



HAL
open science

Machine and deep learning approaches to understand and predict habitat suitability for seabird breeding

Antonio Garcia-Quintas, Amédée Roy, Christophe Barbraud, Hervé Demarcq, Dennis Denis, Sophie Lanco Bertrand

► **To cite this version:**

Antonio Garcia-Quintas, Amédée Roy, Christophe Barbraud, Hervé Demarcq, Dennis Denis, et al.. Machine and deep learning approaches to understand and predict habitat suitability for seabird breeding. *Ecology And Evolution*, 2023, 13 (9), pp.e10549. 10.1002/ece3.10549 . hal-04262865

HAL Id: hal-04262865

<https://hal.umontpellier.fr/hal-04262865>

Submitted on 22 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

Machine and deep learning approaches to understand and predict habitat suitability for seabird breeding

Antonio Garcia-Quintas^{1,2}  | Amédée Roy¹  | Christophe Barbraud³  |
Hervé Demarcq¹  | Dennis Denis⁴  | Sophie Lanco Bertrand¹ 

¹Institut de Recherche pour le Développement (IRD), MARBEC (Université de Montpellier, Ifremer, CNRS, IRD), Sète, France

²Centro de Investigaciones de Ecosistemas Costeros (CIEC), Cayo Coco, Cuba

³Centres d'Etudes Biologiques de Chizé UMR7372, Centre National de la Recherche Scientifique, Villiers en Bois, France

⁴Museo Nacional de Historia Natural, La Habana, Cuba

Correspondence

Antonio Garcia-Quintas, Institut de Recherche pour le Développement (IRD), MARBEC (Université de Montpellier, Ifremer, CNRS, IRD), Sète, France.
Email: agquintas86@gmail.com

Funding information

European Union's Horizon 2020 research and innovation program, Grant/Award Number: 817578; Allocations de Recherche pour une Thèse au Sud (ARTS) program of the French Institut de Recherche pour le Développement (IRD)

Abstract

The way animals select their breeding habitat may have great impacts on individual fitness. This complex process depends on the integration of information on various environmental factors, over a wide range of spatiotemporal scales. For seabirds, breeding habitat selection integrates both land and sea features over several spatial scales. Seabirds explore these features prior to breeding, assessing habitats' quality. However, the information-gathering and decision-making process by seabirds when choosing a breeding habitat remains poorly understood. We compiled 49 historical records of larids colonies in Cuba from 1980 to 2020. Then, we predicted potentially suitable breeding sites for larids and assessed their breeding macrohabitat selection, using deep and machine learning algorithms respectively. Using a convolutional neural network and Landsat satellite images we predicted the suitability for nesting of non-monitored sites of this archipelago. Furthermore, we assessed the relative contribution of 18 land- and marine-based environmental covariates describing macrohabitats at three spatial scales (i.e. 10, 50 and 100km) using random forests. Convolutional neural network exhibited good performance at training, validation and test (F1-scores >85%). Sites with higher habitat suitability ($p > .75$) covered 20.3% of the predicting area. Larids breeding macrohabitats were sites relatively close to main islands, featuring sparse vegetation cover and high chlorophyll-*a* concentration at sea in 50 and 100km around colonies. Lower sea surface temperature at larger spatial scales was determinant to distinguish the breeding from non-breeding sites. A more comprehensive understanding of the seabird breeding macrohabitats selection can be reached from the complementary use of convolutional neural networks and random forest models. Our analysis provides crucial knowledge in tropical regions that lack complete and regular monitoring of seabirds' breeding sites.

KEYWORDS

animal habitat modeling, convolutional neural networks, gulls and terns breeding, remote sensing, seabirds' colonies, selection pattern

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

TAXONOMY CLASSIFICATION

Applied ecology, Community ecology, Spatial ecology

1 | INTRODUCTION

The quality of environment in breeding habitats may greatly affect animal fitness (Danchin et al., 1998). Thus, individuals are under strong selective pressure for optimal breeding habitats (Orians & Wittenberger, 1991; Piper, 2011). This complex selection process involves environmental conditions over a large range of spatial scales and relies on hierarchical and sequential decision-making by animals (Block & Brennan, 1993).

Several factors have been suggested for explaining how seabirds choose a place to breed: geographical features of the nesting area (area, spatial isolation) (Greer et al., 1988; Orians & Wittenberger, 1991); vegetation characteristics (coverage, height, density) (Muzaffar et al., 2015; Raynor et al., 2012); climate variability (temperature, rainfall, wind) (Córdoba-Córdoba et al., 2010; Muzaffar et al., 2015) and socio-ecological factors (competition, territoriality, predation pressure, fidelity to the breeding site, group cohesion, information exchange between individuals, colony recruitment, previous breeding experience) (Córdoba-Córdoba et al., 2010; García Borboroglu & Yorio, 2007; Greer et al., 1988). Commonly, adult seabirds gather information on habitat quality (Doligez et al., 2002) over a range of spatial scales through prospective movements before breeding (Kristan, 2006; Ponchon et al., 2013). A range of oceanographic conditions surrounding the nesting sites may also be assessed by seabirds when selecting a place to nest: water masses characteristics (temperature, salinity), bathymetry and productivity-related variables (chlorophyll-*a* concentration, distance to food sources, prey availability and abundance). In particular, water mass properties and zooplankton abundance have been shown as important factors for this selection process in boobies and auklets (Oppel et al., 2015; Sorensen et al., 2009).

Existing studies on breeding habitat selection by seabirds are mostly focusing either on terrestrial habitats, where nests and colonies are installed, or on the surrounding marine areas, that birds use to forage during the breeding (e.g. García Borboroglu & Yorio, 2007; Raynor et al., 2012). Also, most of these studies focused on a single spatial scale of analysis and were often species or colony specific. More integrative (over land and seascapes), multi-specific and multi-scale approaches should improve our understanding of the breeding habitat selection process by seabirds. In addition to these existing limitations, seabird habitat selection in the tropics is much less understood than that of temperate and polar species. In tropical waters, primary productivity is generally low and seasonally stable compared to the cooler waters of polar and temperate regions (Hockey & Wilson, 2003; Jaquetmet et al., 2008). One might therefore expect key factors for habitat selection to differ between tropical, temperate or polar seabirds, and hypothesize that tropical

seabirds are comparatively more influenced by terrestrial than marine features.

Furthermore, many tropical regions lack a full and regular monitoring of seabirds' breeding sites due to economical and logistical constraints, as well as the scarcity of qualified human resources. For instance, Laridae (gulls, terns and skimmers; Winkler et al., 2020) in Cuba are the most abundant seabird group with 25 species recorded (Navarro, 2021), 36% of them breeding in the archipelago (Jiménez et al., 2009). However, information on their colonies is presently very limited: scarce records of sites, species, number of breeding pairs, and basic habitat features and breeding parameters (e.g. Acosta et al., 2022; Jiménez et al., 2009). In particular, the most important environmental variables affecting breeding habitat selection remain poorly known. In order to prioritize the areas to be monitored, an important prerequisite is to predict potential breeding sites as well as to identify the main drivers of breeding habitat selection at the scale of the entire archipelago. Considering both terrestrial and marine areas should provide a more realistic and eco-functional approach to predict tropical seabirds' breeding sites.

The synergy of the aforementioned characteristics (dependence on terrestrial and marine factors, multi-scale influence, multi-species breeding) makes the study of seabird breeding macrohabitat (i.e. the breeding location such as island, peninsula, beach) complex. Apparently, few tools have enough potential to assess, holistically, the suitability and selection of these macrohabitats. But a solution could be found within the field of machine learning, a family of artificial intelligence tools that aims to learn functional relationships from data (Borowiec et al., 2022; Fincham et al., 2020; Olier et al., 2021). Since their dissemination in the 1990s, machine learning models have shown marked statistical and predictive superiority over classical approaches, such as the maximum likelihood estimation and null hypothesis significance testing (Pichler & Hartig, 2023).

Among the most popular models are the neural networks (such as convolutional neural networks [CNN]), and random forest (RF). CNNs belong to the deep learning subfield and have become a state-of-the-art approach in the field of computer vision and remote sensing (Borowiec et al., 2022; Ghanbari et al., 2021; Ma et al., 2019). CNNs are composed of multiple layers of processing units which can learn from complex features and represent data with a high level of abstraction at multiple scales. They are known for their outstanding ability to segment and classify images within end-to-end learning framework, i.e. without requiring any preliminary feature engineering (Fincham et al., 2020; Kattenborn et al., 2021; Ma et al., 2019). RF (Breiman, 2001) is highlighted for its robustness to heterogeneous predictors, its high accuracy (Ma et al., 2019) and its ability to provide a contribution level or importance of each covariate. CNNs usually outperform RFs for

classification and prediction purposes (Kattenborn et al., 2021; Mahdianpari et al., 2018). Yet, an important advantage of RFs over CNNs is their more explicit understanding of the associations between the response variable and its covariates. In the remote sensing research area, CNNs have been identified as potentially well-suited to prediction of habitat suitability for animals such as birds (e.g. Chilson et al., 2019; Su et al., 2018). “A picture is worth a thousand words” and a satellite image represents an excellent example of that due to its stack of spectral bands with high potential for seabird macrohabitat description. Su et al. (2018) used CNNs (and Support Vector Machine) with satellite images to model the habitat suitability for a migratory geese species. Others, as Chilson et al. (2019) and Wang et al. (2021), identified birds' habitat elements using radar data and photographic images, respectively. Deneu et al. (2021) used CNNs to improve species distribution modeling by capturing complex spatial structures of the environment.

Despite the development of some alternative procedures, the main limitation of CNNs (and deep learning in general) is that they operate as a “black box,” which prevents the ecological interpretation of the processes under study (Borowiec et al., 2022; Pichler & Hartig, 2023). However, considering the main strengths of CNN (high performance for prediction) and RF (assessing of ecological hypotheses through the covariates contribution), the complementary use of both methods could increase our understanding of the patterns and processes involved in macrohabitat selection and be helpful for developing effective management and conservation strategies (Figure 1). Here, we predict the suitability of macrohabitat for the breeding of Laridae in Cuba (using CNN) and investigate the ecological variables driving their habitat selection (using RF) from satellite data. More precisely we (1) predict the breeding macrohabitat suitability of Laridae at the scale of the entire Cuba archipelago using CNN, and (2) assess the selection of the breeding macrohabitat by these seabirds considering the contribution of landscape and seascape covariates, at different spatial scales, using RFs.

2 | MATERIALS AND METHODS

2.1 | Study area

This study focuses on the marine coastal ecosystems of the Cuban archipelago (Figure 2). Cuba is the largest Caribbean island (length=1256.2km, maximal width=191km) and includes four insular groups (Los Colorados, Sabana-Camagüey, Canarreos and Jardines de la Reina) featuring >1600 cays (small, low-elevation, sandy islands on the surface of the coral reef) and islets with large variation in relief, geology and landscapes. Climate is tropical hot and seasonally wet with marine influence and semi-continental traits (www.insmet.cu). Annual mean temperature varies between 24°C in the plains of the main island and >34°C at the eastern coasts. Mean relative humidity in the island is high (~82%–90%) and mean annual precipitation ≈1375mm. Daily weather variations are more important

between November and April while the weather is more stable during May–October due to the influence of a North Atlantic anticyclone (www.insmet.cu). The mean sea surface temperature over the continental shelf varies from ~23 to 28°C in January and from ~29 to 32°C in September, from North to South, with the largest spatial gradients at the vicinity of the shelf break. The mean chlorophyll-*a* varies from ~0.5 to >10mgm⁻³ with the largest values observed between the coast and the northern islands as well as in the Southwest region, with moderate seasonal variations.

2.1.1 | Breeding and available sites

We compiled and filtered (deletion of duplicates, erroneous and imprecise data about species identification, location of colonies and date of breeding) all available informations on observed breeding sites of Laridae (i.e. cay, islet or coastal site) from scientific publications, books, thesis, project reports and unpublished data. A database was built with the names and spatial coordinates of the 49 reported breeding sites (Figure 2), years of observation (1980–2020), breeding species and information sources (Table S1). Observed breeding species of Laridae were Laughing Gull *Leucophaeus atricilla* (LAGU), Brown Noddy *Anous stolidus* (BRNO), Sooty Tern *Onychoprion fuscatus* (SOTE), Bridled Tern *Onychoprion anaethetus* (BRTE), Least Tern *Sternula antillarum* (LETE), Gull-billed Tern *Gelochelidon nilotica* (GBTE), Roseate Tern *Sterna dougallii* (ROST), Royal Tern *Thalasseus maximus* (ROYT) and Sandwich Tern *Thalasseus sandvicensis* (SATE). Breeding records of Common Tern *Sterna hirundo* were treated as ROST due to the misidentification of these species' colonies (Navarro, 2021; Nisbet, 2020). Additionally, we selected (non-randomly) 52 sites distributed along the coast of Cuba where none of the nine species was observed breeding in 2020 (Figure 2, Table S1). These non-breeding sites represented the potential macrohabitat available. Both terrestrial and marine features surrounding the observed breeding sites were considered for predicting suitable breeding macrohabitats.

2.2 | Acquisition and formatting of breeding macrohabitat data

2.2.1 | Satellite images

For each breeding and available site we extracted satellite imagery of Landsat 5 and 7 (Table S2) from EOS Data Analytics platform (<https://eos.com>). The date of the image was matched to the year of the breeding colony presence record, while images from 2020 were used for non-breeding sites. Several images were associated to each site (depending on availability) to ensure a good representation of the natural variability during the breeding period (May–August) and to reduce the influence of clouds in some images. In some cases, we incorporated images of both months of April and September (climatology similar to the May–August period) when none was available

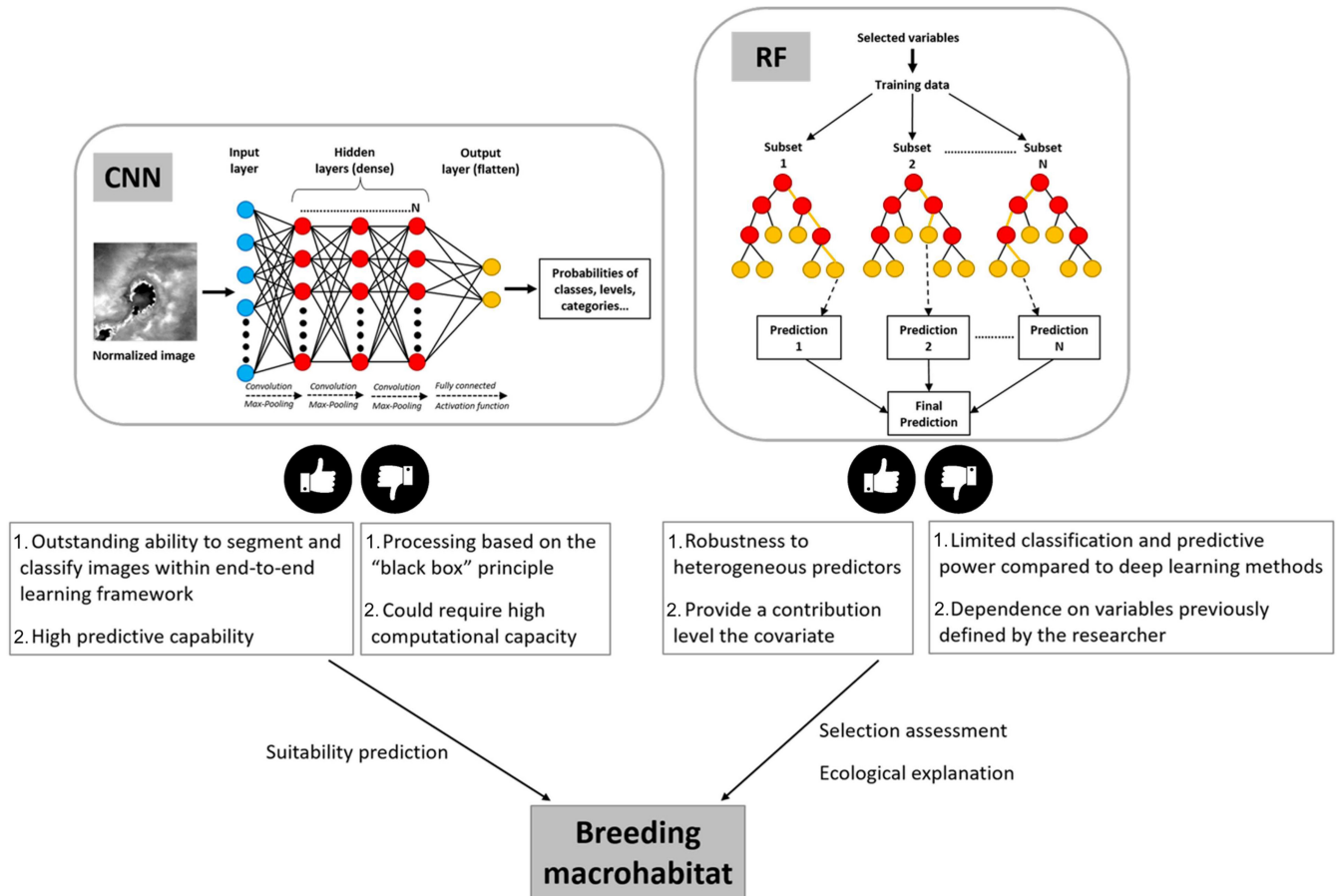


FIGURE 1 Complementarity approach between convolutional neural networks and random forests models to study seabird breeding macrohabitat.

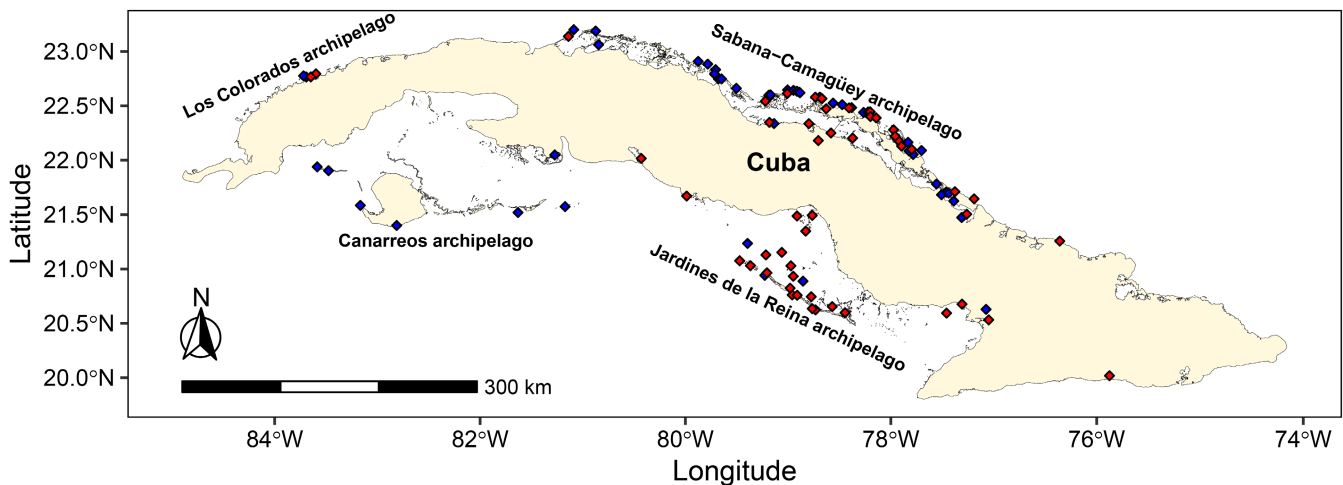


FIGURE 2 Map of the study area and the sites used to train and validate the modeling of the breeding macrohabitat suitability of nine Laridae species. Blue diamonds = breeding sites, red diamonds = non-breeding sites.

from May to August of the current year. In total, we selected 136 satellite images describing the conditions of the study sites (Table S2).

We then resized the Landsat images into scenes (9.0 × 9.0 km square areas centered on the sites) and standardized them through the "Dark Subtraction" (based on the bands minimum

digital number) to apply atmospheric scattering corrections to the imagery data and "SLC Gap-Filled" correction for Landsat 7 imagery since 31 May 2003. In the end, GeoTIFF files (299 × 299 pixels, 30-m-spatial resolution) were created that included the visible and infrared bands. Panchromatic and thermic bands were excluded

because these do not match between Landsat satellites. The definition of scene size followed a balanced criterion: sufficiently large image size that included land and sea components, and sufficiently small to minimize the inclusion of other breeding and available sites in the same image scene (which would affect the prediction quality). We organized the data into two datasets considering the quality of scene images related with cloud cover (Table S3) to control for cloud-related confounding effects. We randomly mixed the images database of both types of sites (i.e. with and without breeding colonies) and then split them into two groups containing 70% (training) and 30% (validation) of the data (Table S3) for the building and selection of the best CNN.

Satellite images from 2021 at 12 breeding and 52 non-breeding sites (verified as such that year) were used as test dataset to assess the predictive performance of the CNN. For predicting breeding macrohabitat suitability, we applied the same preprocessing to Landsat 7 images of 2021 (Table S2) that covered the entire Cuban archipelago. A mosaic was built with these images, masking mainland to retain only the marine-coastal ecosystems up to the insular shelf. Images with open water only or predominance of land of the main two islands (Cuba and Juventud) were excluded since they were irrelevant to Laridae breeding (Jiménez et al., 2009). Finally, this mosaic image was gridded into 793 scenes (63,805.5 km²) with the same format and structure than train and validation datasets. Satellite images were processed using the ENVI 4.7 (ITT VIS Inc) software.

2.2.2 | Physical and geographical covariates

Because of the absence of information on the foraging ranges of Cuban seabirds during their breeding period, we did a bibliographic compilation of all information available on the same species observed elsewhere during breeding, from polar to tropical zones. From this review, we estimated the potential maximum foraging ranges during breeding for each studied species (Table S4). Then, we defined three spatial scales (radius of 10, 50 and 100 km from the breeding site) approximately corresponding to the estimated foraging ranges of our study species, and computed several oceanographic characteristics at each spatial scale.

Twelve potentially important features for the establishment of Laridae breeding colonies, described through 18 metrics, were considered at the defined spatial scales (Table 1): 11 of them described the conditions of the nesting landscape, and seven of them described the conditions of the surrounding seascape. The variables were measured at the date of the colony observation, and in 2020 for the non-breeding sites.

Using the Landviewer product of the EOS Data Analytics platform, we calculated the Normalized Difference Water (Gao, 1996) and Vegetation (Rouse et al., 1973) Indexes (NDWI and NDVI respectively) from satellite images (using Green, Red and Near Infra-red bands) of the Landsat series. Dates of images ranged between May and August, matching the breeding season of Laridae in Cuba (Jiménez et al., 2009). Both spectral indices vary between -1 and 1

with higher numbers corresponding to higher humidity (NDWI) or green vegetation (NDVI). Based on the NDWI we then calculated the drought emerged areas (NDWI range = -1 to 0.2) and the percentage of non-flooding cover (NDWI range = -1 to 0). The NDVI allowed to quantify the total (NDVI range = 0.2-1) moderate and dense vegetation covers (NDVI range = 0.4-1). Thus, the area, non-flooding zone cover and vegetal covers of each nesting site were computed from the satellites images. The perimeter was calculated after vectorization of the imagery scenes.

We calculated the index of shape complexity for islets and cays (Hu et al., 2011) as $SI = P / [2 \times (\pi \times A)^{1/2}]$, where SI = shape index, P = perimeter and A = area of the site. A SI value of 1 indicates an islet or cay with a perfect circular shape and SI increases as the shape becomes more irregular and complex. Isolation variables (minimum distances to Cuba/Isla de la Juventud (IJ) and to nearest cay, and cays/islets number at the three spatial scales) were estimated using Google Earth Pro 7.3.3 software. Minimum distances to the 200-m isobath were measured using a bathymetric shapefile of the exclusive economic zone of Cuba (using information both from GEBCO and Cuban research agencies databases).

Sea surface temperature (SST) and surface chlorophyll *a* concentration (Chl *a*) around the sites were obtained at a spatial resolution of 1 km from level-2 MODIS-Aqua satellite data (<https://ocean.color.gsfc.nasa.gov/data/aqua/>) after sampling and spatial re-projection of the data. For both variables we averaged the monthly values between May and August of the year corresponding to the last register of each breeding colony and for 2020 for available sites. Nevertheless, due to the absence of logistical support for a systematic monitoring of these variables before 2002, and given the small interannual variability of both variables in Cuba, data for breeding colonies observed in that period were estimated from the mean of 2002-2021 period for the same months (Figure S1). A summary of the satellite images used for the study is provided in Table S2.

The eldest colony record, at Rincón del Guanál (Table S1), was excluded from the study as we could not obtain the same variables from Landsat 4 satellite (fewer bands than Landsat 5 and 7). Breeding macrohabitat was characterized considering both mixed and monospecific colonies.

2.3 | Prediction and selection assessment of the breeding macrohabitats

2.3.1 | CNN implementation

A CNN architecture is typically composed by multiple layers of processing units where two main processes occur: convolution and pooling. During convolutions several filters are applied to extract relevant features of data that will be used for calculating the matches in the testing phase. Pooling operations capture large images and reduces the parameters to preserve important information. Kattenborn et al. (2021) and Krishna and Kalluri (2019) offer more details

TABLE 1 Potentially important variables for Laridae breeding macrohabitat selection (breeding site, i.e. cay, islet or coastal site) in Cuba.

Feature	Variable	Unit	Ecological meaning
Site extent	Area	km ²	Available space for nests establishment
Site perimeter	Perimeter	km	Indicator of the availability of potential coastal zones for breeding
Site shape	Shape index based on perimeter/area ratio	–	Related to geographical features (e.g. peninsulas) that could be important for breeding
Isolation degree	Minimum distance from the colony to the nearest cay/islet	km	Indicator of accessibility for predators and other disturbance sources
	Minimum distance from the colony to the nearest main island (Cuba or Isla de la Juventud) of the archipelago	km	Indicator of accessibility for predators and high disturbance sources (higher risk)
	Cays/islets number within 10, 50 and 100 km of the colony	–	Indicator of the number of potential sources (at different spatial scales) of predators, alien species and other disturbances that could affect the colonies
Terrain	Non-flooding zone cover	%	Suitability of the locality for colony establishment based on flooding risk
Vegetation	Vegetal cover _{total}	%	Surface occupied by plants (species-specific suitability for breeding, Burger & Gochfeld, 1981; Raynor et al., 2012)
	Vegetal cover _{moderate+dense}	%	Surface occupied by moderate to dense vegetation that could affect the establishment of colonies (some Laridae tend to avoid high vegetation cover while others are attracted, Bukacinska & Bukacinsky, 1993; Burger & Gochfeld, 1986)
Oceanographic	Sea surface temperature within 10, 50 and 100 km radius of the colony	°C	Reflects thermal conditions that influence primary productivity and prey availability at different foraging ranges
Bathymetric	Minimum distance to the 200-m isobath	km	Indicates the limit of the insular shelf in Cuba and therefore is a proxy for suitable foraging areas for most Laridae (Schreiber & Burger, 2002).
Phytoplanktonic biomass	Chlorophyll <i>a</i> concentration at sea surface within 10, 50 and 100 km radius of the colony	mg m ⁻³	Proxy for phytoplanktonic biomass, primary productivity and prey availability at different foraging ranges

about architectures, parameters and functions of CNNs. Here, we use a CNN with three consecutive layers of convolutions and max pooling, followed by a dense network. Finally, the last layer consists in a sigmoid activation so that the output of the network is a value between 0 and 1. This CNN aims therefore to ingest the selected bands of Landsat satellite images of 299 × 299 pixels (Table S3) as input data and to output the probability of the described habitat to be suitable for breeding seabirds.

Parameters were finally estimated using an Adam optimizer and minimizing a “Sparse categorical crossentropy” loss. In order to prevent overfitting, we also added a L2 regularizer, a Deep Learning technique to get better generalization and predictive properties (Kussul et al., 2017). Finally, models were trained using the “training dataset,” and selected when minimizing accuracy score over the “validation dataset.” This analysis was implemented with the keras R package (v. 2.6.1) (Kalinowski et al., 2021).

We performed this training procedure for distinct parameters and the best CNN (better performance and lowest loss in validation and test datasets) was used to predict the breeding macrohabitat suitability using the satellite images of 2021 along all marine-coastal ecosystems of the archipelago. The most frequently reported

metrics, Overall Accuracy, Precision, Recall and F1-score (Table S5) were used to assess CNN performance. All analysis was implemented in R 4.1.1 (R Core Team, 2021).

2.3.2 | RF implementation and contribution of variables

Variables were compared among sites using Mann–Whitney *U*-tests considering significance at $p < .01$. Breeding macrohabitat selection was analyzed through random classification forests (RFs) models considering the measured variables at three spatial scales. Similar to CNN, we mixed and split the data to create the training and validation datasets (70% vs. 30% proportions) for the building and selection of the best RF. The same training and validation datasets were used in all RFs to compare their classification performances. We implemented three RFs, one that processed physical-geographical variables registered within a radius of 100 km from the breeding colony (RF_100; 300 trees and four variables by split), a second within a 50 km radius (RF_50; 200 trees and four variables by split) and a third one within a 10 km radius (RF_10; 150 trees and three variables by

split). For each RF type we used the 10 best runs in order to get an average and variance of model performance.

Same performance metrics than CNN (Table S5) were used to assess the training, validation and test of RFs. Variables contributions were calculated from the mean decreasing Gini index (values are directly proportional to variable importance) derived from the RF with better performance. The randomForest R package (v. 4.6-14) (Breiman et al., 2018) was used to this analysis.

3 | RESULTS

3.1 | Prediction of macrohabitats breeding suitability

The performance of CNN was high, exhibiting indicators values $\geq 80.0\%$ for the validation datasets (Table 2). Image quality (according to cloud cover) had no significant consequence on the classification power of CNNs (F1-score validation = 85.7% and 84.6% for CNNs that used all and best images, respectively; Table S3). We thus worked with the architecture that used all images. Based on the test dataset, the CNN exhibited good performance test indicators with accuracy = 79.7%, precision = 91.5%, recall = 82.7% and F1-score = 86.9%.

When used for predicting over the entire Cuba archipelago, the CNN estimated 32,184, 12,069, 6598 and 12,954 km² of suitable habitat for breeding within 0–.25, .26–.50, .51–.75 and .76–1 probability ranges, respectively (Figure 3). Thus, the probability ranges $>.50$ and $>.75$ covered 30.6% and 20.3% of the predicted area, respectively. The best areas (high suitability scores) tended to be concentrated in three general types of ecosystems: (1) remote cays/reef islets of all subarchipelagos (Figure 2, with Jardines de la Reina archipelago under-represented), (2) coastal zones with sand, rocks or interior lagoons and (3) interior of bays, gulfs and swamps that contained small islets and sand banks (Figure 3). The southern region of Cuba had less suitable breeding macrohabitats than the northern region (Figure 3).

3.2 | Breeding macrohabitat selection and importance of covariates

The general statistics (median, quartiles) of most covariates were quite similar between breeding and non-breeding sites (Figure 4, Table S6). Only SST at the three spatial scales and Chl *a* at 50 and 100 km radii exhibited significant differences: SST was lower and Chl *a* higher at breeding sites (Figure 4). Additionally, breeding sites tended to have a smaller number of cays/islets within 10 km compared to non-breeding sites ($p = .01$, Figure 4). Performance metrics for RF_50 and RF_100 were very similar and outperformed RF_10 (Table 2). For RF_100, SST had the highest contribution to discriminate breeding from non-breeding sites (Figure 5). A second group of covariates with lower contribution included isolation-related variables (number cays/islets within 10 and 50 km from colonies) and Chl *a* within 100 and 50 km radius (Figure 5). Remaining variables exhibited relatively low contributions and non-flooding area cover had the lowest importance (Figure 5).

4 | DISCUSSION

4.1 | Prediction of breeding macrohabitat suitability

According to our results, prediction of habitat suitability can be successfully obtained by processing satellite images with CNN exclusively. This constitutes a significant advance for habitat ecology studies and expands the applications and perspectives of image analysis via deep learning approaches. Deep architecture of CNNs conveys a high computation cost (Kattenborn et al., 2021) but, at the same time its versatility provides a great generalization capacity with a broad applicability in the remote sensing field (Kattenborn et al., 2021; Ma et al., 2019; Mahdianpari et al., 2018). Our CNN had relatively good performance and provided a map of macrohabitats suitability for Laridae over the whole Cuban archipelago for the year 2021. Predictions were based on physical-geographical suitability of

TABLE 2 Training and validation performance (in %) of a convolutional neural network (CNN) and three random forest (RF) models used to respectively predict breeding site suitability and assess breeding site (macrohabitat) selection by Laridae in Cuba.

Model type	Training accuracy	Training F1-score	By image				By site			
			Validation accuracy	Validation precision	Validation recall	Validation F1-score	Validation accuracy	Validation precision	Validation recall	Validation F1-score
CNN	98.4	98.7	75.7	75.5	76.9	76.2	86.7	80.0	92.3	85.7
RF_10	72.0 ± 2.5	71.2 ± 2.3	–	–	–	–	70.0 ± 1.6	67.4 ± 2.0	69.3 ± 3.4	68.3 ± 1.9
RF_50	81.7 ± 2.4	80.7 ± 2.4	–	–	–	–	77.3 ± 2.1	77.8 ± 3.4	72.1 ± 2.3	74.8 ± 2.0
RF_100	85.1 ± 1.1	84.5 ± 1.4	–	–	–	–	76.7 ± 0 ^a	76.9 ± 0 ^a	71.4 ± 0 ^a	74.1 ± 0 ^a

Note: RF_10, RF_50 and RF_100 indicate models that used physical-geographical variables within 10, 50 and 100 km radius from localities, respectively; CNN = model that used Landsat images with 9 × 9 km square areas. Statistics indicate the mean (± standard deviation) of the best 10 runs for each RF.

^aThere was no variability in the indicator.

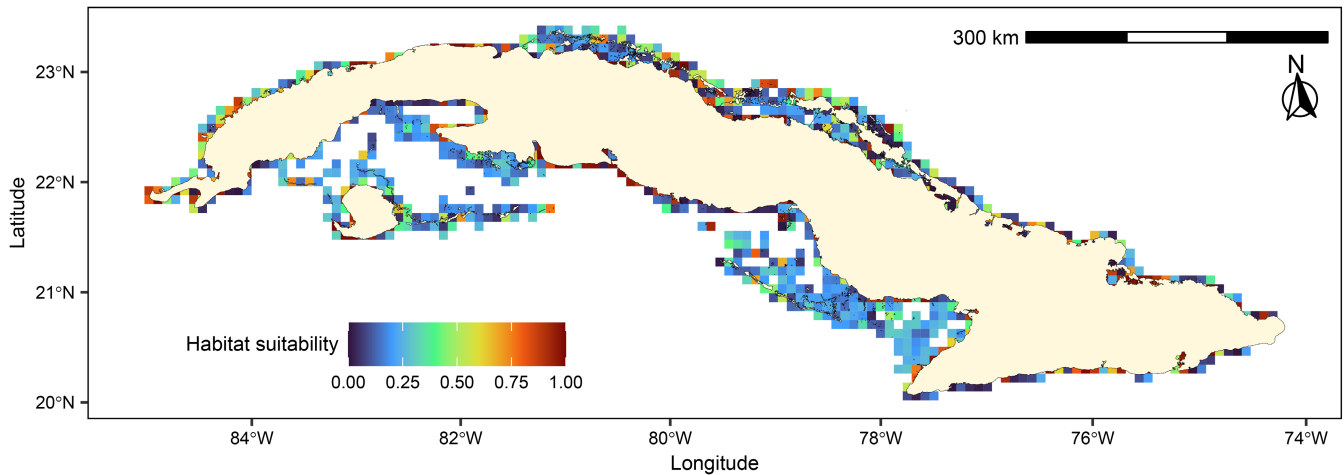


FIGURE 3 Convolutional neural network prediction of macrohabitat suitability for the breeding of Laridae in Cuba, for the 2021 breeding season, using Landsat images.

marine-coastal ecosystems (from the images) but it does not imply per se the existence of breeding colonies at areas with higher habitat suitability. Deep learning algorithms, such as CNNs, do not necessarily “learn” the right causal dependencies between input features and the response (Pichler & Hartig, 2023). However, these areas represent sites with favorable conditions for breeding in 2021 and thus constitute alternatives or additional breeding sites for Laridae around Cuba.

The most suitable breeding sites (probability range $>.50$) for Laridae exhibited a scattered general distribution along the coasts of Cuba. Although a slightly higher concentration of sites occurred in the northern marine-coastal ecosystems. Nevertheless, areas of Los Colorados, Canarreos and Sabana-Camagüey archipelagos showed higher number of favorable breeding sites acting as potential hotspots for Laridae reproduction (Figures 2 and 3). These predictions are relatively consistent with current field knowledge in these areas, and with the location of historical persistent breeding colonies.

In particular, breeding habitats for Laridae such as beaches, rocky platforms and sand banks in distant or difficult-access cays and islets were often true positive predictions. Two of the most scientifically studied regions, the Sabana-Camagüey and Jardines de la Reina archipelagos, had respectively high and low predicted breeding habitat suitability depicted by CNN, a point confirmed by field observations (Figures 2 and 3). Nevertheless, some predicted areas seemed irrelevant for breeding because of the presence of extensive anthropogenic infrastructures (cities, towns, industries, agricultural fields) that might cause disturbance. Although this study focused on the physical-geographical suitability of Laridae breeding macrohabitats, the satellite images processed through our CNN also contained spectral information on some anthropogenic features. This provides a more comprehensive aspect to our predictions.

On the other hand, main false negatives of the predictions were located at Mono Grande, Cinco Leguas, Felipe de Barlovento and Las Salinas breeding sites (Table S1). Prediction scenes (grid cells)

containing these breeding sites were not highly different (from a visual interpretation) from the training sites. Then, this suggests the importance of incorporating oceanographic (e.g. SST, Chl *a*) or ecological variables (e.g. prey availability-related) into the CNN to improve its prediction quality. Scene classification with an emphasis on land cover, vegetation and crop types is one of the most common applications of CNNs (e.g. Kattenborn et al., 2021; Kussul et al., 2017; Mahdianpari et al., 2018). However, these classifications are based on relatively easily distinguishable element classes (e.g. water, bare soil, marsh, fen, forest, grassland, paddy rice) whereas animal habitat suitability is a more complex phenomenon.

It is important to point out that the habitat suitability diversity predicted by the model (that addressed all Cuban Laridae) could result from the different species-specific breeding habitats requirements; especially LETE that has a high dynamic and opportunistic behavior to select its breeding sites (this species may change its nesting sites between consecutive breeding seasons depending on the availability of isolated sand bodies). For this reason, even if CNN obtained relevant predictive performance metrics, the prediction map (Figure 3) should be interpreted carefully, and further studies would be required to improve its accuracy. However, the good global quality of the model highlights its potential for application to other species and regions of the world. Normally, CNNs are excellent tools for capturing complex hierarchical patterns from images (Borowiec et al., 2022).

4.2 | Breeding macrohabitat selection pattern

Simplistic approaches to the study of breeding habitat selection have been criticized decades ago (e.g. Burger & Shisler, 1978) as the environment of many animals, such as seabirds, consists in heterogeneous composition of habitat characteristics along several spatial and temporal scales (Danchin et al., 1998). The abundance of food resources, microclimate and physical-geographical attributes of the

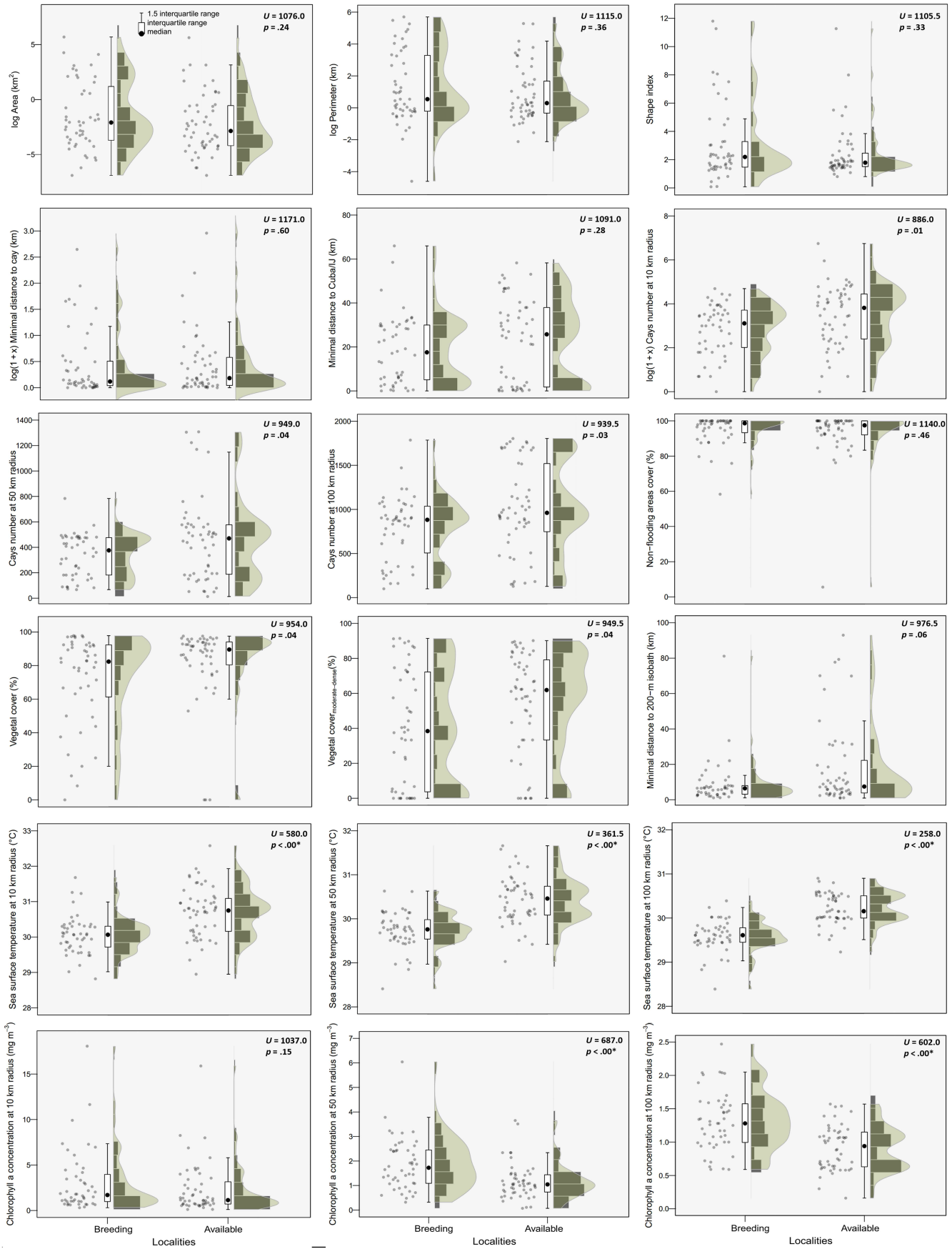


FIGURE 4 Comparison of 18 physical-geographical variables corresponding to 48 breeding and 52 non-breeding sites (available macrohabitats) for Laridae in Cuba. Some variables were log-transformed for visualization purpose exclusively. *Significant differences.

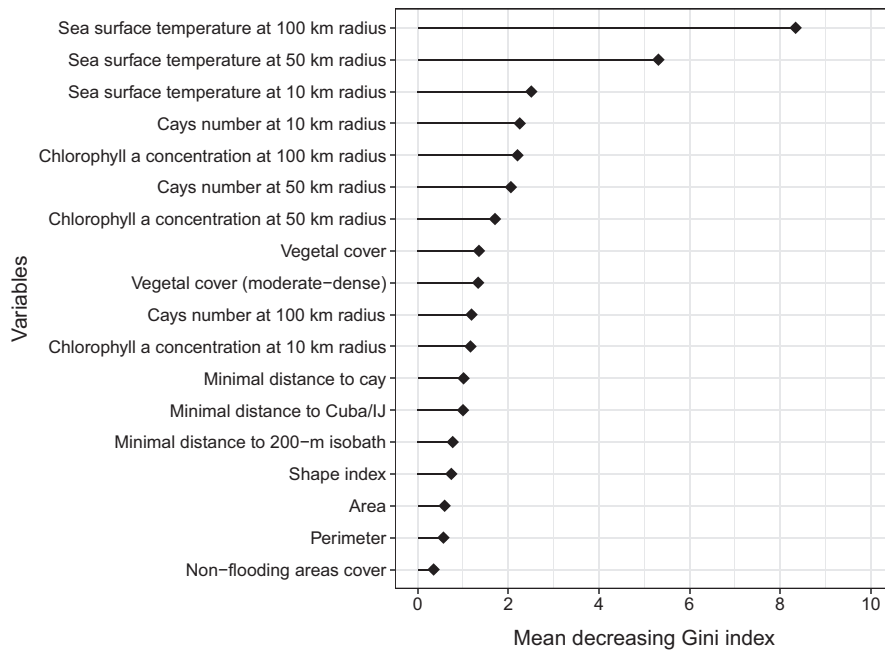


FIGURE 5 Contribution of marine and terrestrial environmental features to the breeding macrohabitat selection pattern of Laridae in Cuba based on random forests classification models (corresponding to the spatial scale of 100 km radius from breeding colonies, but includes the nested scales of 10 and 50 km radius).

local landscape are elements usually relied upon by colonial seabirds for the selection of good breeding habitats (García Borboroglu & Yorio, 2007). Here, we found that breeding macrohabitat selection by Laridae in Cuba could be partly explained through seascape and landscape features of the breeding sites.

Overall, breeding site (macrohabitat) selection by Laridae was mainly explained by lower SST values within 100 and 50 km from colonies. Thus, we show that SST at larger scales played an important role in Cuban Laridae habitat selection, despite its relative seasonal stability throughout tropical waters (Hockey & Wilson, 2003; Jaquetmet et al., 2008). The greater contribution of larger spatial scales for SST probably reflects the role of oceanographic conditions (e.g. thermal fronts) at relatively large distances from breeding habitats, coinciding with the foraging range of most species which often exceeds 30 km from colonies (Table S4). Chl *a* at the same spatial scales was also important, although to a lesser extent, for breeding site selection, highlighting the role of marine productivity for breeding. However, it should be noted that Chl *a* is an index of phytoplanktonic biomass that does not match exactly in space with the maximum of forage fish abundance (e.g. Zavalaga et al., 2010).

Sea surface temperature (indirectly) and Chl *a* could be considered as proxies of marine productivity and food availability, and hence key factors for breeding habitat selection by seabirds (Vilchis et al., 2006). However, mismatching patterns between both variables may occur (e.g. Zavalaga et al., 2010). In several regions of the world these variables have been shown as having important effects on seabirds foraging, demography and population dynamics, with generally cooler SST favoring higher Chl *a*, and hence the foraging and breeding success of seabirds (Barbraud et al., 2012; Carroll et al., 2015). The selective pattern of breeding macrohabitat found for Laridae in Cuba is consistent with this general pattern, although the effect of Chl *a* appears lesser than for SST (Figure S1).

The lower contribution of Chl *a* could be consequence of terrestrial runoff leading to eutrophication processes, as occurs in some marine ecosystems of northwestern Cuba (Rey-Villiers et al., 2021). Thus, such excessive and polluted upwelling of primary productivity may not be attractive for Laridae feeding. This also could be magnified in areas of low marine circulation as in the interior macro-lagoons (water bodies between the cays and the island of Cuba) of the Sabana-Camagüey archipelago. In addition, differences in marine current patterns between regions (e.g. Arriaza et al., 2008) and the influence of extreme climate events such as rainfall and hurricanes (Alvarez-Socorro et al., 2021) should cause differences in local oceanography (e.g. SST, Chl *a*) between coastal areas of Cuba.

Similar to previous studies (e.g. Burger & Gochfeld, 1981, 1986; Greer et al., 1988), areas with moderate to dense vegetation cover were avoided by Laridae for breeding (Figure 4). For seabirds, high cover and dense vegetation cover usually constitutes a barrier to breeding as it limits the visibility and social communication between neighbors at colonies, and hence, may increase predation risk (Bukacinska & Bukacinsky, 1993; Raynor et al., 2012).

4.3 | Complementarity of CNN and RF approaches

After visual inspection of the image mosaic of 2021, the more suitable predicted areas for Laridae in Cuba included heterogeneous terrestrial covers and waters with contrasting depths. Yet, our ecological understanding here is limited as we have no explicit indication on the metrics CNNs finally used to maximize prediction quality. Some technics exist in order to get insights into the metrics automatically extracted by the CNN, such as layer-wise backpropagation, saliency maps, “network dissection” and the explainable Artificial Intelligence (xAI) methods (see Borowiec et al., 2022; Pichler & Hartig, 2023;

Samek et al., 2021). It is however an active field of research, and using these approaches is beyond the scope of this paper.

Random forests are a classic example of so-called ensemble models (interacting sets of simple algorithms or statistical models that make up more complex algorithms), which typically have low prediction errors and weight the contribution of the variables analyzed by the model (Pichler & Hartig, 2023). This last property contributes to the explanation of the process or phenomenon being studied. Thus, based on our RFs results and statistical analysis, we could yet hypothesize that important features such as vegetation, number of cays/islets but also ocean color (related to Chl *a*) were captured through Landsat spectral data. Also, the CNN performance might be probably increased including SST information at the radius of 100km (considering the main results of RFs) and being trained over larger datasets.

This study illustrates the benefits obtained from a complementary analysis of CNN and RF. Here, RF can be seen as an explicative tool relying on features directly related to our a priori ecological hypotheses, while CNN can act as an evaluative tool in order to assess the relevance of habitat spectral features, as well as an efficient predictive tool producing large scale prediction of habitat suitability in a convenient manner. Particularly, CNNs are becoming increasingly important in remote sensing and ecology due to the inclusion of the spatial dimension within their convolutional layers, thus facilitating the identification/characterization of relevant ecological patterns and processes (Hayes et al., 2021). Our prediction of breeding habitat suitability with CNN should also be systematically updated considering the changing dynamics of marine ecosystems and seabird colonies. Finally, we recommend the exploration of building CNNs that use both spectral and relevant ecological data (identified by RF), to produce finer predictions supported ecologically.

4.4 | Management and conservation implications from the complementary approach

The dispersed distribution pattern of suitable breeding sites in Cuba offers to Laridae a wide variety of options for colony establishment. This could buffer the effects of climate change and anthropogenic pressures on breeding macrohabitats due to presumed vulnerability differences (e.g. flood risk, ease of access to predators and humans, local oceanographic anomalies) among these sites. This also provides flexibility for management agencies, considering the existence of a large number of alternative sites for Laridae conservation in Cuba. Also, the potential for legal protection of some important breeding colonies (e.g. predicted hotspots) is increased due to the low risk of spatial overlap of breeding sites with places of socioeconomic interest, under an appropriate marine spatial planning.

From a practical point of view, we recommend a field validation of the effective presence of colonies in sites that are expected to be highly suitable. This could be done through field surveys at these sites, as a way to optimize logistical and economic resources for conservation purposes. Then, confirmed breeding sites could

be considered to update the boundaries of Marine Protected Areas in Cuba, improve the governmental strategy of adaptation to climate change, detect negative effects due to natural and anthropogenic causes (McGowan et al., 2013; Perrow et al., 2015) and set up a sustainable use of marine-coastal ecosystems (tourism, fishing, industry). Effective conservation measures for seabird populations must necessarily include both the establishment sites of colonies and their surrounding waters (Opper et al., 2018). More precisely, according to our complementary CNN-RF approach, conservation and management actions for Laridae breeding macrohabitats in Cuba should include the areas of predicted breeding hotspots. Also, oceanic characteristics (SST and Chl *a*) at mesoscale (50–100km) around the cays, as well as the degree of dense and moderate vegetation cover, should be considered in future management plans.

AUTHOR CONTRIBUTIONS

Antonio Garcia-Quintas: Conceptualization (lead); data curation (equal); formal analysis (lead); investigation (equal); methodology (equal); validation (equal); writing – original draft (lead); writing – review and editing (equal). **Amédée Roy:** Formal analysis (equal); methodology (equal); software (lead); validation (lead); visualization (equal); writing – original draft (lead); writing – review and editing (equal). **Christophe Barbraud:** Methodology (equal); supervision (lead); validation (equal); visualization (equal); writing – review and editing (lead). **Hervé Demarcq:** Data curation (lead); methodology (equal); validation (equal); visualization (equal); writing – review and editing (equal). **Dennis Denis:** Conceptualization (equal); formal analysis (equal); methodology (equal); supervision (lead); validation (equal); writing – review and editing (equal). **Sophie Lanco Bertrand:** Conceptualization (equal); formal analysis (equal); funding acquisition (lead); methodology (equal); supervision (lead); validation (equal); visualization (equal); writing – review and editing (lead).

ACKNOWLEDGEMENTS

This research is part of a A.G.Q. PhD scholarship supported by the Allocations de Recherche pour une Thèse au Sud (ARTS) program of the French Institut de Recherche pour le Développement (IRD) in cooperation with the Centro de Investigaciones de Ecosistemas Costeros (CIEC) of Cuba. A.R. has been supported by TRIATLAS project (European Union's Horizon 2020 research and innovation program – grant agreement No. 817578). We thank two anonymous reviewers for helpful comments to improve an initial version of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data from this study are available in the Supporting Information (Tables S1, S2 and S6). The used satellite images can be downloaded from <https://eos.com>.

ORCID

Antonio Garcia-Quintas  <https://orcid.org/0000-0002-6360-6195>
 Amédée Roy  <https://orcid.org/0000-0003-3047-1463>
 Christophe Barbraud  <http://orcid.org/0000-0003-0146-212X>
 Hervé Demarcq  <https://orcid.org/0000-0003-1995-1183>
 Dennis Denis  <https://orcid.org/0000-0003-4808-7195>
 Sophie Lanco Bertrand  <https://orcid.org/0000-0003-1976-3799>

REFERENCES

- Acosta, M., Mugica, L., Rodríguez-Ochoa, A., González, A., Aguilar, S., & Aguilar, K. (2022). Caracterización de la colonia reproductiva de aves marinas en Cayo Hijo de los Ballenatos, Archipiélago de los Canarreos, Cuba. *Revista Cubana de Ciencias Biológicas*, 10, 8.
- Alvarez-Socorro, G., Fernández-Alvarez, J. C., Sorí, R., Pérez-Alarcón, A., Nieto, R., & Gimeno, L. (2021). Space-time assessment of extreme precipitation in Cuba between 1980 and 2019 from multi-source weighted-ensemble precipitation dataset. *Atmosphere*, 12, 995. <https://doi.org/10.3390/atmos12080995>
- Arriaza, L., Simanca, J., Rodas, L., Lorenzo, S., Hernández, M., Linares, E. O., Milian, D., & Romero, P. (2008). Corrientes marinas estimadas en la plataforma suroriental cubana. *Serie Oceanológica*, 4, 1–10.
- Barbraud, C., Rolland, V., Jenouvrier, S., Nevoux, M., Delord, K., & Weimerskirch, H. (2012). Effects of climate change and fisheries bycatch on Southern Ocean seabirds: A review. *Marine Ecology Progress Series*, 454, 285–307. <https://doi.org/10.3354/meps09616>
- Block, W. M., & Brennan, L. A. (1993). The habitat concept in ornithology: Theory and applications. In D. M. Power (Ed.), *Current ornithology* (Vol. 11, pp. 35–91). Plenum Press.
- Borowiec, M. L., Dikow, R. B., Frandsen, P. B., McKeeken, A., Valentini, G., & White, A. E. (2022). Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, 13(8), 1640–1660. <https://doi.org/10.1111/2041-210X.13901>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Breiman, L., Cutler, A., Liaw, A., & Wiener, M. (2018). *randomForest: Breiman and Cutler's random forests for classification and regression*. R package version 4.6-14. <https://CRAN.R-project.org/package=randomForest>
- Bukacinska, M., & Bukacinsky, D. (1993). The effect of habitat structure and density of nests on territory size and territorial behaviour in the black-headed gull (*Larus ridibundus* L.). *Ethology*, 94, 306–316.
- Burger, J., & Gochfeld, M. (1981). Nest site selection by kelp gulls in southern Africa. *Condor*, 83, 243–251.
- Burger, J., & Gochfeld, M. (1986). Nest site selection in sooty terns (*Sterna fuscata*) in Puerto Rico and Hawaii. *Colonial Waterbirds*, 9(1), 31. <https://doi.org/10.2307/1521141>
- Burger, J., & Shisler, J. (1978). Nest site selection and competitive interactions of herring and laughing gulls in New Jersey. *The Auk*, 95, 252–266.
- Carroll, M., Butler, A., Owen, E., Ewing, S., Cole, T., Green, J., Soanes, L., Arnould, J., Newton, S., Baer, J., Daunt, F., Wanless, S., Newell, M., Robertson, G., Mavor, R., & Bolton, M. (2015). Effects of sea temperature and stratification changes on seabird breeding success. *Climate Research*, 66(1), 75–89. <https://doi.org/10.3354/cr01332>
- Chilson, C., Avery, K., McGovern, A., Bridge, E., Sheldon, D., & Kelly, J. (2019). Automated detection of bird roosts using NEXRAD radar data and convolutional neural networks. *Remote Sensing in Ecology and Conservation*, 5(1), 20–32. <https://doi.org/10.1002/rse.2.92>
- Córdoba-Córdoba, S., Ouyang, J. Q., & Hauck, S. J. (2010). Nesting preferences and population estimates of a new Black Noddy Anous minutus breeding colony on one tree island, great barrier reef, Australia. *Marine Ornithology*, 38, 79–84.
- Danchin, E., Boulinier, T., & Massot, M. (1998). Conspecific reproductive success and breeding habitat selection: Implications for the study of coloniality. *Ecology*, 79(7), 2415–2428. [https://doi.org/10.1890/0012-9658\(1998\)079\[2415:CRSABH\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1998)079[2415:CRSABH]2.0.CO;2)
- Deneu, B., Servajean, M., Bonnet, P., Botella, C., Munoz, F., & Joly, A. (2021). Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. *PLoS Computational Biology*, 17(4), e1008856. <https://doi.org/10.1371/journal.pcbi.1008856>
- Doligez, B., Danchin, E., & Clobert, J. (2002). Public information and breeding habitat selection in a wild bird population. *Science*, 297(5584), 1168–1170. <https://doi.org/10.1126/science.1072838>
- Fincham, J. I., Wilson, C., Barry, J., Bolam, S., & French, G. (2020). Developing the use of convolutional neural networking in benthic habitat classification and species distribution modelling. *ICES Journal of Marine Science*, 77(7–8), 3074–3082. <https://doi.org/10.1093/icesjms/fsaa208>
- Gao, B. (1996). NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58, 257–266.
- García Borboroglu, P., & Yorio, P. (2007). Breeding habitat requirements and selection by Olog's Gull (*Larus atlanticus*), a threatened species. *The Auk*, 124, 1201–1212.
- Ghanbari, H., Mahdianpari, M., Homayouni, S., & Mohammadimanesh, F. (2021). A meta-analysis of convolutional neural networks for remote sensing applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 3602–3613. <https://doi.org/10.1109/JSTARS.2021.3065569>
- Greer, R. D., Cordes, C. L., & Anderson, S. H. (1988). Habitat relationships of island nesting seabirds along coastal Louisiana. *Colonial Waterbirds*, 11(2), 181. <https://doi.org/10.2307/1520999>
- Hayes, M. C., Gray, P. C., Harris, G., Sedgwick, W. C., Crawford, V. D., Chazal, N., Crofts, S., & Johnston, D. W. (2021). Drones and deep learning produce accurate and efficient monitoring of large-scale seabird colonies. *Ornithological Applications*, 123(3), duab022. <https://doi.org/10.1093/ornithapp/duab022>
- Hockey, P., & Wilson, W. (2003). Reproductive traits of marine terns (Sternidae): Evidence for food limitation in the tropics? *Ostrich*, 74(1–2), 110–116. <https://doi.org/10.2989/00306520309485376>
- Hu, G., Feeley, K. J., Wu, J., Xu, G., & Yu, M. (2011). Determinants of plant species richness and patterns of nestedness in fragmented landscapes: Evidence from land-bridge islands. *Landscape Ecology*, 26(10), 1405–1417. <https://doi.org/10.1007/s10980-011-9662-7>
- Jaquemet, S., Potier, M., Cherel, Y., Kojadinovic, J., Bustamante, P., Richard, P., Catry, T., Ramos, J. A., & Le Corre, M. (2008). Comparative foraging ecology and ecological niche of a superabundant tropical seabird: The sooty tern *Sterna fuscata* in the south-west Indian Ocean. *Marine Biology*, 155(5), 505–520. <https://doi.org/10.1007/s00227-008-1049-1>
- Jiménez, A., Rodríguez, P., & Blanco, P. (2009). Cuba. In P. E. Bradley & R. L. Norton (Eds.), *An inventory of breeding seabirds of the Caribbean* (p. 11). University Press of Florida.
- Kalinowski, T., Falbel, D., Allaire, J. J., Chollet, F., Tang, Y., Van Der Bijl, W., Studer, M., & Keydana, S. (2021). *keras: R interface to 'Keras'*. R package version 2.6.1. <https://cran.r-project.org/web/packages=keras>
- Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 24–49. <https://doi.org/10.1016/j.isprsjprs.2020.12.010>
- Krishna, S. T., & Kalluri, H. K. (2019). Deep learning and transfer learning approaches for image classification. *International Journal of Recent Technology and Engineering (IJRTE)*, 7(5S4), 427–432.

- Kristan, W. B., III. (2006). Sources and expectations for hierarchical structure in bird-habitat associations. *The Condor*, 108, 5–12.
- Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778–782. <https://doi.org/10.1109/LGRS.2017.2681128>
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177. <https://doi.org/10.1016/j.isprsjsprs.2019.04.015>
- Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., & Zhang, Y. (2018). Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. *Remote Sensing*, 10(7), 1119. <https://doi.org/10.3390/rs10071119>
- McGowan, J., Hines, E., Elliott, M., Howar, J., Dransfield, A., Nur, N., & Jahncke, J. (2013). Using seabird habitat modeling to inform marine spatial planning in central California's National Marine Sanctuaries. *PLoS One*, 8(8), e71406. <https://doi.org/10.1371/journal.pone.0071406>
- Muzaffar, S. B., Gubiani, R., & Benjamin, S. (2015). Nest location influences hatching success in the Socotra Cormorant (*Phalacrocorax nigrogularis*) on Siniya Island, United Arab Emirates. *Wildlife Research*, 42, 13. <https://doi.org/10.1071/WR14225>
- Navarro, N. (2021). *Annotated checklist of the birds of Cuba* (Vol. 4). Ediciones Nuevos Mundos.
- Nisbet, I. C. T. (2020). Distinguishing between roseate and common terns at breeding colonies in the West Indies. *Journal of Caribbean Ornithology*, 33, 117–118.
- Olier, I., Orhobor, O. I., Dash, T., Davis, A. M., Soldatova, L. N., Vanschoren, J., & King, R. D. (2021). Transformational machine learning: Learning how to learn from many related scientific problems. *Proceedings of the National Academy of Sciences of the United States of America*, 118(49), e2108013118. <https://doi.org/10.1073/pnas.2108013118>
- Oppel, S., Beard, A., Fox, D., Mackley, E., Leat, E., Henry, L., Clingham, E., Fowler, N., Sim, J., Sommerfeld, J., Weber, N., Weber, S., & Bolton, M. (2015). Foraging distribution of a tropical seabird supports Ashmole's hypothesis of population regulation. *Behavioral Ecology and Sociobiology*, 69(6), 915–926. <https://doi.org/10.1007/s00265-015-1903-3>
- Oppel, S., Bolton, M., Carneiro, A. P. B., Dias, M. P., Green, J. A., Masello, J. F., Phillips, R. A., Owen, E., Quillfeldt, P., Beard, A., Bertrand, S., Blackburn, J., Boersma, P. D., Borges, A., Broderick, A. C., Catry, P., Cleasby, I., Clingham, E., Creuwels, J., ... Croxall, J. (2018). Spatial scales of marine conservation management for breeding seabirds. *Marine Policy*, 98, 37–46. <https://doi.org/10.1016/j.marpol.2018.08.024>
- Orians, G. H., & Wittenberger, J. F. (1991). Spatial and temporal scales in habitat selection. *The American Naturalist*, 137, 29–49.
- Perrow, M. R., Harwood, A. J. P., Skeate, E. R., Praca, E., & Eglington, S. M. (2015). Use of multiple data sources and analytical approaches to derive a marine protected area for a breeding seabird. *Biological Conservation*, 191, 729–738. <https://doi.org/10.1016/j.biocon.2015.08.031>
- Pichler, M., & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, 14(4), 994–1016. <https://doi.org/10.1111/2041-210X.14061>
- Piper, W. H. (2011). Making habitat selection more “familiar”: A review. *Behavioral Ecology and Sociobiology*, 65(7), 1329–1351. <https://doi.org/10.1007/s00265-011-1195-1>
- Ponchon, A., Grémillet, D., Doligez, B., Chambert, T., Tveraa, T., González-Solís, J., & Boulinier, T. (2013). Tracking prospecting movements involved in breeding habitat selection: Insights, pitfalls and perspectives. *Methods in Ecology and Evolution*, 4(2), 143–150. <https://doi.org/10.1111/j.2041-210x.2012.00259.x>
- R Core Team. (2021). *R: A language and environment for statistical computing* (4.1.1) [Computer software]. R Foundation for Statistical Computing. <https://www.Rproject.org/>
- Raynor, E. J., Pierce, A. R., Leumas, C. M., & Rohwer, F. C. (2012). Breeding habitat requirements and colony formation by Royal Terns (*Thalasseus maximus*) and Sandwich Terns (*T. sandvicensis*) on barrier islands in the gulf of Mexico. *The Auk*, 129(4), 763–772. <https://doi.org/10.1525/auk.2012.11181>
- Rey-Villiers, N., Sánchez, A., & González-Díaz, P. (2021). Stable nitrogen isotopes in octocorals as an indicator of water quality decline from the northwestern region of Cuba. *Environmental Science and Pollution Research*, 28(15), 18457–18470. <https://doi.org/10.1007/s11356-020-09956-x>
- Rouse, J. W., Jr., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In S. C. Freden (Ed.), *Third Earth Resources Technology Satellite-1 Symposium. Volume I: Technical presentations, section A* (pp. 309–317). NASA.
- Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278. <https://doi.org/10.1109/JPROC.2021.3060483>
- Schreiber, E. A., & Burger, J. (2002). *Biology of marine birds*. CRC Press.
- Sorensen, M. C., Hipfner, J. M., Kyser, T. K., & Norris, D. R. (2009). Carry-over effects in a Pacific seabird: Stable isotope evidence that pre-breeding diet quality influences reproductive success. *Journal of Animal Ecology*, 78(2), 460–467. <https://doi.org/10.1111/j.1365-2656.2008.01492.x>
- Su, J.-H., Piao, Y.-C., Luo, Z., & Yan, B.-P. (2018). Modeling habitat suitability of migratory birds from remote sensing images using convolutional neural networks. *Animals*, 8(5), 66. <https://doi.org/10.3390/ani8050066>
- Vilchis, L., Ballance, L., & Fiedler, P. (2006). Pelagic habitat of seabirds in the eastern tropical Pacific: Effects of foraging ecology on habitat selection. *Marine Ecology Progress Series*, 315, 279–292. <https://doi.org/10.3354/meps315279>
- Wang, Z., Wang, J., Lin, C., Han, Y., Wang, Z., & Ji, L. (2021). Identifying habitat elements from bird images using deep convolutional neural networks. *Animals*, 11(5), 1263. <https://doi.org/10.3390/ani11051263>
- Winkler, D. W., Billerman, S. M., & Lovette, I. J. (2020). Gulls, terns, and skimmers (Laridae), version 1.0. In S. M. Billerman, B. K. Keeney, P. G. Rodewal, & T. S. Schulenberg (Eds.), *Birds of the world*. Cornell Lab of Ornithology. <https://doi.org/10.2173/bow.larida1.01>
- Zavalaga, C. B., Halls, J., & Dell’Omo, G. (2010). Marine habitat use of Peruvian boobies: A geographic and oceanographic comparison between inshore and offshore islands. *ICES Journal of Marine Science*, 67(5), 940–951. <https://doi.org/10.1093/icesjms/fsp298>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Garcia-Quintas, A., Roy, A., Barbraud, C., Demarcq, H., Denis, D., & Lanco Bertrand, S. (2023). Machine and deep learning approaches to understand and predict habitat suitability for seabird breeding. *Ecology and Evolution*, 13, e10549. <https://doi.org/10.1002/ece3.10549>