International
Journal of
**Biological
and Natural
Sciences**

BAT FALCON (*Falco rufigularis*) DISTRIBUTION SPATIAL PATTERNS IN MEXICO

Salvador Garcilita Arguello
Department of Computer Science.
Autonomous Technological Institute of
Mexico, Mexico City, Mexico

Daniela Treviño Díaz
Faculty of Sciences, National Autonomous
University of Mexico, Mexico City, Mexico

Guillermo Raguer González López
Faculty of Sciences, National Autonomous
University of Mexico, Mexico City, Mexico

Abstract: Despite the importance of birds of prey in ecosystems, knowledge about them is limited for most species, especially in tropical regions. On the other hand, there are efforts to collect massive amounts of data, but it is a challenge to convert them into information. This work makes use of the eBird database, updated with Citizen Science, to contribute to the state of knowledge about birds of prey by analyzing the current distribution of the bat falcon. (*Falco rufigularis*) in Mexico using two complementary approaches; 1) species distribution modeling and 2) population regionalization (records clustering) applying supervised and unsupervised machine learning algorithms. Results show bat falcon has a wide distribution in Mexico, with an area of environmental suitability of 19,720 km2. Its most abundant and extensive populations are found in the lowlands of the southeast of the country, which is a region with strong anthropogenic pressures. On the other hand, unsupervised clustering models turned out to be a good tool to analyze spatial patterns in distribution, identifying 8 groups of records that could be interpreted as subpopulations, presenting greater continuity in the states of Veracruz, Tabasco and Chiapas, as well as in the Yucatan peninsula.

Keywords: Birds of prey; Species distribution modeling; machine learning; Biogeography

INTRODUCTION

The bat falcon (*Falco rufigularis*) is a species of bird of prey belonging to the Falconidae family (Order: Falconiformes) widely distributed in America; historically, it is found from southern Sonora and Tamaulipas (with Mexico being the northernmost part of its distribution), distributed throughout the country in tropical and subtropical lowlands of up to 1500 meters above sea level; the rest of its distribution covers almost all of Central America and extends to Ecuador, Peru, Bolivia, Paraguay, Argentina and Brazil; however, it is likely that its current distribution and abundance have changed considerably, considering the intense change in land use and habitat destruction to which the Neotropics has been subjected, in addition to the indiscriminate use of pesticides such as Dichloro Diphenyl Trichloroethane (DDT) (Cade, 1982; Mindell et al. 2018).

The existence of three subspecies has been proposed in *F. rufigularis*; *F. r. petoensis, F. r. rufigularis* and *F. r. ophryophanes*; the subspecies belonging to Mexico and Central America is *petoensis* and, like the others, it is not considered migratory, with individuals often established in pairs in the same place throughout the year, and in the case of juveniles, dispersing from breeding areas and perhaps carrying out local nomadic movements (Global Raptor Information Network, 2022).

However, the Mexican population, even though it belongs to the same subspecies, given the heterogeneity of the environment with respect to biomes and mountain axes, can probably be divided into different subpopulations if the appropriate tools are used and sufficient data is available. *F. rufigularis*, given its wide distribution and abundance in suitable habitats, is listed in the Least Concern category of the Red List of the International Union for Conservation of Nature (IUCN); however, its population trend is estimated to be decreasing (IUCN, 2022).

One of the most notable peculiarities of birds as a study group is the important contribution made by members of society who are not specialized in data collection due to the interest they generate among people; these data recorded by amateurs have great ecological value and must be considered to update and improve the state of knowledge of the species, as well as to strengthen the ties between the academic sector and society

(Berlanga et al. 2015). In particular, eBird is a citizen science initiative of the Cornell Laboratory of Ornithology and the National Audubon Society created in 2002, which allows for the massive collection of data collected by bird watchers (citizen scientists) using standardized protocols (Sullivan et al. 2009). This source of data can be and has been used to evaluate and adapt bird conservation strategies in different parts of the world, since it offers the possibility of improving our understanding of birds, the habitats they require and how to protect them (Sullivan et al. 2017).

Species distribution models relate data on either presence or abundance at known sites with information on the environment (both biotic and abiotic factors) that have or could have an impact on their distribution; in order for them to perform well in the projection, 1) a good sampling design is needed (quantity and quality of data), 2) choosing correct predictor variables (that have an impact on distribution) and 3) choosing an appropriate computational model (for example, non-parametric models that are flexible and resilient to the possible correlation between predictor variables) (Elith and Leathwock, 2009). In birds, bioclimatic variables have been successfully used as distribution predictors, allowing to identify suitable habitat for the species within the established range based on specific environmental requirements (Mi et al. 2016).

How to study and protect birds of prey in a changing environment is a non-trivial question; as predators, they have specific characteristics (relatively low population density and elusive habits) that make it a challenge to have solid information about their biology, however, the development of successful strategies demands more work to guide future sampling, study and conservation efforts. The objective of this article is to reveal the current distribution of *F. rufigularis* in Mexico based on records collected through citizen science using a replicable methodology for spatial modeling of the distribution of other species, allowing us to know details in the patterns of its distribution that are only possible thanks to 1) the existence of a large amount of data collected collectively and systematically over long periods of time and 2) the use of statistical techniques and machine learning algorithmsthat allow us to process this data and find patterns that might otherwise go unnoticed (Kelling et al. 2013).

MATERIALS AND METHODS

DATA ACQUISITION AND PRE-PROCESSING

The data used for the analysis were obtained from the eBird database (eBird Basic Dataset, 2021), filtering the records of *F. rufigularis* for the years 2010-2020 throughout the country. The data processing in this work was carried out using the programming languages R (R Core Team, 2021) and Python (Van Rossum and Drake, 1995), choosing one or the other based on the availability and usability of packages to perform the analyses.

Once the database was obtained, the R package AUK was used to process the data and eliminate duplicate records due to lists shared by observers and keeping only records corresponding to the specified period (Astriñas-Mackey and Hochachka, 2018; Strimas-Mackey et al. 2020). Furthermore, given the potential bias in the distribution of observation records in eBird (sampling bias) and in order to generate a statistically more reliable model, spatial filtering was performed using the R package spThin, reducing observations to one point per five kilometers, thus decreasing spatial autocorrelation. In the end, 668 records were obtained (Aiello-Lammens et al. 2015).

DISTRIBUTION MODELING

A distribution model for the species was developed using the 19 bioclimatic variables from WorldClim (http://www.worldclim.org/) as predictors: (BIO1) mean annual temperature; (BIO2) mean diurnal range; (BIO3) isothermality; (BIO4) temperature seasonality; (BIO5) maximum temperature of the warmest month; (BIO6) minimum temperature of the coldest month; (BIO7) annual temperature range; (BIO8) mean temperature of the wettest quarter; (BIO9) mean temperature of the driest quarter; (BIO10) mean temperature of the warmest quarter, (BIO11) mean temperature of the coldest quarter; (BIO12) annual precipitation; (BIO13) precipitation of the wettest month; (BIO14) precipitation of the driest month; (BIO15) precipitation seasonality; (BIO16) precipitation of the wettest quarter; (BIO17) precipitation of the driest quarter; (BIO18) precipitation of the warmest quarter, and (BIO19) precipitation of the coldest quarter. These variables were obtained using the R package raster version 3.6-11 at a spatial resolution of 10 minutes (approximately 340 km2) (Hijmans, 2022).

Once the climatic data for the coordinates with records of the species were extracted, a matrix was generated in which each record stores the numerical value of each of the 19 predictor variables for the target presence/absence variable. Since the data downloaded from eBird are only for sightings (positive presence records), there is no absence data (negative records), so in order to use supervised machine learning models, a simulation of 3000 random absence data (background data) was performed, which establish the domain of the bioclimatic variables in the study area (Mexico) and would represent the records if the distribution of the species were not related to the predictor variables, that is, if the species had no preference for the different environmental parameters considered as variables (Philips et al. 2009). This data processing was performed with the specialized R libraries maptools (version 1.1-5), raster and dismo (version 1.3-9) (Bivand & Lewin-Koh, 2022; Hijmans et al. 2022).

The Breiman random forest algorithm was used as a predictive model because it has demonstrated superior performance to traditional regression techniques and other predictive algorithms in species distribution modeling, particularly those for which relatively few records are available and the distribution territory has been sampled in a limited way, in addition to being resilient to the correlation between predictor variables (Mi et al. 2016; Mi et al. 2017). The model was developed with the R package randomForest (Liaw and Wiener, 2002).

The random forest is a supervised machine learning algorithm based on decision trees, however, it prevents the overfitting that trees usually imply by determining the final prediction as the vote (the average) of the trees that make up the ensemble (wisdom of the crowds) (Breiman, 2001). The model was trained with 75% of the records, using the remaining 25% to evaluate it with a balanced stratification. The number of m predictor variables randomly selected for each tree (hyperparameter mtry) was three, thus seeking a balance between the correlation of the model trees and the variance (expressed in overfitting), which reflected an out-of-bag error of 0.127 (Figure 1). The random forest was set to regression mode, thus prioritizing the discriminatory capacity of the model between presence and random absence data (Zhang et al. 2019). The assembly was adjusted with 500 trees since from there the error (average of the squared residuals; 0.1242157) did not decrease significantly, explaining 16.61% of the model variance.
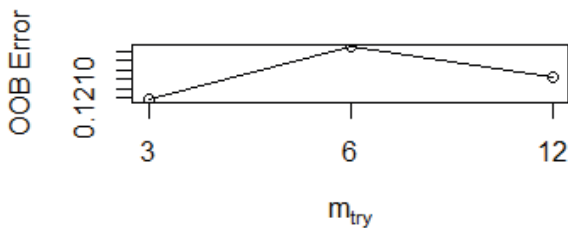
Figure 1: Out-of-bag error values with different values of m predictor variables.

For greater interpretability, the measure of the importance of the variables in the prediction of the model was obtained as the average decrease in the impurity of the node (being the residual sum of squares because it is a regression model) by dividing according to the variable averaged in all the trees; in turn, the partial dependence graph of each variable with the response variable was obtained, which shows the marginal effect of the predictors on the target variable (Liaw and Wiener, 2002).

The model was evaluated by computing a confusion matrix, deriving from it the *Receiver Operating Characteristic Curve* (ROC) and obtaining the Area Under the ROC Curve (AUC) metric. The AUC has been used extensively in species distribution modelling, whose probabilistic interpretation in this context would be the probability that the model classifies a randomly chosen site of presence of the species higher than a randomly chosen site of absence (Pearce and Ferrier, 2000; Liu et al. 2010).

GENERATION OF GROUPS

To understand the distribution of the species in the country and its potential grouping into subpopulations, a DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm was applied, an unsupervised machine learning clustering technique based on point density that allows classifying unlabeled data into groups in an unsupervised manner and that has an appropriate performance with spatial data, in addition to being able to detect records that naturally under the criteria used do not belong to any group (noise) (Ester et al. 1996). DBSCAN requires three parameters (in the context of Machine Learning, hyper parameters) that were defined by us;

1) an epsilon value, which is the minimum distance criterion between two points to be considered neighbors and form a group (subpopulation in the context of this work), 2) minimum samples, which is the minimum number of grouped records to be considered a group and not outliers or noise (isolated records) and 3) a metric to calculate the distance between points.

To choose the most appropriate value for epsilon, a Nearest Neighbors model was first created, another machine learning algorithm that groups the data based on a number of groups determined by the researcher, allowing the distance to the n closest points of each record to be calculated, ordered and then graphed to observe the optimal epsilon value, which is reflected as the point of maximum curvature (Figure 2) (Rahmah and Sukaesih, 2016).
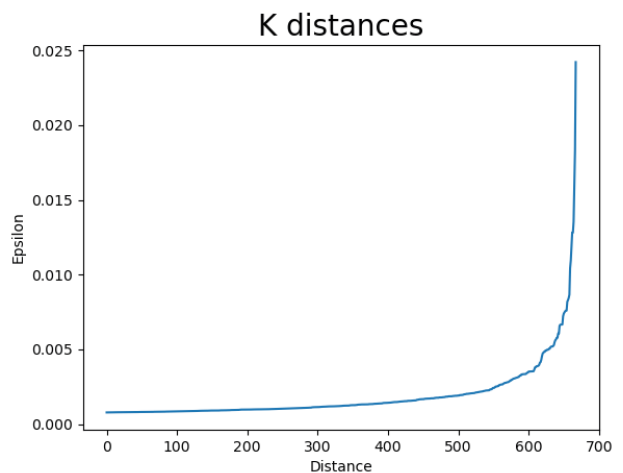

Figure 2: Distances plotted from smallest to largest, with the inflection point observed at approximately 0.010.

Finally, the hyperparameters chosen for the model were 0.010 for epsilon and 4 for the minimum number of samples, applying the semiverse formula as a metric to calculate the distance between instances. The analysis was performed in Python using the Scikit-learn library (Pedregosa et al. 2011).

The silhouette coefficient was obtained as a metric for evaluating clustering, which considers both the distance from each point (record) i to all points in the closest cluster (intercluster distance) and the distance from each point i to all points in the same cluster (intracluster distance), being a point value in the range from -1, which means a poor generation of clusters, to 1, which means well-formed clusters with sufficient distance between them (Rousseeuw, 1987).

RESULTS

SPECIES DISTRIBUTION MODEL

Figure 3 shows the continuous prediction of the random forest model projected onto the Mexico raster; the area under the curve (AUC) value is 0.825 (Figure 4).
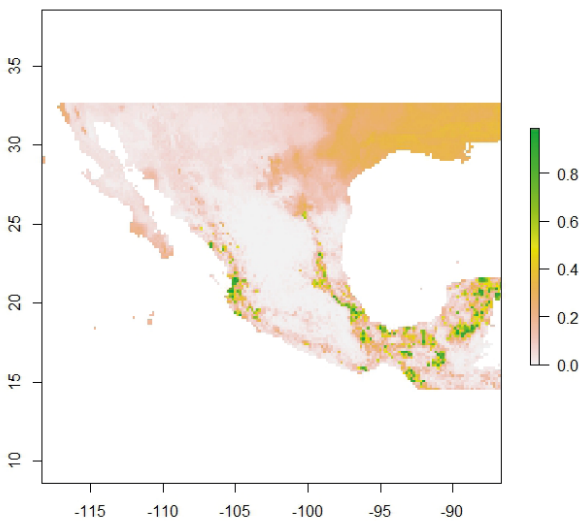

Figure 3: Original prediction of the model.

The importance of each of the bioclimatic variables used in the model is shown in Figure 5, with bio7 (annual temperature range), bio 18 (precipitation in the warmest quarter) and bio 12 (annual precipitation) being the variables with the greatest influence on the prediction.
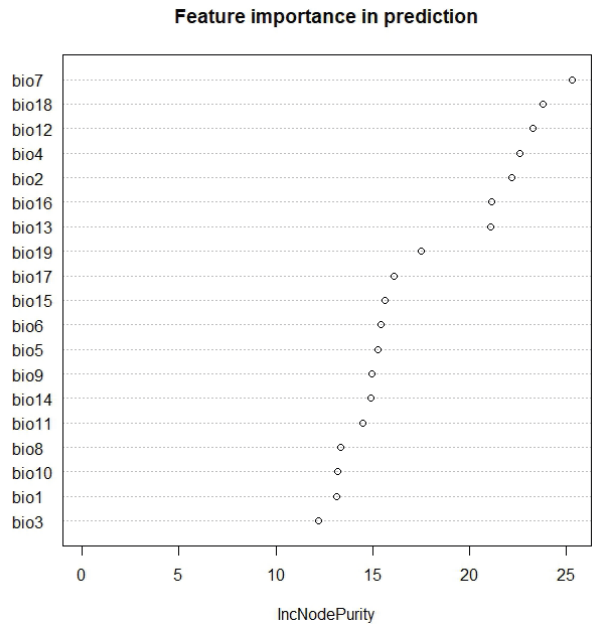

Figure 5: Importance of predictor variables considering the average decrease in node impurity.

The partial dependence of each of the variables used in the model is shown in Figure 6; in it, we can see the effect that each variable has marginally on the target variable, that is, the value of the prediction generated by the model.

For a better interpretation, we transformed the original continuous distribution map made with the model into a binary one, whose interpretation would be a presence-absence map with a threshold of .85 (Figure 7), which we considered appropriate to reduce false positives in the result; thus, the current prediction of suitable area for the presence of the species in Mexico is 19720 km^2.
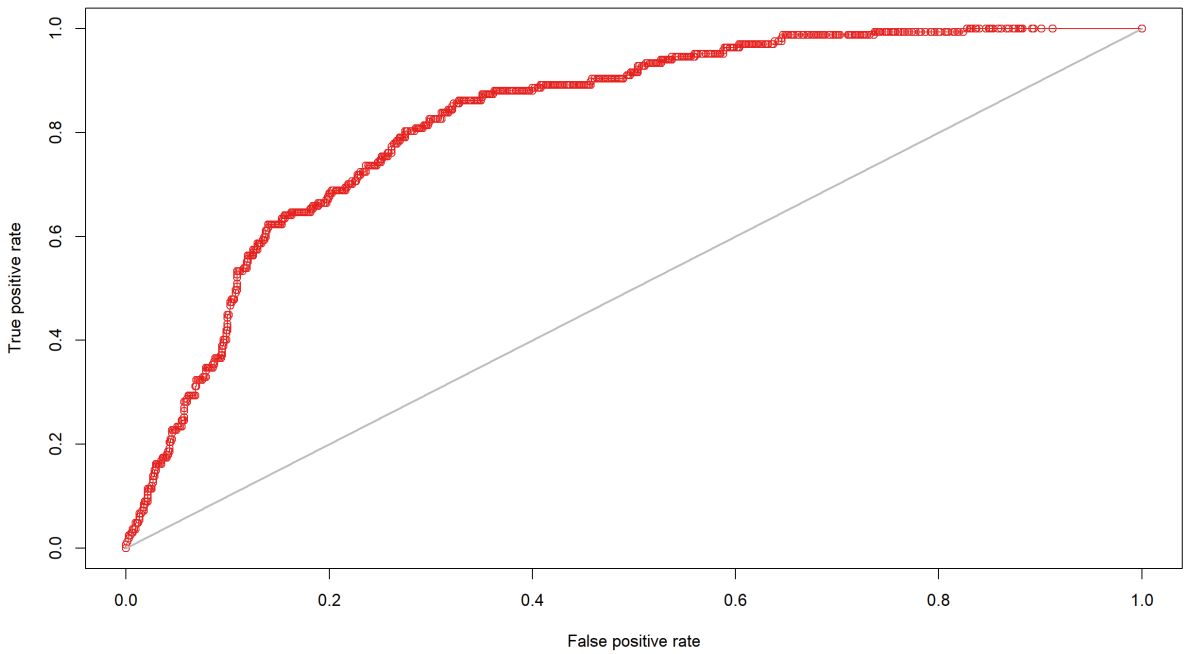
AUC= 0.825

Figure 4: ROC (receiver-operating characteristic) curve of the model; the area covered (AUC) reflects a model with good performance.
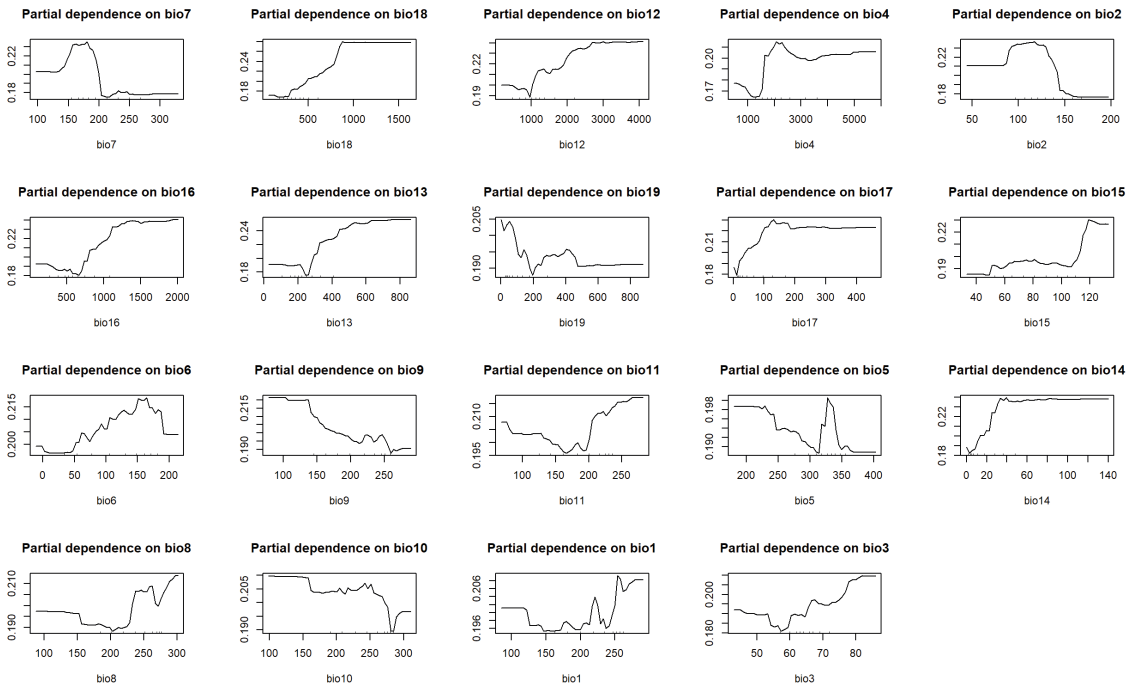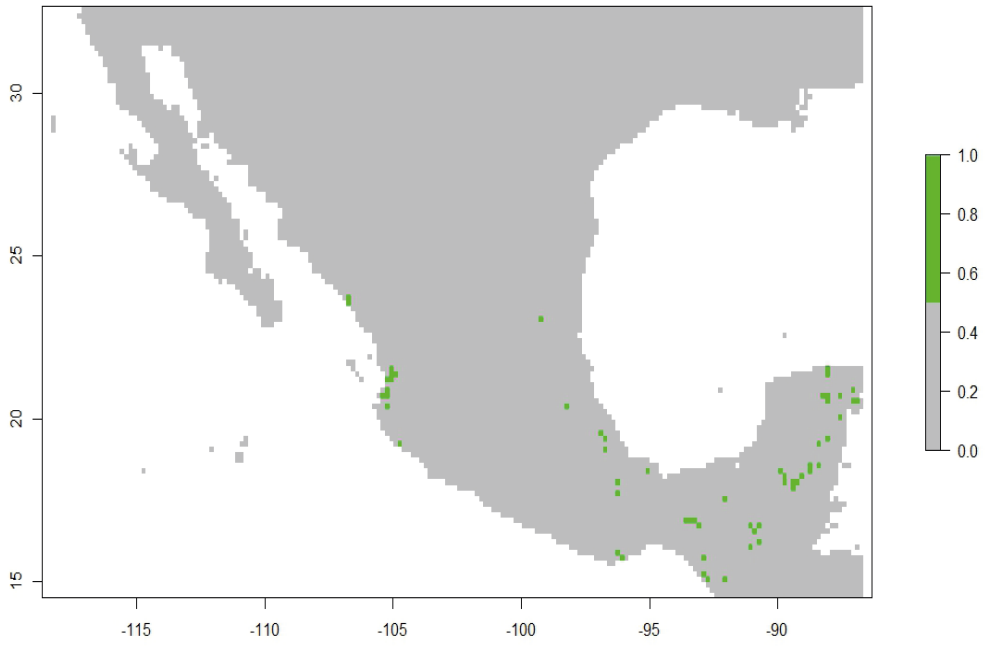

Figure 6: Partial dependency graphs.
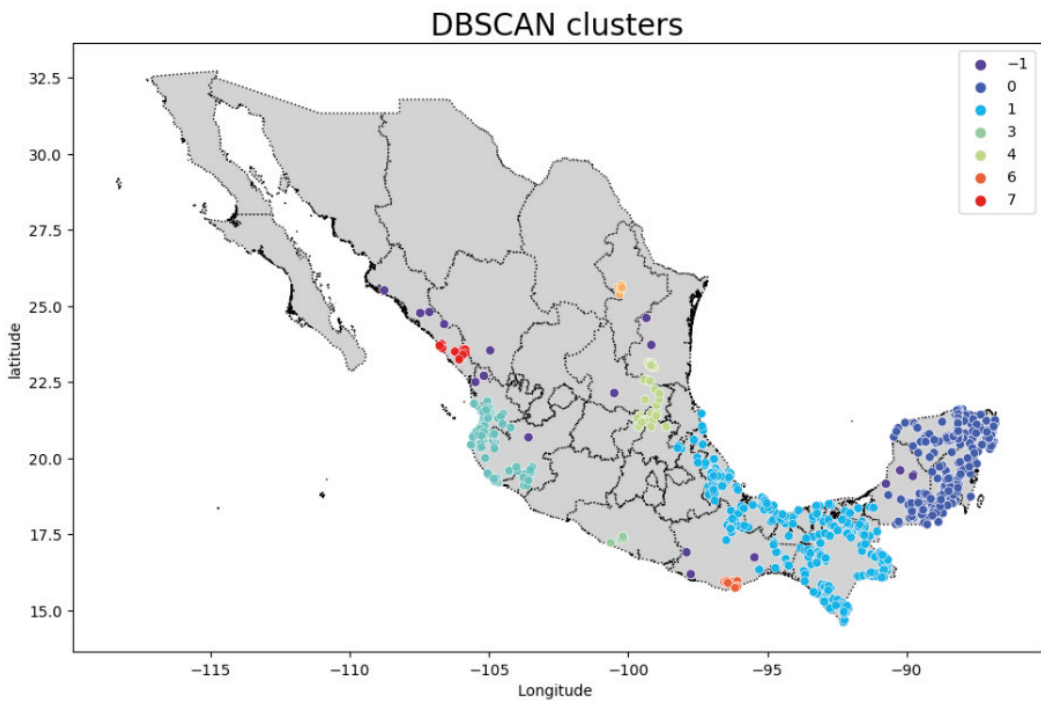
Figure 7. Binary raster map with pixels of 340 km^2.

Figure 8. Representation of the groups assigned by the DBSCAN algorithm for the records of the species in Mexico.

GROUPING INTO SUBPOPULATIONS

Regarding the grouping of records, the DBSCAN algorithm labeled eight groups (Figure 8). In purple and with a label of -1, the atypical data (or noise, in the context of the DBSCAN algorithm) are represented, which are records that did not meet the criterion of belonging to any group by proximity. The value of the silhouette coefficient is 0.344.

DISCUSSION

The area under the curve (AUC) value of 0.825 reflects that the model adequately captures the relationship between the presence of the species and the values of the climatic variables used as predictors (Figure 4); for most applications, values higher than 0.75 mean that the model has good performance and predicts considerably better than if the values were assigned randomly (Pearce and Ferrier, 2000; Mi et al. 2016). We can understand the selection of areas that the model proposed as highly suitable after making a joint analysis of the importance of variables in the prediction together with the partial dependence graphs; the most relevant predictor variables (annual temperature range, precipitation in the warmest quarter and annual precipitation, Figure 5) allow us to understand why the species is found mostly in areas of low altitude above sea level and mostly in the southeast of the country, maintaining a Neotropical distribution.

The annual temperature range (calculated as the maximum temperature of the warmest month minus the minimum temperature of the coldest month) corresponds to the highest prediction values between 16 and 19 °C; in the case of precipitation in the warmest quarter and annual precipitation, the highest prediction value is achieved from approximately 800 mm to 2750 mm.

The highest density of records for *F. rufigularis* is concentrated in the south and

southeast of the country, where we can identify two large populations (recognized as groups by the DBSCAN algorithm), one in the Gulf of Mexico, Tabasco and Chiapas (group one, figure 8), and one in the Yucatan Peninsula (group zero, figure 8). Both regions present high environmental suitability for the species, however, like many other tropical birds, bat falcons are sensitive to changes in the environment, and it is precisely in these regions where greater deforestation and land use change are recorded; although it has been seen that falcons have the capacity to adapt to new environments, isolated tropical forests that do not regenerate represent a temporary habitat that does not necessarily satisfy the ecological requirements of the species (Cade, 1982; Prieto-Torres et al. 2021). For example, Estrada-Contreras et al. (2015) modeled the potential effect of climate change on vegetation types in Veracruz, which is a region with a high density of records for this species; the results show a reduction of the evergreen tropical forest by 53% by 2050, being the most affected vegetation type, and precisely the one inhabited by the bat falcon.

In the West of the country, the most important population is found in Nayarit and Jalisco (group 2, figure 8), in an area with a mostly semi-warm sub-humid climate, so the tropical sub-deciduous and deciduous forest is the majority of the landscape; however, the area is also subject to strong pressures from direct or indirect development actions, such as land use change, unsustainable livestock farming and unplanned intensive tourism (Delgadillo and Cupul, 1999). At higher latitudes, records are rather isolated and do not represent considerably connected and abundant populations, which is supported by the environmental suitability model, so records obtained there that do not belong to a group can be interpreted as outliers (or noise, group -1 figure 8). The same applies to the first

and only record of this species in the United States, which was announced in February 2022, involving an individual observed in the Santa Ana National Wildlife Refuge in southern Texas (Fieldstadt, 2022). It is possible that the isolated records correspond mostly to juvenile individuals dispersing from their parents' territory and looking for new sites with suitability to establish themselves and not to a constant expansion of their distribution (Global Raptor Information Network, 2022). The geographic separation between the country's populations, particularly between those on the Atlantic and Pacific coasts, could result in biological divergence due to different ecological and environmental pressures, although complementary studies would be needed to identify and understand this process (Báez, 2019). In any case, the loss and fragmentation of the tropical forest affects their community of diurnal birds of prey, where a high dependence on the natural environment and a low capacity to use human matrices have generally been observed (Kattan et al. 1994; Renjifo, 2001; Thiollay 1989, 1996).

In contrast, raptors in temperate ecosystems do not seem to constitute a group particularly sensitive to habitat replacement and fragmentation, probably due to their capacity to adapt to and use agroecosystems (Filloy and Bellocq 2007; Rodríguez-Estrella et al. 1998). In particular, *F. rufigularis* has been reported to have a high sensitivity to deforestation and fragmentation, with a low capacity to cross human matrices (Cadena, 2012); Furthermore, like other birds of prey, it is especially sensitive to the indiscriminate use of pesticides such as DDT (Dichloro Diphenyl Trichloroethane), which could be used in regions of Veracruz, Tabasco and other parts of southeastern Mexico (regions of great importance in the distribution of the species in the country), and affects birds at a metabolic level resulting in thinner egg shells, increasing reproductive failure (Cade, 1982; Kiff et al., 1981). Due to its abundance and distribution in the Neotropical region of the country, as well as its sensitivity to disturbance, it is possible that this species is a good environmental indicator to know the impact of anthropogenic activities on other tropical birds of prey present in the region but due to lower population density and more elusive habits are difficult to study and monitor.

THANKS

REFERENCES

Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B. and Anderson, R.P. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography, 38: 541-545. https://doi.org/10.1111/ecog.01132.

Báez O. 2019. El origen de nuevas especies. Universidad Central del Ecuador. Vol.1. 377.

Berlanga, H., H. Gómez de Silva. V. M. Vargas-Canales, V. Rodríguez-Contreras. L. A. Sánchez-González, R. Ortega-Álvarez y R. Calderón-Parra. 2015. Aves de México: Lista actualizada de especies y nombres comunes. CONABIO. México D.F.

Bivand R, Lewin-Koh N (2022). _maptools: Tools for Handling Spatial Objects_. R package version 1.1-5, <https://CRAN.R-project.org/package=maptools>.

Breiman L. (2001). Random forests. Machine Learning 45(1):5-32.

Cade, T. J. 1982. The falcons of the world. Cornell Univ.Press. N.York.

Cadena, H., Carrión, O., Bahamonde, D. y Vries, T. 2012. Observaciones etológicas del Halcón Cazamurciélagos Falco rufigularis en Ecuador (Falconiformes: Falconidae). Avances, Vol. 4, No. 2 Pag. B12-B16.

Delgadillo, R. R. y Cupul-Magaña, F. G. 1999. Contribución al conocimiento de la flora de la Bahía de Banderas, Nayarit-Jalisco, México. Ciencia Ergo Sum, vol. 6, núm. 2.

eBird Basic Dataset. 2021. Version: EBD_relOct-2021. Cornell Lab of Ornithology, Ithaca, New York. Oct 2021.

Elith, J., Y Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics, 40(1), 677–697. doi:10.1146/annurev.ecolsys.110308.120159.

Ester, M., Kriegel Hans-Peter, Sander, J. y Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 226-231.

Estrada-Contreras, I., Equihua, M., Castillo-Campos, G., y Rojas-Soto, O. 2015. Climate change and effects on vegetation in Veracruz, Mexico: an approach using ecological niche modelling. Acta botánica mexicana, (112), 73-93.

Fieldstadt, E. l. 2022. Bat falcon spotted for the first time in the U.S. nbcnews. https://www.nbcnews.com/news/us-news/bat-falcon-spotted-first-time-us-rcna16691.

Filloy, J. y Bellocq, M. I. 2007. Respuesta de las aves rapaces al uso de la tierra: un enfoque regional. Hornero 22:131-140.

Global Raptor Information Network. 2022. Species account: Bat Falcon Falco rufigularis. Downloaded from http://www.globalraptors.org on 28 Apr. 2022.

Hijmans R. (2022). _raster: Geographic Data Analysis and Modeling_. R package version 3.6-11, <https://CRAN.R-project.org/package=raster>.

Hijmans RJ, Phillips S, Leathwick J, Elith J. (2022). _dismo: Species Distribution Modeling. R package version 1.3-9, <https://CRAN.R-project.org/package=dismo>.

IUCN. 2022. The IUCN Red List of Threatened Species. Version 2021-3. https://www.iucnredlist.org.

Kattan, G. H., Alvarez-López, H. y Giraldo, M. 1994. Forest fragmentation and bird extinctions: San Antonio eighty years later. Biological Conservation 8:138-146.

Kelling, S., Lagoze, C., Wong, W., Yu, J., Damoulas, T., Gerbracht, J., Fink, D. y Gomes, Carla. 2013. eBird: a Human/Computer Learning Network to Improve Biodiversity Conservation and Research. AI Magazine.

Kiff, L. F., Peakall, D. B. y Hector, D. P. 1981. Eggshell thinning and organochlorine residues in the Bat and Aplomado Falcons in Mexico. Proceedings of the International Ornithological Congress (Berlin). pp.949-952.

Liaw, A. and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

Liu, C., White, M., Y Newell, G. (2010). Measuring and comparing the accuracy of species distribution models with presence-absence data. Ecography, 34(2), 232–243. doi:10.1111/j.1600-0587.2010.06354.

Mi C, Huettmann F, Guo Y, Han X, Wen L. (2017). Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. PeerJ 5:e2849 .

Mi C., Falk H. Y Guo Y. (2016). Climate envelope predictions indicate an enlarged suitable wintering distribution for Great Bustards (Otis tarda dybowskii) in China for the 21st century. PeerJ 4:e1630.

Mindell, D. P., Fuchs, J. y Johnson, J. 2018. Phylogeny, Taxonomy, and Geographic Diversity of Diurnal Raptors: Falconiformes, Accipitriformes, and Cathartiformes. 10.1007/978-3-319-73745-4_1.

Pearce, J. Y Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling. 2000;133(3):225–245. doi: 10.1016/S0304-3800(00)00322-7.

Pearce, J. Y Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Modelo. 133: 225-245.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. y Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. JMLR 12, pp. 2825-2830.

Phillips, S.J., M. Dudik, J. Elith, C.H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19: 181-197.

Prieto-Torres, D. A., Sánchez-González, L. A., Ortiz-Ramírez, M. F., y Ramírez-Albores, J. E. 2021. Climate warming affects spatio-temporal biodiversity patterns of a highly vulnerable Neotropical avifauna. Climatic Change 165, 57.10.1007/s10584-021-03091-3.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Rahmah, N. y Sukaesih, I. 2016. Determination of optimal Epsilon (Eps) value on DBSCAN Algorithm to clustering data on Peatland Hotspots in Sumatra. IOP Conf. Ser.: Earth Environ. Sci. 31 012012.

Renjifo, L. M. 2001. Effect of natural and anthropogenic landscapes matrices on the abundance of subandean bird species. Ecological Applications 11:14-31.

Rodríguez-Estrella, R., Donázar, J. A e Hiraldo, F. 1998. Raptors as indicators of environmental change in the scrub habitat of Baja California sur, Mexico. Conservation Biology 12:921-925.

Rousseeuw, Peter. (1987). Rousseeuw, P.J.: Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Comput. Appl. Math. 20, 53-65. Journal of Computational and Applied Mathematics. 20. 53-65. 10.1016/0377-0427(87)90125-7.

Strimas-Mackey, M. Miller, E. y Hochachka, W. 2018. auk: eBird Data Extraction and Processing with AWK. R package version 0.3.0, https://cornelllabofornithology.github.io/auk/.

Strimas-Mackey, M., Hochachka, W. M, Ruiz-Gutierrez, V., Robinson O. J., Miller E. T., Auer T., Kelling S., Fink D. y Johnston A. 2020. Best Practices for Using eBird Data. Version 1.0. https://cornelllabofornithology.github.io/ebird-best-practices/. Cornell Lab of Ornithology, Ithaca, New York. https://doi.org/10.5281/zenodo.3620739

Sullivan, B. L., Phillips, T., Dayer, A. A., Wood, C. L., Farnsworth, A., Iliff, M. J. y Kelling, S. 2017. Using open access observational data for conservation action: A case study for birds. Biological Conservation, 208, 5–14.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., y Kelling, S. 2009. eBird: A citizen-based bird observation network in the biological sciences. Biological Conservation, 142(10), 2282–2292.

Thiollay, J. M. 1989. Area requirements for the conservation of rain forest raptors and game birds in French Guiana. Conservation Biology 3:128-137.

Thiollay, J. M. 1996. Distributional patterns of raptors along gradients in the northern Andes and effects of forest fragmentation. Journal of Tropical Ecology 12:535-560.

Van Rossum, G., y Drake Jr, F. L. 1995. Python tutorial. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

Zhang, L., Huettmann, F., Liu, S., Sun, P., Yu, Z., Zhang, X., Y Mi, C. (2019). Classification and regression with random forests as a standard method for presence-only data SDMs: A future conservation example using China tree species. Ecological Informatics. doi:10.1016/j.ecoinf.2019.05.003.