Assessing the uncertainty arising from standard land-cover mapping procedures when modelling species distributions

Miguel Cánibe | Nicolas Titeux | Jesús Domínguez | Adrián Regos

Abstract
Aim: Habitat loss and degradation is one of the main threats to biodiversity worldwide. Ecological modelling community has been increasingly incorporating habitat changes in the species distribution models (SDMs). However, the effect that the uncertainty arising from the standard procedures of land-use/land-cover (LULC) mapping can have on SDMs has been overlooked. Remote sensing offers a great source of information to assess the habitat changes at different spatiotemporal scales. However, despite the great variety of satellite imagery preprocessing and supervised classification procedures currently available, their effects on SDMs remain largely unknown.

Location: Baixa Limia–Serra do Xurés Natural Park, NW Spain.

Methods: We developed SDMs for 27 bird species with 4 modelling algorithms using 12 sets of the LULC predictors derived from the combination of 3 image preprocessing levels and 4 supervised classification algorithms.

Results: The preprocessing level, classification method and modelling algorithm had a significant impact on the SDM performance and on the spatial and temporal predictions, with the considerable variation between species. Even if the effects of satellite image processing and classification procedures on the performance and transferability of SDMs were lower than that of the modelling algorithms themselves, the spatial and temporal predictions of SDMs were overall strongly affected by these LULC mapping procedures, sometimes overriding the effect of modelling algorithms.

Main conclusions: We argue that the uncertainty arising from the standard LULC mapping procedures when predicting species distribution dynamics should receive the same level of attention as the variability due to the different types of climate simulations and models often used in the ecological research. We also highlight the need to report accuracy of the satellite-based LULC maps for each LULC class since some of these classes can be of key importance for certain types of the ecological applications.

Keywords: ecological niche modelling, land-use/land-cover mapping, model performance, remote sensing, species distribution models, transferability
Species distribution models (SDMs) are a widely used tool to assess and predict the impacts of the environmental changes on species distributions (Brotons, 2014). These models relate the occurrence of species (i.e. their geographical distribution) to environmental variables describing their habitat, in the case of correlative SDMs (Guisan & Zimmermann, 2000), or to the demographical and physiological processes, in the case of process-based SDMs (Briscoe et al., 2019). It has been proven that the SDMs are effective tools with a wide spectrum of applications such as management of invasive species (Kolby, 2014; Pearson, 2015), discovery of new populations and new species (Alizpurua et al., 2015; Blair et al., 2013; Pearson et al., 2007; Raxworthy et al., 2007) or the identification of important areas for conservation (Kremen et al., 2008). The use of SDMs to predict the effects of environmental change on species distributions has, however, some limitations. One of the most important drawbacks is the uncertainty of SDM predictions when extrapolated beyond the range of environmental conditions used for their calibration (i.e. SDM transferability; see review in Werkowska et al., 2017). To address this limitation, we need in-depth assessment of the sources of uncertainty affecting model predictions. Another option to overcome the challenge that transferability presents is process-based SDMs, and these kind of models implement demographical and physiological processes in the prediction of species. Several types of process-based models exist, including different processes and each with their own limitations (Briscoe et al., 2019), although they might be better than correlative SDMs when predicting species distributions outside of the calibration context (Higgins et al., 2020). However, process-based SDMs imply an increased difficulty to the user than correlative SDMs and require higher processing power and population data that is difficult to obtain. These requirements constrain their accessibility, so they remain underused when compared with correlative SDMs, although this might change in coming years, thanks to increasing data availability and computing power.

The modelling process also involves various sources of uncertainty, which can affect the predictions, such as the sampling design (Tessarolo et al., 2014), scale of the analysis, selection of environmental variables or modelling algorithm (Wiens et al., 2009). In the particular case of the climate change impact assessment, the scenario and climate model used to represent climatic conditions were shown to considerably affect the SDM predictions (Thuiller et al., 2019). A common way to manage this source of variability is to make predictions using various modelling algorithms, sets of variables or RCPs/GCMs (Buisson et al., 2010) and then exploring the range of variation in the predictions. An alternative option is to integrate these predictions through ensemble models that account for this variability and uncertainty (Araújo & New, 2007).

Another important source of variability is the choice of environmental variables used as predictors (Petitpierre et al., 2017). The model assumes that variables included as predictors in the model represent all the ecological requirements that determine the geographical distribution of the target species (Wiens et al., 2009). Although it is not possible to include all the variables that determine this distribution (Scherrer & Guisan, 2019), most SDMs are based solely on the climatic variables—mostly because of their increasing accessibility (Titeux et al., 2016). Over the last years, the number of SDMs calibrated with both climate and land-cover predictors has increased, although not much attention is paid to the uncertainty associated with the standard procedures applied to develop such predictor variables (but see Bede–Fazekas & Somodi, 2020; Lembrechts et al., 2019). The recent advances in remote sensing and the increasing availability of Earth observation data have made possible to explore the potential of new predictor variables in the SDMs (He et al., 2015; Randin et al., 2020). This new type of predictors allows for a better description of the ecological niche of the species, since they provide detailed information that is closely related to the conditions experienced by organisms at the ground level (Amiri et al., 2020). This better characterization of the ecological niche of the species usually comes with an increase in the predictive ability of the models when used at regional or local scales in combination with the climatic predictors (Regos et al., 2019; Tuanmu et al., 2011).

Satellite imagery is a source of information at global scale, constantly updated and usually free of charge, making it possible to obtain standardized environmental information for the last 40 years (e.g. from NASA’s Landsat mission) (Wulder et al., 2016). Satellite images store information in various spectral bands, each of them encoding data for reflectance in a given wavelength (U.S.G.S., 2019). To work with Landsat imagery, a series of preprocessing steps are required, which depend on the kind of product selected. The first correction applied to the satellite image is the geometric correction, which involves the steps of georeferencing (alignment to the correct geographical location) followed by orthorectification (which accounts for the effects of view direction and terrain relief) (Young et al., 2017). Landsat level-1 products are already terrain-corrected and usually do not require a geometric correction by the user (USGS, 2020). After geometric correction, a series of processes (depending on the needs for the desired analysis) known as absolute radiometric corrections are applied. The purpose of the absolute radiometric correction is to obtain the true reflectance values that make data comparable across images from the same or different sensors that have gone through the same level of correction (Young et al., 2017). In contrast to absolute radiometric corrections, relative radiometric corrections can be applied; these corrections make data across different satellite images comparable by bringing values from the input images to the same scale as a designated reference image. The first absolute correction usually applied is the conversion of the digital numbers (DNs) back to radiance, also known as sensor calibration. Landsat sensors collect data on radiance, which is stored in the form of DNs as unsigned integers. These DNs can be converted back to radiance using the rescaling factors stored in the metadata of the image, which are specific for each band of each sensor (Goslee, 2011). The next absolute correction is the solar correction, which accounts for the influence of the sun on pixel values and converts at sensor radiance to top of the atmosphere (TOA) reflectance. These values can also be retrieved from the metadata of the image and included parameters such as exoatmospheric solar irradiance,
Earth–Sun distance and solar elevation angle, which vary with date, time and latitude. After solar correction, the data will be in the form of TOA reflectance (USGS, 2021a,b). To convert TOA reflectance to surface reflectance, another preprocessing step known as atmospheric correction is required (Goslee, 2011). Atmospheric correction allows to correct for the effects of atmospheric conditions at the time the image was taken. Depending on the desired application for the satellite imagery, a given preprocessing level might be good enough, while following the subsequent preprocessing steps might help to correct additional errors or artefacts in the data. Nevertheless, selecting the appropriate preprocessing level can become a difficult task for certain applications, especially for the nonexpert users (Young et al., 2017).

Land-use/land-cover (LULC) maps can then be derived from preprocessed satellite imagery by means of image classification, which can be unsupervised or supervised. In unsupervised image classification, the clustering-based algorithms sort the spectral image into a number of spectral classes based exclusively on the statistical information contained within the image, and the user is then responsible for giving a meaning to the resulting classes (Lu & Weng, 2007). Meanwhile, supervised classification incorporates training samples from reference data, which can be from the image itself or from other comparable images; here, the user is responsible for previously defining the spectral signature of land-cover classes (Bakker et al., 2013). Whether the classification method is supervised or unsupervised, different algorithms exist for the process, and depending on the classifier selected, the resulting classification can vary significantly (Briem et al., 2002; Hubert-Moy et al., 2001; Regos & Domínguez, 2018). Ensemble procedures can be also applied to the LULC maps, where a consensus classification can be made to account for the variation between different classification algorithms. LULC maps are finally used to generate environmental predictors that will feed into the SDMs. Therefore, the whole chain of LULC mapping initiatives is affected by several sources of variability, and this may have a potential impact on the SDMs that are based on these products. This issue remains, however, largely unexplored as there is a lack of studies assessing the uncertainties in SDM predictions resulting from the different preprocessing levels and image classification procedures.

In this study, our aim is to assess the uncertainty in SDM predictions arising from the standard procedures applied for LULC mapping—namely the satellite imagery preprocessing levels and supervised classification procedures—relative to other well-known sources of variation in the SDM performance and predictions such as the modelling technique.

2 | METHODS

2.1 | Study area

The study was conducted in the Baixa Limia–Serra do Xurés Natural Park (29,762 ha) located in the Galician province of Ourense (NW Spain; Figure 1). The natural park is included in the Gerês-Xurés transboundary UNESCO Biosphere Reserve. The study area is also protected by European legislation as it is included in the Natura 2000 network. The study area is a topographically diverse mountainous region, with an altitude ranging from 323 to 1529 m above the sea level. The main LULC classes are shrubland and sparsely vegetated areas (69%), forest areas (24%) and, in lesser extent, farmlands (Regos et al., 2015). Bioclimatically, the area is located in the transition zone between the Mediterranean and the Eurosiberian regions (Martínez Cortizas & Pérez Alberti, 1999). The area is strongly affected by land abandonment and wildfires, making it a highly dynamic and unstable system.

2.2 | Conceptual framework

With the aim of assessing the impacts of the image preprocessing levels and supervised classification procedures on SDMs, we generated a range of LULC maps derived from Landsat satellite images based on three different preprocessing levels and four classification methods. From the LULC maps, we then generated several environmental variables and used them to describe habitat conditions for 27 birds in SDMs. We used species associated with various ecological traits, and we applied four frequently used modelling algorithms, thereby accounting for the well-known sources of uncertainty in SDMs. To evaluate the SDM performance and transferability, we validated our models with a traditional cross-validation method (i.e. partitioning available data into one set for training and another set for validation) and against two independent datasets: (1) one dataset obtained from the fieldwork carried out in the same year of the data used for model calibration but with a different methodology and at different locations within the study area (i.e. spatially independent dataset) and (2) another dataset derived from fieldwork that has been previously carried out by the same observers 10 years ago (i.e. temporally independent dataset). These independent datasets allowed us to measure the transferability of the SDMs and thus to evaluate their predictive capacity within and beyond the range of environmental conditions used to calibrate them. In addition, the SDMs were used to predict the habitat change for birds over the 10-year period based on the changes in the LULC maps and to evaluate the effect of different procedures for LULC mapping on the spatial and temporal predictions of the SDMs.

2.3 | Satellite imagery and preprocessing

Satellite images used in this study were captured by Landsat 5 TM (Thematic Mapper) for years 2000 (September 4) and 2010 (July 30 and October 18) and Landsat 7 ETM+ (Enhanced Thematic Mapper) for year 2000 (June 24) and were obtained from the USGS (United States Geological Survey, available at https://earthexplorer.usgs.gov/). We selected Landsat collection 1 level-1 imagery with the purpose of being able to compare the different preprocessing levels. Collection 1 level-1 images are already geometrically corrected
with a consistent georegistration and image-to-image tolerances of ≤12 m radial root mean square error (RMSE) (USGS, 2020). Three versions of each Landsat image were computed from raw Landsat imagery using standard preprocessing procedures:

1. Nonpreprocessed images, i.e. using the DN (digital number) values, hereafter “DN”.
2. Images absolutely corrected by applying sensor calibration and solar correction using the apparent reflectance method (Goslee, 2011), available from the ‘RStoolbox’ R package (Leutner et al., 2019), which converts the DN values to TOA reflectance using the parameters included in the image metadata, hereafter “TOA”.
3. This images incorporate sensor calibration and solar correction as well as atmospheric correction following the Dark Object Subtraction methodology (Chavez, 1996) (also available in the ‘RStoolbox’ package), which assumes that the radiometric minimum for the image is located in pixels completely covered by shadow or underwater. The method assumes that the reflectance in those pixels is exclusively due to the atmospheric effects and uses these values to correct the whole image, converting the values from TOA reflectance to surface reflectance, hereafter “sRef”.

Due to the revisit time (16 days) of Landsat imagery, many images have gaps introduced by clouds and their shadows (Zhu & Woodcock, 2012). Cloud removal algorithms are available in many software, including the ‘RStoolbox’ R package (Leutner et al., 2019). However, we deemed this step unnecessary since cloud-free images were available for the periods of interest for this study. All images have a spatial resolution of 30 m.

2.4 | Supervised classification

To generate the LULC maps from each image, we applied different supervised classification methods. This step implies the creation of training and validation areas that were created for both years with the ArcMap software (version 10.5; ESRI). Six LULC classes were identified in our study site: (i) water bodies, including rivers, lakes and dams; (ii) deciduous forests, mainly riparian forests and oak
evaluation was important because the cross-validation procedures of our models with independent observations allowed us to test model transferability (i.e. the ability of a model to predict species distribution beyond the range of data used for calibration). This evaluation was important because the cross-validation procedures tend to overestimate model predictive capacity (Araújo & Guisan, 2006).

Censuses were performed during breeding season (May–June), always during the first 4 hours after sunrise, which is a period of peak vocal activity for birds, and always under suitable weather conditions (i.e. days without marked rainfall or wind). In the calibration data, the point counts were resampled during the 4 h before sunset, which is another period of peak vocal activity.

2.6 | Species distribution models

For each of the 24 LULC maps (obtained from 2 years, 3 levels of image preprocessing and 4 supervised classification algorithms; see Figure 2 for the workflow), we calculated the percentage of occupied area for each of the six LULC classes at 200-m spatial resolution, resulting in 12 sets of 6 environmental predictors for each year. Elevation, slope and aspect were added to every dataset and calculated from the digital elevation model ASTER Global DEM v2 (ASTER Science Team, 2009) using the ‘raster’ package in R (Hijmans, 2021). All layers were aggregated from the original 30-m to a 200-m resolution to match the grain size at which birds were sampled (point counts with ~80-m distance; roughly 160-m spatial resolution). To do so, each LULC class was separated in a different raster layer at the original 30-m resolution, with values of 100 if a grid cell had the corresponding coverage and a value of 0; otherwise, these layers were then resampled from a 30-m to a 200-m resolution by using the bilinear interpolation method available in the ‘raster’ package in R (Hijmans, 2021), allowing us to estimate percentage coverage of each LULC class in every 200-m grid cell. After this process, the predictor dataset used to calibrate and validate the model contained the following information at 200-m resolution grid-cell level: percentage area occupied by (1) water, (2) deciduous forest, (3) evergreen forest, (4) shrubland, (5) cropland and grassland and (6) sparse vegetation; plus (7) altitude, (8) slope and (9) aspect.

Species distribution models were developed for each of the 27 bird species using separately the 12 sets of predictors for the year 2010 (i.e. based on 3 levels of preprocessing and 4 supervised classification algorithms). Models were calibrated using the ‘biomod2’ (Thuiller et al., 2020) R package using four widely used algorithms available in the package: artificial neural networks (ANN) (Rammer & Seidl, 2019), gradient boosting model (GBM) (Elith et al., 2008), random forest (RF) (Prasad et al., 2006) and generalized linear model (GLM) (Guisan et al., 2002). For each algorithm, we used the default settings (see Appendix S2 for detailed modelling parameters) since the packages is optimized for the elaboration of the SDMs. Calibration and validation point counts for the cross-validation procedure were randomly selected 20 times always maintaining a 70:30 ratio. The 20 cross-validation runs were performed with each algorithm using these randomly selected points.

We then tested the predictive capacity of the models for each cross-validation run with the nonindependent validation data using the ‘Biomod2’ R package. We also assessed the transferability of the

2.5 | Bird data

We used three bird occurrence (presence/absence) datasets for 27 bird species surveyed in years 2000 and 2010, available from Regos et al. (2019) (Figure 1; see Appendix S1). The first dataset consists of 384 20-min point counts where birds were surveyed within a limited distance of 80 m (roughly 160-m spatial resolution) twice during the spring of 2010; this dataset was used for model calibration and standard cross-validation procedures (hereafter: ‘Calibration data’). In addition, two independent datasets were used to test model transferability: (1) a dataset of 344 5-min point counts with unlimited distance (roughly 200-m spatial resolution) carried out in year 2010 in locations that differed from the calibration data (i.e. spatially independent dataset; hereafter: ‘Splnd’ dataset); and (2) a second dataset obtained from fieldwork carried out by the same observers in 204 5-min point counts in the year 2000 (i.e. spatially and temporally independent dataset; hereafter: ‘SpTemplnd’ dataset). The validation of our models with independent observations allowed us to test model transferability (i.e. the ability of a model to predict species distribution beyond the range of data used for calibration). This evaluation was important because the cross-validation procedures
models resulting from each run by testing their predictive capacity with the spatially independent dataset ‘SpInd’ and with the spatially and temporarily independent dataset ‘SpTempInd’ (Werkowska et al., 2017).

We used AUC-ROC (area under the curve for the receiver-operating characteristic) for the subsequent statistical analysis, together with TSS (true skill statistic) and Cohen’s Kappa statistic to assess model performance and transferability (see Appendix S1).

2.7 | Effects of LULC mapping on model predictive capacity and predictions

To explore the effects of the satellite image preprocessing levels, supervised classification procedures and modelling algorithms on SDM predictive capacity (i.e. AUC values), we fitted linear mixed models (LMM) for each species in which the SDM cross-validation run and validation dataset were incorporated as random effects (i.e. as grouping factors, see Bolker et al., 2009; Harrison et al., 2018). The image preprocessing level, the supervised classification procedure and the modelling algorithm were incorporated as fixed effects.

We calculated marginal $R^2$ (which accounts for variability explained only by fixed factors) for all models allowing us to estimate the relative importance of these factors on AUC variability. We also report significative differences between factor levels for $p$-values lower than .05. To disentangle the amount of variability explained by each of the fixed factors, we ran separate models including only one of the fixed factors at a time, thus getting an $R^2$ value accounting for the variability explained by that factor. All models were developed using the ‘glimmTMB’ R package (Bolker, 2019; Brooks et al., 2017). Likelihood ratio tests were also performed by comparing the full models to restricted models where either modelling algorithm, preprocessing level or classification algorithm were not included in the models, thus allowing us to know if the full model performed significatively better than the restricted models and estimate the importance of the removed factor.

To test the impact of the preprocessing level, supervised classification algorithm and modelling algorithm on model predictions, we conducted a pixel-level analysis using the raw-predicted values (Table S4.5). We sampled 100 random pixels from each individual projection, gathering around 5 million data entries, on which we then performed 100 GLMMs (with a Poisson link function) on 1000 random data entries each. Due to the extremely large database and the unreliable results coming from both the GLMMs and the subsequent likelihood ratio tests, we decided to evaluate the model predictions on the relative change in suitable area. To do so, we obtained the difference in suitable area provided by the model binary predictions for each year, then divided by the suitable area for the year 2000 and multiplied by 100 to give a percentage value. For the LMMs, we used the same fixed and random effects as in LMMs for the AUC values, except for ‘validation dataset,’ which was not applicable in this case.
3 | RESULTS

3.1 | Impact of preprocessing and classification procedures on LULC mapping

For the year 2010, LULC maps with the higher accuracies (up to 85%) were obtained using the RF algorithm for the three-image preprocessing level (Figure 3), while the other algorithms had lower accuracies. For the year 2000, LULC maps with the higher accuracies (up to 91%) were obtained using the GBM algorithm for the three levels of preprocessing (Figure 3).

Considering sensitivity and positive predictive power (PPP) for each LULC class separately, most of them were mapped with a good accuracy (Appendix S3), although certain LULC classes showed some problems. In the year 2000, sensibility and PPP values for the LULC class ‘Croplands and Grasslands’ were highly variable across the different maps, resulting in poor accuracy under some classification methods such as neural networks (Appendix S3). In the year 2010, classifiers had difficulties to correctly classify the two forest classes (i.e. ‘Evergreen Forest’ and ‘Deciduous Forest’) resulting in values for sensibility lower than 0.5 in both classes. These differences in class-specific accuracies result in some level of variation in the estimates of occupied surface for each LULC class, especially for the classes that showed the lower accuracy. This underestimation or overestimation of the estimated occupied area for the different LULC classes generated large differences on the change analysis when comparing the different maps generated for both years (Figure 4).

3.2 | SDM predictive capacity and transferability

Our results showed that 37.6% of the SDMs had an AUC value higher than 0.7 (mean AUC of 0.668 ± 0.098; 0.766 ± 0.05 for acceptable models only) when tested over the remaining 30% of data used for calibration (i.e. cross-validation). However, the number of models considered acceptable (i.e. AUC ≥ 0.7) decreased up to 28.7% when tested against the spatially independent dataset ‘SpInd’ (mean AUC of 0.644 ± 0.089; 0.749 ± 0.04 for acceptable models only) and up to 17.5% (mean AUC of 0.601 ± 0.098; 0.743 ± 0.037 for acceptable models only) when validated with the spatially and temporally independent dataset ‘SpTempInd’. Our results also showed a large variability in the model predictive capacity depending on the modelled species (Appendix S1).

3.3 | Influence of different factors on model predictive capacity

Large differences in the variability of AUC values were explained by fixed effects (Marginal R-square) across the different modelled species, ranging from 0.01 to 0.417 (Figure 5; Table S4.1). LMMs including preprocessing levels only performed significatively better than LMMs missing this factor for 8 of the 27 species (Table S4.3), and variability explained by this factor was very low (i.e. lower than 1% on all species). Supervised classification algorithm significatively improved LMMs on 22 of the 27 modelled species (Table S4.3), but variability explained was also very low, although higher than for the preprocessing. On the other hand, modelling algorithm significantly improved models for all modelled species (Table S4.3) and captured most of the variability in AUC values explained by the models.

3.4 | Influence of different factors on habitat change predictions

The likelihood ratio tests for LMMs on the relative change of suitable area showed significative improvements of LMMs including image preprocessing levels for all species, classification algorithm for 26 of them, and the modelling algorithms for 25 species (Table S4.4). Large differences in the amount of variability explained by fixed factors were found among modelled species, ranging from 0.112 to 0.505 (see $R^2$ for the full model in Table S4.2). The factor explaining the largest part of the variability depended on the modelled species, and in most cases, it was either the modelling algorithm (mean $R^2 = 0.089 ± 0.019$ with values ranging from 0.003 to 0.469) or the classification algorithm (mean $R^2 = 0.089 ± 0.014$ with values ranging from 0.003 to 0.258). Preprocessing level was the most important factor for 4 species, had a mean $R^2$ of 0.057 and ranged from 0.014 to 0.193 (Table S4.2; Figure 5).

4 | DISCUSSION

The present study compares, for the first time, the impact that standard procedures for developing LULC maps from Landsat imagery can have on SDM accuracy and predictions across species with different life-history traits. Here, we provide insight on how different satellite
imagery preprocessing levels and supervised classification procedures that significantly influence LULC mapping (Paolini et al., 2006; Regos & Domínguez, 2018) propagate uncertainty into correlative SDMs outputs when these maps are used to predict the geographical distribution of bird species.

Regarding SDM predictive capacity, all models performed better than random (AUC > 0.5), and most of them showed acceptable or good predictive capacity. The ability of the SDMs to predict species distributions across space (spatial transferability) was higher for most species than their ability to predict distributions in another spatiotemporal context (spatiotemporal transferability), even though our spatially independent dataset is not totally independent since data were collected in the same study area (although at different points and using a different methodology). A possible explanation for the low transferability is that our models were solely based on topographic and LULC information while part of the change in species distribution over the last 10 years could also be due to climate change—which was not accounted for in our models. Results showed that the different preprocessing levels of the satellite images together with the subsequent image classification procedures had low, although significant effects on the predictive capacity and transferability of the SDMs (see Tables S4.1 and S4.3; Appendix S5).

Regarding the modelling algorithm used to build the SDM, we found highly significant effects on the predictive capacity and transferability (AUC values) and on predicted relative change in suitable area for the bird species. These results are in line with the general consensus among the ecological modelling community that the algorithm used to develop SDMs is one of the main sources of variability in model outcomes, as it has been previously reported by other studies (Carvalho et al., 2011; Thuiller et al., 2019). In our analysis, SDM algorithm was found to have a much greater overall effect on modelling predictive capacity than the preprocessing level and supervised classification algorithms. Nevertheless, the effects of the preprocessing levels and supervised classification algorithms on AUC values were significant (although very low) for many of the modelled species (see Tables S4.1 and S4.3). However, the effects of preprocessing levels and supervised classification algorithms were much higher when considering model predictions, as happened for the relative change in suitable area. For model predictions, the modelling algorithms and the supervised classification algorithms were
the main factors driving variations in estimated change in suitable area, being the most important factor for 11 and 12 species, respectively, while preprocessing level generally had a lower influence and was the most important factor for 4 species (see Table S4.2). These results support the idea of Grimmett et al. (2020) that SDMs should be evaluated using not only performance metrics, but also their predictions, especially when testing different methods or environmental datasets for developing models.

Our results therefore demonstrate that the standard procedures of LULC mapping have significant effects on the spatial and temporal predictions of SDMs (i.e. on the predicted relative change in suitable area for the species between two time periods; see Appendix S6). SDM predictions were significantly affected by preprocessing and supervised classification for almost all modelled species (Tables S4.2 and S4.4). We argue that this source of variability in LULC maps production should be accounted for when using such maps to extract environmental variables that are fed into SDMs to predict the distribution of species. Using a single LULC map to build SDMs would therefore neglect the range of uncertainty around the whole production chain of such LULC information. A possible approach to deal with this variability is through the use of ensemble approaches (Du et al., 2012), by combining the information provided by different image classifiers and thus better reflecting the range of outcomes of the different classification procedures (Lu & Weng, 2007). Another way to account for this variability would be to test the individual performance of different algorithms and to select the best one(s) from the specific perspective of the ecological modelling application. Algorithm selection is an important issue since the performance of an algorithm can depend on each specific LULC class (Ge et al., 2020; Henriques et al., 2010), and the different LULC classes are not similarly important for the development of relevant SDMs. Algorithm selection should therefore be applied from the perspective of both remote sensing (i.e. algorithm performance) and ecological applications (i.e. depending on the environmental variables that best represent the habitat requirements of the modelled species). Unfortunately, most of the SDM studies integrating LULC information do not pay much attention to this issue and instead use readily available maps without carefully considering the methods used for their production and without analysing whether some specific LULC classes are documented in a reliable way, which can strongly impact the model predictive capacity and predictions for the target species (Fernandes et al., 2019; Fournier et al., 2017; Kiss et al., 2020; Maiorano et al., 2019).

Apart from the sources of variation considered in this study (i.e. modelling algorithm, preprocessing level and supervised classification algorithm), there are other sources of variation common in remote sensing not covered in this research, such as methods for gap filling of Landsat 7 imagery collected after the SCL failure (Chen et al., 2011), for cloud masking (Zhu & Woodcock, 2012) or for data standardization when using data from different sensors, among other possibilities. However, the implementation of these steps usually require a high degree of expertise in the field of remote sensing, so ecological modellers will often gravitate towards tools that simplify some of these steps like the ‘RStudio’ package in R (Leutner et al., 2019).

Throughout the analysis, we found considerable differences among species in the factors driving the variation on modelling predictive capacity and relative change in suitable area. We argue that this variation could be attributed to species traits, for example, habitat preference, habitat specialization, biogeographical origin or phenology. Several studies have reported that the ecological traits affect the predictive capacity and predictions of SDMs based on climate and/or LULC variables (Dobrowski et al., 2011; Pöyry et al., 2008; Regos et al., 2019; Wogan, 2016). In this study, species like the Eurasian skylark (Alauda arvensis), the common stonechat (Saxicola torquata) or the Iberian chiffchaff (Phylloscopus ibericus) were among the species whose model predictions were most affected by supervised classification algorithms (see Table S4.2 and Appendix S4). These three species are considered habitat specialists (Keller et al., 2020), although the Iberian chiffchaff prefers forest habitats, while the other two prefer more open habitats. We argue that uncertainty in the classification of the LULC classes that broadly represent species’ habitat conditions can greatly affect model prediction. The
LULC class can be overpredicted or underpredicted in the study area, which will then translate into an overestimation or underestimation of the suitable area for the species, thus generating a great variability in their predicted range of distribution. From an ecological modelling perspective, this sheds light on the importance of considering (and reporting) not only the overall accuracy of the LULC mapping method applied, but also the mapping accuracy for each individual LULC class. Our results emphasize the need for considering the quality of remote sensing products at the LULC class level when using such data for representing habitat conditions for species in distribution models. Based on this detailed information, ecological modellers will be in a better position to evaluate and compare the relevance of different options for LULC mapping and select the ones that best suit their needs.

In conclusion, this study shows for the first time, to our knowledge, that standard procedures for land-use/land-cover (LULC) mapping can affect the predictive capacity of species distribution models (SDMs) and their ability to be transferred in space and over time. Even if satellite image processing and classification procedures had a lower effect on the predictive capacity and transferability of SDMs than the algorithms used to model the distribution of the species, SDM predictions were strongly affected by these standard LULC mapping procedures. When selecting environmental variables from available remote sensing products and using them as predictors in such models, ecological modellers should therefore carefully consider the variability arising from the satellite image preprocessing levels and classification procedures, either by using ensemble techniques to reflect this range of uncertainties or by selecting the most appropriate procedures. Our results also show that, to guide such a methodological choice, it is key to evaluate not only the overall accuracy of the mapping procedures but also their ability to represent adequately particular LULC classes; some classes are indeed more important than others when building SDMs to predict the distributions of species across space and time—especially for species with a high degree of habitat specialization. This study is a first glance at how standard LULC mapping procedures affect the SDM outputs, but more in-depth analyses are now required to learn more about this issue and exploit the full potential of remote sensing-based environmental predictors in SDMs. We call for the collaboration between remote sensing specialists and ecological modellers to better identify the challenges and opportunities to build ecologically relevant LULC products for ecological applications.

**ACKNOWLEDGEMENTS**

This work received funding from Xunta de Galicia through the grant to structure and consolidate the competitive research groups of Galicia (ED431B 2018/36). This work was supported by PORBIOTA—E-Infraestrutura Portuguesa de Informação e Investigação em Biodiversidade (POCI-01-0145-FEDER-022127). A.R. and M.C. were funded by the Xunta de Galicia, Spain (postdoctoral fellowship ED481B2016/084-0). A.R is currently supported by Juan de la Cierva fellowship program funded by the Spanish Ministry of Science and Innovation (IJC2019-041033-I).

**CONFLICT OF INTEREST**

The authors declare no conflict of interest that might be perceived as influencing the author's objectivity with this work.

**PEER REVIEW**

The peer review history for this article is available at https://publo ns.com/publon/10.1111/ddi.13456.

**DATA AVAILABILITY STATEMENT**

Bird data available from supplementary material of the article entitled “Effects of species traits and environmental predictors on performance and transferability of ecological niche models” https://www.nature.com/articles/s41598-019-40766-5. Landsat data obtained from the USGS (United States Geological Survey, available at https://earthexplorer.usgs.gov/). R-Scripts and relevant data layers available on https://github.com/CanibeMiguel/LULC_SDM_Uncertainty.git

**ORCID**

Miguel Cánibe https://orcid.org/0000-0001-8196-9048
Nicolas Titeux https://orcid.org/0000-0002-7719-7417
Jesus Dominguez https://orcid.org/0000-0002-8669-022X
Adrián Regos https://orcid.org/0000-0003-1983-936X

**REFERENCES**

Aizpurua, O., Cantú-Salazar, L., San Martin, G., Biver, G., Brotons, L., & Titeux, N. (2015). Reconciling expert judgement and habitat suitability models as tools for guiding sampling of threatened species. *Journal of Applied Ecology, 52*(6), 1608–1618. https://doi.org/10.1111/1365-2664.12515
Amiri, M., Tarkesh, M., Jafari, R., & Jetschke, G. (2020). Bioclimatic variables from precipitation and temperature records vs. remote sensing based bioclimatic variables: Which side can perform better in species distribution modeling? *Ecological Informatics, 57*(January), 101060. https://doi.org/10.1016/j.ecoinf.2019.101060
Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography, 33*(10), 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x
Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. Trends in Ecology and Evolution, 22(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010
ASTER Science Team (2009). ASTER global digital elevation model. U.S. Department of the Interior | U.S. Geological Survey. https://doi.org/10.5067/ASTER/ASTGTM.002
Bakker, W. H., Feringa, W., Gieske, A. S. M., Gorte, B. G. H., Grabmaier, K. A., Hecker, C. A., Horn, J. A., Huurneman, G. C., Janssen, L. L. F., Kerle, N., Van Der Meer, F. D., Parodi, G. N., Pohl, C., Reeves, C. V., Van Ruitenbeek, F. J., Schetselaar, E. M., Tempfli, K., Weir, M. J. C., Westinga, E., & Woldai, T. (2013). Principles of remote sensing: an introductory textbook. In K. Tempfli, N. Kerle, G. C. Huurneman & L. L. F. Janssen (Eds.), *ITC Educational Textbook series* (4th ed., pp. 280–314). The International Institute for Geo-Information Science and Earth Observation (ITC).
Bede-Fazekas, Á., & Somodi, I. (2020). The way bioclimatic variables are calculated has impact on potential distribution models. *Methods in Ecology and Evolution, 11*(12), 1559–1570. https://doi.org/10.1111/2041-210X.13488
Blair, M. E., Sterling, E. J., Dusch, M., Raxworthy, C. J., & Pearson, R. G. (2013). Ecological divergence and specification between lemur
Brotons, L. (2014). Species distribution models and impact factor growth. *Ecology and Evolution*, 4(12), 2473–2483. https://doi.org/10.1002/ece3.1701

Buisson, L., Thuiller, W., Casajus, N., Lek, S., & Grenouillet, G. (2015). Drivers of species range dynamics with process-explicit models: matching methods to applications. *Ecology Letters*, 18(11), 1315–1327. https://doi.org/10.1111/ele.12522

Carvalho, S. B., Brito, J. C., Crespo, E. G., Watts, M. E., & Possingham, H. P. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation* (Vol. 1(1), 1–4. https://doi.org/10.1002/2015RS000012

Chavez, P. S. (1996). Image-Based Atmospheric Corrections - Revisited and Improved. *Photogrammetric Engineering & Remote Sensing*, 62(9), 1025–1036. https://doi.org/10.1137/1.9781572891001

Chen, J., Zhu, X., Vogelmann, J. E., Gao, F., & Jin, S. (2011). A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. *Remote Sensing of Environment*, 115(4), 1053–1064. https://doi.org/10.1016/j.rse.2010.12.010

Dobrowski, S. Z., Thorne, H. J., Greenberg, J. A., Safford, H. D., Mynsberge, A. R., Crimmins, S. M., & Swanson, A. K. (2011). Modeling plant ranges over 75 years of climate change in California, USA: Temporal transferability and species traits. *Ecological Monographs*, 81(2), 241–257. https://doi.org/10.1890/10-1325.1

Du, P., Xia, J., Zhang, W., Tan, K., Liu, Y., & Liu, S. (2012). Multiple classifier system for remote sensing image classification: A review. *Sensors*, 12(4), 4764–4792. https://doi.org/10.3390/s120404764

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Fernandes, R. F., Horrondo, J. P., Guisan, A., Roxo, A., Alves, P., Martins, J., & Vicente, J. R. (2019). Species distribution models support the need of international cooperation towards successful management of plant invasions. *Journal for Nature Conservation*, 49, 85–94. https://doi.org/10.1016/j.jfc.2019.04.001

Fournier, A., Barbet-Massin, M., Rome, Q., & Courchamp, F. (2017). Predicting species distribution combining multi-scale drivers. *Global Ecology and Conservation*, 12, 215–226. https://doi.org/10.1016/j.gecco.2017.11.002

Ge, G., Shi, Z., Zhu, Y., Yang, X., & Hao, Y. (2020). Land use/cover classification in an arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms. *Global Ecology and Conservation*, 22, e00971. https://doi.org/10.1002/gecco.2020.e00971

Goslee, S. C. (2011). Analyzing remote sensing data in R: The landsat package. *Journal of Statistical Software*, 43(4), 1–25.

Grimmett, L., Whithed, R., & Horta, A. (2020). Presence-only species distribution models are sensitive to sample prevalence: Evaluating models using spatial prediction stability and accuracy metrics. *Ecological Modelling*, 431(January), 109194. https://doi.org/10.1016/j.ecolmodel.2020.109194

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157(2–3), 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 2018(5), 1–32. https://doi.org/10.7717/peerj.4794

He, K. S., Bradley, B. A., Cord, A. F., Rocchini, D., Tuanmu, M. N., Schmidtlein, S., Turner, W., Wegmann, M., & Pettorelli, N. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation* (Vol. 1(1), 4–18. https://doi.org/10.1002/2015RS000012

Henriques, A. P. M., Dória Neto, A. D., & Amaral, R. F. (2010). Classification of multispectral images in coral environments using a hybrid of classifier ensembles. *Neurocomputing*, 73(7–9), 1256–1264. https://doi.org/10.1016/j.neucom.2010.01.003

Higgins, S. I., Larcombe, M. J., Beeton, N. J., Conradi, T., & Nottebrock, H. (2020). Predictive ability of a process-based versus a correlative species distribution model. *Ecology and Evolution*, 10(20), 11043–11054. https://doi.org/10.1002/ece3.6712

Hijmans, R. J. (2021). raster: Geographic Data Analysis and Modeling. R package version 3.4.. https://CRAN.R-project.org/package=raster

Hubert-Moy, L., Cottenec, A., Le Du, L., Chardin, A., & Perez, P. (2001). A comparison of parametric classification procedures of remotely sensed data applied on different landscape units. *Remote Sensing of Environment*, 75(2), 174–187. https://doi.org/10.1016/S0304-4265(00)00165-6

Keller, V., Herrando, S., Voříšek, P., Franch, M., Kipson, M., Milanesi, P., Marti, D., Anton, M., Klaňová, A., Kályakin, M. V., Bauer, H.-G., & Foppen, R. P. B. (2020). *European breeding bird atlas 2: Distribution, abundance and change*. European Bird Census Council & Lynx Edicions.

Kiss, O., Tokody, B., Nagy, K., & Véghvári, Z. (2020). Potential enlargement of the European roller’ breeding range in the Carpathian Basin. *Journal for Nature Conservation*, 56, 125841–https://doi.org/10.1016/j.jnc.2020.125841

Kolby, J. E. (2014). Ecology: Stop Madagascar’s toad invasion now. *Nature*, 509(7502), 563. https://doi.org/10.1038/509563a

Kremen, C., Cameron, A., Mollanen, A., Phillips, S. J., Thomas, C. D., Beentje, H., Dransfield, J., Fisher, B. L., Glaw, F., Good, T. C., Harper, G. J., Hijmans, R. J., Lees, D. C., Lewis, E., Nussbaum, R. A., Roxworthy, C. J., Razafimpanahana, A., Schatz, G. E., Vences, M., ... Zhra, M. L. (2008). Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science*, 320(5873), 222–226. https://doi.org/10.1126/science.1155193

Lembrchts, J. J., Lenoir, J., Roth, N., Hattab, T., Milbau, A., Haider, S., Pellissier, L., Pauchard, A., Roff, H., Di Marco, R. D., Nuñez, M. A., Aalto, J., & Nijs, I. (2019). Comparing temperature data sources for use in species distribution models: From in-situ logging...
BIOSKETCH

Miguel Cánibe is currently a PhD student at the Universidade de Santiago de Compostela and IPB (Instituto Politecnico de Bragança). His thesis is focused on (1) the combined effects of global change drivers on biodiversity in fire-prone landscapes, and (2) how remote sensing can contribute to this goal. In his thesis, he combines scenario planning tools, landscape simulations and biodiversity modelling platforms to predict shifts in the spatial patterns of vertebrate communities under future climate and land-cover scenarios.

Author Contributions: AR developed the main ideas of this study. MC and AR designed the analyses. JD provided bird data. MC analysed the data and interpreted results with assistance from all authors. All authors contributed significantly to the writing of this paper.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher’s website.

How to cite this article: Cánibe, M., Titeux, N., Domínguez, J., & Regos, A. (2022). Assessing the uncertainty arising from standard land-cover mapping procedures when modelling species distributions. Diversity and Distributions, 28, 636–648. https://doi.org/10.1111/ddi.13456