**The mismatch between location of protected areas and suitable habitat for the Vulnerable taruka *Hippocamelus antisensis***

Cristina Mata, Nicolás Fuentes-Allende, Juan E. Malo, André Vielma and Benito A. González

Supplementary Material 1 Detailed methodology.

Database construction

Taruka occurrences come from earlier studies (Sielfeld & Guzmán, 2011; Fuentes-Allende et al., 2016) and our own data (BAG & NFA, unpubl. data), comprising a total of 155 occurrences during 2004–2015, 90% of which are the outcome of post-2012 intensive monitoring throughout the study area. In addition, 115 absence records resulted from intensive sampling in 2014 covering 525.5 km (484.7 km by vehicle and 40.7 km on foot) searching for tarukas within the study area. All the absence locations correspond to 1-km grid cells where, apart from regular deer sampling along roads and tracks, stops for intensive searches were made. In all cases with both presence and absence records in the same 1-km grid cell, presence took precedent. For all records, both presence and absence, 28 environmental variables were compiled using the Spatial Analyst tool in *ArcGIS* v. 10.1 (ESRI, Redlands, USA). The six topographic variables were derived from the ASTER GDEM (METI & NASA) digital terrain model: altitude, gradient and roughness, each defined by their mean and standard deviation. The original resolution of the digital terrain model was 30 m, but this was reduced so that all variables would have the same spatial resolution of 1 km x 1 km. Thus, each model used the means and standard deviations as summary variables for each of the 1 km2 pixels. A total of 19 climatic variables were derived from the WorldClim database. The normalized difference vegetation index (NDVI) was obtained from MODIS-Terra (MODIS13A3/Terra Vegetation Indices Monthly L3 Global 1 km SIN Grid V005) images. A monthly index was constructed from all images obtained during April 2012. The distance of observations to the nearest ravine was also included, as determined from a ravines map generated by GEP Laboratory, Chile University. The distance of each observation to the nearest settlement was obtained from the official Geospatial Data of Chile (IDE Chile, 2016).

Variable selection for modelling

From a correlation analysis of the 19 climate variables eight were selected after discarding those with high correlations (r > 0.7). To reduce the number of variables further, we used Boosted Regression Trees with the remaining 17 variables (eight of which were the selected climate variables). This model required presence and absence records and we thus used the whole database (155 presence and 115 absence records). The model chosen by means of the fitted regression trees and resampling to predict taruka presence gave a good fit, with 750 trees (AUC = 0.992 ± SE 0.015). Eight of the 17 variables included in the model were chosen for the subsequent construction of the distribution model (Supplementary Table 2.), following Harrell's rule of thumb (Harrell, 1984) to minimize overfitting.

Modelling the distribution of the taruka

We used *MaxEnt v. 3.3.3.k* to model the distribution of the taruka. This method only works with presence records. The final number of presence records used for model construction was reduced to the 76 one-km UTM grid cells with taruka records, to reduce sampling bias and spatial autocorrelation (Wellenreuther et al., 2012). Different models were first compared using *ENMTools*, varying some of the algorithm parameters, including the regularization parameter (Radosavljevic et al., 2014) together with the environmental variable functions (Merow et al., 2013). We compared two different sets of feature classes using the default autofeatures option (which allows all possible features), allowing only linear, quadratic and product (LQP) features. The best fitting model was selected using *ENMTools*, which enables comparisons via Akaike Information Criterion (AIC) models with different parameters (Warren et al., 2010, Warren & Seifert, 2011). The best selected model had a *β* regularization parameter of 1 and LQP features (Supplementary Table 2) and it was constructed via 4-fold cross-validation. Once the model was built we applied two cut-off thresholds to establish the areas in which the species potentially occurs: the maximum value of the sensitivity–specificity sum (*MaxSS*), the point on the curve at which the sum of sensitivity and specificity is maximized (Manel et al., 2001), and the average of the values of all the pixels of the prediction (*Averprob*) which is the mean predicted value for actual occurrence points.

**References**

Harrell, F.E. (1984). Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*, 3, 143–152.

IDE Chile (2016) *Chilean Geospatial Data. Ministerio de Bienes Nacionales*. Http://www.ide.cl [accessed 22 December 2016].

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, 921–931.

Merow, C., Smith, M.J. & Silander, J.A. (2013) A practical guide to MaxEnt for modeling species’ distributions: what it does, and why inputs and settings matter. *Ecography*, 36, 1058–1069.

Radosavljevic, A. & Anderson, R.P. (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, 41, 629–643.

Warren, D.L., Glor, R.E. & Turelli, M. (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography*, 33, 607–611.

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.

Supplementary Table 1 Chilean protected areas located within the study area, and our proposal for a new protected area (for locations see Fig. 1).

|  |  |  |  |
| --- | --- | --- | --- |
| Protected Area | Province | District | Area (km2) |
| 1. Quebrada de Cardones Natural Monument | Arica | Arica-Parinacota | 113.26 |
| 2. Salar de Surire Natural Monument | Parinacota | 113.40 |
| 3. Las Vicuñas National Reserve | 2028.51 |
| 4. Lauca National Park | 1372.47 |
| 5. Pampa del Tamarugal National Reserve\* | Tamarugal | Tarapacá | 996.28 |
| 6. Salar de Huasco National Park | 1109.62 |
| 7. Volcan Isluga National Park | 1698.71 |
| 8. Proposal for a new protected area | Parinacota | Arica-Parinacota | c. 490 |

\*Not considered in any computation because the location is out of the taruka’s altitudinal range.

Supplementary Table 2 Main results of the model selection process carried out with *ENMtools* and based on the Akaike Information Criteria (AIC). Features for each model include the regularization parameters and functions of environmental variables. The selected *Maxent* model is shown in bold.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model features | Log likelihood | Parameters | Sample Size | AIC | AICc |
| **β=1 and linear, quadratic, and product (LQP) features** | **-580.82** | **14** | **76** | **1189.64** | **1196.52** |
| β=0.5 and LQP | -576.38 | 18 | 76 | 1188.76 | 1200.76 |
| β=2 and Auto-features (AUTO)\* | -581.71 | 20 | 76 | 1203.42 | 1218.69 |
| β=3 and AUTO | -592.47 | 15 | 76 | 1214.94 | 1222.94 |
| β=2 and LQP | -597.97 | 15 | 76 | 1225.94 | 1233.94 |
| β=3 and LQP | -612.07 | 12 | 76 | 1248.14 | 1253.09 |
| β=1 and AUTO | -561.93 | 41 | 76 | 1205.86 | 1307.15 |
| β=0.5 and AUTO | -553.69 | 49 | 76 | 1205.39 | 1393.85 |

Supplementary Fig. 1 Presences and absences of taruka *Hippocamelus antisensis* within the study area. Chilean Protected Areas are also represented.



Supplementary Fig. 2 Potential distribution of taruka based on two threshold criteria: *Averprob* (>0.562) taking the average predicted probability of the model-building data as the threshold, and *MaxSS* (>0.105), in which the sum of sensitivity and specificity is maximized as the threshold, and protected areas (Supplementary Table 1).



Supplementary Fig. 3 Response probability of taruka presence according to the variables included in the model (Table 1). To facilitate interpretation, figures correspond to *Maxent* models created using only the corresponding variable. (a) NDVI index, (b) temperature seasonality (ºC \*10; Bio4), (c) precipitation seasonality (Bio15), and (d) distance to settlements. The curves show mean responses of the four *Maxent* replicates (red) ± SD (blue), and the shaded area correspond to the range of values with taruka presences.

