**Supplementary Material**

The effects of spatial survey bias and habitat suitability on predicting the distribution of threatened species living in remote areas

LAURA CARDADOR, JOSÉ A. DÍAZ-LUQUE, FERNANDO HIRALDO, JAMES D. GILARDI and JOSÉ L. TELLA

**Contents**

Appendix S1: Model predictions according to occurrence locations with evidences of reproduction.

Appendix S2: Evaluation of the relationship between distribution of roads and rivers and habitat characteristics in the study area.

Appendix S3: Accessibility to humans at sampled (presences and absences) and random locations.

**Appendix S1**

Model predictions according to occurrence locations with evidences of reproduction

We re-built SDMs to estimate the probability distribution of blue throated macaw using a subset of the original dataset, that is, occurrences where evidences of reproduction were detected. Model predictions were compared to those obtained using the complete dataset using Pearson correlations in ENMTools (Warren et al. 2010).

Average values of AUC and TSS of models using different predictor sets was high (>0.7 and >0.4, respectively, Table A1), indicating overall good ability to predict the distribution of the species in the breeding season. Model predictions (Fig. A1) were highly correlated to those obtained using the dataset with all occurrences. Pearson correlation coefficients were 0.67 for habitat, 0.98 for accessibility and 0.76 for the habitat+accessibility model. Thus, for the purpose of this study we retained Maxent models based on the complete data set.

|  |
| --- |
| **Table S1.** Model performance of Maxent models based on different sets of variables and occurrences with evidence of reproduction using AUC, TSS and test gain (TG) values. Note that for each set of variables, AUC, TSS and Test gain values are averaged values across 10 replicate models calibrated using different randomly selected subsamples of total data (N = 33 records). Model significance was tested using threshold-dependent binomial probability tests, using the 10 percentile training presence (10p TP) and the maximum sensitivity plus specificity values (MSPS) as thresholds. The number of significant replicate models is provided. |
| Variable sets | AUC | TSS | TG | 10p TP | MSPS |
| Habitat | 0.85 | 0.61 | 0.86 | 8 p≤0.01; 2 p>0.05 | all p<0.01 |
| Accessibility | 0.71 | 0.40 | 0.37 | 7 p≤0.01; 3 p>0.05 | 6 p≤0.01; 4 p>0.05 |
| Habitat+Accessibility | 0.85 | 0.50 | 0.87 | all p<0.01 | 8 p<0.01, 2 p≥0.05 |

**Figure S1**. Predicted distributions of the blue-throated macaw in Bolivia. Predicted distributions are based on Maxent models using occurrence data with evidences of reproduction (N=33) and different sets of predictors: habitat, accessibility and habitat+accessibility. Note that models developed for each set of predictors were calibrated using 10 different randomly selected subsamples of total data. Averaged predictions are shown.



**Reference**

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

**Appendix S2**

Evaluation of the relationship between distribution of roads and rivers and habitat characteristics in the study area

We evaluated whether biases in occurrence localities towards more accessible areas could lead to biases in sampled habitats in the study area. With this aim, we compared the frequency distribution of distance to main and secondary roads and rivers across different habitat types at the local and landscape level using Kolmogorov-Smirnov tests. We also analysed the relationship between distance to main and secondary roads and rivers and distance to palms using Spearman rank correlations. Analyses were based on 10,000 randomly selected points across the study area. Analyses were conducted using IBM SPSS Statistics 19.

Differences in the proximity to main and secondary roads and rivers were found across different habitat types at the local scale (i.e, forest *vs*. other habitats). The presence of forest was more frequent at lower distance to rivers and higher distances to main and secondary roads than other habitat types (Figure S2). Differences in the occurrence of different habitat types at the landscape scale with respect to rivers and main and secondary roads were also found (Figure S3 & Table S2). At this scale, the presence of forest was also more frequent at higher distances to main and secondary roads than other habitat types (but see exceptions with respect to shrubland and urban habitats, Table S2). Correlation between accessibility variables and distance to palms was very low (Pearson correlation coefficients ranged -0.187 – 0.108), although significant (Figure S4).

**Figure S2.** A comparison of accessibility characteristics of forest and other habitat types in the study area. Comparisons are based on 10,000 randomly selected points across the study area. Results from Kolmogorov-Smirnov tests are provided.



**Figure S3**. A comparison of accessibility characteristics of dominant habitat types at the landscape level (~1km) in the study area. Comparisons are based on 10,000 randomly selected points across the study area.



**Figure S4**. Relationship between distance to rivers and main and secondary roads and palms in the study area. Results are based on 10,000 randomly selected points across the study area. Pearson correlation test are provided.



|  |
| --- |
| **Table S2.** Results for Kolmogorov-Smirnov tests for comparisons of accessibility characteristics of dominant habitat types at a landscape scale (~1km) in the study area. Comparisons are based on 10,000 randomly selected points across the study area. In bold, significant tests after Bonferroni corrections (i.e. at α < 0.001). |
|   | Forest |   | Grassland |   | Shrubland |   | Wetlands |   | Urban |
|   | *D* | *P* |  | *D* | *P* |  | *D* | *P* |  | *D* | *P* |  | *D* | *P* |
| *Distance to rivers* |  |  |  |  |  |  |  |  |  |  |  |  |
| Cultivated land | 0.179 | 0.542 |  | 0.136 | 0.854 |  | 0.094 | 0.999 |  | 0.202 | 0.399 |  | 0.500 | 0.228 |
| Forest |  |  |  | **0.118** | **< 0.001** |  | 0.120 | 0.251 |  | 0.059 | 0.006 |  | 0.616 | 0.045 |
| Grassland |  |  |  |  |  |  | 0.122 | 0.237 |  | **0.152** | **< 0.001** |  | 0.522 | 0.132 |
| Shrubland |  |  |  |  |  |  |  |  |  | 0.144 | 0.117 |  | 0.534 | 0.138 |
| Wetlands |  |  |  |  |  |  |  |  |  |  |  |  | 0.644 | 0.032 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Distance to main roads* |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Cultivated land | **0.601** | **<0.001** |  | **0.572** | **<0.001** |  | **0.681** | **<0.001** |  | **0.629** | **<0.001** |  | 0.750 | 0.022 |
| Forest |  |  |  | **0.065** | **<0.001** |  | 0.162 | 0.047 |  | **0.107** | **<0.001** |  | **0.985** | **<0.001** |
| Grassland |  |  |  |  |  |  | 0.188 | 0.012 |  | **0.165** | **<0.001** |  | **0.967** | **<0.001** |
| Shrubland |  |  |  |  |  |  |  |  |  | 0.125 | 0.238 |  | **1.000** | **<0.001** |
| Wetlands |  |  |  |  |  |  |  |  |  |  |  |  | **0.994** | **<0.001** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Distance to secondary roads* |  |  |  |  |  |  |  |  |  |  |  |  |
| Cultivated land | 0.436 | 0.001 |  | 0.206 | 0.366 |  | 0.201 | 0.548 |  | 0.285 | 0.082 |  | 0.350 | 0.711 |
| Forest |  |  |  | **0.394** | **<0.001** |  | **0.285** | **<0.001** |  | **0.224** | **<0.001** |  | 0.692 | 0.017 |
| Grassland |  |  |  |  |  |  | 0.169 | 0.033 |  | **0.252** | **<0.001** |  | 0.300 | 0.759 |
| Shrubland |  |  |  |  |  |  |  |  |  | 0.195 | 0.011 |  | 0.438 | 0.330 |
| Wetlands |  |  |  |  |  |  |  |  |  |  |  |  | 0.549 | 0.100 |

**Appendix S3**

Accessibility to humans at sampled (presences and absences) and random locations.

|  |
| --- |
| **Table S3.** Differences in accessibility to humans at sampled locations (presences and absences) and random locations in the Beni Department, Bolivia. Results of Kruskal-Wallis tests are provided. For significant Kruskal-Wallis tests, letters indicate values that are not significantly different according to pairwise comparisons using Mann–Whitney U-tests. |
| Variable | Presences | Absences  | Random  | Chi-square | P |
|   | (N=79) | (N=105) | (N=10 000) |   |   |
| Distance to rivers | 3 ± 3 | 5 ± 5 | 4 ± 4 | 3.1 | 0.21 |
| Distance to main roads | 28 ± 23 | 31 ± 25 | 38 ± 46 | 1.1 | 0.58 |
| Distance to secondary roads | 4 ± 3a | 4 ± 3a | 12 ± 15b | 57.7 | <0.001 |