DISTRIBUTION MODELING OF THE LONG-TAILED MARMOT (*MARMOTA CAUDATA*) FOR OBJECTIVES OF DIRECTING FIELD SURVEYS AND GROUND VALIDATION OF THE SNOW LEOPARD (*PANTHERA UNCIA*) HABITAT QUALITY

Volodymyr Tytar[1], Matthias Hammer[2], Tolkunbek Asykulov[3]

[1]*Schmalhausen Institute of Zoology, NASU (Kyiv, Ukraine)*
[2]*Biosphere Expeditions (Dublin, Ireland)*
[3]*Kyrgyz National University (Bishkek, Republic of Kyrgyzstan); Der Naturschutzbund Deutschland e. V. NABU (Bishkek, Republic of Kyrgyzstan)*

Distribution modeling of the long-tailed marmot (*Marmota caudata*) for objectives of directing field surveys and ground validation of the snow leopard (*Panthera uncia*) habitat quality. — V. Tytar, M. Hammer, T. Asykulov. — Marmots form a part of the diet of some endangered species such as the snow leopard (*Panthera uncia*), therefore the knowledge on their distribution and habitat preferences are crucial to the interest of the conservation and management of carnivores at high altitudes. Considering this, within a snow leopard project run by Biosphere Expeditions and NABU (Kyrgyzstan), surveys were carried out in summer field seasons of 2014–2019 to assess the distribution of the long-tailed marmots (*Marmota caudata*) in an area centered around the Karakol Mountain Pass (polygon centroid 74.83° E, 42.37° N) in the Kyrgyz Ala-Too Range. The presence of occupied marmot burrows was recorded using the location (cell) given by a grid, the code of which was displayed in a GPS. Using cells allows examination of data at a wider scale, so information is collected from different cells that are spread from each other, avoiding data autocorrelation. Environmental factors that may affect the spatial distribution of burrow systems were considered: land surface temperature (LST) in winter and summer, summer normalized difference vegetation index (NDVI), a Digital Elevation Model (DEM), and soil type data. The relationship between environmental factors and burrow records was analyzed using ecological niche models (Maxent) to predict the distributions of marmot burrows. The models performed well with average test AUC values of 0.939. The contribution orders of the variables in the models were summer NDVI and DEM, winter LST, summer LST, and soil type. The distribution of the suitable areas was largely (up to 38 % permutation importance) affected by summer NDVI. NDVI is an indicator of the feeding conditions of marmots and most of the records were distributed in areas with NDVI in summer ranging from 0.5 to 0.7. According to the prediction maps, suitable marmot habitat (> 0.5 predicted probabilities of occurrence) can occupy up to 40 % of study area. These maps are used to direct sampling efforts to areas on the landscape that tend to have greater predicted probabilities of occurrence and accomplish ground validation of snow leopard habitat quality.

K e y w o r d s : *Marmota caudata*, *Panthera uncia*, Kyrgyz Ala-Too, species distribution models, *Maxent*.

C o r r e s p o n d e n c e t o : Volodymyr Tytar; Schmalhausen Institute of Zoology NASU, B. Khmelnytsky St. 15, Kyiv, 0160 Ukraine; e-mail: vtytar@gmail.com; orcid: 0000-0002-0864-2548

Introduction

Marmots on the whole form a part of the diet of some endangered species such as the snow leopard (*Panthera uncia*). The snow leopard is, generally, distributed at higher elevations and its range is limited to the Asian continent only. Snow leopards normally inhabit rugged ranges and are associated through most of the range with arid and semi-arid shrub lands, grasslands or steppes. They commonly occur at elevations ranging between 3.000 and 4.500m, which may occasionally go up to 5.500m in the Himalayas. However, the species may also occur at much lower elevations such as from 560m to 1.500m. Across much of its range, snow leopards are dependent on ibex (as "primary" wild prey), which constitutes a substantial portion of its wild prey in its diet composition. Because marmots hibernate for up to eight months of the year, depending on species, latitude, and altitude, they are available for the predator for less than half of the year between emerging from their winter hibernation between May and September. For this reason, in comparison to other prey species upon which snow leopards heavily depend, marmots are considered "secondary" wild prey. How-

ever, their role in sustaining snow leopards during the summer season should not be underestimated or neglected. Therefore, knowledge about the distribution and habitat preferences of marmots is crucial to the interest of the conservation and management of carnivores at high altitudes (Ahmed et al., 2016).

Protected areas play a vital role in long-term nature conservation with the associated ecosystem services and cultural values. The snow leopard is considered 'Vulnerable', with an estimated 4.000 left in the wild, and protected areas have been created to safeguard its habitat. But of the 170 protected areas in the global range of the snow leopard, 40 % are smaller than the home range of a single adult male and only 4–13 % are large enough for containing 15 or more adult females. Because the animals range over much larger areas, there is a need not only for establishing greater numbers of large protected areas, but also by linking protected areas, corridors or ecological networks, including areas of traditional land use. Such conservation networks could spatially distribute the risk of extinction and address the life-history needs of a vagarious species, and enhance successful co-existence with humans. For top carnivores, for which the loss of habitat has often contributed towards a population decline, an important factor to consider is the conservation of the prey species and ensure its availability within the developing conservation network.

Unfortunately, knowledge on the availability of the wild prey, especially across a sizeable area, is often absent or insufficient. In such a case habitat suitability for prey species may serve as a proxy for availability, assuming that areas of high suitability can accommodate larger numbers of prey.

A common approach in this field is to model the suitability of habitat for a given species or group of species using habitat suitability indices based on an assessment of habitat attributes. Habitat suitability indices are indices in the sense that they usually combine many different variables (such as elevation, soil type, and land cover etc.) into a single composite measure. Predicting the distribution of wildlife species from habitat data is frequently perceived to be a useful technique and cost-efficient. However, species are not always present in areas where high probabilities of habitat suitability are predicted. This mismatch between modeled predictions and field observations may result from an array of issues, including conceptual errors, performance of modelling algorithms, insufficient understanding of factors driving species distribution, unallocated anthropogenic pressure etc. For these reasons, ground validation of predictions is recommended. This approach is only the first step toward identifying suitable habitat and is useful in directing subsequent field surveys.

Species conservation usually focuses on the most suitable habitat for the species of concern, but the challenge is to identify high-quality habitat across large areas. Among the various tools used in conservation planning to protect biodiversity, species distribution models, also known as climate envelope models, habitat suitability models, and ecological niche models provide a way to identify the potential habitat of a species in an ecoregion and their applications have greatly increased.

Species distribution models are based on the concept of the "ecological niche" (Hutchinson, 1957), which can be defined as the sum of the environmental factors that a species needs for its survival and reproduction. Many niche models are based on climate variables because these data are readily available, covering large spatial scales (Hijmans et al., 2005). Species distribution models predict the potential distribution of a species by interpolating identified relationships between presence/absence or presence-only data of a species on one hand and environmental predictors on the other hand across an area of interest. From the array of various applications, Maxent (Phillips et al., 2006) stands out because it has been found to perform best among many different modeling methods (Elith et al., 2006).

Five species of marmots, including the long-tailed marmot, *Marmota caudata*, are found across the global range of the snow leopard. Long-tailed marmots occur in the Hindu Kush, Karakoram, and Tien Shan mountains of Central Asia. They are most common in the mountain meadows which are often grazed by domestic sheep, goats, and yaks, and are found from elevations of 1.400 to 5.500 m, meaning there is an overlap with snow leopard habitat.

Here, we addressed the following questions:

(1) Which environmental predictors are best associated with marmot occurrence?

(2) How well do models predict the occurrence of this species?

(3) What is the potential distribution of highly suitable habitat for long-tailed marmot in the study area and how this is spatially related to records of the snow leopard?

Material and methods

Study area

The chosen study area is located in the southern Kyryz Ala-Too, away from the main cities in the densely populated north. Surveys were centered around the Karakol Mountain Pass (3.452 m) and encompassed areas in the upper catchment of the West and East Karakol rivers. Data were collected during annual citizen science expedition's, run by wildlife conservation NGO Biosphere Expeditions and NABU (Kyrgyzstan), and lasting between four to eight weeks during the summer months 2014–2019. The main access route into the area was the Suusamyr plateau, a high steppe plateau (2.200 m). Although only some 160 km from the capital city of Bishkek, it is also one of the most remote and rarely visited regions of Kyrgyzstan. The expedition camp was located next to the Suusamyr-Kochkor road approximately in the middle of the study area (42.359535° N, 74.737829° E, 3002 m a.s.l.). From this base camp mostly one-day surveys, but also some two-day/one-night surveys were conducted to various parts of the Kyrgyz Ala-Too Range and the neighboring Jumgal-Too Range, reaching altitudes of up to 4.000 m.

In Soviet times the place was one of the major sheep breeding areas in the country. Up to four million sheep a year were driven over the mountain passes in spring to graze on the grasses of the steppe. Today the grazing pressure has reduced, but in the summer hundreds of people still live in yurts and graze their livestock here. The commonest violations of land use are unsystematic livestock grazing, leading to the disruption of normal growth and evolution of pasture vegetation, and poaching (including the hunting of marmots).

The area was divided into 2 x 2 km cells following the methodology manual developed for citizen science expeditions by Mazzolli & Hammer (2013). The corresponding grid covering the study area was uploaded into the expedition's GPS units to aid navigation and data collection.

Species records

Marmot presence was identified by characteristic and easily identifiable whistle and sightings. Other records included scat and tracks. Marmot scat is easily recognized because it is dark green when fresh and malodorous character. Active burrows often have fresh scat at the entrance, and vegetation does not protrude across the opening, whereas inactive burrows typically have vegetation growing into the entrance.

The presence of active marmot burrows was recorded using the location (cell) given by a grid, the code of which was displayed in a GPS. Using cells allows examination of data at a wider scale, so information is collected from different cells that are spread from each other, avoiding data autocorrelation. The extracted points were georeferenced using a Garmin eTrex 10 GPS receiver in combination with Google Earth (www.google.com/earth). All the coordinates were expressed in decimal degree and converted to a point vector file for modeling the distribution of the species. Only spatially unique ones (n=30), corresponding to a single grid cell were used.

Environmental data

To relate the occurrence records of marmots with environmental conditions, the following environmental factors, following Lu et al. (2016), were included: land surface temperature (LST) in winter and summer, normalized difference vegetation index (NDVI) in summer, derived from the Moderate resolution Imaging Spectroradiometer (MODIS) satellite, Digital Elevation Model (GDEM), and soil type data. All data adopted in this area were resampled to 1 km spatial resolution.

MODIS is a key instrument aboard the Terra and Aqua satellites. Eight-day composite MODIS LST and 16-day composite MODIS NDVI, both at a resolution of 1 km, were used to represent the

thermal environment and feeding conditions for the long-tailed marmot, respectively. NDVI values vary between −1 and +1; the higher the NDVI value, the denser the green vegetation (Haque et al., 2010), and a zero means no vegetation. These two remote sensing variables were obtained from the MODIS website (https://modis.gsfc.nasa.gov/). To characterize the thermal and vegetation conditions more reliably, the LST and NDVI values were averaged for the years of the survey time.

The DEM was aggregated from the 30 seconds (~30m) NASA Shuttle Radar Topographic Mission (SRTM) DEM (http://srtm.csi.org). Soil data with 1km spatial resolution was available from SoilGrids (https://soilgrids.org), a system for global digital soil mapping that uses state-of-the-art machine learning methods to map the spatial distribution of soil properties across the globe (Hengl et al., 2014).

SAGA (System for Automated Geoscientific Analyses) GIS software (v. 2.2.7), a free and open source geographic information system, was used for processing and editing spatial data, and geostatistical analysis of the study area (Conrad, 2006).

Statistical modeling

Factor analysis in JASP statistical software (https://jasp-stats.org/) was used to examine the contributions and the main patterns of inter-correlation among the potential environmental controls. JASP is a free and open-source graphical program for statistical analysis supported by the University of Amsterdam. Principal component was used as the extraction method. By rotating the factors a factor solution was found that is equal to that obtained in the initial extraction but which has the simplest interpretation, and for this purpose the Varimax normalized type of rotation was applied. Usually a solution that explains 75−80 % of the variance is considered sufficient. Variables with the highest factor loadings are those most strongly correlated with the corresponding principal components and regarded the best single-dimensional descriptors of the dataset.

Species distribution models obtained from a data set of associated environmental covariates often inherently result in multi-collinearity, a statistical problem defined as a high degree of correlation among covariates. Factor analysis is among the statistical procedures proposed to solve or to reduce multi-collinearity because the obtained factors are independent of each other, that is, they are orthogonal. Accordingly, this permits them to be used as independent, non-correlated variables in analyses of modeling potential species distribution (Cruz-Cárdenas et al*.,* 2014) instead of the raw data of environmental variables.

Maxent distribution model

Maxent is a machine learning model that uses presence-only data (occurrence records of burrows in this research) and environmental variables to build relationships based on the principle of maximum entropy (Phillips et al., 2006). The basic principle of the Maxent model is to estimate the potential distribution of a species by determining the distribution of the maximum entropy (i.e., closest to uniform), with constraints imposed by the observed spatial distributions of the species and the environmental conditions. Maxent computes a probability distribution that describes the suitability of each grid cell (varying from 0 to 1, indicating the lowest suitability and the highest suitability, respectively) as a function of the environmental variables at the known occurrence locations and then produces a map of the species' potential geographical distribution by projecting into the geographic space.

The Maxent software[1] version 3.3.3 k was used in this study to predict the potential distribution of burrows of the long-tailed marmot. To evaluate the model performance, the data were split into two parts: 75 % for training and 25 % for test sets. A 10-fold cross-validation was used to perform the model training and testing to assess the performance of our model. The test gain and test area under the receiving operator curve (AUC) were used to evaluate the model's goodness-of-fit. The AUC is an effective indictor of model performance. The larger the AUC, the higher is the sensitivity

---

1 Available from http://www.cs.princeton.edu/~schapire/maxent/

rate, and the lower is the 1-specificity rate. In general AUC values > 0.9 are considered to have 'very good', > 0.8 'good' and > 0.7 'useful' discrimination abilities (Swets, 1988).

Maxent also allows the construction of response curves to illustrate the effect of selected variables on habitat suitability (consequently, on the probability of occurrence and giving an idea of where for each variable, under the constraints and conditions of the modelling situation, the focal species has it's optimum). These response curves, obtained for separate predictors, e.g. elevation, consist of the specific environmental variable as the *x*-axis and, on the *y*-axis, the predicted probability of suitable conditions as defined by the logistic output.

We used 10 percentile training presence because this threshold value is considered to provide a better ecologically significant result when compared with more restricted thresholds values (Phillips, Dudík, 2008)

Results

The factor analysis provided a comprehensive way to analyze the niche of the long-tailed marmot in the study area and captures various aspects of marmot ecology and environmental requirements of the species. Factors 1–3 extracted from all the variables explained ~87 % of the variance. A path diagram giving a visual representation of the direction and strength of the relation between the environmental variable and factor is shown in Fig. 1.

Fit of the model was tested by the Tucker-Lewis index, which yielded a value of 1.25; by convention a value higher than 0.9 indicates a good fit.

The first factor heavily correlated with the NDVI for July and inversely with elevation (49.4 % contribution), emphasizing the high importance of summer vegetation growth, acting on the availability and quality of food resources, which ultimately determine the amount of marmot fat reserves, which are needed for hibernation survival. Evidently, better foraging opportunities are found at some intermediate elevations, as brought out by the corresponding response curve (Fig. 2), where better suitability (> 0.5) appears between altitudes of around 3200 and 3600 m.
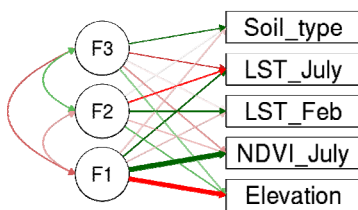

Fig. 1. Path diagram. The factors in the model are represented by the circles. The variables are represented by the boxes (from top to bottom: soil type, land surface temperature for July, land surface temperature for February, normalized difference vegetation index for July, elevation). Arrows going from the factors to the variables are representing the loading from the factor on the variable. Red indicates a negative loading, green a positive loading. The wider the arrows, the higher the loading.

Рис. 1. Шляхова діаграма. Фактори в моделі представлені кругами. Змінні представлені полями (зверху вниз: тип ґрунту, температура поверхні землі для липня, температура поверхні землі за лютий, нормалізований диференційний вегетаційний індекс за липень, висота над рівнем моря). Стрілки, що йдуть від факторів до змінних, представляють навантаження від фактора на змінну. Червоний колір позначає негативне навантаження, зелений — позитивне навантаження. Чим ширші стрілки, тим більше навантаження.
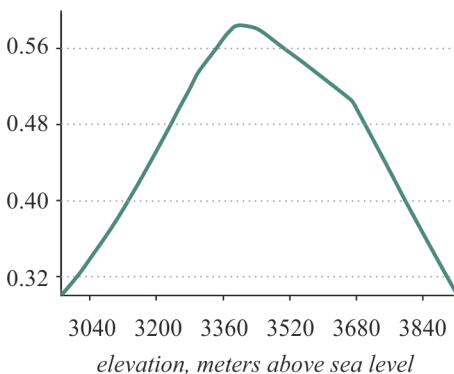

elevation, meters above sea level

Fig. 2. The response curve, obtained for elevation; the specific environmental variable is represented on the *x*-axis (in m), on the *y*-axis are the predicted probability of suitable conditions as defined by the logistic output.

Рис. 2. Крива відгуку, отримана для висоти над рівнем моря; ця змінна представлена на горизонтальній осі *x* (в м), на вертикальній осі *y* відкладена прогнозована ймовірність відповідних умов, визначених логістичним результатом.

The second factor is solely related to temperature conditions (20.2 % contribution). The stronger positive correlation of this factor with the LST for February may suggest that marmots preferred warm areas for hibernation burrows (Lu et al., 2016). In these areas, the snow cover melts early in spring, and therefore the survival and reproduction rate of the populations in these areas are likely to be higher than those of populations in other areas.

Soil type is attached to the third factor (17.9 % contribution) by a fairly loose correlation of 0.42. This may be due to the dominance in the area of one soil type, namely Haplic Cambisols, for which erosion and deposition cycles account for their widespread occurrence in mountain regions.

Gridded data layers produced in SAGA GIS representing the first three factors were then used for modelling species distribution.

From the 25 model runs, the average AUC was 0.921 (i.e. 'very good' discrimination abilities of the final model), with little variation in AUC between runs (SD = 0.085). The averaged output from these model runs is shown in Fig.3, where a dashed contour line is drawn to delimit areas of predicted probability of long-tailed marmot occurrence exceeding the 10 percentile training presence threshold value of 0.35.

In terms of nature conservation planning and setting snow leopard research priorities these areas of high predicted probability of long-tailed marmot occurrence are of prime interest. At a minimum, the model identified areas for more focused study and survey.

More specifically, close attention should be drawn to areas where for both the predator and prey species there is a potential that their home ranges may overlap or be in close neighbourhood. In our opinion, these could be located between (or nearby) the boundary for suitable marmot habitat and the 3500 m elevation isoline, below which there has been no sign of the snow leopard (the lowest record made throughout the entire survey was at 3512 m).

With additional model validation, such as field validation, the model could support the delineation of suitable habitat for the long-tailed marmot, especially in and around areas where snow leopards have been recorded (as depicted by arrows in Fig. 3).
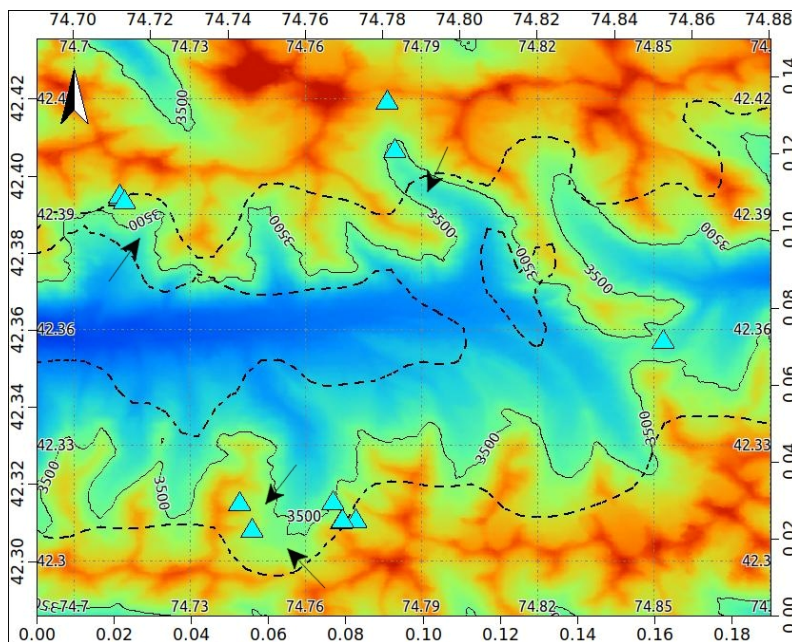

Fig. 3. Digital elevation map of the Kyrgyz Ala-Too study area (range from red at highest elevation to blue at lowest). The dashed line contours areas of predicted probability of long-tailed marmot occurrence exceeding the 10 percentile training presence threshold value; the solid line contours elevation at 3500 m; triangles indicate places where records were made of the snow leopard; arrows point to priority areas.

Рис. 3. Карта цифрової моделі висот досліджуваного району Киргизького Ала-Тоо (червоний колір вказує на найвищі висоти, а блакитний — на найнижчі). Пунктирна лінія обведена навколо областей передбачуваної ймовірності появи довгохвостого бабака, що перевищує 10-процентильне порогове значення присутності; суцільна лінія вказує на горизонталь у 3500 м; трикутники позначають місця, де були зроблені записи перебування снігового барса; стрілки вказують на пріоритетні території.

As such, we propose to incorporate this niche model as a reference model into the development of a conservation strategy, which can guide the protection of the snow leopard in this specific area of the Kyrgyz Ala-Too Range.

Acknowledgements

References

Ahmed, T., M. Shoeb, P. Chandan, A. Khan. 2016. On the status of the Long-tailed Marmot Marmota caudata (Mammalia: Rodentia: Sciuridae) in Kargil, Ladakh (Indian Trans-Himalaya). *Journal of Threatened Taxa*, **8** (9), 9171–9176.

Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, J. Böhner. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development Discussions*, **8** (2): 2271–2312.

Cruz-Cárdenas, G., L., J. L. López-Mata, L. Villaseñor, E. Ortiz. 2014. Potential species distribution modeling and the use of principal component analysis as predictor variables. *Revista Mexicana de Biodiversidad*, **85** (1): 189–199.

Elith, J., C. H. Graham, R.P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion., C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A.T. Peterson, S.J. Phillips, K. Richardson, R. Schachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. Wisz, N.E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29** (2): 129–151.

Haque U., M. Hashizume, G. E. Glass, A. M. Dewan, H. J. Overgaard, T. Yamamoto. 2010. The role of climate variability in the spread of malaria in Bangladeshi highlands. PLoS One. **5** (12): e14341.

Hengl T., J. M. de Jesus, R. A. MacMillan, N. H. Batjes, G. B. M. Heuvelink, et al. 2014. SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLoS ONE*, **9** (8): e105992.

Hijmans, R.J., S. E. Cameron, J. L. Parra, P.G. Jones, A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25** (15): 1965–1978.

Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, No. 22: 415–427.

Lu L., Z. P. Ren, Y. J. Yue, X. T. Yu, S. Lu, G. C. Li, H. L. Li, J. C. Wei, J. L. Liu, Y. Mu, [et al.]. 2016. Niche modeling predictions of the potential distribution of Marmota himalayana, the host animal of plague in Yushu County of Qinghai. *BMC Public Health*, **16**: 183.

Mazzolli, M., M. Hammer. 2013. Sampling and analysis of data for large terrestrial mammals during short-term volunteer expeditions. *Biosphere Expeditions*, 1–23.

Phillips, S. J., M. Dudik. 2008. Modelling of species distributions with MAXENT: new extensions and a comprehensive evaluation. *Ecography*, **31** (2): 161–175.

Phillips, S. J., R. P. Anderson, R. E. Schapire. 2006. Maximum entropy modelling of species geographic distributions. *Ecological Modeling*, **190** (3–4): 231–259.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, [et al.]. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29** (2): 129–151.

Swets, J. 1988. Measuring the accuracy of diagnostic systems. *Science*, **240** (4857): 1285–1293.