**Supplementary Material**

Habitat modelling locates nesting areas of the Endangered Black-capped Petrel *Pterodroma hasitata* on Hispaniola and identifies habitat loss

YVAN G. SATGÉ, ERNST RUPP, ADAM BROWN and PATRICK G. R. JODICE

**Contents**

Appendix S1. Detailed modelling methods.

Figure S1. Georeferenced file of Black-capped Petrel habitat suitability on Hispaniola, at a resolution of 90-m pixel; format GEOTIFF.

Figure S2. Maps of Black-capped Petrel habitat suitability on Hispaniola, showing only suitable habitat.

Appendix S2. Detailed discussion of location of suitable habitat on Hispaniola.

**Appendix S1: Detailed modelling methods**

**Petrel activity sites and available sites**

Using a stratified-random selection at each nesting area, we partitioned the 81 nest sites used in this study into two datasets: 80% of the nest sites formed the training dataset (n = 64), and a minimum of 20% at each area formed the validation dataset (n = 17: La Visite: n = 6; Morne Vincent: n = 1; Loma del Toro: n = 8; Loma Quemada: n = 1; Valle Nuevo: n = 1). Black-capped Petrels often nest in clusters of 5-10 burrows within 100-200 m. Therefore, to avoid risks of pseudo-replication and to better quantify environmental data at nest sites, we created buffers of 50 m radius around each nest, and combined overlapping buffers into presence polygons. To inform the presence/available framework, we randomly located 500 available (i.e. pseudo-absence) sites (see Barbet-Massin *et al.* 2012 for a discussion of sample size for generalized linear models) in all areas in Hispaniola above 1200 m in elevation (Simons *et al.* 2013) at least 500 m away from presence sites (see Olivier and Wotherspoon 2005 for a discussion of buffer size). This buffer distance was wide enough to avoid pseudo-absence sites overlapping with known presence sites but still allowed for random sampling in areas ecologically relevant to the study (e.g. narrow canyons). We also ensured that no pseudo-absence sites were located within a 1500 m radius of locations where recent radar studies have shown the presence of suspected (but unconfirmed) nesting Black-capped Petrels (Brown 2017). We created pseudo-absence polygons from pseudo-absence sites, following the same procedure used for creating presence polygons. We partitioned the dataset of pseudo-absence polygons into two datasets: 80% of the polygons formed the training dataset used to build the model, and the remaining 20% formed the validation dataset.

**Environmental variables**

Unless described otherwise, each dynamic biotic variable was averaged over the period from 1 November 2017 to 31 July 2018. This period provides the most complete datasets for environmental predictor variables and corresponds to the typical nesting period of Black-capped Petrel. Static geographic variables included elevation (HydroSHEDS Digital Elevation Model, DEM); distance to coast, calculated as the projected distance between the center of DEM cells and the coastline (from the Global Self-consistent, Hierarchical, High-resolution Shoreline v2.3.7; Wessel and Smith 2017); distance to closest ridgeline, calculated following methods in Troy *et al.* (2014, Text S1; on the basis of HydroSHEDS DEM); slope steepness, calculated using the Slope Tool in ArcGIS 10.1 (on the basis of HydroSHEDS DEM); aspect (i.e. slope orientation), calculated using the Aspect Tool in ArcGIS 10.1 (on the basis of HydroSHEDS); and flow accumulation, calculated using the Flow Accumulation Tool in ArcGIS 10.1 (on the basis of HydroSHEDS DEM). Dynamic environmental variables included 16-day enhanced vegetation index (EVI; MOD13A1), which is responsive to canopy structure and maintains sensitivity over dense vegetation; 4-day leaf-area index (MCD15A3H) which is sensitive to canopy types and can detect differences between broadleaf and conifer canopies; 8-day evapotranspiration (MOD16A2); yearly primary productivity (MOD17A3H); percent tree cover (Global Forest Watch 2019a; thereafter, tree cover); aboveground live woody biomass density for the year 2000 (Global Forest Watch 2019b; thereafter, woody biomass); monthly mean wind speed for 1970-2000 (November to July; Fick and Hijmans 2017); and monthly average radiance corrected for stray light (NOAA VIIRS; median; Elvidge *et al.* 2017). We chose to include average radiance because it is an observation of light density and a useful proxy for human settlements, and therefore potential anthropogenic disturbance. The monthly mean value was influenced by temporary forest fires therefore we chose to use the monthly median value instead, which only retained permanent lighting. Unless noted otherwise, environmental datasets were obtained via Google Earth Engine (sample code available at <https://code.earthengine.google.com/ec32e28acb5ef1d9b5e62934b9641d5c> ).

**Logistic regression**

For each presence and pseudo-absence polygon, we calculated the mean value of each environmental variable as the mean of values of the raster cells whose centers were covered by the polygon. We then compared by logistic regression the distributions of values between presence and pseudo-absence polygons for each environmental variable. To reduce the effect of differences in sample size between the two groups, we weighted to simulate an equal number of presence and pseudo-absence sites such that the total weight of the presence data was the same as the total weight of the absence data (Barbet-Massin *et al.* 2012).

We sought to retain only the variables showing a strong significant difference between confirmed and available sites and therefore selected variables with individual regression coefficients significant at p ≤ 0.01 (Table 2). We then checked for collinearity among these retained variables using a generalized variance-inflation factor for generalized model. MODIS vegetation variables showed strong collinearity among themselves and with the composite dataset tree cover-EVI. Therefore, we retained only tree cover-EVI because it best represented actual field conditions and had the second most significant regression coefficient (after EVI). Altitude and mean wind speeds also showed collinearity: we rejected the wind variable because it had the least significant regression coefficient (pMean Wind = 0.05 vs. pElevation < 0.005). Variance-inflation factors for the remaining variables were <1.8, which suggested that these variables could be included in the GLM without large risk of collinearity (Quinn and Keough 2002). The retained environmental predictors were: elevation, distance to coast, and tree cover-EVI.

**Habitat suitability model**

We compared separate candidate logistic regression models, which included all possible single-level combinations of the significant predictors (level 1), using a GLM framework with a binomial distribution (presence vs pseudo-absence) and a logit link function. We ranked candidate models on the basis of their AICc: a single model had a ΔAICc value <2 and an Akaike weight >0.9 therefore we considered it the top performing model (Table 3). We used several measures to estimate the discrimination ability of the top model. First, the area under the receiver operating curve (AUC) is not recommended for presence only or presence-pseudo-absence frameworks (Lobo *et al.* 2008); therefore, we estimated model predictions with a continuous Boyce Index (Hirzel *et al.* 2006) using a moving window of width 0.1 and a resolution of 100 focals (function *ecospat.boyce* in R, package *ecospat*; Di Cola *et al.* 2017). The Boyce Index characterizes the predictive capability of presence-only models and is a reliable measure of their predictions (Hirzel *et al.* 2006). Following Hijmans (2012), we also assessed the possibility that high spatial autocorrelation biases the model’s discrimination capabilities. To address this particular issue, we calculated the amount of spatial sorting bias, which ranges from zero (highly correlated dataset) to 1 (uncorrelated dataset; Hijmans 2012). As advised by Lobo *et al.* (2008), we also calculated the sensitivity (i.e., true positive rate, or proportion of instances of presence correctly predicted as presence), specificity (i.e. true negative rate, or proportion of instances of absence correctly predicted as absence), and the threshold of specificity-sensitivity (maximum of the sum of the sensitivity and the specificity; Hijmans 2012).

We computed the habitat suitability model in a GIS framework by composing the model’s regression equation into the inverse logit function and applying it to digital rasters of the environmental variables included in the top performing model. This produced a final habitat suitability model in the form of a digital raster with cell values ranging from zero (habitat not suitable for Black-capped Petrel nesting) to one (habitat highly suitable for Black-capped Petrel nesting).

**Model validation**

We used two different methods to test the validity of the habitat suitability model. First, we calculated the suitability value at presence and pseudo-absence sites earlier set aside for validation and compared the distribution of suitability values between both groups using an F-test. Secondly, we used an independent dataset of positive results of acoustic monitoring (i.e., locations where petrel activity was recorded by acoustic monitors) carried out between 2015 and 2017 in the Cordillera Central of the Dominican Republic (McKown *et al.* 2016, Fleishman and McKown 2017). We created buffers of 500 m radius around each location where we had placed automated sound recorders, and combined overlapping buffers of similar results into acoustic polygons. We chose a value of 500 m because it is the likely maximum detection distance for the acoustic sensors (McKown *et al.* 2016). We also used a dataset of radar surveys conducted from 2012-2017 throughout Hispaniola (Brown 2017): we created buffers of 1,500 m radius around each location where we observed Black-capped Petrels flying in circles and suspected (but did not confirm) breeding activity. For each acoustic and radar polygon, we then calculated the mean suitability value as the mean of values of the suitability raster cells whose centers were covered by the polygon. Finally, for the validation dataset and the acoustic and radar datasets, we calculated the proportion of sites correctly classified as suitable by the model, with binned suitability levels ranging from ≥0.5 to ≥0.9 (in increments of 0.1).

**References**

Barbet‐Massin, M., Jiguet, F., Albert, C.H. and Thuiller, W. (2012) Selecting pseudo‐absences for species distribution models: how, where and how many? *Meth. Ecol. and Evol. 3*:327-338.

Brown, A. (2017) *Radar Surveys, Nest Monitoring and Conservation of the Black-capped Petrel on Hispaniola: February 2017*. Green Cove Spring, Florida, USA: Environmental Protection In the Caribbean.

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F.T., d'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A. and Pellissier, L. (2017) ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, 40:774-787.

Elvidge, C.D., Baugh, K., Zhizhin, M., Hsu, F.C. and Ghosh, T. (2017) VIIRS night-time lights. *Int. J. Remote Sens.* 38:5860-5879.

Fick, S.E. and Hijmans, R.J. (2017) WorldClim 2: new 1‐km spatial resolution climate surfaces for global land areas. *Int. J. Climatol*. 37:4302-4315.

Fleishman, A.B., McKown, M. (2017) *Acoustic Surveys for Black-capped Petrel Valle Nuevo, Hispaniola – 2016-17 – Final Report*. Santa Cruz, California, USA: Conservation Metrics, Inc.

Global Forest Watch. (2019a) Tree cover (2000) [on-line]. Updated 2 April 2019. Downloaded from http://data.globalforestwatch.org on 5 April 2019.

Global Forest Watch, (2019b) Aboveground live woody biomass density [on-line]. Updated 2 April 2019. Downloaded from http://data.globalforestwatch.org on 5 April 2019.

Hijmans, R.J., 2012. Cross‐validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology.* 93:679-688.

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C. and Guisan, A. (2006) Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Model.* 199:142-152.

Jiménez‐Valverde, A. (2012) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecol. Biogeogr.* 21:498-507.

Lobo, J.M., Jiménez‐Valverde, A. and Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.* 17:145-151.

McKown, M., Fleishman, A.B., Earl, A.D. (2016) Acoustic Surveys for Black-capped Petrel on Hispaniola and Dominica – 2016. Santa Cruz, California, USA: Conservation Metrics, Inc.

Olivier, F. and Wotherspoon, S.J. (2005) GIS-based application of resource selection functions to the prediction of snow petrel distribution and abundance in East Antarctica: comparing models at multiple scales. *Ecol. Model.* 189:105-129.

Quinn, G.P. and Keough, M.J. (2002) Experimental design and data analysis for biologists. Cambridge University Press.

Simons, T.R., Lee, D.S. and Haney, J.C. (2013) Diablotin Pterodroma hasitata: a biography of the endangered Black-capped Petrel. *Marine Ornithology.* 41:S3-S43.

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science.* 240:1285-1293.

Troy, J.R., N.D. Holmes, J.A. Veech, A.F. Raine, and M.C. Green. (2014) Habitat suitability modelling for the Newell’s shearwater on Kauai. *J. Fish Wildl. Manag.* 5: 315-329.

Wessel, P. and Smith, W.H. (2017) A global, self‐consistent, hierarchical, high‐resolution shoreline database [on-line]. Updated June 15, 2017. Downloaded from <https://www.soest.hawaii.edu/pwessel/gshhg/> on [1 June 2018]. **Figure S1. Georeferenced file of Black-capped Petrel habitat suitability on Hispaniola, at a resolution of 90-m pixel; format GEOTIFF.**

This file is available at <https://doi.org/10.5066/P9FWJPBD>.



**Figure S2. Maps of Black-capped Petrel habitat suitability on Hispaniola, showing only suitable habitat; format PNG. a) Map of suitable habitat (suitability > 0.65); b) map of most highly suitable habitat (suitability > 0.9).**

These files are available at <https://doi.org/10.5066/P9FWJPBD>.

a)



b)



**Appendix S2. Detailed discussion of location of suitable habitat on Hispaniola**

**Haiti**

In the Massif de la Hotte the suitable areas are located on both slopes of the main ridgelines of Pic Macaya, and on the north and west slopes of a lower escarpment to the west. In the Massif de la Selle, the suitable areas are located in four main regions. From west to east: at the western end of the escarpment, where suitable areas are located along both north and south slopes of that plateau, and on the steep banks of the cirque to the west; in the remaining forests of Hispaniolan pine on the La Visite plateau; along the steep north-facing slopes of the escarpment; and to the north and west of Pic de la Selle. The more gentle southern slopes of the La Visite escarpment are actively farmed and not suitable for Black-capped Petrels. Moreover, although forest remnants on the La Visite plateau are categorized as suitable based on environmental characteristics, villages are present under the forest canopy and all habitat is occupied by dwellings and cultivations (Anderson Jean, Jeunes en Action pour la Sauvegarde de l’Ecologie en Haïti, 2020, pers. com.) thus effectively preventing Black-capped Petrels from nesting in this area. North- and west-facing slopes of steep valleys descending from Pic de la Selle to the south and north-east are categorized as suitable but the drier and rockier south-eastern slopes of the massif are not. Finally for Haiti, the top ridges and north-facing patches of foothills between the Massif de la Selle and the capital Port au Prince are categorized as suitable but this habitat appears too fragmented to effectively host nesting populations of petrels.

In the Massif de la Hotte, close inspection of satellite imagery suggests that most of the forest damage associated with Hurricane Matthew in the fall of 2016 extended into the following petrel breeding season (Global Forest Watch 2019) and may have prevented nesting in 2017; fortuitously, Hurricane Matthew affected an area that is suspected to host only a small population of petrels (Brown 2013; Brown 2014; U.S. Fish and Wildlife Service 2018).

**Dominican Republic**

In the Dominican Republic, most of the suitable habitat is divided between Sierra de Bahoruco and Cordillera Central. In the Sierra de Bahoruco, the suitable habitat covers most of the wide ridgeline and the slopes on both north and south sides. Historically burnt areas in the center of the massif, and the drier savannas leading to the geological formation of Hoyo de Pelempito to the east, do not appear to be suitable for Black-capped Petrels. In the Cordillera Central, only the southwestern corner of the range is suitable. There, the most suitable habitat blankets the higher ridgeline of the peak Alto de la Bandera, and most of the forested plateau extending to the southeast. The higher elevations of a separate massif to the south, where the known nest sites in the area have been located, are also highly suitable. The suitable habitat covers roughly a fifth of the Cordillera Central; the remaining part, like Sierra de Neiba, appears to not be suitable for Black-capped Petrel. Finally for the Dominican Republic, fragments of suitable habitat are available on the extreme east of Sierra de Bahoruco, on coastal foothills of lower elevation and dense broadleaf vegetation.

**References**

Brown A. (2013*) Tracking the Devil: Radar Surveys for Black-capped Petrels*. Oral presentation. Regional Meeting of the Society for the Conservation and Study of Caribbean Birds, Grenada.

Brown, A. (2014) Radar Surveys for Black-capped Petrels on Hispaniola: January – March 2014. Green Cove Spring, Florida, USA: Environmental Protection In the Caribbean.

Global Forest Watch. (2019) Tree cover loss 2001-2018 (Version 1.6) [on-line]. Accessed online at http://data.globalforestwatch.org on [15 November 2019].

U.S. Fish and Wildlife Service. (2018) *Species status assessment report for the Black-capped Petrel (Pterodroma hasitata). Version 1.1. June 2018.* Atlanta, Georgia, USA: U.S. Fish and Wildlife Service.