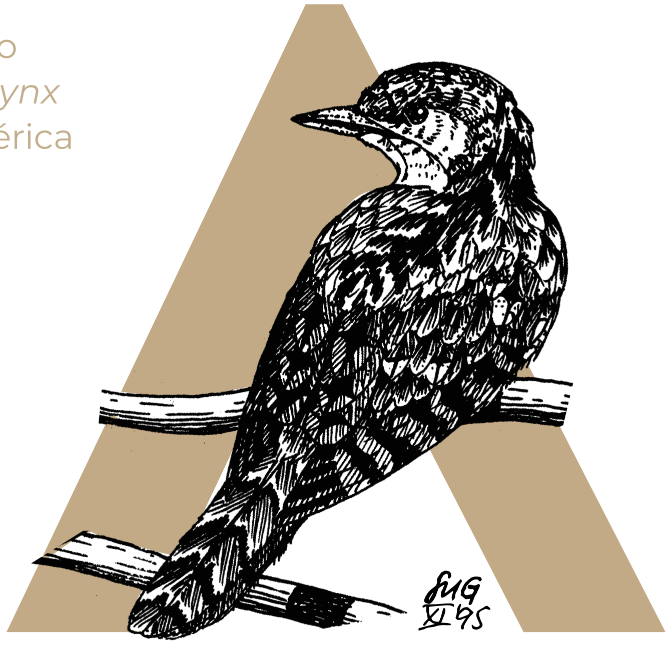
Modelling the winter distribution of the Eurasian Wryneck *(Jynx torquilla)* in Iberia

Modelação da distribuição de inverno do Torcicolo *(Jynx torquilla)* na Península Ibérica



Jesús Pinilla[1]

1 SEO/BirdLife, Delegación Territorial de Andalucía. Universidad Pablo de Olavide, Biblioteca – Despacho 25.1.11. Autovía A-376, km 1. 41013 Sevilla, Spain.

Corresponding author: jpinilla@seo.org

ABSTRACT

The Eurasian Wryneck (*Jynx torquilla*) has been considered an irregular and scarce winter species in Iberia which, together with its low detectability outside the breeding season, may have lead to inaccurate distribution patterns in the Portuguese and Spanish winter bird atlases. Citizen science programmes can offer large coverage data sets for improving the knowledge of understudied or elusive species. To analyze such data sets, species distribution modelling (SDM) is a very useful method that predicts species occurrences, using environmental data. In this paper, I tested different algorithms for modelling the winter distribution of the Wryneck in Iberia, using the records in the GBIF database together with climate factors, topography and land cover variables. The best performance was obtained with the "random forest" model that was subsequently used to determine which variables had the most predictive power. These variables were annual mean temperature, mean temperature of the coldest quarter, annual precipitation and altitude. Finally, with that model, I produced a map which predicts the main potential distribution areas for Wryneck in Iberia, which are the Southwestern quadrant, the main river basins, the coastal areas (especially the Mediterranean) and the Balearic Islands. These areas can be characterized as warmer locations, presumably with higher food availability.

Keywords: Citizen Science, Iberia, *Jynx torquilla*, SDM, winter distribution.

RESUMO

Durante o inverno, o Torcicolo (*Jynx torquilla*) tem sido considerado uma espécie irregular e escassa na Península Ibérica o que, juntamente com a sua baixa detectabilidade fora da época de reprodução, pode ter levado à identificação de padrões de distribuição imprecisos nos atlas das aves invernantes de Portugal e Espanha. Os projectos de ciência cidadã podem gerar grandes conjuntos de dados de cobertura que permitem melhorar o conhecimento de espécies pouco estudadas ou pouco conspícuas. Para tal, os modelos de distribuição de espécies (SDMs) são métodos muito úteis que permitem prever as ocorrências de espécies individuais usando dados ambientais. Neste artigo, testei diferentes algoritmos para modelar a distribuição de inverno do Torcicolo na Península Ibérica, usando os registos da base de dados GBIF juntamente com fatores climáticos, topografia e variáveis de cobertura do solo. O melhor desempenho foi obtido com o modelo "random forest" que foi posteriormente usado para determinar quais as variáveis com maior poder preditivo. Estas variáveis são a temperatura média anual, a temperatura média do trimestre mais frio, a precipitação anual e a altitude. Finalmente, com este modelo, produzi um mapa que prevê as principais áreas de distribuição de Torcicolo na Península Ibérica, e que são o quadrante sudoeste, as principais bacias hidrográficas, as zonas costeiras (especialmente o Mediterrâneo) e as Ilhas Baleares. Estas áreas podem ser caracterizadas como locais mais quentes, provavelmente com maior disponibilidade de alimento.

Palavras-chave: Ciência cidadã, Península Ibérica, *Jynx torquilla*, SDM, distribuição inverno.

Introduction

Bird atlases have become increasingly popular ways of documenting species' status and distributions since the first large-scale efforts were initiated in the 1960s (Gibbons et al. 2007). Expanding networks of amateur surveyors have enabled the completion of bird atlas projects covering geographic scales ranging from regions or countries, to continents (e.g., Hagemeijer & Blair 1997).

Both in the Spanish and Portuguese atlases of wintering birds (SEO/BirdLife 2012, Equipa Atlas 2018), fieldwork was carried out by volunteers and, since available resources were not enough to cover the whole territories, predictive methodologies were applied to produce distribution maps. These methods, known as species distribution models (SDMs), have become an essential tool in ecology and conservation (Guisan & Thuiller 2005, Elith & Leathwick 2009), and several techniques have been developed in order to predict (based in correlations) the occurrences of individual species using a varied array of environmental data (Zimmermann et al. 2010).

The Eurasian Wryneck (*Jynx torquilla*) is a migratory species whose European population winters mainly in sub-saharan Africa, but also in small numbers or irregularly in the Mediterranean basin and the Middle East (Cramp 1985). Wintering in Iberia has been traditionally considered scarce (only 22 records from 1981 to 1993) and irregularly distributed (Díaz et al. 1996), but this scarcity added to the low detectability outside the breeding season may have driven the production of maps that do not accurately reflect its actual distribution in the aforementioned atlases (Carbonell 2012, Reino 2018). However, incidental observation records, collected through other citizen science programs, can offer larger coverage both in space and time, useful for unveiling distribution patterns of understudied or elusive species (Bonney et al. 2014).

Thus, my goal in this study was to improve knowledge and obtain a more accurate picture of the actual winter distribution of the Eurasian Wryneck in the Iberian Peninsula and the Balearic Islands, building an SDM with the data collected through citizen science programs.

Methods

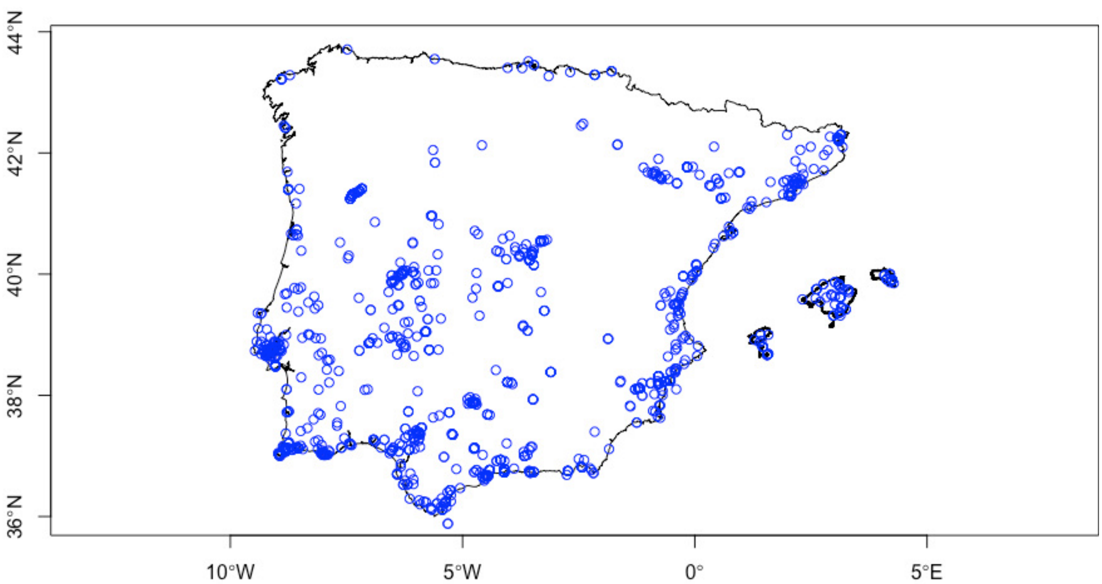
The Global Biodiversity Information Facility (GBIF) is the largest repository of species distribution data in the world and it also provides easy combination with SDMs (Beck et al. 2013). I requested all GBIF records of the Eurasian Wryneck in the Iberian Peninsula and Balearic Islands, recorded between November and February, from 1990 to 2020, since this three-decades period would allow obtaining a larger set of records and thus improving the coverage obtained in the atlases.

The query yielded a dataset of 1,480 records (Gbif.org 2020) but, in order to avoid pseudo-replication i.e. different records for the same bird, only one record was considered for each location on the same date. The final recordset comprised a total of 1,361 observations, as represented in Figure 1.

Environmental data, as distribution predictors, were obtained from two on-line sources: WorldClim.org (Fick & Hijmans 2017) and Copernicus.eu (Corine Land Cover 2012; EU-DEM 2016; Smets et al. 2018). Selected variables, presented in Table 1, included climate factors, topography, vegetation and land cover, that are considered the main drivers of bird distribution (adapted from Milanesi et al. 2017). The Normalized Difference Vegetation Index (NDVI), an index that describes the difference between visible and near-infrared reflectance of vegetation cover, can be used to monitor the canopy structure and phenology (Weier & Herring, 2000).

Figure 1- Point location of the winter records of Eurasian Wryneck (*Jynx torquilla*) in the Iberian Peninsula and Balearic Islands, stored in the GBIF database from 1990 to 2020.

Figura 1 - Localização dos registos de inverno de Torcicolo (*Jynx torquilla*) na Península Ibérica e nas Ilhas Baleares, armazenados na base de dados do GBIF de 1990 a 2020.



Corine Land Cover codes were re-coded in a sequential order for analitical purposes, and vegetation indexes were selected for the November-February period (see captions in Table 1).

The non independence of the selected predictor variables (collinearity) could be a problem because it inflates the variance of regression parameters and hence potentially leads to the wrong identification of relevant predictors. However, as long as the models are only used for interpolation, i.e. for predicting species occurrence under currently prevailing conditions, the selection of a strong correlate instead of the causal driver is of little concern, since similar outcomes can be expected (Braunisch et al. 2013).

For the modelling process I tested the most common algorithms: MaxEnt, radial basis function (RBF) and polynomial (P) Kernel, Generalized Linear Models (GLM), k-nearest neighbour (KNN), gradient boosting machine (GBM) and random forests (see Guisan et al. 2017). In order to evaluate the performance of each of these algorithms, I used the area under the curve (AUC) of the receiver operator characteristic (ROC) evaluation procedure (Hanley & McNeil 1982), the most widespread method for this purpose since it is largely independent of sampling prevalence when applied to presence-absence data (Manel et al. 2001).

Table 1 - Environmental variables used in the species distribution models (SDMs).

Tabela 1 - Selecção de variáveis ambientais utilizadas nos modelos de distribuição de espécies (SDMs).

TYPE	REFERENCE	VARIABLE	ABBREVIATION
Climate	Fick & Hijmans (2017)	Annual mean temperature	ann_mean_temp
		Minimum temperature of the coldest month	min_temp_cold_month
		Mean temperature of the coldest quarter	mean_temp_cold_qar
		Annual precipitation	ann_prec
		Mean precipitation of the wettest month	prec_wet_month
		Mean precipitation of the coldest quarter	prec_cold_qar
Topography	EU-DEM (2016)	Altitude	elevR
Land cover	Corine Land Cover (2012)	Land use (44 classes* with minimum mapping unit of 25 hectares)	usosR
Vegetation	Smets et al. (2018)	Normalized Difference Vegetation Index**	ndvi

* CLC class descriptions can be found at https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/html/.

** NDVI has been calculated as the mean values from November to February along the 1999-2017 period.

* As descrições das classes CLC podem ser encontradas em https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/html/.

** O NDVI foi calculado como os valores médios de novembro a fevereiro ao longo do período de 1999-2017.

The best model was used to evaluate the importance of each variable used in the model. As a complementary analysis, I tested de impact of individual predictors on the distribution model, plotting the response of the model to the variation of each environmental variable. Finally, I implemented the model with the highest fit on the stack of predictors, in order to create a map of winter habitat suitability for the species in the Iberian Peninsula and the Balearic Islands.

Data analysis and map generation were carried out using R software (R Core Team 2020), and the R packages dismo (Hijmans et al. 2017), caret (Kuhn 2020), pROC (Robin et al. 2011), gbm (Greenwell et al. 2019), randomForest (Liaw & Wiener 2002) and maptools (Bivand & Lewin-Koh 2019). For those algorithms based on presence-absence points, I divided the presence data into training (75%) and validation (25%) sets, and performed a 10-fold cross-validation for model selection (see Arlot & Celisse 2010).

Results

As shown in Table 2, every technique applied produced AUC values over 0.8, considered as acceptable results according to Swets (1988). Nevertheless, the highest accuracy is obtained with the "random forest" predictive model. This model shows that annual mean temperature, mean temperature of the coldest quarter, annual precipitation and elevarion are the most important variables for predicting the probability of presence of the species in the specified area and period (Table 3).

Table 2 - Performance of the applied modelling techniques according to the area under the curve (AUC) evaluation procedure.

Tabela 2 - Desempenho das técnicas de modelação aplicadas de acordo com o procedimento de avaliação da área sob a curva (AUC).

ALGORITHM	AUC
GLM	0.8672
KNN	0.8702
P Kernel	0.8709
RBF Kernel	0.8730
MaxEnt	0.8739
GBM	0.9246
RandomForest	0.9414

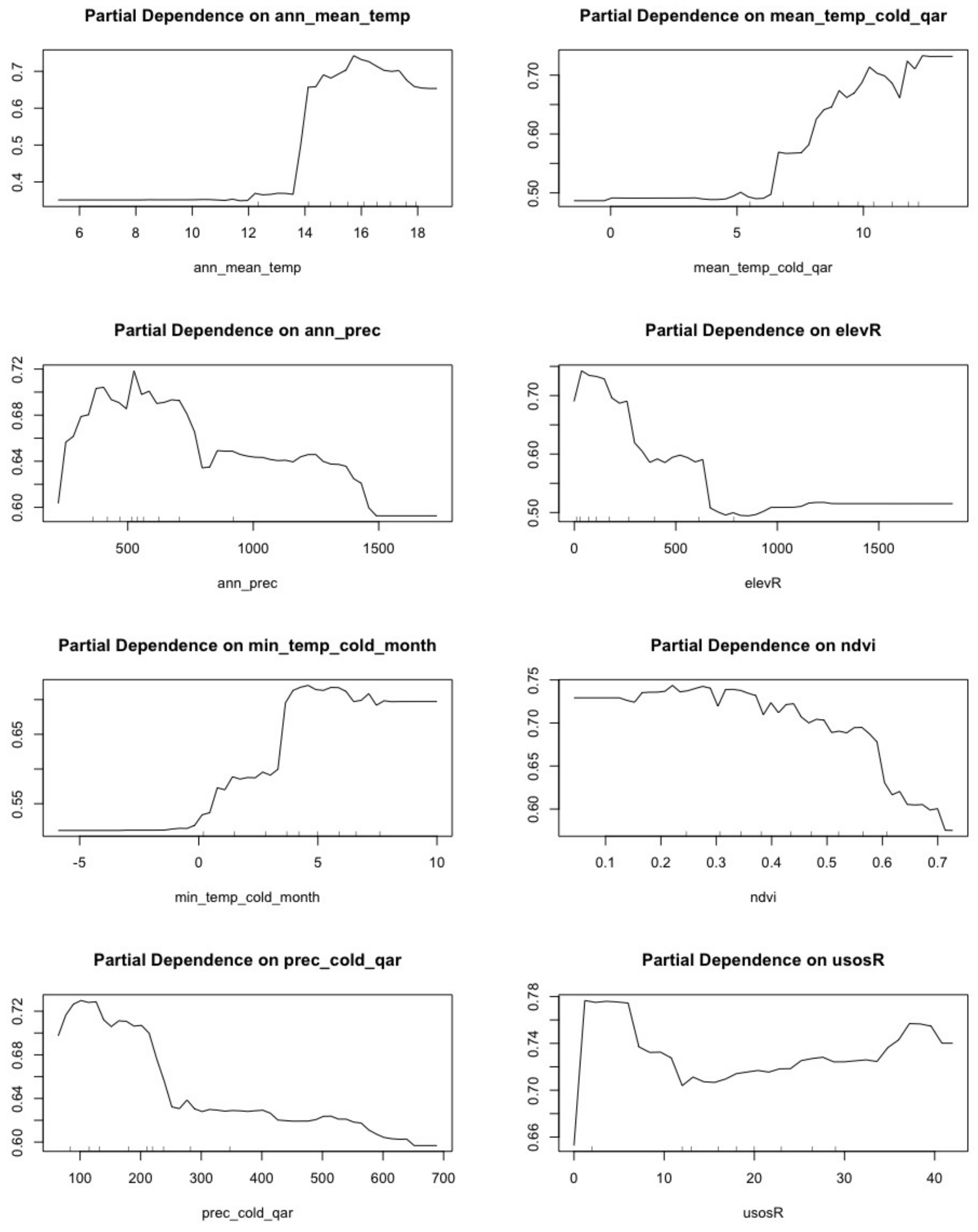
Table 3 - Variable importance in the "random forest" model. The measure of importance is scaled to have a maximum value of 100.

Tabela 3 - Importância das variáveis no modelo "random forest". A medida de importância foi dimensionada para ter um valor máximo de 100.

VARIABLE	OVERALL
Ann_mean_temp	100.00
mean_temp_cold_qar	54.03
ann_prec	46.44
elevR	46.14
min_temp_cold_month	39.84
ndvi	28.58
prec_cold_qar	24.94
usosR	18.64
prec_wet_month	0.00

Figure 2- Impact of individual predictors on the "random forest" distribution model. Values of the impact (y axis), ranging from 0 to 1, show how each predictor (x axis) influence the habitat suitability and species potential presence if the rest of the variables remained constant. Temperature variables are depicted in degrees Celsius (ºC), altitude in meters (m), and rainfall in millimeters (mm). The Corine Land Cover 44 classes codes have been re-coded in a sequential order, from 0 (no data or sea waters) to 43 (the original order can be consulted in https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/html/).

Figura 2 - Impacto dos preditores individuais no modelo de distribuição "random forest". Os valores do impacto (eixo y), variam entre 0 e 1, e demostram como cada preditor (eixo x) influencia a adequabilidade do habitat e a presença potencial da espécie se o resto das variáveis permaneceram constantes. As variáveis de temperatura são representadas em graus Celsius (ºC), altitude em metros (m) e precipitação em milímetros (mm). Os códigos de classes do Corine Land Cover 44 foram recodificados em ordem sequencial, de 0 (sem dados ou águas do mar) a 43 (a ordem original pode ser consultada em https://land.copernicus.eu/user-corner/technical -library / corine-land-cover-nomenclature-guidelines / html /).



The results also indicate that the Wryneck winter habitat suitability reaches its highest values in areas with annual mean temperatures over 14 °C, mean temperatures of the coldest quarter over 10 °C, annual precipitation between 300 and 700 mm, and altitudes below 300 m a.s.l. Also, habitat suitability is predicted to be slightly higher with minimum temperatures of the coldest month over 4 °C, NDVI values below 0.4, precipitations in the coldest quarter below 200 mm and land cover categories 2 to 7, corresponding to artificial surfaces (Figure 2).

Finally, I implemented the "random forest" model on the stack of predictors, in order to create a map of winter habitat suitability for the species in the Iberian Peninsula and the Balearic Islands (Figure 3). This map predicts that the main potential distribution areas are the Southwestern quadrant, the main river basins, the coastal areas (especially the Mediterranean) and the Balearic Islands.

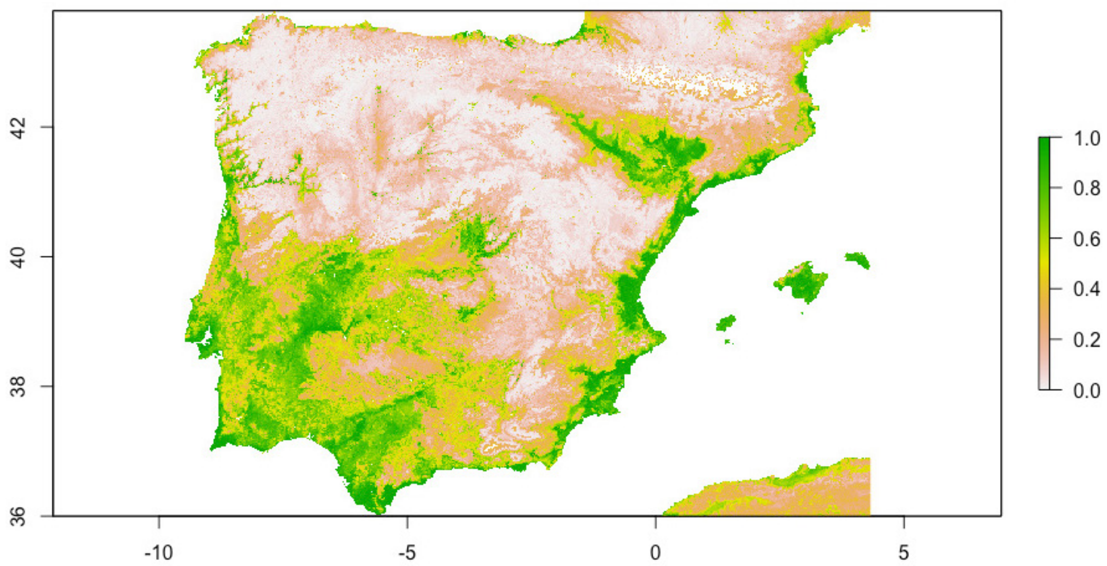
Discussion

Traditionally, SDMs predict the geographic distribution of suitable climatic space for a species by relating species occurrence records to long-term average climate variables. Such models are generally a good representation of a species' broad range (Pearson & Dawson 2003). This approach has nonetheless its drawbacks, related to the variation experienced in the environmental conditions along the considered period of time and consequently the species distribution itself. However, they are still useful for modelling general distribution patterns (Kearney & Porter 2009).

The results presented here suggest that temperature, precipitation and altitude are the most important factors for predicting the Wryneck winter occurrence in Iberia, and it seems to rely much less on land cover. This could be explained because, outside the breeding season, the species occurs in a wide variety of habitats,

Figure 3 - Predictive map of the winter presence of the Eurasian Wryneck (*Jynx torquilla*) in Iberia using the random forest model. Values range from 0 (where the modelled presence probability is lowest) to 1 (highest probability of presence).

Figura 3 - Mapa preditivo da presença no inverno do Torcicolo (*Jynx torquilla*) na Península Ibérica usando o modelo "random forest". Os valores variam entre 0 (onde a probabilidade de presença modelada é a menor) e 1 (probabilidade de presença máxima).



as long as they provide enough food (mainly ants and other invertebrates~) which are more dependent on higher temperatures that ensure their activity and availability (Santos & Tellería 1985, Mermod et al. 2009). Nevertheless, the vegetation variable used (NDVI) does not provide detailed information on the understorey of wooded areas and the land cover data, based on categories, only provide a static picture. Thus, the current analyses do not exclude the possibility that variations in, for exampleplethe scrub cover, may be a driving factor of distribution, as it has been suggested for the breeding season (e.g. Gordinho 2008). Finally, it must be also pointed out that citizen science data are usually spatially biased (Johnston et al. 2020), so the results here presented may well be affected by this circumstance. All these limitations, however, do not seem to produce unexpected results, although they may actually have some predictive value that should be tested in future studies.

Acknowledgement

Javier Seoane, Luís de Oliveira Gordinho and an anonymous referee provided insightful comments and directions that helped improve this paper.

References

Arlot, S. & Celisse, A. 2010. A survey of cross-validation procedures for model selection. Statistics Surveys 4: 40-79.

Beck, J., Ballesteros Mejia, L., Nagel, P. & Kitching, I.J. 2013. Online solutions and the 'Wallacean shortfall': what does GBIF contribute to our knowledge of species' ranges? Diversity Distribution 19: 1043-1050.

Bivand, R. & Lewin-Koh, N. 2019. maptools: Tools for Handling Spatial Objects. R package version 0.9-9. https://CRAN.R-project.org/package=maptools

Bonney, R., Shirk, J.L., Phillips, T.B., Wiggins, A. & Ballard, H.L. 2014. Next steps for citizen science. Science 343: 1436-1437.

Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H. & Bollmann, K. 2013. Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. Ecography 36: 971–983.

Carbonell, R. 2012. Torcecuello euroasiático Jynx torquilla. In: SEO/BirdLife 2012: Atlas de las aves en invierno en España 2007-2010, pp 350-351. Ministerio de Agricultura, Alimentación y Medio Ambiente – SEO/BirdLife, Madrid.

Corine Land Cover 2012. Version 2020_20u1, http://land.copernicus.eu/pan-european/corine-land-cover/clc-2012/view

Cramp, S. 1985. Handbook of the birds of Europe, the Middle East and North Africa. Vol. 4: Terns to woodpeckers. Oxford University Press, Oxford, UK.

Díaz, M., Asensio, B. & Tellería, J.L. 1996. Aves ibéricas I. No paseriformes. Reyero, Madrid.

Elith, J. & Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution and Systematics 40: 677–697.

Equipa Atlas 2018. Atlas das Aves Invernantes e Migradoras de Portugal 2011-2013. Sociedade Portuguesa para o Estudo das Aves, LabOr- Laboratório de Ornitologia – ICAAM - Universidade de Évora, Instituto da Conservação da Natureza e das Florestas, Instituto das Florestas e Conservação da Natureza (Madeira), Secretaria Regional da Energia, Ambiente e Turismo (Açores) e Associação Portuguesa de Anilhadores de Aves. Lisboa.

EU-DEM 2016 European Digital Elevation Model, version 1.1, http://land.copernicus.eu/pan-european/satellite-derived-products/eu-dem/eu-dem-v1.1/view

Fick, S.E. & Hijmans, R.J. 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37 (12): 4302-4315.

GBIF.org 3 April 2020 GBIF Occurrence Download https://doi.org/10.15468/dl.wmt8tx.

Gibbons, D. W., Donald, P. F., Bauer, H.G., Fornasari, L., & Dawson, I. K. 2007. Mapping avian distributions: The evolution of bird atlases. Bird Study 54: 324–334.

Gordinho L 2008 Torcicolo Jynx torquilla. p. 304 in Equipa Atlas. Atlas das Aves Nidificantes em Portugal (1999-2005). Instituto da Conservação da Natureza e da Biodiversidade, Sociedade Portuguesa para o Estudo das Aves, Parque Natural da Madeira e Secretaria Regional do Ambiente e do Mar dos Açores. Assírio & Alvim, Lisboa.

Greenwell B, Boehmke B, Cunningham J and GBM Developers 2019. gbm: Generalized Boosted Regression Models. R package version 2.1.5. https://CRAN.R-project.org/package=gbm

Guisan, A. & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993–1009.

Guisan, A., Thuiller, W. & Zimmermann, N.E. 2017. Habitat Suitability and Distribution Models. Cambridge University Press. Cambirdge, UK.

Hagemeijer, W. J. M., & Blair, M. J. (Eds.) 1997. The EBCC atlas of European Breeding Birds: Their distribution and abundance. London: T & A Poyser.

Hanley, J.A. & McNeil, B.J. 1982. The Meaning and Use of the Area under a Receive Operating Characteristic (ROC) Curve. Radiology 143: 29-36.

Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. 2017. dismo: Species Distribution Modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo

Johnston, A., Moran, N., Musgrove, A., Fink, D. & Baillie, S.R. 2020. Estimating species distributions from spatially biased citizen science data. Ecological Modelling 422, 108927. https://doi.org/10.1016/j.ecolmodel.2019.108927

Kearney, M. & Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecology Letters 12: 334–350.

Kuhn, M. 2020. caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret

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

Manel, S., Williams, H.C. & Ormerod, S. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38: 921-931.

Mermod, M., Reichlin, T. S. & Arlettaz, R. 2009. The importance of ant-rich habitats for the persistence of the Wryneck Jynx torquilla on farmland. Ibis 151: 731-742.

Milanesi, P., Herrando, S., Pla, M., Villero, D. & Keller, V. 2017. Towards continental bird distribution models: environmental variables for the second European breeding bird atlas and identification of priorities for further surveys. Vogelwelt 137: 53–60.

Pearson, R.G. & Dawson, T.P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology & Biogeography 12: 361–371.

R Core Team 2020. R: a Language and Environment for Statistical Computing. R foundation for statistical computing, Vienna, Austria. URL https://www.R-project.org/.

Reino, L. 2018. Torcicolo Jynx torquilla. In: Equipa Atlas 2018. Atlas das Aves Invernantes e Migradoras de Portugal 2011-2013, pp 366-367. Sociedade Portuguesa para o Estudo das Aves, LabOr- Laboratório de Ornitologia – ICAAM - Universidade de Évora, Instituto da Conservação da Natureza e das Florestas, Instituto das Florestas e Conservação da Natureza (Madeira), Secretaria Regional da Energia, Ambiente e Turismo (Açores) e Associação Portuguesa de Anilhadores de Aves. Lisboa.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F, Sanchez, J.C. & Müller, M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12: 77. DOI: 10.1186/1471-2105-12-77 http://www.biomedcentral.com/1471-2105/12/77/

Santos, T. & Tellería, J.L. 1985. Patrones generales de la distribucion invernal de passeriformes en la península Ibérica. Ardeola 32: 17-30.

SEO/BirdLife 2012. Atlas de las aves en invierno en España 2007-2010. Ministerio de Agricultura, Alimentación y Medio Ambiente – SEO/BirdLife, Madrid.

Smets, B., Jacobs, J., Swinnen, E., Toté, C. & Wolfs, D. 2018. Gio Global Land Component - Lot I, "Operation of the Global Land Component", Framework Service Contract N° 388533 (JRC), Product User Manual, Normalized Difference Vegetation Index (NDVI), Collection 1 km. Version 2.2, Issue I2.31. Copernicus Service Information (www.copernicus.eu).

Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. Science, 240: 1285–1293.

Tryjanowski, P., Sparks, T.H., Biaduń, W., Brauze, T., Hetmański, T., Martyka, R. et al. 2015 Winter Bird Assemblages in Rural and Urban Environments: A National Survey. PLoS ONE 10(6): e0130299. https://doi.org/10.1371/journal.pone.0130299

Weier, J. & Herring, D. 2000. Measuring Vegetation (NDVI & EVI). NASA Earth Observatory, Washington DC.

Zimmermann, N. E., Edwards Jr., T. C., Graham, C. H., Pearman, P. B. & Svenning, J. C. 2010. New trends in species distribution modelling. Ecography 33: 985–989.