et al.*, 2006; Pauwels *et al.*, 2008). Trade of *E. notaeus* has been considered among the most extensive of any Neotropical species (Waller & Miccuci, 1993) with snakes sold for their skins, as pets or for traditional medicine (Alves & Filho, 2007). Up to 320,000 units were traded world-wide from 1980 to 1999 and Paraguay and Argentina represented the first and second largest respective suppliers, until most trade was effectively banned in 1999 (Waller *et al.*, 2007). Recent research has indicated that *E. notaeus* exhibits particular ecological attributes, such as high reproductive rate and short generation time, which enabled it to withstand the high and unregulated harvest levels of the past (Waller *et al.*, 2007).

However, genetic evidence of strong population structure within northern Argentina suggests the presence of barriers to dispersal (Mendez *et al.*, 2007; McCartney-Melstad *et al.*, 2012). Specifically, genetic distance measures using three distinct regions of the mitochondrial genome suggest the existence of several demographically distinct populations within the Provinces of Formosa and Corrientes in northern Argentina (Mendez *et al.*, 2007; McCartney-Melstad *et al.*, 2012; Fig. 1). The observed population structure is attributed to the relative autonomy of different wetland systems in this region, absence of suitable habitat throughout the wide latitudinal range of the species and patterns of historic colonization (McCartney-Melstad *et al.*, 2012). It is likely, however, that some dispersal connections have been altered relatively recently due to the dynamic nature of the wetland habitat (Ginzburg *et al.*, 2005; McCartney-Melstad *et al.*, 2012). In

light of the management concerns raised by the evidence of demographically independent populations in the region, improved understanding of the biogeography, distribution, habitat suitability and potential barriers to dispersal, is of key interest to the conservation and management of this species.

Ecological niche models offer valuable insights into the environmental drivers of species biogeography and distribution and can therefore be used to infer potential barriers to dispersal (Walteri & Guralnick, 2008; Burbrink *et al.*, 2012). When combined with other sources of information, niche models represent powerful tools for informing species conservation and management (Raxworthy *et al.*, 2003; Marni *et al.*, 2009; Urbina-Cardona & Flores-Villela, 2010). However, the usefulness of such models is often constrained by a lack of validation by species experts and limited sample sizes (Pearson *et al.*, 2007). Particularly for lesser known species (Guisan *et al.*, 2006), there is a need to maximize sample size by developing integrated data sets comprising information from multiple sources. Likewise, extensive expert input throughout the modelling process is necessary to ensure enhanced biological realism and therefore increased confidence in the interpretation and application of the model (Anderson *et al.*, 2003).

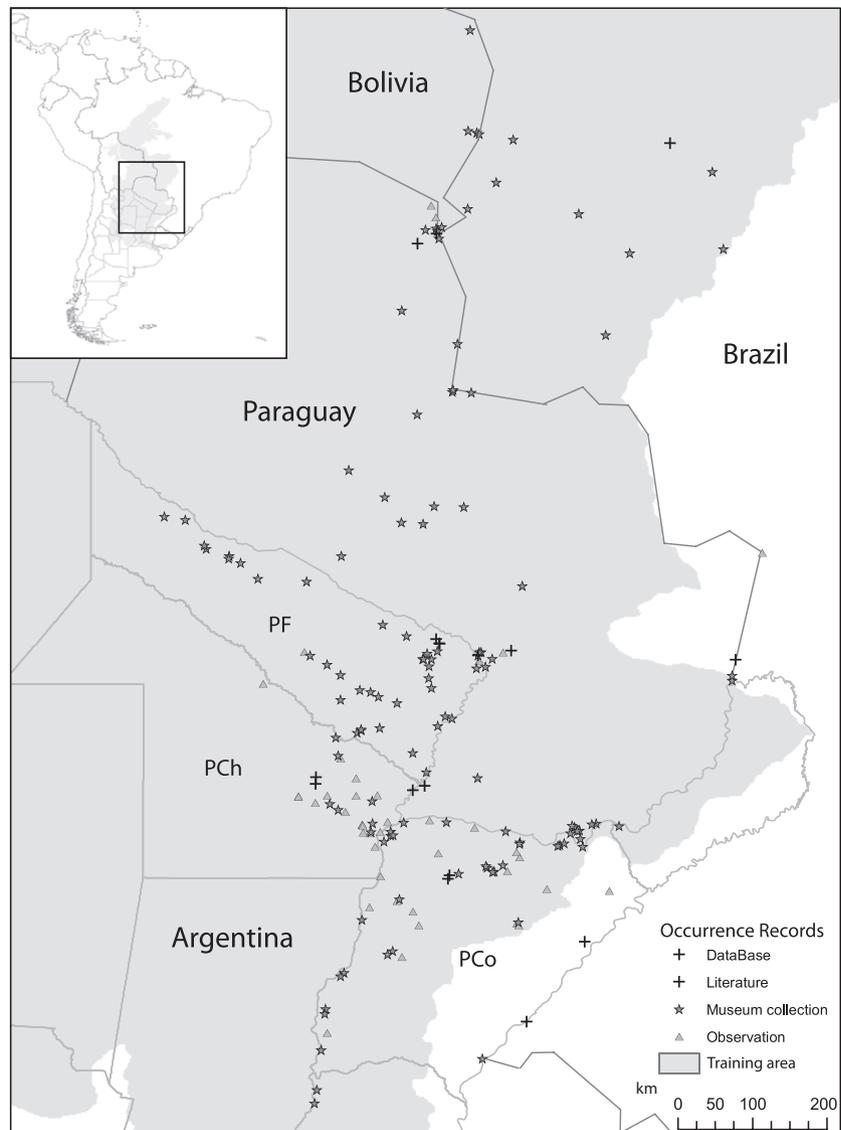
Here, we develop an ecological niche model for the yellow anaconda using an expert-validated, integrated species occurrence data set derived from multiple data sources. We then use the model to characterize the ecological niche of the yellow anaconda and to explore environmental barriers to dispersal, in order to better delineate species management units. Finally, we quantify the geographic and environmental overlap of habitat suitability predictions originating from models built using the different sources of occurrence data to assess the value of using integrated data sets in ecological niche modelling.

## METHODS

### Study region and species occurrence

The yellow anaconda is an aquatic, generalist snake species restricted mainly to river floodplains and wetlands (Henderson *et al.*, 1995). The known range of the species encompasses about 15 degrees in latitude throughout the Paraguay and lower Paraná River drainages from Bolivia and Central Brazil (ca. 15°S) in the north, to north-eastern Argentina (ca. 30°S) in the south (Henderson *et al.*, 1995). This vast region exhibits the largest wetland systems of South America, from the Pantanal area in Brazil and Bolivia, and continuing south through the Wet Chaco ecoregion, in Paraguay and Argentina.

We were able to collate a species occurrence data set comprising 204 records. Of these, 19 records originated from peer-reviewed literature and databases, 57 from recent field observations and 128 from expert-validated museum records. All records were rechecked for taxonomy and georeferencing



**Figure 1** Distribution of species occurrence records for *Eunectes notaeus* ( $n = 204$ ). Provinces of Argentina are indicated: PF, Province of Formosa; PCh, Province of Chaco; PCo, Province of Corrientes. Symbols indicate the original source of the records: 'Database/Literature' ( $n = 19$ ); 'Museum' ( $n = 128$ ); 'Observations' ( $n = 57$ ). Museum and observation records underwent expert-validation prior to being used in this study. Shaded area indicates the 'training area' derived from terrestrial and freshwater ecoregions of the world that was used to define the study extent for the Maxent model.

precision. Any museum records purporting localities counter to the best available knowledge of the distribution of *E. notaeus* were not included. Four extreme locality outliers were removed due to possible errors in the data, and redundancies in the database were eliminated ensuring that only one record was included for each Lat-Lon position (Fig. 1).

### Environmental variables

The selection of environmental variables for the final model was based on expert guidance and followed a three-stage approach. Initially, variables were selected based on their biological relevance to *E. notaeus*. Second, a correlation matrix was constructed for variables comprising continuous data, and when a pair of variables were highly correlated (Pearson correlation coefficient value  $> 0.7$ ), we removed one of the variables from the analysis. Finally, variables were subjected to the jackknife procedure implemented in Maxent (Phillips *et al.*, 2006) for assessing variable importance and

were retained or removed as a result of their relative contribution to the model. During stages two and three, no variables were removed if they were deemed to be of singular biological importance to the species.

Initially, 22 environmental parameters were considered for inclusion in the ecological niche model. Yellow anacondas follow the subtropical and temperate climates accompanying the extensive floodplains of the Paraná and Paraguay Rivers (Giraud & Arzamendia, 2003), possibly favouring the thermal refuge that these large water masses provide during the local winter (Waller *et al.*, 1994). The freshwater ecosystems of the Paraguay River drainage, as most wetlands, result from the interplay between terrain slope, soil type and water balance. However, the hydrological dynamics of the drainage are also tightly influenced by the great latitudinal and longitudinal variation in precipitation and temperature and the characteristic north to south orientation of the principal rivers connecting it. To account for the broad influence of temperature and rainfall, we compiled climate data from the

Worldclim database (Hijmans *et al.*, 2005), comprising a set of 19 global climate coverages with a spatial resolution of one square kilometre. To further characterize wetland habitat, we generated a 'distance to water body' parameter based on coverage of inland water features for South America (DCW, 1992; ISCGM, 2006) and included data on soil type derived from the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012) and freshwater ecoregions (TNC, 2009).

Finally, a subsection of the overall study area was selected to train the model. This 'training area' (Fig. 1) represents the geographic area from which background pixels are chosen at random for calibrating the Maxent model (Philips *et al.*, 2006). These presence and pseudoabsence data are also used to establish measures of model performance (Philips *et al.*, 2006). We defined the training area through expert selection of both terrestrial and freshwater ecoregions (TNC, 2009) deemed to encompass the area accessible to *E. notaeus* (Anderson & Raza, 2010).

### Ecological niche model

A maximum-entropy approach for ecological niche modelling was undertaken using Maxent version 3.3.3e (Philips *et al.*, 2006; Philips & Dudik, 2008). Maxent is a machine learning method that estimates a species' ecological niche by finding the probability distribution of maximum entropy (i.e. that closest to uniform), subject to constraints representing incomplete information about the distribution (Philips *et al.*, 2006). The model evaluates the habitat suitability of each grid cell as a function of the environmental variables within that cell. Recent comparative analysis of ecological niche models has shown Maxent to be among the best performing methods, demonstrating good performance with presence-only data over a range of data set sizes (Elith *et al.*, 2006; Pearson *et al.*, 2007; Wisz *et al.*, 2008).

A multistage modelling approach was employed in order to ensure the highest possible confidence in the final model given the data available. Each round of model outputs was provided to an expert in anaconda natural history and biogeography (T. Waller) for an accuracy assessment and for recommendations on the next set of variables to be tested. For all model runs, the logistic output of Maxent was used, with predictions ranging from 0 (unsuitable habitat, low probability of occurrence) to 1 (suitable habitat, high probability of occurrence). Duplicate records in the same 1-km grid cell were removed meaning that 155 of the original 204 presence records were included in the final model (Database/Literature = 12; Observations = 43; Museum = 100). There were no instances of occurrence records originating from different data sources co-occurring in the same 1-km grid cell. Default linear, quadratic, product, threshold and hinge features were selected with a default convergence threshold ( $10^{-5}$ ) and a maximum number of 500 iterations under five-fold cross-validation (20% records withheld as test data, 80% records used as training data) across five replicates. We used

AUC (the area under the receiver operating characteristic curve; Manel *et al.*, 2001) scores as a threshold independent measure for model validation.

The model's regularization parameter penalizes the coefficients to values that balance model fit and model complexity. Varying this parameter therefore enables an exploration of the trade-off between these measures. A higher regularization value results in a model less dependent on the empirical means of the sampling localities and which is therefore more suited to extrapolation. However, such 'relaxed' models may not adequately fit the data. Conversely, a lower regularization value results in a model highly dependent on the conditions at sampling localities and therefore causes the model to be more vulnerable to over-prediction due to high model complexity. As the regularization parameter can strongly influence the final model output, we used AUC scores to assess model performance across a range of model fit parameters: regularization was equal to 1 in initial model iterations, and the final variable set was run using five regularization values (0.5, 1, 2, 5, 10). The AUC score was highest for the final model with a regularization parameter value of 0.5. Therefore, this regularization value was used in all subsequent models.

### Identifying barriers to dispersal

The Maxent probability surface was reclassified to binary presence/absence grids to more clearly visualize unsuitable habitat representing potential barriers to dispersal. Because there are several alternative thresholds that may be selected for reclassifying continuous data to binary predictions (Freeman & Moisen, 2008), we calculated multiple thresholds for the five replicates: (i) minimum training presence (MTP), for which the suitability associated with the least suitable training presence record is used as the threshold. MTP uses all training points and so is preferred if data quality is high; (ii) 10th percentile training presence (TPTP), for which the suitability threshold associated with the presence record that occurs at the 10th percentile of presence records is used as the threshold. TPTP allows 10% of presence records to be omitted and therefore can be useful if errors in the data set are suspected; (iii) maximum training sensitivity plus specificity (MXTR), describes the threshold that maximizes the sum of sensitivity (rate of false negatives) and specificity (rate of false positives) and therefore minimizes the mean of the error rate for positive and negative observations; (iv) equal training sensitivity and specificity (EQTR); and (v) equal test sensitivity and specificity (EQTE), which describes the threshold at which false negatives are equally as likely as false positives in the training and test data sets, respectively.

To further investigate the effect of the threshold value on barrier presence, a sensitivity analysis was conducted by reclassifying the final model output using threshold values at 0.02 intervals. The interval size was selected as appropriate following preliminary sensitivity tests. Subsequently, jackknifing of environmental variables, as implemented in the Maxent software, was undertaken. These 'leave-one-out models'

were run under final model settings and across the multiple thresholds described above (with the exception of MTP, for which the predicted distribution was too broad to adequately characterize fine-scale biogeographic barriers) to explore the contribution of each environmental variable to the presence of the observed barrier.

### Contribution of occurrence records

The contribution of each data type to the model was assessed by comparing the overlap of the geographic and environmental space sampled by each of the data subsets. The spatially unique occurrence records (i.e. records that fell in distinct grid cells at our resolution of analysis) that were used in the final model were partitioned into three subsets: 'Database/Literature' ( $n = 12$ ), 'Observations' ( $n = 43$ ) and 'Museum' ( $n = 100$ ). Subsampling without replacement was used to standardize sample size across data subsets for comparison. Four sets of 12 random samples were generated for the 'Observations' data, and eight sets of 12 random samples were generated for the 'Museum' data. These randomly selected sets were then run once under final model conditions and reclassified to binary grids of suitable/unsuitable habitat using the tenth percentile training presence threshold (TPTP). Although application of multiple thresholds was necessary for identifying barriers to dispersal (previous section), a single threshold was sufficient here to test the contribution of different data types.

To assess the amount of geographic overlap between models built using alternative data sets, pairwise measurements of the percentage overlap of the habitat classed as suitable between each data subset were conducted. Results were then averaged across all random subsets to summarize the degree of similarity between the models built using the different data types. Overlap in the environmental space sampled by each data type was assessed for the two environmental variables that were found to contribute most to the final model: precipitation seasonality and soil type. Values for each of these variables were extracted to the original spatially unique occurrence records of each data type used in the final model (i.e. Database/Literature = 12; Observations = 43; Museum = 100). 1000 random localities were sampled within the range of habitat classed as 'suitable' by the Maxent model, and values for precipitation seasonality and soil type were extracted at these random localities. Differences in the mean and range of precipitation seasonality for each data subset were explored using boxplots and Student's  $t$ -tests. The range of soil types sampled by each data subset was explored using histograms. All statistical analyses were implemented using the R statistical package.

## RESULTS

### Niche characterization

Nine of the original 22 environmental variables were included in the final model (see Table S1 in Appendix).

Cross-validation suggested that the Maxent model is relatively robust as the AUC calculated for the test data was high (mean: 0.926; range: 0.893–0.954) and similar to that of the training data (mean: 0.962; range: 0.958–0.967). While all variables were deemed relevant for defining the ecological niche of *E. notaeus*, the model summary statistics comprising the average values over the five replicates (Table S1 in Appendix) demonstrate that the three most influential variables are precipitation seasonality (31.1% contribution to the model), soil type (13.8%) and mean temperature of the coldest quarter (12.5%).

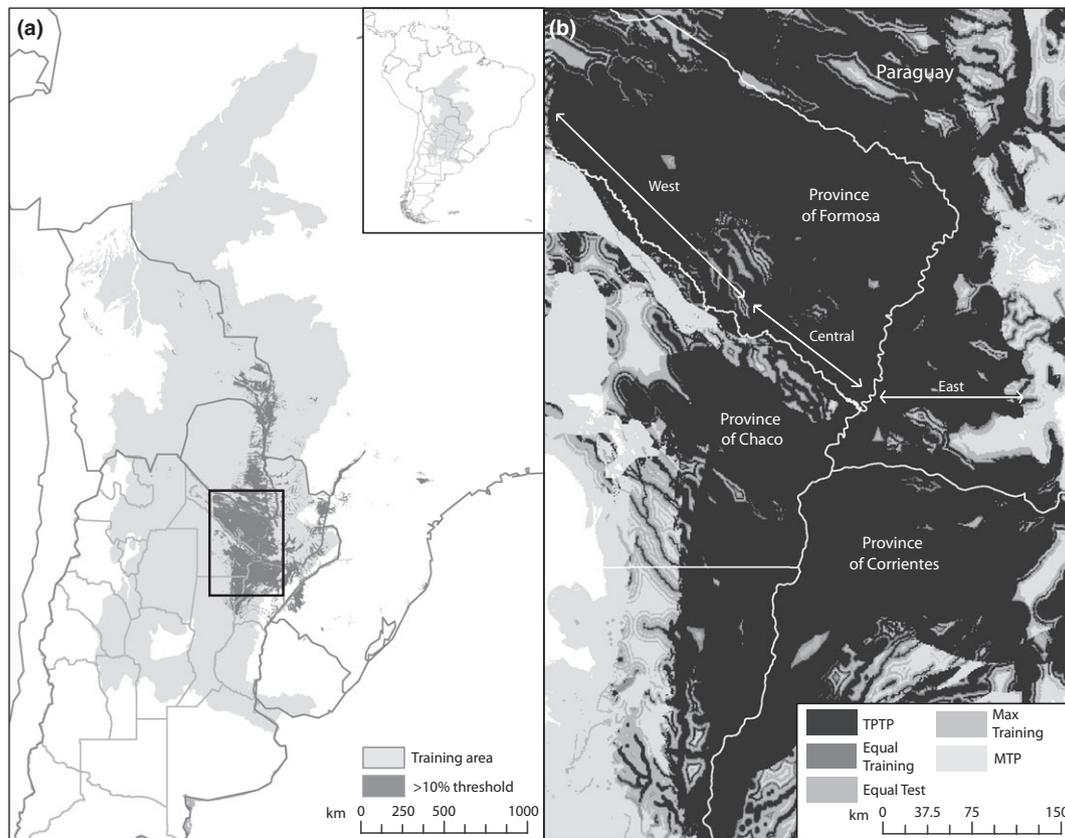
The habitat suitability surface indicates a number of highly suitable areas for *E. notaeus*, particularly in the north-easternmost region of Argentina, within the vicinity of the Paraguay River in central Paraguay, and northwards towards the Pantanal region of Brazil (Fig. 2a; Fig. S1 in Appendix). The ecological niche in the northernmost regions is not as well characterized by the model (Fig. 2a). For example, yellow anaconda is a common species in most of the Pantanal region of Brazil (Strüssmann & Sazima, 1993; Strüssmann, 1997), yet suitable habitat was only identified in the westernmost part of this region. This is most likely due to a relative lack of occurrence records in northern parts of the species' known range.

### Barriers to dispersal

Given that the model performs less well in the northern portion of the species range, we focus on exploring potential dispersal barriers in northern Argentina (Fig. 2b). Within Argentina, areas of low suitability identified by the model correspond to habitat where *E. notaeus* is known to be absent (Waller, T., unpublished data). Notably, the boundary representing the known southern extent of the species range between a high suitability area in central areas of the Formosa Province and a low suitability area directly to the south emulates the transition from wetlands to the extensive dry savannas and forested uplands found in the Province of Chaco south of Formosa (Fig. S1 in Appendix). The model is therefore showing good ability to identify barriers to dispersal within this limited region.

We identified a new potential barrier to dispersal within Argentina that was persistent across reclassification thresholds (Fig. 2b; Table S2 in Appendix). The barrier represents a linear feature that runs parallel to the Bermejo River and continues to the mid-Paraná River, interrupted only by the descending Paraguay–Paraná axis. The barrier therefore separates the highly suitable habitat found in the eastern region of Formosa Province from Chaco Province to the south, as well as south-eastern Paraguay from Corrientes Province in Argentina. Sensitivity analysis showed the barrier to be resolved from west to east as threshold value increases, and is present in some form from a threshold value of 0.06 and is fully present at 0.24 (Fig. 2b; Table S2 in Appendix).

Jackknife tests revealed that a single environmental variable, soil type, was responsible for the presence of the barrier



**Figure 2** (a) Distribution of suitable habitat defined by the 10th percentile training presence (TPTP) threshold superimposed on the model training area. Suitable habitat is distributed primarily in northern Argentina and Paraguay where the majority of species occurrence records were concentrated. (b) Effect of threshold value on presence of ecological barrier. Average Maxent thresholds for the final model are shown: TPTP, 10th percentile training presence; Equal training sensitivity and specificity; Maximum (Max) training and sensitivity plus specificity; Equal test sensitivity and specificity; MTP, Minimum training presence. Areas (West, Central, East) used to qualitatively assess the presence of the barrier for the sensitivity analysis (Table S2 in Appendix) are indicated by the arrows.

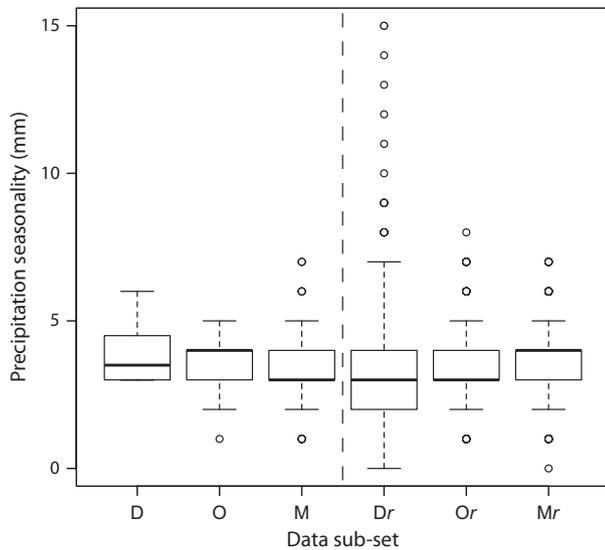
(Table S3 in Appendix). The soil type was identified as the FAO soil group Cambisols, which are characterized by a texture of sandy loam and lack a layer of accumulated clay, humus, soluble salts or iron and aluminium oxides (Driessen *et al.*, 2001). This contrasts with the soil types deemed suitable habitat by the model, such as Fluvisols, which are found typically in river floodplains and deltas consistent with the freshwater habitat preferences of *E. notaeus*. As such, the presence of dry and sandy Cambisols soil appears to represent an ecologically meaningful barrier to dispersal for the species in northern Argentina.

### Comparison of data partitions

Geographic overlap of suitable habitat defined by the TPTP threshold was generally low and varied between the models built with each of the data subsets: 'Database/Literature' versus 'Observations' = 20.05% (range = 15.91–24.99, four samples); 'Database/Literature' versus 'Museum' = 16.76% (range = 3.33–22.29, eight samples); and 'Observations' versus 'Museum' = 23.35% (range = 3.27–39.49, 32 samples).

The environmental space sampled by each of the data subsets also differed. No significant differences were found between the mean precipitation seasonality values sampled by each of the subsets of raw occurrence data when pairwise comparisons were made using Student's *t*-tests ('Database/Literature': mean = 3.92, range = 3.00–6.00; 'Observations': mean = 3.56, range = 1.00–5.00; 'Museum': mean = 3.47, range = 1.00–7.00;  $P > 0.05$  for all pairwise Student *t*-tests; see Fig. 3). In contrast, for values sampled using 1000 random points within the area classified as 'suitable' habitat, highly significant differences were observed in mean precipitation seasonality sampled by each data subset ('Database/Literature': mean = 3.06, range = 0.00–15.00; 'Observations': mean = 3.46, range = 1.00–8.00; 'Museum': mean = 3.60, range = 0.00–7.00;  $P < 0.001$  for all pairwise Student's *t*-tests; see Fig. 3).

Differences in the sampling of soil type were also observed (Fig. 4). For the raw occurrence records, the numbers of soil types sampled by each data subset were: 'Database/Literature' = 6; 'Observations' = 10; and 'Museum' = 11. In pairwise comparisons, 'Database/Literature' and 'Observations'



**Figure 3** Boxplot showing the mean and range of precipitation seasonality sampled by each of the data subsets. Data sets are denoted as follows: D, 'Database/Literature' ( $n = 12$ ); O, 'Observations' ( $n = 43$ ); M, 'Museum' ( $n = 100$ ). Results from the analysis of the raw occurrence data included in the Maxent model are presented to the left of the dashed line. To the right of the dashed line,  $r$  indicates results from random locations ( $n = 1000$  for all data subsets) sampled within the area defined as 'suitable' habitat by the tenth percentile training presence (TPTP) threshold.

shared four soil types; 'Database/Literature' and 'Museum' shared six soil types; and 'Observations' and 'Museum' shared eight soil types (Fig. 4a–c). Of these comparisons, the 'Museum' subset included all soil types sampled by the 'Database/Literature' and 'Observations' subsets, and one additional soil type. The number of soil types sampled was increased when 1,000 randomly generated locations were used: 'Database/Literature' = 24; 'Observations' = 17; and 'Museum' = 17. There was also more redundancy in pairwise comparisons of the soil type sampled, with all data subsets sharing the same seventeen soil types (Fig. 4d–f). The 'Database/Literature' subset sampled an additional seven soil types.

## DISCUSSION

The use of an integrated species occurrence data set employed within an ecological niche modelling framework and validated by expert opinion has enabled the characterization of suitable habitat for yellow anaconda in northern Argentina and the Paraguay River drainage, and the identification of a meaningful ecological barrier to dispersal for this species. These findings are concordant with published genetic evidence of effective migration and population structure of yellow anacondas in northern Argentina (Mendez *et al.*, 2007; McCartney-Melstad *et al.*, 2012). Insights into the environmental factors potentially driving population subdivision in this species are essential for delineating appropriate

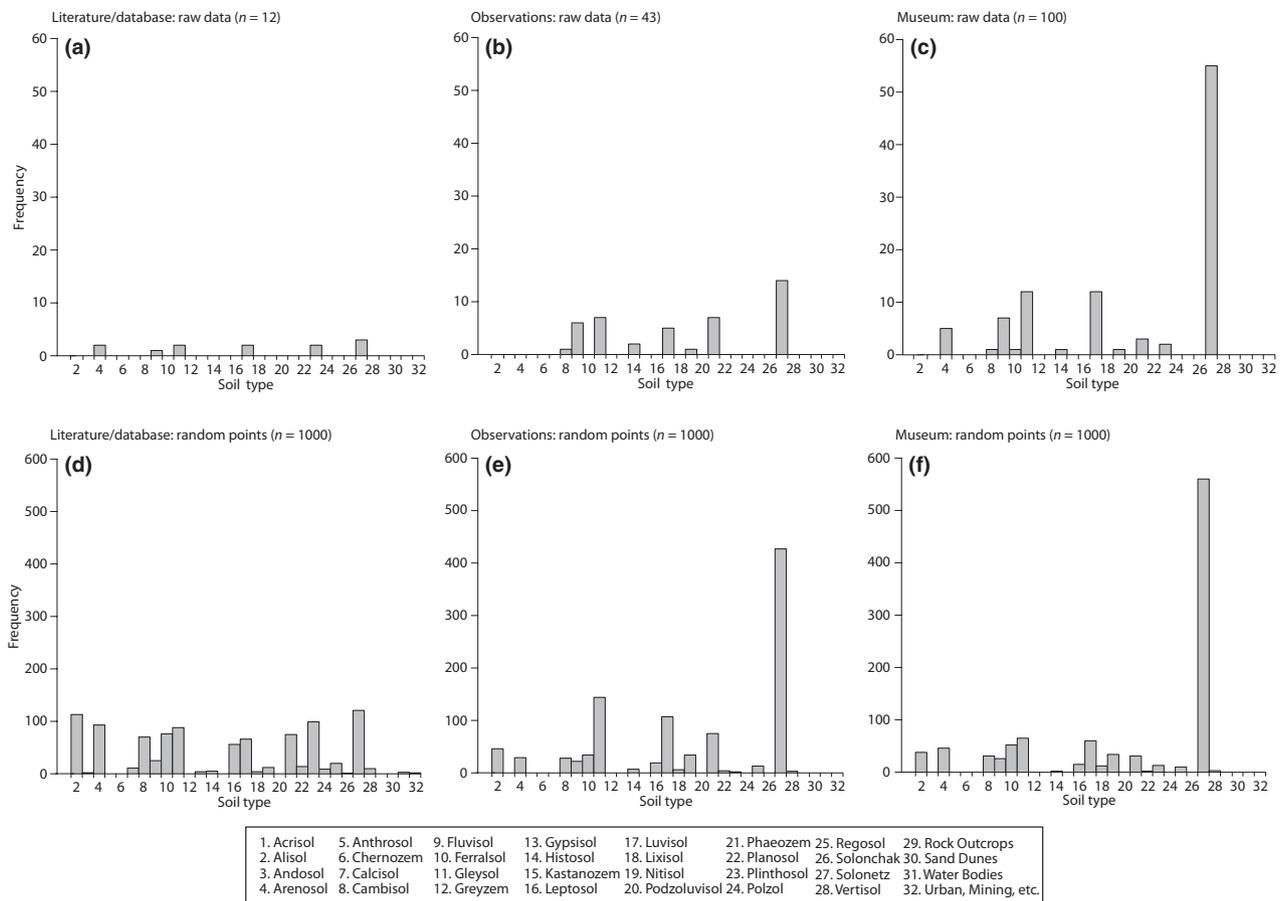
management units and have direct relevance to the sustainable harvest programme currently being implemented in northern Argentina (Micucci & Waller, 2007).

## Environmental drivers of population subdivisions

Our finding that precipitation seasonality was the most influential model variable indicates its importance in driving the hydrology of the study region. Soil type was the second most important variable and appears to represent a suitable proxy for the presence of freshwater habitat (e.g. low porosity, clay soils) or its absence (e.g. high porosity, sandy soils). The high contribution of soil type indicates the importance of selecting environmental variables based on their relevance to the specific study species and considering the inclusion of data sources outside of the 19 bioclimatic variables provided by WorldClim. Mean temperature of the coldest quarter was the third most influential variable in our model and suggests the anaconda's physiological intolerance to low temperature is a key determinant of the distribution.

Within northern Argentina, potential environmental barriers to dispersal were evident. Of particular interest is the linear barrier running west to east between Formosa Province and Chaco Province in north-west Argentina, and between Corrientes Province in north-east Argentina and Paraguay (Fig. 2b). Our finding that this barrier corresponds to the presence of dry, sandy Cambisols soil (Table S3 in Appendix) is consistent with our hypothesis that habitats which impede or prevent dispersal are likely to correspond with dry and upland environments given the species' reliance on freshwater. The barrier is therefore likely to be ecologically meaningful as the preferred wetland habitat of *E. notaeus* is absent in these areas and therefore is likely to be of utility in the delineation of a species management unit in this region.

Consistent with the barrier described here, population genetic studies have elucidated significant genetic divergence between yellow anaconda populations in Formosa relative to those in Corrientes (Mendez *et al.*, 2007; McCartney-Melstad *et al.*, 2012). This would be expected if levels of gene flow between the two areas were limited in some way, and the presence of the barrier provides a potential mechanism for explaining the observed population structure. While the presence of such dry areas is likely to impede dispersal for freshwater obligate species, the ability of yellow anaconda to disperse along rivers both up and downstream (Waller, 1986; Waller & Micucci, 1993) may mean that this barrier is not absolute, with snakes bypassing dry areas by dispersing along the Paraguay–Paraná River axis. Indeed, recent research has shown that gene flow between anaconda populations positively correlates with distance along the rivers connecting them, rather than Euclidean distance, and that migration occurs both up and downstream (McCartney-Melstad *et al.*, 2012). Yet the presence of dry areas is still likely to limit dispersal between the two regions and may reduce the number of successful migrants to an extent capable of driving genetic differentiation.



**Figure 4** Histograms showing the different soil types sampled by each subset of raw occurrence data used in the final Maxent model (top row: a–c) and from random locations ( $n = 1000$  for all data types) sampled within the area defined as ‘suitable’ habitat by the tenth percentile training presence (TPTP) threshold (bottom row: d–f). The key for the numbered soil types is positioned below the histograms (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012).

As the riverine system of the Paraguay–Paraná River drainage is extremely dynamic both spatially and temporally, with past evidence of dramatic landscape reconfiguration (Castellanos, 1965; Giraud & Arzamendia, 2003) that involved several changes in the river courses and even reconnections with other major rivers in the region (i.e. the Amazon and Uruguay Rivers), populations of *E. notaeus* have likely experienced relatively recent fragmentation due to shifts in habitat that either promote or constrain dispersal. For instance, the Andes-originated Pilcomayo River, that has historically affected most of the Formosa Province with its changing floodplain, is thought to have driven the progress of the species to the west, as new habitat forms by progressive flooding of the forests of the Dry Chaco ecoregion (Ginzburg *et al.*, 2005; Waller *et al.*, 2007), and the reactivation of old dry river beds creates new potential migratory connections. Understanding how such habitats have changed through time will assist in the interpretation of biogeographic patterns as well as facilitate the forecasting of future habitat suitability, enabling proactive and adaptive management of the species (Jetz *et al.*, 2007; Fouquet *et al.*, 2010). The ability to assess both persistence and variance in habitat suitability under sub-

optimal conditions is particularly relevant in light of the impacts of climate change on wetlands in some parts of the world (McMenamin *et al.*, 2008). Further work is needed to elucidate how the riverine system interacts with the surrounding landscape to influence habitat suitability and population structure of the yellow anaconda in the region.

### Support for integrated data sets

Our findings support the hypothesis that an integrated data set comprising expert-validated museum records, field observations and records extracted from the peer-reviewed literature provides the best predictive ability. Each type of data was found to sample different geographic and environmental spaces, and so contribute different information to the predictions of the range of *E. notaeus*.

In terms of geographic space, we found low and almost equivalent overlap (~20%) of ‘suitable’ habitat for each of the data types. This suggests that the inclusion of observational and museum data contributes substantial additional information to the model. As the sample size and distribution of each data type differed across the study region (Fig. 1), it is to be

expected that they would capture different information that would influence the final model outcome. For example, the 'Museum' occurrence records are not only the greatest in number ( $n = 100$ ), but also sample a broader latitudinal gradient, with a greater number of locations present in Paraguay and Brazil than the other two data types (Fig. 1).

Findings were more mixed when environmental space was explored. For precipitation seasonality (Fig. 3), no significant differences in the mean and range of environmental space sampled were observed when the raw occurrence data were considered. In contrast, strong significant differences were found between the 1000 data points randomly sampled from the area classified as 'suitable' habitat ( $P$ -values of all pairwise  $t$ -tests  $< 0.001$ ). The boxplot (Fig. 3) suggests that it may be the much larger range of values that are sampled, rather than variation in the mean, that is driving these statistical differences. Similarly, the raw and random data points of each of the data subsets sample different soil types (Fig. 4). For soil type, there is a greater degree of redundancy among data subsets (e.g. complete overlap in soil types sampled between the 'Observations' and 'Museum' subsets). This finding is likely in part due to soil type representing a categorical variable rather than the continuous variable, meaning large geographic areas are attributed a single value. Overall, the 'Database/Literature' data subset contributes most additional environmental information, which may be a function of this model being the least constrained due to small sample size ( $n = 12$ ).

In summary, these findings indicate the value of integrating multiple data types when modelling a species' ecological niche. We therefore caution against assumptions that field observations and museum records should be dismissed due to fears of low accuracy (Tyre *et al.*, 2003; Newbold, 2010), providing they undergo a prior procedure of expert validation, such as that demonstrated by our study.

### Management implications

The identification of population subdivisions, and the ecological and environmental agents driving such barriers to dispersal, is crucial for the accurate delineation of management units (Avice, 2000). Because demographically isolated populations are less resilient to anthropogenic pressures, such as those imposed by harvesting, it is essential that management approaches take population discreteness and environmental barriers to dispersal into account (Avice, 2000; Amato *et al.*, 2009). This will guard against evolutionary unique populations being lost and the genetic diversity of the species being permanently degraded (Mendez *et al.*, 2007; Amato *et al.*, 2009; Hekkala *et al.*, 2011). We therefore recommend that the findings of this study be used to inform the ongoing development and future decisions of the yellow anaconda management programme in northern Argentina (Micucci & Waller, 2007). Additional collection and compilation of georeferenced *E. notaeus* occurrence data throughout the Pantanal in Brazil, in Bolivia, and throughout the Paraguay River drainage, will be important to improve future model predictions in those regions.

More generally, we support recommendations that ecological niche models are employed as a support tool alongside other approaches when developing species conservation and management plans (Peterson & Robins, 2003; Chefaoui *et al.*, 2005; Papeş & Gaubert, 2007). As demonstrated by our study, integrating different and complementary types of data and cross-referencing analytical findings will lead to better informed and therefore more effective conservation strategies.

### ACKNOWLEDGEMENTS

We acknowledge the Ministerio de la Producción y Ambiente of the Province of Formosa, the Dirección de Recursos Naturales of the Province of Corrientes and the Fundación Biodiversidad – Argentina for providing institutional support to our work. The Office vétérinaire federal of Switzerland and the Yellow Anaconda Management Program provided financial assistance to this research. E. McCartney-Melstad provided useful personal communications regarding *E. notaeus* biology and genetics, and commented on the draft manuscript. E. Sterling and M. Blair provided guidance on developing the models and commented on early drafts of the manuscript.

### REFERENCES

- Alves, R. & Filho, G. (2007) Commercialization and use of snakes in North and Northeastern Brazil: implications for conservation and management. *Biodiversity & Conservation*, **16**, 969–985.
- Amato, G., DeSalle, R., Ryder, O.A. & Rosenbaum, H.C. (2009) *Conservation genetics in the age of genomics*. Columbia University Press, New York.
- 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., Lew, D. & Peterson, A.T. (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, **162**, 211–232.
- Avice, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, Massachusetts.
- Bennet, A.F. (1990) Thermal-dependence of locomotor capacity. *American Journal of Physiology*, **259**, 253–258.
- Burbrink, F.T., Ruane, S. & Pyron, R.A. (2012) When are adaptive radiations replicated in areas? Ecological opportunity and unexceptional diversification in West Indian dipsadine snakes (Colubridae: Alsophiini). *Journal of Biogeography*, **39**, 465–475.
- Butchart, S.H.M., Walpole, M., Collen, B. *et al.* (2010) Global biodiversity: indicators of recent declines. *Science*, **328**, 1164–1168.
- Carvalho, S.B., Brito, J.C., Crespo, E.J. & Possingham, H. (2010) From climate change predictions to actions – conserving vulnerable animal groups in hotspots at a regional scale. *Global Change Biology*, **16**, 3257–3270.

- Castellanos, A. (1965) *Estudio fisiográfico de la provincia de Corrientes*. Instituto de Fisiografía y Geología (Universidad Nacional del Litoral). *Publicaciones*, **49**, 222.
- Chefaoui, R.M., Hortal, J. & Lobo, J.M. (2005) Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian Copris species. *Biological Conservation*, **122**, 327–338.
- Cox, N., Chanson, J. & Stuart, S. (2006) *The status and distribution of reptiles and amphibians of Mediterranean Basin*. IUCN, Gland, Switzerland.
- DCW (1992) Digital Chart of the World. Environmental Systems Research Institute, Inc. (ESRI). Available at: <http://princeton.edu/~geolib/gis/dcw.html> (accessed 27 March 2011).
- Driessen, P., Deckers, J. & Nachtergaele, F. (eds) (2001) *Lecture notes on the major soils of the world. World soil resources reports*. FAO, Rome, Italy.
- 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.
- FAO/IIASA/ISRIC/ISSCAS/JRC (2012) *Harmonized World Soil Database (version 1.2)*. FAO, Rome, Italy and IIASA, Laxenburg, Austria.
- Fouquet, A., Ficetola, G.F., Haigh, A. & Gemmel, N. (2010) Using ecological niche modelling to infer past, present and future environmental suitability for *Leiopelma hochstetteri*, an endangered New Zealand native frog. *Biological Conservation*, **143**, 1375–1384.
- Freeman, E.A. & Moisen, G.G. (2008) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, **217**, 48–58.
- Ginzburg, R., Adámoli, J., Herrera, P. & Torrella, S. (2005) Los humedales del Chaco: clasificación, inventario y mapeo a escala regional. *INSUGEO, Miscelánea*, **14**, 121–138, Tucumán.
- Giraud, A.R. & Arzamendia, V. (2003) ¿Son las planicies fluviales de la Cuenca del Plata corredores de biodiversidad? Los vertebrados amniotas como ejemplo. *Humedales de Iberoamérica*, 380 pp (ed. by J.J. Neiff), pp. 157–170. Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo, Subprograma XVIII, Red Iberoamericana de Humedales (RIHU).
- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A. & Zimmerman, N.E. (2006) Using niche-based models to improve the sampling of rare species. *Conservation Biology*, **20**, 501–511.
- Hekkala, E., Shirley, M.H., Amato, G., Austin, J.D., Charter, S., Thorbjarnarson, J., Vliet, K.A., Houck, M.L., DeSalle, R. & Blum, M.J. (2011) Splitting an ancient icon: mummy DNA resurrects a cryptic Nile crocodile. *Molecular Ecology*, **20**, 4199–4215.
- Henderson, R.W., Waller, T., Micucci, P.A., Puorto, G. & Burgeois, R.W. (1995) Ecological correlates and patterns in the distribution of Neotropical boas (Serpentes: Boidae): a preliminary assessment. *Herpetological Natural History*, **3**, 15–27.
- 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.
- ISCGM (2006) Instituto Geográfico Militar/International Steering Committee for Global Mapping. Ciudad Autónoma de Buenos Aires. Available at: <http://gis.columbia.edu/data.html> (accessed 27 March 2011).
- Jetz, W., Wilcove, D.S. & Dobson, A.P. (2007) Projected impacts of climate and land-use change on the global diversity of birds. *PLoS Biology*, **5**, e157.
- Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 291–331.
- Marni, M.A., Barbet-Massin, M., Lopes, L.E. & Jiguet, F. (2009) Predicting climate-driven bird distribution changes and forecasted conservation conflicts in a neotropical savanna. *Conservation Biology*, **23**, 1558–1567.
- McCartney-Melstad, E., Waller, T., Micucci, P.A., Barros, M., Draque, J., Amato, G. & Mendez, M. (2012) Population structure and gene flow of the yellow anaconda (*Eunectes notaeus*) in northern Argentina. *PLoS ONE*, **7**, e37473.
- McConnachie, S., Greene, S.N. & Perrin, M.R. (2011) Thermoregulation in the semi-aquatic yellow anaconda, *Eunectes notaeus*. *Journal of Thermal Biology*, **36**, 71–77.
- McMenamin, S.K., Hadly, E.A. & Wright, C.K. (2008) Climatic change and wetland desiccation cause amphibian decline in Yellowstone National Park. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 16988–16993.
- Mendez, M., Waller, T., Micucci, P.A., Alvarenga, E. & Morales, J.C. (2007) Genetic population structure of the yellow anaconda (*Eunectes notaeus*) in northern Argentina: management implications. *Biology of boas and pythons*. (ed. by R. W. Henderson and R. Powell), pp. 210–221. Eagle Mountain Publishing, LC, Utah.
- Micucci, P.A. & Waller, T. (2007) The management of yellow anacondas (*Eunectes notaeus*) in Argentina: from historical misuse to resource appreciation. *Iguana*, **14**, 161–171.
- Mitchell, N.J. & Janzen, F.J. (2010) Temperature-dependent sex determination and contemporary climate change. *Sexual Development*, **4**, 129–140.
- Newbold, T. (2010) Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, **34**, 3–22.
- Noss, R.F., O'Connell, M. & Murphy, D. (1997) *The science of conservation planning: habitat conservation under the Endangered Species Act*. Island Press, Washington, D.C.
- Papeş, M. & Gaubert, P. (2007) Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. *Diversity and Distributions*, **13**, 890–902.

- Pauwels, O.S.G., Wallach, V. & David, P. (2008) Global diversity of snakes (Serpentes: Reptilia) in freshwater. *Hydrobiologia*, **595**, 599–605.
- 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. & Robins, C.R. (2003) Using ecological-niche modelling to predict Barred Owl invasions with implications for Spotted Owl conservation. *Conservation Biology*, **17**, 1161–1165.
- Philips, S.J. & Dudik, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.
- Philips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Pizzatto, L., Marques, O.A.V. & Martins, M. (2007) Ecomorphology of Boine snakes, with emphasis on South American forms. *Biology of boas and pythons*. (ed. by R. W. Henderson and R. Powell), pp. 35–48. Eagle Mountain Publishing, LC, Utah.
- Raxworthy, C.J., Martinez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A. & Peterson, A.T. (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, **426**, 837–841.
- Strüssmann, C. (1997) Hábitos alimentares da sucuri-amarela, *Eunectes notaeus* Cope, 1862, no Pantanal matogrossense. *Biociencias*, **5**, 35–52.
- Strüssmann, C. & Sazima, I. (1993) The snake assemblage of the Pantanal of Poconé, western Brazil: faunal composition and ecological summary. *Studies on Neotropical Fauna and Environment*, **28**, 157–168.
- Svenning, J.-C., Kerr, J. & Rahbek, C. (2009) Predicting future shifts in species diversity. *Ecography*, **32**, 3–4.
- TNC (2009) *Global ecoregions, major habitat types, biogeographical realms and the nature conservancy terrestrial assessment units as of December 14, 2009*. The Nature Conservancy, Arlington, Virginia.
- Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. & Possingham, H.P. (2003) Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications*, **13**, 1790–1801.
- Urbina-Cardona, J.N. & Flores-Villela, O. (2010) Ecological-niche modelling and prioritization of conservation-area networks for Mexican Herpetofauna. *Conservation Biology*, **24**, 1031–1041.
- Waller, T. (1986) *La boa curiyú*. Fauna Argentina 103, Centro Editor de América Latina (CEAL), Buenos Aires, Argentina.
- Waller, T. & Micucci, P.A. (1993) Estado Actual del Conocimiento de la Biología, Comercio y Conservación de las Boas Neotropicales: Géneros *Boa* L. (1758) y *Eunectes* Wagner (1830) (Reptilia: Serpentes: Boidae). Technical report prepared for TRAFFIC Sudamérica for the CITES Secretariat. Buenos Aires, Argentina.
- Waller, T. & Micucci, P.A. (1994) Situación del Género *Caiman* en Argentina. Apuntes sobre su distribución, situación poblacional y comercio histórico. *Memorias del IV Workshop sobre Conservación y Manejo del Yacaré Overo (Caiman latirostris)*. (ed. by A. Larriera, A. Imhof, M. C. von Finck, A. L. Costa and S. C. Tourn), pp. 40–49. “La Región”, Fundación Banco Bica, Santo Tomé, Santa Fe, Argentina.
- Waller, T., Micucci, P.A. & Alvarenga, E. (2007) Conservation biology of the yellow anaconda (*Eunectes notaeus*) in northeastern Argentina. *Biology of boas and pythons*. (ed. by R.W. Henderson and R. Powell), pp. 340–362, Eagle Mountain Publishing, LC, Utah.
- Walteri, E. & Guralnick, R.P. (2008) Ecological niche modelling of montane mammals in the Great Basin, North America: examining past and present connectivity of species across basins and ranges. *Journal of Biogeography*, **36**, 148–161.
- Williams, J.W., Jackson, S.T. & Kutzbach, J.E. (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5738–5742.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H. & Guisan, A. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.

## SUPPORTING INFORMATION

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

**Appendix S1** Additional model summary statistics and sensitivity analyses.

## BIOSKETCH

**Francine Kershaw** is a PhD student interested in developing methodologies that integrate behavioural, environmental and genetic data to quantify the ecological processes that drive evolutionary patterns. She is also interested in how this information can be applied within the policy arena to develop effective mechanisms for conservation and management.

Author contributions: F.K. and M.M. conceived the study; T.W., P.M., J.D., M.B. and E.B. collected the data; F.K. analysed the data under the guidance of R.P. and T.W.; F.K. led the writing, T.W. contributed the sections on yellow anaconda biology and distribution, and all authors reviewed the manuscript.

---

Editor: Mark Robertson