



HAL
open science

Correcting detection bias in mapping the abundance of marine megafauna using a Mediterranean seabird as an example

Charlotte Lambert, Jacopo Cecere, Federico de Pascalis, David Grémillet

► To cite this version:

Charlotte Lambert, Jacopo Cecere, Federico de Pascalis, David Grémillet. Correcting detection bias in mapping the abundance of marine megafauna using a Mediterranean seabird as an example. *ICES Journal of Marine Science*, 2024, 10.1093/icesjms/fsae058 . hal-04573237

HAL Id: hal-04573237


<https://hal.science/hal-04573237v1>

Submitted on 9 Oct 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.

Correcting detection bias in mapping the abundance of marine megafauna using a Mediterranean seabird as an example

Charlotte Lambert ^{1,2,*}, Jacopo G. Cecere³, Federico De Pascalis³, David Grémillet^{2,4}

¹ Littoral ENvironnement et Sociétés (LIENSs, UMR 7266), CNRS - La Rochelle Université, 17000 La Rochelle, France

² Centre d'Études Fonctionnelles et Évolutives (CEFE, UMR 5175), Univ Montpellier - CNRS - EPHE - IRD, 34000 Montpellier, France

³ Area Avifauna Migratrice, Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), 40050 Ozzano dell'Emilia, Italy

⁴ Department of Biological Sciences, FitzPatrick Institute of African Ornithology, University of Cape Town, Rondebosch 7701, South Africa

*Corresponding author. LIENSs UMR 7266 CNRS-La Rochelle Université, 17000 La Rochelle, France. E-mail: charlotte.lambert@univ-lr.fr

Abstract

Distance sampling surveys are extensively used to estimate the abundance of wide-ranging species but are prone to detection biases. This may be particularly acute for strip-transect protocols, which assume perfect detection. We examined this assumption by quantifying the detection probability of a declining seabird (Scopoli's shearwater, *Calonectris diomedea*), with particular attention to time of day and observation conditions at sea. We found detection probability was negatively affected by sun glare but positively by cloud cover and considerably dropped during mid-day hours due to circadian changes in behaviour (reduced detectability while resting). This result urges for systematically assessing and correcting detection bias when using strip-transect data to derive abundance information. Here, we did so by building a detection-corrected presence-absence ensemble model and combining it with a compilation of colony sizes and locations. A Monte-Carlo simulation ensured uncertainty propagation within and across data sources. The corrected abundance map showed shearwaters were largely prevalent in the central Mediterranean, Tunisia hosting most of the population both at sea and at colonies (45% of the global population; 79% of breeding pairs). This first accurate map is an essential conservation tool, emphasizing the importance of transnational actions for such species, that know no political boundaries.

Keywords: aerial survey; detection probability; detection bias; hour; distance sampling; *Procellariidae*; ACCOBAMS Survey Initiative; *Calonectris diomedea*; species distribution models

Introduction

Abundance estimates are paramount for conservation, but counting individuals is often not enough. We also need to know where they live, how they distribute, and use the geographic space available to them (Truchy et al. 2019). The spatial distribution of abundance is nonetheless challenging to estimate for wide-ranging and elusive marine megafauna (mammals, birds, reptiles, cephalopods, and fishes; Buckland et al. 2015). For these, large-scale distance sampling surveys (Buckland et al. 2015) based on visual observations from plane or boat platforms are the only reliable way to estimate population sizes and distributions. Biologging can also provide information on the spatial ecology of marine biota (Pereira et al. 2022), but they are often limited by sample size, accessibility, or restricted to specific population subsets, leading to biased estimates (Phillips et al. 2017), while observation-based surveys provide a homogeneous and representative assessment of entire populations.

In this context, the reliability of observation-based abundance estimates and associated species distribution models (SDMs) is critical but survey data are sensitive to detection biases (Lahoz-Monfort et al. 2014, Buckland et al. 2015). The detection of an animal present within the area surveyed by an observer depends upon the availability of the animal to detection (i.e. a diving individual is not visible) and its perception by the observer, which depends on observation conditions (sea

state, sea turbidity, sun glare, and cloud cover; Lambert et al. 2019).

Sighting an item at the sea surface, in flight or under the water, is also strongly dependent upon the sun angle relative to the sea surface (the closer the sun to the zenith, the farther we see under the sea surface). To accommodate for this during aerial surveys, flights are constrained to daylight hours when the sun is above a certain angle. Given this, the effect of time of day is generally neglected, but still may significantly impact perception. Whether this impact is stronger than that of other observation conditions (listed above) remains an open question.

Most current large-scale aerial surveys use multi-species protocols to provide valuable data about a wide variety of ecosystem components (from wildlife to anthropogenic activities; Panigada et al. 2021, Rogan et al. 2018, Pettex et al. 2014, Laran et al. 2017). Although such protocols do not hinder the capacity of observers to detect some taxa relative to others (Lambert et al. 2019), for practical reasons (mostly dependent on the frequency of observations), during a single survey, some taxa may be surveyed with line transects (e.g. cetaceans) but others with strip transects (e.g. seabirds).

The latter considers that detection is perfect, with a 100% probability of detecting an individual when it is present (Buckland et al. 2015). Yet, in the case of seabirds, this assumption may be violated since their detectability is also affected by

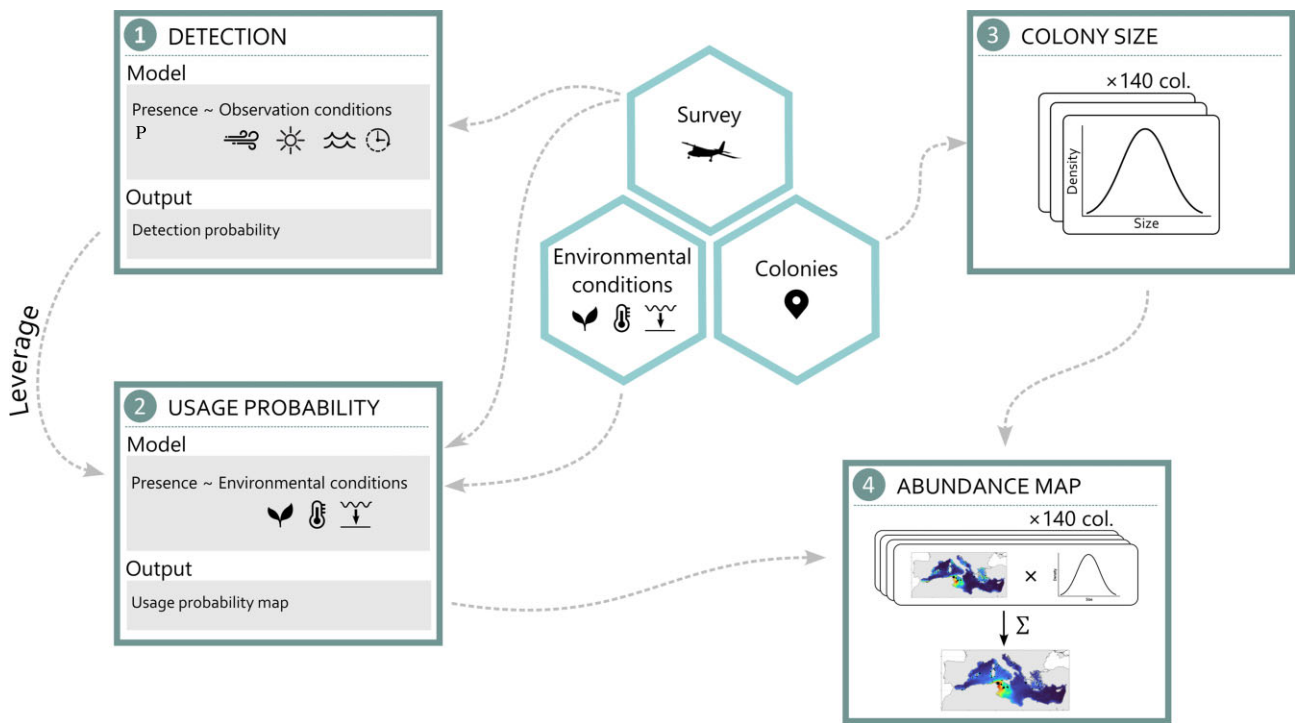


Figure 1. The study is based on three data sources and on a four-step analysis: (1) we modelled detection probability based on observation conditions experienced during the survey; (2) we modelled usage probability based on environmental conditions, leveraging by the detection probability; (3) we constructed the statistical distribution of colony sizes for each colony separately based on best estimates; and (4) we built colony-specific abundance maps by combining colony size distributions with usage probability maps predicted for each colony. By summing these colony-specific maps, we built the final Mediterranean-wide abundance map of Scopoli's shearwaters during the breeding season.

observation conditions. Failing to account for such detection bias hinders reliable abundance estimation from strip-transect surveys, and considerably challenges their relevance for population status assessment or spatial planning.

The Scopoli's shearwater (*Calonectris diomedea*) is a medium-sized migratory procellariiform species and a Mediterranean endemic, where it is one of the most abundant seabird species in summer, albeit severe bycatch is inducing population declines (García Barcelona et al. 2010, BirdLife International 2023). Yet, we still lack reliable basin-scale mapping of the species at-sea abundance during breeding. The last census of known colonies estimated the breeding population to 177 000 [141 000; 223 000] breeding pairs (mean [min; max]), which corresponds to 853 000 [679 000; 1 073 000] individuals roaming the Mediterranean (Defos du Rau et al. 2015). In the 2018 summer, the ACCOBAMS Survey initiative (ASI) was conducted over the entire Mediterranean. Although focusing on cetaceans, the survey implemented a multi-specific protocol and recorded all marine taxa and human activities. Based on this unprecedented information, two estimates of the population size were provided for Scopoli's shearwater, using different methods (Panigada et al. 2021): (1) the design-based estimate (219 333 individuals) simply scaled the observed density of individuals to the total area of the survey (following the strip-transect methodology), and (2) the model-based estimate (128 194 individuals; for the complete population, comprising both breeders and adult/immature non-breeders), derived from the abundance map predicted from a density surface model (DSM), incorporating the relationship of the species to its environment. The population estimates derived from ASI data are thus

surprisingly much lower than those derived from colony censuses.

Such a discrepancy questions the reliability of estimates based on aerial observation. Preliminary exploration of ASI data revealed this underestimation may originate from detection issues, with the suspicion of a significant effect of the time of the day. Such a pattern has to be thoroughly verified if we are to generate reliable distribution maps for species surveyed with strip-transect protocols.

Here, we first quantify detection bias for Scopoli's shearwaters during the ASI, assessing the effect of the time of day. We know from a recent study that although detection-biased SDMs cannot reliably estimate abundance, they still correctly estimate spatial patterns (Lambert et al. In review). We built upon these theoretical insights to propose a corrective method for SDMs based on biased strip-transect data (Fig. 1), using a Monte Carlo approach (Rubinstein and Kroese 2016) to ensure the reliable propagation of uncertainty into the final abundance map: (1) we estimated the detection probability based on observation conditions and time of the day; (2) we modelled the usage probability based on environmental conditions while correcting for detection bias, within an ensemble modelling approach; (3) we exhaustively censused colony locations and sizes and (4) we built colony-specific abundance maps by combining the colony size with the usage probability map predicted for the focal colony, which we summed to build a Mediterranean-wide abundance map of Scopoli's shearwaters during the breeding period. We make the abundance and the colony compilation available to be used for conservation, management, and other research purposes. Finally, we assessed the responsibility of Mediterranean border

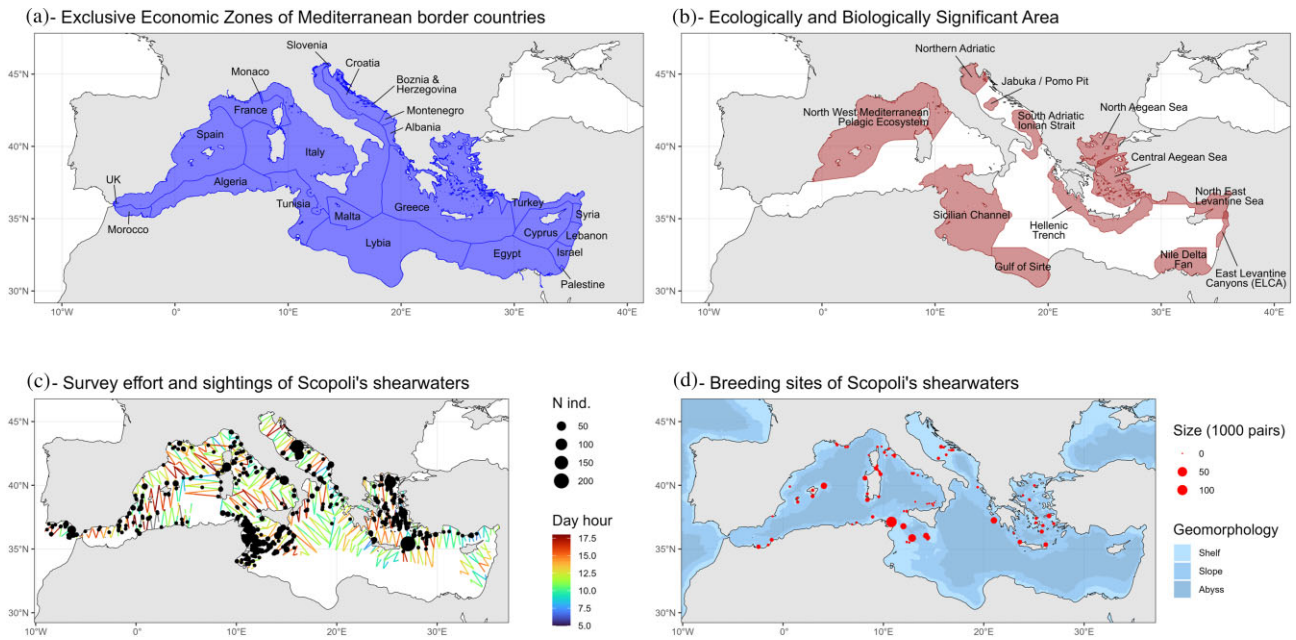


Figure 2. (a) Exclusive Economic Zones of the Mediterranean Sea. (b) Ecologically and biologically significant areas recognized under the Convention for Biological Diversity. (c) ASI survey effort colour-coded according to the time of day at which each segment was sampled, alongside Scopoli's shearwater sightings (point size proportional to the number of individuals, expressed as thousands of individuals). (d) Locations of breeding sites for Scopoli's shearwaters as compiled in the present study, where point size is proportional to the mean estimated number of breeding pairs in each site (in thousands of pairs). The maps are displayed in WGS 1984.

countries and MPAs for the conservation of the Scopoli's shearwater by quantifying the proportion of the shearwater population present in each EEZ or MPA, contrasting when the birds are at colony or at sea.

Mediterranean wildlife faces multiple threats from anthropogenic activities (overfishing, bycatch, pollution, and ship strikes; Coll et al. 2012), and a better understanding of where the species distribute while at sea will considerably help in implementing relevant and efficient mitigation strategies to reduce these threats. Our original approach, blending colony counts with at-sea observations, offers a generalist statistical toolkit that can be applied to similar datasets across all marine systems, with the aim of providing improved abundance maps and key spatial information for spatial planning and conservation purposes.

Materials and methods

All analyses described hereafter were performed in R version 4.2.3 (R Core Team 2023).

ASI data

The ASI was conducted from June to August 2018 (Fig. 2; Panigada et al. 2021, Lambert et al. 2020). High-wing aircraft equipped with bubble windows were used to scan the sea surface and sub-surface below the aircraft on the transect line (constant speed $\approx 167 \text{ km h}^{-1}$ (90 knots), height $\approx 183 \text{ m}$ asl). Two observers surveyed at all times during transect observations, one per side, while a third recorded data using the SAMMOA software (Observatoire Pelagis LRUniv-CNRS, Code Lutin, version 1.1.2. 2018).

Observation conditions (sea state, turbidity, cloud cover, glare severity, and glare orientation) were continuously recorded during survey effort, for each side separately, along-

side an overall observation conditions index. Observers recorded all shearwaters sighted within a strip of 200 m from the transect line (strip-transect methodology; Buckland et al. 2015).

Legs of homogeneous detection conditions were subsequently subdivided into 10 km segments, of which each side was considered separately (Lambert et al. 2020). The number of Scopoli's shearwaters detected was tallied per segment side, and the presence was considered one whenever at least one shearwater was sighted.

Compilation of colony locations

The locations of colonies were compiled from the literature and from expert knowledge, building a comprehensive database across the species breeding range and retrieving the most recent estimations of breeding pair numbers. The locations were compiled at the island scale: when several colonies were present on a single island, the whole island was registered as a breeding site.

Environmental data

We used 20 environmental variables known to affect seabird distribution (Wakefield et al. 2009). Bathymetry was retrieved from the GEBCO 2020 database; slope was derived from it ('terrain' function, 'raster' package; Hijmans et al. 2014). We downloaded dynamic variables from the Copernicus Marine Service (<https://data.marine.copernicus.eu>; Mediterranean Sea Physics Reanalysis and Mediterranean Sea Biogeochemical Reanalysis products), namely the monthly climatologies (June–August) for temperature (SST), sea surface height above geoid (SSH), sea surface salinity (SSS), mixed layer depth (MLD), chlorophyll *a* concentration (CHL), phytoplankton concentration, and net primary production (NPP). Current Speed ($\sqrt{U^2 + V^2}$) and Eddy Kinetic Energy (EKE;

$0.5 \times (U^2 + V^2)$) were estimated from northward (V) and eastward (U) sea water velocities. All variables were provided as mean (hereafter, avg) and standard deviation (sd). Environmental conditions were extracted at segment centroids for the corresponding month, and the Distance to the Nearest Colony was computed for each survey segment and each cell of the prediction grid. Variables with outliers were saturated to facilitate model fitting without discarding data: values below 1% or 2% and/or above the 98% or 99% quantiles were set to this quantile value where necessary (following visual inspection of data).

Species distribution modelling

Detection model

We first constructed a detection model to test for the presence of detection bias in ASI data and built a correction factor to be used in the usage probability model. We tested the effect of the time of the day, other observation conditions and segment length, as the latter influences the final number of sighted individuals per sampling unit (the longer the segment, the higher the number of sighted individuals).

We constructed the model using Generalized Additive Models (GAMs) relating either the presence (presence model) or the number of sightings of shearwater (the number of times the species was detected on a sampling unit; count models) to the time of day, segment lengths, observation condition index, glare severity, and cloud cover ('gam' function, 'mgcv' package; Wood 2011). The residuals were modelled with a Binomial and a Tweedie distribution, respectively (the Tweedie distribution is characterized by a mass at zero while being continuous on other positive real numbers). For both models, the splines were constrained to four degrees of freedom to avoid overfitting the data. We retrieved the explained deviances and RMSE (measuring the predictive capacity of the models; 'qpcR' package; Spiess 2018) and plotted the partial effects of each covariate with associated residuals. We compared each model based on these parameters; the best was used to predict a detection correction factor for each segment based on its time of day, length, and observation conditions.

We also tested whether the observed circadian variation in shearwater detectability was linked to behavioural changes, by collating the GPS data from a subset of Mediterranean colonies (French, Italian, and Tunisian colonies) during the breeding season and computing the activity budget over the course of a day (see Supplementary Material S1). Our GPS-based analysis showed shearwaters rested more on the water and flew less during mid-day hours, suggesting the low detectability of birds at this period of the day was behaviour-mediated.

Usage probability map estimation

We built a presence-absence GAM relating shearwater presence to environmental conditions, using the detection correction factor as a weight. We implemented a selection procedure that tested every combination of up to four variables in the model, discarding all combinations of variables whose Spearman's correlation coefficients were higher than 50%. Splines were bounded to four degrees of freedom maximum, models were fitted with the REML method, and residuals were assumed to follow a Binomial distribution.

It must be noted here that because shearwaters' movement ranges are larger than sampling units (they move in and out of

sampling units along the study period), the presence probability must be interpreted as the probability a location is used by at least one individual (Steenweg *et al.* 2018), i.e. our presence-absence model provides a usage probability map.

Explained deviances and RMSE were retrieved for of the 4391 tested models, which were sorted based on AIC. Delta AIC and Akaike weights were computed for all models ('akaike.weights', 'qpcR' package), and we selected the top-ranking models to be incorporated into the ensemble based on these metrics.

A Monte Carlo simulation approach ensured the propagation of uncertainty throughout the ensemble (Rubinstein and Kroese 2016). For each of the top-ranking models, the usage probability was predicted for every month separately. These were converted into distributions by drawing 500 samples, using mean and standard deviation predicted from the model, resulting in an array of 500 maps for each monthly prediction (for one cell, the 500 values follow a normal distribution with its mean and standard deviation matching the prediction from the model). The three monthly arrays were then averaged to obtain the final distribution of the usage probability map (an array of 500 samples of the map) for the considered model. Each sample was then normalized by its sum.

The usage probability maps of all single models were averaged together to obtain the final ensemble prediction (500 samples of the map), whose distribution was summarized as mean, standard deviation, 10%, and 90% quantiles of the final usage probability map.

Conversion to abundance

The usage probability map was converted to abundance by multiplying it by population size, following commonly used methods (Carneiro *et al.* 2020). We re-ran the presence-absence SDM for each colony separately to get colony-specific usage maps. Shearwaters being central-place foragers when breeding, all individuals of the population cannot use the entire basin and are spatially constrained within a distance to their colony. To account for such physical barriers, we replaced the Distance to Nearest Colony from the original SDM by the least-cost distance to the colony of interest ('costDist', 'terra' package; Hijmans 2023).

These usage probability maps were subsequently multiplied by the distribution of the number of breeders in the considered colony. This number of breeders was incorporated as a continuous process to account for the uncertainty in the estimate (Fig. 1), and was built by drawing 500 values from a normal distribution with the mean and standard deviation of the estimated number of breeders. We considered the lower and upper boundaries for colony size gathered in the colony compilation to be the 95% confidence interval to derive the mean and standard deviation used to build the size statistical distribution. Where the compilation only provided a single value, we applied a 10% standard deviation.

The obtained map of the abundance of breeders was divided by 0.485 to scale it up to the total number of individuals for that colony (using the tool developed by Carneiro *et al.* (2020) with demographic parameters from Jenouvrier *et al.* (2008) and Igual *et al.* (2009), breeding individuals were estimated to represent 48.5% of the population). Finally, we summed the abundance maps of all colonies and summarized the resulting arrays to retrieve the mean, standard deviation, 10%, and 90% quantiles of the final abundance map, which

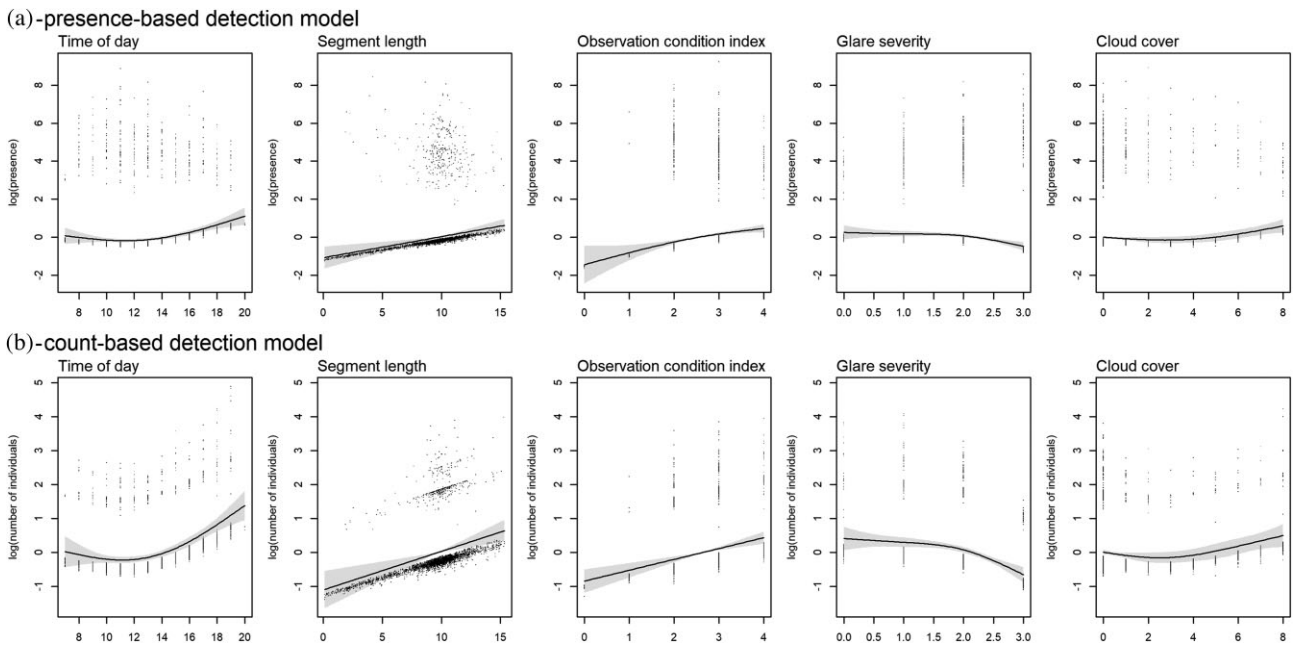


Figure 3. Partial effects of the five considered variables on (a) the presence of Scopoli’s shearwater on sampled segments and on (b) the number of Scopoli’s shearwater sighted per segment.

must be interpreted as the average number of shearwaters using each cell over the study period.

National responsibility and MPA relevance

To estimate country-specific population numbers, we: (1) measured the proportion of colonies and breeding pairs per country (number of colonies/breeding pairs per country divided by the total number of colonies/breeding pairs in the population); (2) we measured the proportion of the total Mediterranean surface encompassed in each country’s Economic Exclusive Zone (EEZ) and the proportion of the population using each country’s EEZ by summing the abundance of shearwaters in the EEZ and dividing it by the total population abundance over the entire basin. We finally built a responsibility score by leveraging the proportion of the population in each EEZ by the proportion of the basin this EEZ represents. These extractions were performed for mean estimates, lower 10%, and upper 90% estimates. EEZ contours were retrieved from the Flanders Marine Institute (Fig. 2a; Flanders Marine Institute 2019).

A similar analysis was conducted with the Ecologically and Biologically Significant Areas (EBSA) recognized under the Convention on Biological Diversity (<https://www.cbd.int/ebsa/>; Twelfth Conference of the Parties to the Convention on Biological Diversity 2014). Twelve areas have been recognized as meeting the EBSA criteria in 2022 (Fig. 2b), and we extracted the proportion of the population encompassed by each of these sites, following the same method as with EEZs.

Results

Survey and colony locations

The entire survey effort summed up to 50 935 km, with 467 sightings totalling 645 individuals. Large numbers of sightings occurred in the Gulf of Gabès and the Aegean Sea (Fig. 2c), as well as in the Tyrrhenian and Balearic Seas. The species was

also present in the central Adriatic Sea, Gulf of Lion, Alboran Sea, and Gulf of Cadiz. This distribution largely reflected that of breeding sites across the Mediterranean (Fig. 2d, Supplementary Material S2). The largest colony is located on Zembra island (113 720–176 750 pairs), off Tunisia, from where shearwaters mostly forage in the Tunisian plateau (Gulf of Gabès), alongside conspecific breeding in Malta (2665–3605 pairs), Pelagie Archipelago (10 070–10 120 pairs), and Pantelleria (500–5000 pairs).

Detection model

The presence model for detection explained 3.4% of the deviance, the count model 6.1%. The RMSE of the presence model was lower (59%) than that of the count model (68%). With the exception of cloud cover which showed a lower effect on the response variable ($P = 0.004$ for the presence model, $P = 0.01$ for the count model), all other variables had a highly significant effect ($P < 4.10^{-4}$) in both models: detection was better in late afternoon, with shearwaters virtually not sighted during mid-day hours (Fig. 3); detection increased with segment length and with the quality of observation conditions; detection was highest when the glare severity was the lowest and the cloud cover larger. Partial effect plots indicated residuals slightly larger for the presence model, but some patterns were present in those of the count model. Given these results, we chose the presence model to estimate the correction factor to be used in the SDM.

Usage probability map

Statistical metrics steadily increased along sorted models (Supplementary Material S3), except Akaike weights which identified the first three models to share most of the statistical weight (59%, 33%, and 7%, respectively) and the remaining models having weights near zero. However, RMSE and explained deviances indicated similar performances for the seven top models, which were thus chosen to perform ensemble predictions.

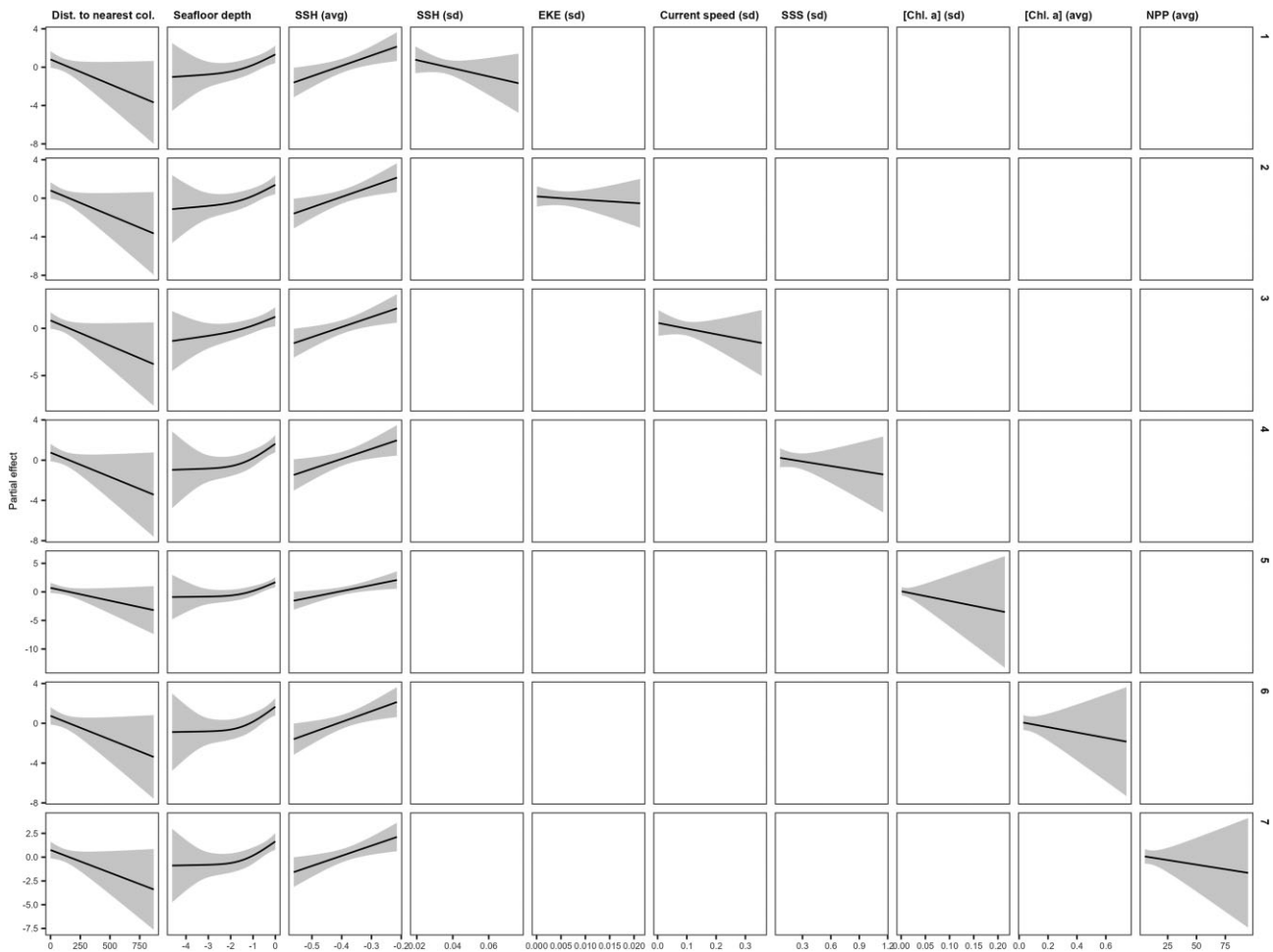


Figure 4. Partial effects of each covariate selected in the seven top-ranking models of Scopoli's shearwater spatial distribution of usage probability.

They all included seafloor depth, Distance to Nearest Colony and SSH (avg) (Fig. 4). Only the fourth covariate changed across models: SSH (sd); EKE (sd); Current Speed (sd); SSS (sd); CHL (sd and avg); NPP (avg). The seven models achieved an explained deviance of 20% (20.5%–20.3%).

The ensemble-based usage probability was predicted to be highest over shelf waters (Fig. 5a), especially on the Tunisian plateau and the Sicily Channel, the southern Alboran Sea, and the Aegean and Balearic Seas. The shelves surrounding the Tuscan Archipelago and along the mainland coast of the Tyrrhenian Sea, in the central Adriatic Sea, in the Gulf of Lion, and along the eastern Spanish mainland were also predicted to be extensively used by shearwaters. These high-use areas are in the vicinity of breeding sites, but the southern Mediterranean (Libyan and Egyptian coasts) and the Turkish shelf also appeared to be suitable for the species despite being farther from colonies.

Abundance map

When usage probability was converted into abundance and leveraged by colony size (Fig. 5b), the large prevalence of the central Mediterranean colonies was strongly visible: the Tunisian plateau and the Sicily Channel hosted by far the largest number of shearwaters. The Tyrrhenian Sea and the Algerian basin, hosting large-sized colonies, also held large abundances despite lower usage probability. Conversely,

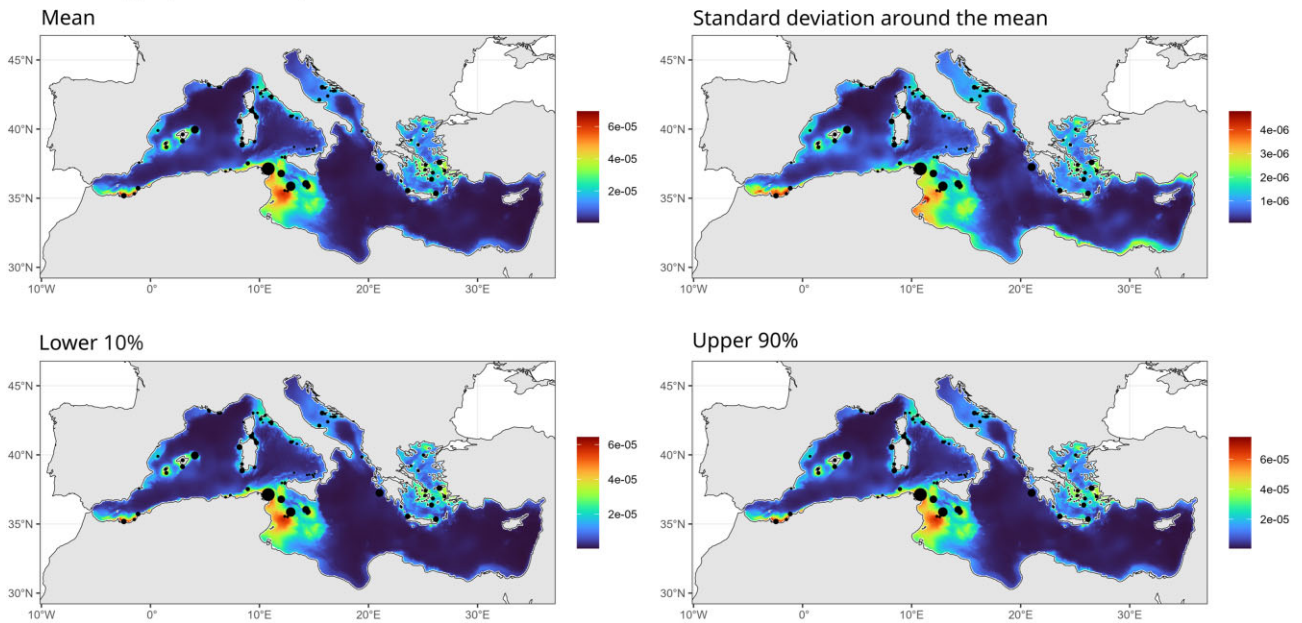
the Alboran Sea, despite having large usage probabilities, hosted only a small number of shearwaters due to the small size of the colonies there.

National responsibility

Tunisia stood out as the most important Mediterranean nation for the conservation of breeding Scopoli's shearwater, with 79.2% of global breeding pairs occurring there (Fig. 6a), mostly on Zembra (>110 000 pairs), with a few hundred pairs in La Galite. Italy, hosting the largest number of colonies ($n = 68$, 48.6% of colonies), of whom the second largest (Pelagie Archipelago) came second with 9.3% of pairs. Greece (5.3% of pairs in 10.7% of colonies [$n = 15$]), Spain (2.9% of pairs in 5.7% of colonies [$n = 8$]), and Malta (1.7% of pairs in 2.9% of colonies [$n = 4$]) also stood out. Algeria, Croatia, and France hosted the remaining 1.4% of the population, despite hosting 30.7% of colonies (17 colonies in Croatia, 22 in France, and 4 in Algeria).

These figures are, however, largely different when we consider the proportion of the population using EEZs (Fig. 6b): 44.8% of the population uses Tunisian waters, 22.3% Italian waters, 14.8% Libyan waters, and 6.2% Maltese waters. Algerian, Greek, and Spanish waters are each used by 3.2% of the population. The EEZs of France, Egypt, Croatia, Morocco, Turkey, Albania, Cyprus, and Montenegro are all used by <0.5% of the population.

(a) - Usage probability



(b) - Abundance distribution

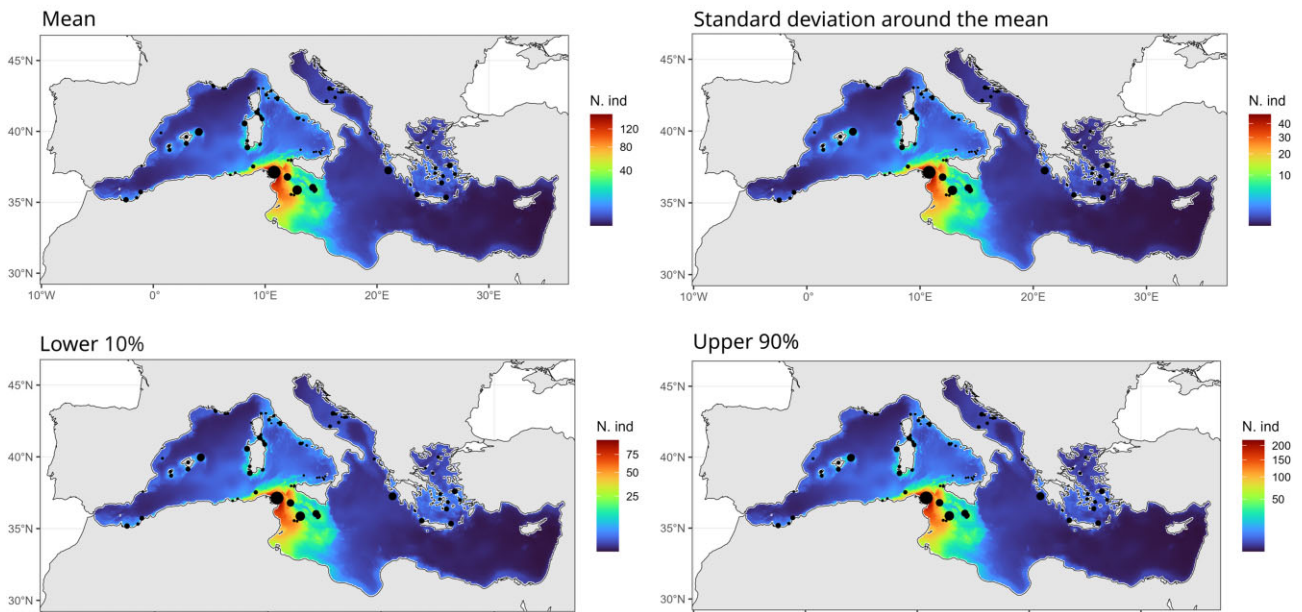


Figure 5. (a) Ensemble-based predicted usage probability map and (b) abundance map for the total population of Scopoli’s shearwaters in the Mediterranean Sea, expressed as numbers of individuals. For each is displayed the mean estimation, standard deviation around the estimation, lower 10%, and upper 90% of the estimated distribution. The colour ramp was square-rooted for readability in (b). The maps are displayed in WGS 1984.

Because Tunisia hosts the largest proportion of the population on a reduced portion of the Mediterranean (Fig. 6c), the country stood out as having by far the highest relative responsibility score, followed by Malta, Italy, and Libya.

4.6% [4.6%; 4.7%], the Gulf of Sirte with 2.3% [2.4%; 2.3%], and the Central Aegean Sea with 1.1% [1.4%; 0.9%] of the population. The remaining sites each encompassed <0.3% of the population.

EBSAs relevance to Scopoli’s shearwaters

All EBSAs but the East Levantine Canyon encompassed at least part of the shearwater population. The most relevant EBSAs were the Sicilian Channel with 73.8% [72.5%; 74.2%] (mean, [lower 10%; upper 90%]) of the population, followed by the North West Mediterranean Pelagic Ecosystem with

Discussion

Correcting for detection bias in species distribution models

The present work confirms the gap found between the ASI-based abundance estimates and the colony censuses is due to

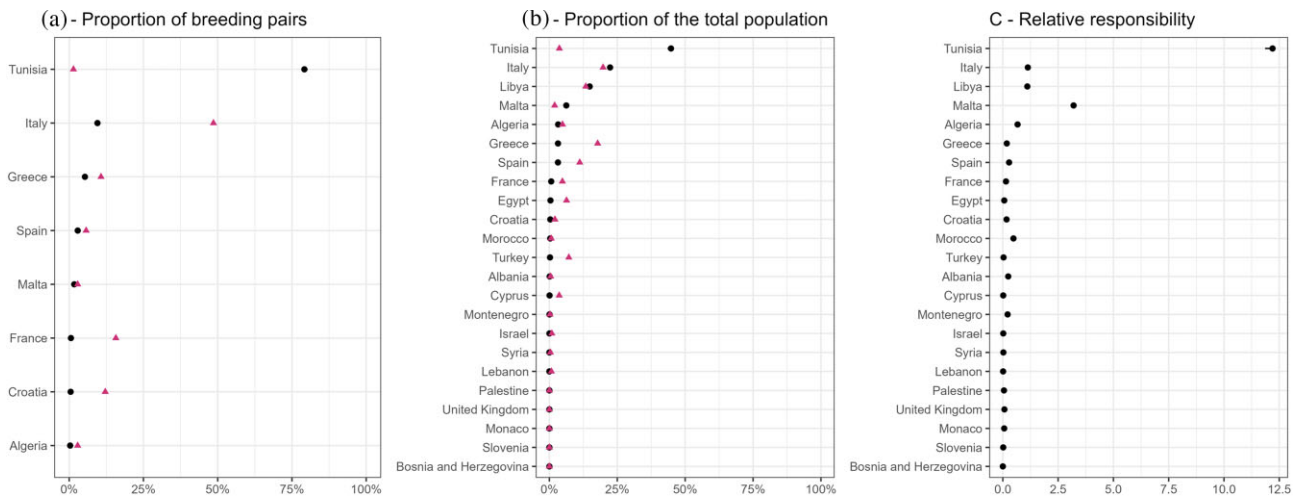


Figure 6. (a) Proportion of the total number of Scopoli's shearwater breeding pairs hosted per country (dots) and proportion of the total number of colonies hosted by each country (triangles). (b) Proportion of the total shearwater population using each country's EEZ (dots) and proportion of the Mediterranean surface covered by each country's EEZ (triangles). (c) Relative responsibility of each country regarding the shearwaters at sea, calculated as the ratio between the proportion of the population using the EEZ and the proportion of the Mediterranean this national EEZ represents. In all panels, the dots depict the mean size estimation and the whiskers the lower 10 and upper 90% around the mean (the uncertainty is extremely small, so that the whiskers are masked by the points, except for Tunisia in panel c).

an important detection bias: Time of the day and other observation conditions significantly affected the detection of shearwaters during aerial surveys, thereby violating the perfect detection assumption.

The effects of observation condition index and glare severity are consistent with most studies based on line-transect protocols: The worse the observation conditions, the less probable it is to detect an individual. In marine surveys, sun glare is one of the most impactful factors, hindering observers from seeing above or under the sea surface. On the contrary, cloud cover induces a sky glint on the sea surface, preventing observers from seeing animals below the sea surface but facilitating the detection of dark objects over it. Shearwater detectability is promoted by their characteristic flight, with fast-paced alternation of dark and light colours while birds glide over the waves. Such behaviour is more easily detected when the sea surface loses transparency, thereby leading to improved detection when cloud cover is larger.

The most impactful factor for shearwater detection probability was the time of the day. Interestingly, this major constraint had never been taken into account in previous studies, neither for seabirds nor for other marine megafauna. Here, we demonstrate that it may yet strongly affect abundance estimations. In our case, the elusiveness of the species to detection during mid-day hours is behaviourally mediated, with shearwaters spending more time resting on the sea surface at this time (Supplementary Material S1). Because lone and dark birds resting at the surface are especially difficult to detect, this circadian pattern in behaviour lowers the species detectability from plane during this period. This is consistent with the detectability of shearwaters being mostly mediated by their flight characteristics and with highest detection probabilities when cloud cover is important (facilitating the detection of scattered dark objects at the surface).

Whether such a pattern occurs in other species remains an open question, since circadian activity budgets are lacking

for most seabird species. The two other shearwater species present in the Mediterranean could also be affected by such a time-dependent detectability, as they are functionally similar to Scopoli's shearwaters. Still, different effects could be expected for other groups of seabirds (lighter-coloured species, or with different flight patterns), and an opposite effect might be expected for underwater species, whose detectability might be optimal when the sun is at its highest, with sunlight penetrating the farthest into the water column and facilitating the perception of animals below the surface.

Our results thus call for urgent consideration of detection probability variation at the daily scale, whatever the species and the observation protocol. Where possible, line-transect protocols should be implemented and the time of the day modelled in detection functions, alongside other factors. However, using line-transect protocol for all species during a survey is logistically complex, or even impossible when some taxa are highly abundant. Fortunately, various solutions may be envisioned to correct this bias when strip-transect protocol cannot be avoided.

Here, we combined detection-corrected presence-absence SDM with colony-based abundance estimations within a Monte-Carlo framework, ensuring uncertainty propagation (Rubinstein and Kroese 2016). This approach successfully overcame the violation of the perfect detection assumption and made the most out of the best available data. Our approach would nonetheless not be transferable to species for which observation surveys are the only reliable source of abundance estimation. In this case, however, correcting for detection bias is still achievable through different means, depending on the objectives.

Where the interest lies in identifying spatial patterns, predicting presence or usage probability distribution while correcting for detection bias should provide reliable results, as demonstrated here. However, results must be interpreted with caution, since a presence or usage probability map

provides a different picture of distribution than an abundance map.

An alternative, precautionary approach would be to model abundance through count-based approaches (DSMs, for example; Miller et al. 2013), corrected or not for detection bias, but only considering predictions as relative abundance by normalizing them. Previous simulation studies indicate that such models might be robust enough to correctly identify spatial patterns despite being biased for absolute predicted abundance (Lambert et al. In review), thus still providing reliable information about where to find most of a species population. Such an approach has previously been used successfully to assess the relevance of marine protected areas for specific populations (Lambert et al. 2017).

Abundance map and responsibilities for each Mediterranean country

We obtained a robust estimation of Scopoli's shearwater distribution throughout the Mediterranean Sea during its breeding period. The species was most abundant in shelf waters, particularly in the central Mediterranean along the Tunisian plateau and the Sicily Channel, where the largest colonies are located.

GPS-tracking studies of shearwaters conducted during this period in the Western Mediterranean (Balearic and Chafarinas Islands, Tyrrhenian Sea, Adriatic Sea, Sicily Channel; Louzao et al. 2009, Afán et al. 2014, Cecere et al. 2013, 2014, Grémillet et al. 2014, Péron et al. 2018) all identified space use patterns corroborating the predictions from our model. On the other side of the Mediterranean, our model predicted favourable areas along the African and Turkish coasts. Although previously not described, these favourable areas were recently confirmed as used by shearwaters breeding in Crete (Raül Ramos, unpublished data).

Our study also highlighted the Aegean Sea as a high-use area. Given this sea is largely impacted by maritime traffic, pollution, and bycatch, the large abundances there may hint to conservation issues and calls for dedicated regional research (Portolou et al. 2022).

The at-sea distribution of the species strongly underlines the responsibility of some Mediterranean countries towards its conservation. Specifically, although Tunisia hosts 79.2% of the breeding population (Defos du Rau et al. 2015), because the species distribute largely at sea, the Tunisian EEZ is used by 44.8% of the total population. Conversely, some countries whose colony-based responsibility is low have a high at-sea responsibility, their EEZ being used by many individuals (Malta, Algeria). Libya stands out as one of the few countries without any referenced colony but still hosting a significant part of the population (14.8%), probably mainly composed of birds associated with the colonies of Zembra, La Galite, Pelagie Archipelago, and Malta (Cecere et al. 2013, Grémillet et al. 2014). A similar pattern was found in Morocco, whose EEZ is probably used by birds breeding in Chafarinas and Las Palomas (Spanish territories; Afán et al. 2014).

Such a huge concentration of individuals in the relatively reduced space that is the Tunisian plateau confirms the outstanding value of the area at the Mediterranean scale (Coll et al. 2012). Being situated at the junction between Western and Eastern Mediterranean basins, the area is a particularly rich system for both benthic (corals, mærl) and

pelagic realms (seabirds, bluefin tunas, cetaceans, elasmobranchs, and sea turtles) and has been recognized as an Important Shark and Ray Area and an Important Marine Mammal Area.

When the area was recognized as meeting EBSA criteria during the Twelfth Conference of the Parties (Twelfth Conference of the Parties to the Convention on Biological Diversity 2014), it was considered that the area might host at least 30% of the Scopoli's shearwater population, potentially up to 50%. We show in the present study that this responsibility was in fact largely underestimated: the Sicilian Channel EBSA encompasses about half a million Scopoli's shearwaters during the breeding period (551 700 [347 100; 761 000]), which corresponds to 73.8% [72.5%; 74.2%] of the global population.

Our work provides a novel and robust analytical framework to estimate abundance maps of highly mobile marine species, ensuring the correction of behaviour-mediated temporal detection bias. These results add a new piece of evidence towards the necessity for transnational cooperation in achieving conservation efforts and management actions, and highlight the relevance of converting such international large-scale areas into effective marine protected areas (Bustamante et al. 2014).

Acknowledgements

We warmly thank Pep Arcos, George Karris, Sven Kapelj, and Ben Metzger for their help in compiling the colony sizes and locations. We thank Clara Péron, Nicolas Courbin, Simona Imperio, Giorgia Gaibani, Carlo Catoni, Andrea Benvenuti, Ivan Maggini, Danilo Pisu, and Davida Pala for their involvement in the GPS tagging activity of Scopoli's shearwaters at French and Italian colonies.

Author contributions

Charlotte Lambert and David Grémillet lead the study conception and design. Jacopo G. Cecere and Federico De Pascalis collected data on colony sizes and locations. Jacopo G. Cecere, Federico De Pascalis, and David Grémillet collected GPS data. Charlotte Lambert lead the compilation of colony locations, designed the methodology and carried out all the analyses. The first draft of the manuscript was written by Charlotte Lambert and all authors commented on previous versions of the manuscript, read and approved the final manuscript

Supplementary data

Supplementary material is available at *ICES Journal of Marine Science* online.

Funding

This work was supported by the European Horizon 2020 Programme EcoScope project under the grant agreement number 101000302.

Data availability

Availability of data and materials: ASI data can be freely accessed from the ACCOBAMS website (<https://accobams.org/asi-data-presentation/>). The R scripts to reproduce the work

are given in Supplementary Material 4. The GIS layer of collated colonies, the usage probability maps and the abundance maps are available on dedicated repositories on the ODATIS platform (<https://doi.org/10.12770/24f5da71-d49b-4ccc-ae5a-3fac604b2551> for the colonies; <https://doi.org/10.12770/d5bd5c63-b32b-4c55-83f0-9bd678a2ea76> for the emaps).

References

- Afán I, Navarro J, Cardador L *et al.* Foraging movements and habitat niche of two closely related seabirds breeding in sympatry. *Mar Biol* 2014;161:657–68.
- BirdLife International. Species factsheet: *Calonectris diomedea*. 2023. <http://datazone.birdlife.org/species/factsheet/scopolis-shearwater-calonectris-diomedea>. Accessed July 2023.
- Buckland S, Rexstad E, Marques T *et al.* Distance sampling: methods and applications. New York: Springer, 2015.
- Bustamante G, Canals P, Di Carlo G *et al.* Marine protected areas management in the Caribbean and Mediterranean seas: making them more than paper parks. *Aquat Conserv Mar Fresh Ecosyst* 2014;24:153–65.
- Carneiro AP, Pearmain EJ, Opper S *et al.* A framework for mapping the distribution of seabirds by integrating tracking, demography and phenology. *J Appl Ecol* 2020;57:514–25.
- Cecere JG, Catoni C, Maggini I *et al.* Movement patterns and habitat use during incubation and chick-rearing of Cory's shearwaters (*Calonectris diomedea diomedea*) (Aves: Vertebrata) from Central Mediterranean: influence of seascape and breeding stage. *Ital J Zool* 2013;80:82–89.
- Cecere JG, Gaibani G, Imperio S *et al.* Effects of environmental variability and offspring growth on the movement ecology of breeding scopolis's shearwater *Calonectris diomedea*. *Curr Zool* 2014;60:622–30.
- Coll M, Piroddi C, Albouy C *et al.* The Mediterranean Sea under siege: spatial overlap between marine biodiversity, cumulative threats and marine reserves. *Global Ecol Biogeogr* 2012;21:465–80.
- Defos du Rau P, Bourgeois K, Thévenet M *et al.* Reassessment of the size of the scopolis's shearwater population at its main breeding site resulted in a tenfold increase: implications for the species conservation. *J Ornithol* 2015;156:877–92.
- Flanders Marine Institute. Maritime Boundaries Geodatabase, version 11. 2019. <https://www.marinerregions.org/>. Accessed June 2023.
- García Barcelona S, Ortiz de Urbina JM, de la Serna JM *et al.* Seabird bycatch in Spanish Mediterranean large pelagic longline fisheries, 2000–2008. *Aquat Living Resour* 2010;23:363–71.
- Grémillet D, Péron C, Pons JB *et al.* Irreplaceable area extends marine conservation hotspot off Tunisia: insights from GPS-tracking Scopoli's shearwaters from the largest seabird colony in the Mediterranean. *Mar Biol* 2014;161:2669–80.
- Hijmans RJ. terra: spatial data analysis r package version 1.7-18. 2023. <https://CRAN.R-project.org/package=terra>.
- Hijmans RJ, Etten Jv, Mattiuzzi M *et al.* raster: geographic data analysis and modeling. 2014. <http://cran.r-project.org/web/packages/raster/index.html>.
- Igual JM, Tavecchia G, Jenouvrier S *et al.* Buying years to extinction: is compensatory mitigation for marine bycatch a sufficient conservation measure for long-lived seabirds? *PLoS One* 2009;4:e4826.
- Jenouvrier S, Tavecchia G, Thibault JC *et al.* Recruitment processes in long-lived species with delayed maturity: estimating key demographic parameters. *Oikos* 2008;117:620–8.
- Lahoz-Monfort JJ, Guillera-Arroita G, Wintle BA *et al.* Imperfect detection impacts the performance of species distribution models. *Glob Ecol Biogeogr* 2014;23:504–15.
- Lambert C, Authier M, Dorémus G *et al.* The effect of a multi-target protocol on cetacean detection and abundance estimation in aerial surveys. *R Soc Open Sci* 2019;6:190296.
- Lambert C, Authier M, Dorémus G *et al.* Setting the scene for Mediterranean litterscape management: the first basin-scale quantification and mapping of floating marine debris. *Environ Pollut* 2020;263:114430.
- Lambert C, Bonnet-Lebrun AS, Grémillet D. Bridging the gap between Lagrangian and Eulerian species distribution models for abundance estimation - A simulation experiment. In review.
- Lambert C, Virgili A, Pettex E *et al.* Habitat modelling predictions highlight seasonal relevance of Marine Protected Areas for marine megafauna. *Deep Sea Res 2 Top Stud Oceanogr* 2017;141:262–74.
- Laran S, Authier M, Van Canneyt O *et al.* A comprehensive survey of pelagic megafauna: their distribution, densities, and taxonomic richness in the tropical Southwest Indian Ocean. *Front Mar Sci* 2017;4:139.
- Louzao M, Bécarea J, Rodríguez B *et al.* Combining vessel-based surveys and tracking data to identify key marine areas for seabirds. *Mar Ecol Prog Ser* 2009;391:183–97.
- Miller DL, Burt ML, Rexstad EA *et al.* Spatial models for distance sampling data: recent developments and future directions. *Methods Ecol Evol* 2013;4:1001–10.
- Observatoire Pelagis LRUniv-CNRS—Code Lutin, version 1.1.2. SAM-MOA, Software dedicated for aerial survey of marine megafauna. 2018.
- Panigada S, Boisseau O, Cañadas A *et al.* Estimates of abundance and distribution of cetaceans, marine mega-fauna and marine litter in the Mediterranean Sea from 2018–2019 surveys. ACCOBAMS - ACCOBAMS Survey Initiative Project, 2021, <https://accobams.org/wp-content/uploads/2021/11/ASI-Med-Report-updated.pdf>.
- Pereira JM, Paiva VH, Krüger L *et al.* Tracking seabirds for conservation and marine spatial planning in Seabird Biodiversity and Human Activities. Boca Raton; CRC Press, 2022, 59–74.
- Péron C, Authier M, Grémillet D. Testing the transferability of track-based habitat models for sound marine spatial planning. *Divers Distrib* 2018;24:1772–87.
- Pettex E, Lambert C, Laran S *et al.* Suivi Aérien de la Mégafaune Marine en France métropolitaine—Rapport Final. Technical report Observatoire Pelagis UMS 3462 Université de La Rochelle—CNRS, 10.13140/2.1.2698.5287, 2014, https://www.researchgate.net/publication/268388832_Suivi_Aerien_de_la_Megafaune_Marine_en_France_metroropolitaine_-_Rapport_final.
- Phillips RA, Lewis S, González-Solís J *et al.* Causes and consequences of individual variability and specialization in foraging and migration strategies of seabirds. *Mar Ecol Prog Ser* 2017;578:117–50.
- Portolou D, Fric J, Evangelidis A *et al.* Seabirds of the Aegean. Berlin: Springer, 2022.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2023. <https://www.R-project.org/>.
- Rogan E, Breen P, Mackey M *et al.* Aerial surveys of cetaceans and seabirds in Irish waters: Occurrence, distribution and abundance in 2015–2017. Technical report, Dublin: Department of Communications, Climate Action & Environment and National Parks and Wildlife Service (NPWS), Department of Culture, Heritage and the Gaeltacht, 2018.
- Rubinstein RY, Kroese DP. Simulation and the Monte Carlo method, Wiley Series in Probability and Statistics. Hoboken: John Wiley & Sons, 2016.
- Spieß AN. qpcR: Modelling and Analysis of Real-Time PCR Data r package version 1.4-1. 2018. <https://CRAN.R-project.org/package=qpcR>.
- Steenweg R, Hebblewhite M, Whittington J *et al.* Sampling scales define occupancy and underlying occupancy–abundance relationships in animals. *Ecology* 2018;99:172–83.

- Truchy A, Göthe E, Angeler DG *et al.* Partitioning spatial, environmental, and community drivers of ecosystem functioning. *Landscape Ecol* 2019;**34**:2371–84.
- Twelfth Conference of the Parties to the Convention on Biological Diversity. Decision adopted by the Conference of the Parties to the Convention on Biological Diversity. 2014. UNEP/CBD/COP/DEC/XII/22. <https://www.cbd.int/doc/decisions/cop-12/cop-12-dec-22-en.pdf>. accessed Oct 2023.
- Wakefield ED, Phillips RA, Matthiopoulos J *et al.* Quantifying habitat use and preferences of pelagic seabirds using individual movement data: a review. *Mar Ecol Prog Ser* 2009;**391**: 165–82.
- Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Series B Stat Methodol* 2011;**73**: 3–36.

Handling Editor: Stephen Votier